DISS. ETH No. 27982

Understanding Asset Price Dynamics

in the Presence of Option Hedging, Circuit Breakers and Cash Flow Growth

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Florian Michael Till Ulmann
M.Sc. in Physics, University of Basel
born on *13.03.1989*
citizen of *Basel-Stadt, Switzerland*

accepted on the recommendation of

Examiner: Prof. Dr. Didier Sornette, ETH Zurich
Co-examiner: Prof. Dr. Stefano Battiston , University of Zurich

2021

Abstract

This dissertation is a compilation of four self-contained research papers that contribute to the literature of financial economics. All papers are co-authored by Professor Didier Sornette, my thesis supervisor, and further, depending on the project, by other authors. While three papers presented in this thesis were submitted to leading, peer-reviewed journals and are either awaiting review or are on revise and resubmit status, one paper is work in progress (the respective status is shown in the references as well as on the paper title page). The papers as presented in the thesis can include additional material that goes beyond the versions submitted.

Understanding the formation and dynamics of asset prices is essential for many regulatory, policy and investment decisions. Chapter 2 and 3 of this dissertation aim to cast light into the dynamics of asset prices in the presence of two feedback effects, option markets and circuit breakers, using relative pricing models. Chapter 4 presents an absolute asset pricing model that provides a framework on how to correctly value firms in the presence of high clash flow growth.

First, chapter 2 presents the discussion of a feedback effect between markets, namely from the options market onto the underlying's market. We show that if the option market makers (OMM) hedge a larger share of their position compared to option market takers (OMT), this leads to a net delta hedge demand in the spot market. And due to liquidity effects (price impact), this mechanically leads to an increase or decrease in the underlying's price volatility and a distortion of the underlying's drift, depending on the positioning of the option market. We describe this dynamic with a relative pricing model, relative to the price without option hedging. Then we focus on the impact of option hedging on the spot markets volatility in the foreign exchange (FX) market. We use trade repository data to estimate the net positioning of the OMM in the FX options market and with it her delta hedge demand in the spot market. We find the OMM to be short gamma on average (implying she has a short option positioning on average) and thus the volatility to be increased by delta hedging of the OMM. Next, we analyze the impact of option hedging on the price stability. Using theoretical arguments as well as simulations, we show that the probability for drift bursts, i.e. explosive directional price moves, can be increased or decreased due to option positioning, depending on the option parameters and delta of the OMM. Using option positioning data for the GameStock stock, we document that there are indeed instances where explosive trends due to hedging were more likely in the first quarter of 2021, during which the stock showed exceptional price instability.

Second, chapter 3 presents the discussion of a feedback effect that stems from the market architecture of

exchanges, i.e. the price dynamics of so called circuit breakers (CB). A price-based CB for example is a trading halt imposed by an exchange after the price drops below (or rises above) a pre-specified value within a trading session. The rational behind a CB mechanism is that it gives traders time to calm down, interact with their counterparts and rethink their trading decision. Given that investors are continuously forming anticipations, their tactical positioning in the presence of the CB may lead to feedback of the CB on the price itself. We model the price dynamic under a CB using a relative pricing model, relative to the price without a CB, in a closed-form setup, in which the investors' anticipation of the probability of a halt feedbacks on the price. In a general stochastic financial framework, this leads to coupled integral and stochastic differential equations. Our theory confirms two observations of this field for price-based CBs, namely an increased price volatility prior to the trigger and a negative effect of attracting the price to the CB level (known as the magnet effect). As our framework generalises to other CB mechanisms, we can then suggest a new CB design which minimises the adverse effects of a price-based CB. We e.g. show that a trading halt that is conditioned on a pre-specified volatility level, instead of a price level, mitigates the side effects.

Third, chapter 4 contributes to the discipline of absolute asset pricing. We aim to refocus the discussion around what could be loosely referred to as an expected return factor. Motivated by economic intuition and empirical insights from the consumption-based capital asset pricing model we show that the expected return on a firm level has a linear relationship to the firm's cash flow growth level. With that we argue that the discount rate should reflect this linear relationship and should thus have the same term structure as the assumed cash flow growth rate process. Based on that we first introduce a discounted cash flow (CDF) model with a time-varying discount rate matching the assumed cash flow growth rate process. This important feature greatly enhances the value of a DCF model, as the valuation is less sensitive to the cash flow growth rate process assumptions, as the discount rate is proportional to the assumed cash flow growth rate term structure, such that these two naturally complementary variables interact to massively reduce the range of reasonable valuation outcomes. Second, using the same time-varying discount rate in an internal cost of capital (ICC) model, we elaborate that the corresponding time-varying internal rate of return allows for a straightforward interpretation as investment return of a buy and hold strategy, compared to e.g. a constant ICC model, which by construction is a constant compound return. We show that our expected return proxy is superior in explaining realised returns compared to a constant internal rate of return, and even renders well known linear factors, which can be motivated from theory, insignificant. Last, the economic significance is confirmed in a trading strategy with yearly rebalancing.

Zusammenfassung

Diese Dissertation ist eine Zusammenstellung von vier in sich abgeschlossenen Forschungsarbeiten, die einen Beitrag zur Literatur der Finanzwirtschaft leisten. Alle Forschungspapiere wurden von Professor Didier Sornette, meinem Doktorvater, und, je nach Projekt, von anderen Autoren mitverfasst. Drei der in dieser Arbeit vorgestellten Forschungspapiere wurden bei führenden Fachzeitschriften eingereicht und befinden sich entweder im Begutachtungsprozess oder sind im Status der Überarbeitung, während eines eine laufende Arbeit ist (der jeweilige Status ist in den Referenzen sowie auf der Titelseite der Arbeit angegeben). Die in der Dissertation vorgestellten Forschungspapiere können zusätzliches Material enthalten, das über die eingereichten Versionen hinausgeht.

Das Verständnis der Bildung und Dynamik von Asset Preisen ist für viele regulatorische, politische und investitionsbezogene Entscheidungen von wesentlicher Bedeutung. In den Kapiteln 2 und 3 dieser Dissertation wird die Dynamik von Vermögenspreisen bei Vorhandensein von zwei Rückkopplungseffekten, Optionsmärkten und Circuit Breakern, mit Hilfe von relativen Asset-Pricing Modellen beleuchtet. In Kapitel 4 wird dann ein Bewertungsmodell vorgestellt, das eine korrekte Bewertung von Unternehmen mit hohem Cashflow-Wachstum zulässt.

Kapitel 2 widmet sich zunächst dem Rückkopplungseffekt zwischen dem Optionsmarkt und dem Preis des Basiswerts. Wir zeigen, dass wenn der Options Market Maker (OMM) einen grösseren Anteil ihrer Position absichert im Vergleich zum Options Market Taker (OMT), dies zu einer Netto-Delta-Hedging Nachfrage auf dem Kassamarkt führt. Aufgrund von Liquiditätseffekten führt dies mechanisch zu einer Erhöhung oder Verringerung der Preisvolatilität des Basiswerts und zu einer Verzerrung des Drifts des Basiswerts, je nach Positionierung des Optionsmarktes. Wir beschreiben diese Dynamik mit einem relativen Asset-Pricing Modell, bezogen auf den Basispreis ohne Einfluss von Optionsmärkten. Anschliessend konzentrieren wir uns auf die Auswirkungen der Optionsabsicherung auf die Volatilität der Preise am Devisenmarkt. Wir schätzen die Nettopositionierung der OMM auf dem Devisenoptionsmarkt und damit ihre Delta-Hedging Nachfrage auf dem Kassamarkt anhand von Transaktionsdaten ab. Wir stellen fest, dass der OMM im Durchschnitt eine Short-Gamma-Positionierung hat (was bedeutet, dass sie im Durchschnitt eine Short-Optionspositionierung hat) und somit die Volatilität durch ihr Delta-Hedging erhöht. Als nächstes analysieren wir die Auswirkungen der Optionsabsicherung auf die Preisstabilität. Wir zeigen anhand von theoretischen Argumenten und Simulationen, dass die Wahrscheinlichkeit von Drift-Bursts, d.h. explosiven direktionalen Preisbewegungen, durch die Optionspositionierung erhöht oder

verringert werden kann, abhängig von den Optionsparametern und dem Delta der Optionen des OMM. Zudem belegen wir anhand von Optionspositionierungsdaten für die GameStock-Aktie, dass im ersten Quartal 2021, in dem die Aktie eine aussergewöhnliche hohe Preisinstabilität aufwies, es tatsächlich einige Situationen gab, in denen explosive Trends aufgrund von Delta-Hedging Aktivität sehr wahrscheinlich waren.

Kapitel 3 widmet sich einem Rückkopplungseffekt, der sich aus der Marktarchitektur von Börsen ergibt, nämlich die Preisdynamik unter sogenannten Circuit Breakers (CB). Ein preisbasierter CB zum Beispiel ist ein Handelsstopp, der von einer Börse verhängt wird, nachdem der Preis innerhalb einer Handelssitzung unter einen vorher festgelegten Wert fällt. Die Idee von CBs ist es den Händlern Zeit zu geben sich zu beruhigen, sich mit anderen Marktteilnehmern auszutauschen und ihre Handelsentscheidung zu überdenken. Da Anleger ständig versuchen die Zukunft zu antizipieren, kann ihre taktische Positionierung in Anwesenheit eines CB zu einer Rückkopplung auf den Preis selbst führen. Wir modellieren die Preisdynamik mit Hilfe eines relativen Asset-Pricing Modell, relativ zum Preis ohne CB, in einer geschlossenen Form, in der die Antizipation der Wahrscheinlichkeit eines Stopps auf den Preis zurückwirkt. In einem allgemeinen stochastischen Modell führt dies zu gekoppelten Integralgleichungen und stochastischen Differentialgleichungen. In Bezug auf preisbasierte CBs bestätigt unsere Theorie dabei zwei Beobachtungen aus diesem Bereich, erstens eine erhöhte Preisvolatilität vor dem CB Auslösepunkt und zweitens einen Effekt, der den Preis auf das CB-Niveau runterzieht (bekannt als Magneteffekt). Da unser Modell sich einfach auf andere CB Mechanismen generalisieren lÃ¤sst, schlagen wir dann ein neues CB-Design vor, das die negativen Effekte eines preisbasierten CBs minimiert. Wir zeigen z.B., dass ein Handelsstopp, der von einem vorher festgelegten Volatilitätsniveau ausgelöst wird, statt von einem Preisniveau, die diskutierten Nebenwirkungen abschwächt.

Kapitel 4 leistet letztlich einen Beitrag zur Disziplin der absoluten Asset-Pricing Modelle. Unser Ziel ist es, die unstrukturierte Faktormodell Diskussion auf einen einzigen simplen Faktor zurückzuführen, den man grob als "erwartete Rendite" Faktor bezeichnen könnte. Auf der Grundlage von ökonomischer Intuition und empirischen Erkenntnissen aus dem konsumbasierten Capital Asset-Pricing Modell zeigen wir, dass die erwartete Rendite auf Unternehmensebene eine lineare Beziehung zum Cashflow-Wachstum des Unternehmens aufweist. Daher argumentieren wir, dass der Diskontierungssatz diese lineare Beziehung stets widerspiegeln und somit die gleiche Laufzeitstruktur wie der angenommene Cashflow-Wachstumsprozess haben sollte. Auf dieser Grundlage führen wir zunächst ein Discounted-Cashflow-Modell (DCF) mit einem zeitlich variablen Diskontierungssatz ein, der die gleiche Laufzeitstruktur wie der angenommene Cashflow-Wachstumsprozess hat. Dieses wichtige Merkmal erhöht den Wert eines DCF-Modells erheblich, da die Bewertung weniger empfindlich auf die Annahmen des Cashflow-Wachstumsprozess reagiert, da der Diskontsatz proportional zur angenommenen Laufzeitstruktur der Cashflow-Wachstumsrate ist, so dass diese beiden sich natürlich ergänzenden Variablen zusammenwirken und den Bereich der Bewertungsergebnisse auf ein vernünftiges Mass reduzieren. Weiter zeigen

wir unter Verwendung desselben zeitvariablen Diskontsatzes in einem internen Kapitalkostenmodell (ICC) auf, dass unsere zeitvariable ICC eine einfache Interpretation als Investitionsrendite einer "Buy and Hold"-Strategie hat, verglichen z. B. mit einem konstanten ICC-Modell, das konstruktionsbedingt diese Interpretation verfehlt. Wir zeigen, dass unser Proxy für die erwartete Rendite die realisierten Renditen besser erklärt als ein konstantes ICC-Modell und sogar bekannte lineare Faktormodelle, die sich aus der Theorie ableiten lassen, unbedeutend macht. Schliesslich wird der Wert unserer Idee in einer Handelsstrategie mit jährlichem Rebalancing bestätigt.

Acknowledgments

This dissertation has benefited from the input of various people and several institutions, who all have accompanied my PhD over the last years. Without their support, guidance, inspiration, critical comments, material support and infrastructure, the current thesis could not have been conducted.

First and foremost, I express my profound gratitude and sincere respect to Professor Didier Sornette, my PhD supervisor. Choosing a PhD supervisor is a big and difficult decision. And with all private and professional engagements of this kind, only time reveals if it was a good decision. Looking back, I can say that I got very lucky. First, simply because I always felt complete support in all my academic, professional and even private endeavours, which accompanied the past four years of my PhD journey. And this support was carried out in a very calm, pragmatic and empathic manner, which I highly appreciated. Second, Didier showed an incredible open-mindedness and enthusiasm toward all kind of research questions, which allowed me to explore a diverse set of topics in my PhD. And third, Didier constantly pushed our projects to high quality, with his seemingly endless source of energy, ideas, knowledge and critical thinking, for which I am also very grateful. Didier became a role model for me in various ways, and I very much look forward to keep working with him in the future.

I am also deeply grateful to Professor Stefano Battiston for being the co-examiner of my thesis, and Professor Michael Ambühl for presiding over the examination committee. And also to Judith Holzheimer to have provided smooth and efficient administrative support. Moreover, I want to thank all former and current members of Professor Sornette's Chair of Entrepreneurial Risks, a truly unique community of enthusiastic, talented and interdisciplinary people. I am fortunate and grateful to have been part of this community.

Next I would like to thank all my co-authors, i.e. Benjamin Anderegg, Sandro Lera, David Solo, Didier Sornette and Alexander Wehrli. It is their ideas, endurance and knowledge that made the presented research possible. Thereby a special "thanks" goes to Alexander Wehrli. We both worked together as quantitative analysts at the SNB Singapore Branch, while pursuing our PhDs aside at the same chair. Alex combines exceptional intellectual, technical and project management skills, and I was fortunate enough to observe his ways during our shared time in Singapore and to soak in as much as I can. Our countless insightful discussions thereby resulted, among other things, in one of the papers presented in this thesis. But more importantly, working with him also resulted in a friendship, which I am very grateful for.

The research conducted as part of this thesis was kindly supported by the Swiss National Bank. I am in particular grateful to Benjamin Anderegg, Thomas Maag and Martin Schlegel for their support of my doctoral studies. Further I want to thank the whole FX unit (among others Fabian, Simon, Lukas, Marco, Michael, Jan, Tanja, Kerstin, Christian) as well as the SNB Singapore branch (among others Alex, Björn, Marc, Alex) for the great time I could spend with them. It was a real honour to have worked alongside such smart and passionate people and to have contributed to the great purpose of implementing the SNB's monetary policy mandate.

Going back some years, I would also like to thank Professor Corey O'Hern. He accepted me as Master student in his group at Yale University and during this time introduced me to the beautiful and powerful approach of using modern data science methods to answer interdisciplinary research questions. His impact on this thesis cannot be underestimated.

I am also grateful to the Studycenter Gerzensee, a foundation of the Swiss National Bank, for allowing me to participate on the "Swiss Program For Beginning Doctoral Students in Gerzensee". My PhD would not have been the same without this fantastic experience. Spending more than 10 weeks in a castle with a private lake, surrounded by the Swiss Alps, and learning from the top Professors of their respective field about economics is already a pretty good deal in itself. But it was the participants that transformed this great deal into a truly unforgettable year. So, a big, big "thank you" to my fellow classmates of the year 2018/2019 (in alphabetic order): Anis, Anja, Beatrice, Benjamin, Buihamy, Chen, Di, Elio, Evert, Fabienne, Felix, Fintan, Flavia, Giulia, Hyacinthe, Jeremy, Juliette, Kyungbo, Laura, Lorenz, Lucas, Matthias, Maxime, Nicola, Pia, Radu, Sarah, Stephanie, Tabea. Not only did I greatly enjoy the intellectual exchange with all of you, but it is the various friendships that formed during this time that I am super grateful for! I very much look forward to all future alumni events!

And besides the various friendships, I also walked away from Gerzensee with newfound love. Juliette Cattin, I am deeply indebted to you for your unconditional support throughout these last years. It is your love, kindness as well as intellectual and emotional support, that made the last years such a light and happy time. And you turned out to be the best home-office buddy I could have ever wished for.

Last but not least I would like to thank my parents, Beatrice Ulmann and Hans-Peter Ulmann, and my sister, Laila Ulmann. Without your unconditional love, empathic support, but also critical feedback, I would not have been able to achieve anything that lead up to this thesis as well as the thesis itself. It is our family's foundation that gives me perspective in difficult times and keeps me grounded in good times.

Zurich, 2021 *Florian Ulmann*

To Beatrice, Hans-Peter and Laila Ulmann, as well as Juliette Cattin

Contents

1 Introduction and Motivation 1

2 Implications of Option Hedging on Asset Prices 7

The Impact of Option Hedging on the Spot Market Volatility 10

On the Directional Destabilizing Feedback Effects of Option Hedging 43

3 Price Dynamics under a Circuit Breaker 75

4 Dynamical Cost of Capital Driven by Dynamical Cash Flows 107

5 Conclusion 147

1 Introduction and Motivation

Asset pricing is the science, or art, of valuing an asset under careful consideration of all internal and external factors. Understanding asset price dynamics thereby is not only highly relevant for private and institutional investors, to efficiently allocate capital, but also for policy makers, to regulate financial markets appropriately. The 2013 Nobel price in economics, that was given to Eugene Fama, Lars Peter Hansen and Robert J. Shiller for their research towards a better understanding of asset prices, confirmed this relevance.

Describing asset price dynamics is challenging by nature. Prices are influenced by countless exogenous and endogenous interconnected variables. And the underlying complexity probably even increased throughout the years, as the globalization of markets and the spill-over effects between these global markets additionally complicates an accurate description of an asset value.

At intraday time scales, financial price properties have been documented to deviate significantly from what a random walk hypothesis would predict. This disaccord suggests the presence of frictions, and naturally raises the question of how such effects can be explained. This is where this thesis aims to contribute to. It discusses asset price dynamics in the presence of three different effects. And even though the discussed effects are not connected with each other in nature, all came to the centre of the attention in one or the other form over past few years, as each is at the core of some of the most extreme price dynamics observed in recent years: The first is the effect that option markets have on the price of the underlying. The second is the effect that circuit breakers have on asset prices. And the third is the effect that cash flow growth has on the price of a stock.

Chapter 2 starts with the discussion of the impact of option hedging on the price dynamics of the underlying asset price. This effect was powerfully displayed in the GameStop share price in the beginning of 2021. An army of retail traders bought large amounts of options, for which the market makers had the other side of the transaction. The delta hedging of the market maker then, among others, presumably contributed to extreme volatility in the GameStop share price.

We present a theoretical framework that models this feedback effect. Based on an economy of two traders, an option market maker (OMM) and an option market taker (OMT), which both have different delta hedge ratios, we present a model that links option positioning to the underlying's price dynamics via liquidity frictions from the net delta hedging demand. Our model predicts that the spot market volatility is increased (decreased) by a negative (positive) gamma exposure of the OMM, whereby the amount of the increase depends on the net delta hedge amount executed in the market of the underlying asset. Further, it predicts that delta hedging can even increase or decrease the probability for explosive directional price moves, so called drift bursts. While **Anderegg** et al. (2021) discusses option positioning with a focus on the underlying's volatility, validated in the FX market, **Sornette** et al. (2021) focuses more on the impact of option positioning on the price stability (i.e. directional price moves), validated in the GameStop stock market. In both markets we can validate our model with very high confidence and find clear indication that the volatility is increased due to the short option positioning of the option market maker. And for the GameStop stock market we find that option positioning indeed suggested explosive price behaviour in the first quarter of 2021.

Next, chapter 3 discusses the feedback effect of circuit breakers (CBs, i.e. trading halts a certain trigger mechanism is activated). In early 2016, a new price-based CB rule was implemented on the CSI 300 financial index which replicates the top 300 stocks traded in the Shanghai and Shenzhen stock exchanges. This new design supposedly lead to a period of very high volatility. On the very first day at which the new CB design was implemented, as well as two days later, the CB got triggered twice per day, as the price on those days dropped -5% and -7% respectively. The Chinese authorities subsequently abandoned the CB mechanism from their index, as it obviously did more harm than it helped. This regulatory nightmare motivated the work presented in chapter 3 of this thesis. The aim was to understand how price barriers might influence the price dynamics and to provide a framework that allows the study of different CB designs.

In **Lera** et al. (2021) we thus model the dynamics of an asset price in the presence of a CB. In short, using a phenomenological approach, we model how the investors' anticipation of the probability of a trading halt feedbacks on the price process. In line with prior theoretical and empirical work for price-based CBs, our theory predicts an increased price volatility prior to the trigger point and shows that the sole existence of a circuit breaker leads to the negative effect of attracting the price to the circuit breaker level. But more importantly, we show how our framework can be generalised to other CB trigger mechanisms and thus be used to design more effective CBs without the described adverse side-effects of a price-based CBs, e.g. by basing the trigger mechanism on a volatility level rather than on a price level.

Last, chapter 4 presents a valuation model ideally suited for high growth firms. Valuing high growth firms is a difficult endeavour, as one can see in the high volatility of stock prices associated with high growth expectations, e.g. technology stocks. This high volatility could be related to the hype surrounding these firms, but also stems from the high sensitivity of standard valuation models to a change in the assumed cash flow growth rate path. With the NASDAQ Composite at its all-time high, the question of how to properly value high growth firms is a very relevant one to ask, as potential large-scale mispricing can have severe economic consequences.

In **Solo** et al. (2021) we present a model that values high growth firms in a consistent manner which is less sensitive to the cash flow growth rate assumptions. First we use economic intuition as well as the consumption based asset pricing model to analyze how the expected return of a firm is related to the cash flow growth rate level. Motivated by the clear linear relationship between the two variables, we suggest a time-varying expected return process which has the same structural form as the assumed cash flow growth rate process. Based on this we present a valuation model as well as an internal cost of capital model (ICC) that uses our time-varying expected return structure, and therefore correct conceptual drawbacks that come along with constant discount rate models and time-varying discount rate models not directly linked to the growth rate path. Specifically, our valuation is less sensitive to the cash flow growth rate process assumptions, as the discount rate is proportional to the assumed cash flow growth rate term structure, such that these two naturally complementary variables interact to massively reduce the range of reasonable valuation outcomes. And the corresponding time-varying ICC model allows for the interpretation as investment return of a buy and hold strategy, compared to e.g. a constant ICC model, which by construction fails this interpretation, as it is a constant compound return. Empirically, using the expected return proxy of our time-varying ICC model as control in a Fama-MacBeth regression of the profitability, the investment and the value factor, both the profitability and the value factor become insignificant in explaining future realised returns. The superiority and economic significance is also confirmed in a trading strategy.

In summary, this thesis has the purpose to contribute to a better understanding of price dynamics under the influence of options markets, circuit breakers and cash flow growth. While chapter 2 about the influence of option positioning as well as chapter 3 about the influence of circuit breakers are probably of more interest to policy makers, the third chapter about valuing stocks in the presence of high cash flow growth hopefully can help private and institutional investors to make more efficient capital allocation decisions.

In the following three chapters the three self-contained research topics are presented. Each chapter covers one of the above discussed effects. Each chapter is proceeded by a general introduction and then presents the corresponding papers [1]. Chapter 5 subsequently concludes, discusses the challenges and shortcomings of the thesis from a broader perspective than adopted in each paper and gives direction of possible future research.

---

[1] All papers are co-authored by Professor Didier Sornette, my thesis supervisor, and further, depending on the project, by other authors from ETH Zurich and the Swiss National Bank. Thereby the research question was formulated jointly, the articles were written jointly and all authors were involved in theoretical formulation, data analysis, simulation studies as well as literature reviews.

Research Papers Presented in this Thesis

B. **Anderegg**, D. **Sornette**, and F. **Ulmann**. The Impact of Option Hedging on the Spot Market Volatility. *Revise and resubmit at the Journal of International Money and Finance*, 2021

S. **Lera**, D. **Sornette**, and F. **Ulmann**. Price Dynamics with Circuit Breakers. *Revise and resubmit at the Journal of Banking and Finance*, 2021

D. **Solo**, D. **Sornette**, and F. **Ulmann**. Internal Cost of Capital Driven by Dynamical Cash Flows. *Reprint submitted to the Journal of Financial Economics*, 2021

D. **Sornette**, F. **Ulmann**, and A. **Wehrli**. On the Destabilizing Feedback Effects of Option Hedging. *Working Paper*, 2021

2 Implications of Option Hedging on Asset Prices

The effect of options markets on the price dynamics of the underlying asset was powerfully displayed in the equity market in the beginning of 2021. The price of the shares of the company GameStop showed erratic price moves and high volatility. The assumed reason was the positioning of the market makers in the options market of the GameStop shares, which lead to extreme price dynamics due to their presumed hedging behaviour. According to the **Financial Times** (2021) "small investors can influence share prices by buying large amounts of call options, [...]. Typically, this means that wholesale brokers such as banks need to purchase shares on the open market to hedge themselves [...]. This hedging can have an outsized impact on prices [...]." This recent extreme example of the implications of option hedging on price dynamics gives this chapter great relevance.

However, the effect is not new and has actually been observed in different markets at different times. Already in 1993 the **Group of Ten** (1993) reported that "the existence of options and related dynamic hedging could increase volatility especially in the smaller and less liquid currency segments, as the spot exchange rate approaches the strike price. When strike prices and or option maturities are highly concentrated, a large volume of one-way hedging could occur in a short period. Market participants reported that sharp movements in spot prices were frequently observed as a result of such concentrations." Thus, understanding price dynamics under option hedging has been and still is of great interest and naturally also also lead to a vast body of academic literature over the last years. While the two papers of this chapter each discuss the content-specific literature in more detail, a high-level embedding of the topic is presented in the following paragraphs.

A crucial assumption of the Black-Scholes model is a frictionless and elastic market, i.e., a market in which traders have no price impact (**Gueant** and **Pu**, 2015). However, research concerned with market microstructure clearly shows that one can reject the assumption of a frictionless market (**Farmer** et al., 2006). In the context of dynamic hedging strategies, this implies a feedback effect on market price dynamics: the dynamic hedging strategy requires the investor who sells (buys) an option to sell (buy) the asset in the spot market when its price declines and to buy (sell) it when its price goes up. From the nature of such a trading pattern, one can expect an increase (decrease) in spot market volatility, given that these hedging trades have a price impact. And indeed, spot market volatility has been reported to have changed due to the introduction of futures and options (**MacKenzie**, 2008).

Among the first, **Frey** and **Stremme** (1997) and **Sircar** and **Papanicolaou** (1998) explain the increase in the volatility using a sample economy with two agents with different hedging behaviour in an equilibrium model. The framework used in both papers presented in this chapter is similar to these models in nature. Nevertheless, these existing absolute pricing models are vastly complicated and due to their complicated setup, their practical usability is limited and a proper empirical validation is prevented. Indeed, most empirical papers do not present their results on theoretical grounds, but only look at empirical proxies like gamma imbalances when discussing effects on price dynamics (**Barbon** and **Buraschi**, 2020). Further, prior models mostly only allow for the discussion of the impact of option hedging on the price volatility, instead of the impact of option hedging on the general price dynamics (i.e. on the volatility and the drift).

In a collaboration between ETH Zurich and the Swiss National Bank (Benjamin Anderegg, Head Foreign Exchange and Alex Wehrli, Senior Quantitative Analyst) we thus aimed to propose a theory that not only accurately accounts for the implications of option hedging on the general price dynamics, but which is also simple enough to allow for straightforward empirical testing. We thus propose a relative pricing model motivated only by basic (option) market properties. We thereby relate the net option hedging demand of the total market with the price dynamics of the underlying through only observable or measurable quantities and without assuming anything else but a linear price impact (in line with the seminal work of **Kyle** (1985), but also clearly supported by modern dynamical theory of liquidity and price impact from **Donier** et al. (2015)).

Based on our model, we then present the following two main contributions: First, our model is simple and general enough to allow for straightforward empirical testing, as only observable or measurable quantities are used. Thus, we can actually statistically validate our theory for different markets, and thus empirically quantify the effect on the grounds of a theoretical framework, which is a novel contribution (to the best of our knowledge). Second, our model setup also allows us to go beyond the implications of option hedging on the price volatility, to the implications of option hedging on price (in)stability, i.e. explosive directional price moves. To the best of our knowledge, this has not been explored on theoretical grounds by prior research. In the following two paragraphs I shortly discuss the key points of each paper.

In **Anderegg** et al. (2021) we focus on the impact of delta hedging on the underlying asset price volatility. We show that the volatility can be significantly impacted by the position of the options market maker. Using options trade repository data for both EURUSD and USDJPY we reconstruct the option market makers position and with it validate our proposed volatility model with very high confidence for the period from 21st October 2017 to 30th June 2018 on daily as well as intraday data. We find the option market maker to have a net short position and show that this leads to an absolute volatility increase of around 1% on average.

In addition, **Sornette** et al. (2021) theoretically discuss the potential of option positioning to generate directional price moves. Using theoretical arguments as well as simulations, we show that the price can exhibit unusual explosive or stable dynamics, depending on the parameters and the delta of the option positioning of the option market maker. Our volatility model as well as the model predictions with respect to the price (in)stability are then empirically validated for the GameStop stock market. Again, we validate our volatility model with very high confidence. And we further find that in the first quarter of 2021 there indeed were times when the option positioning suggested an explosive price behaviour of the GameStop stock market.

In summary, I hope to provide policy makers and investors with a framework on how option markets can influence the prices of the respective assets.

References

B. **Anderegg**, D. **Sornette**, and F. **Ulmann**. The Impact of Option Hedging on the Spot Market Volatility. *Revise and resubmit at the Journal of International Money and Finance*, 2021

A. **Barbon** and A. **Buraschi**. Gamma fragility. *School of Finance Research Paper Series*, 05, 2020

J. **Donier**, J. **Bonart**, I. **Mastromatteo**, and J. P. **Bouchaud**. A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15(7):1109–1121, 2015

J. **Farmer**, A. **Gerig**, F. **Lillo**, and S. **Mike**. Market efficiency and the long memory of supply and demand: Is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6(2):107–112, 2006

Financial Times. GameStop and BlackBerry shares soar on amateur traders' fervour. https://on.ft.com/3oipqrG, 2021. Accessed: 2021-04-23

R. **Frey** and A. **Stremme**. Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4):351–374, 1997

O. **Gueant** and J. **Pu**. Option pricing and hedging with execution costs and market impact. *Mathematical Finance*, 27(3):803–831, 2015

A. S. **Kyle**. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985

D. **MacKenzie**. *An Engine, Not a Camera: How Financial Models Shape Markets*. MIT Press, 1 edition, 2008. ISBN 0262134608

K. **Sircar** and G. **Papanicolaou**. General Black - Scholes models accounting for increased market volatility from hedging strategies. *Applied Mathematical Finance*, 5(1):45–82, 1998

D. **Sornette**, F. **Ulmann**, and A. **Wehrli**. On the Destabilizing Feedback Effects of Option Hedging. *Working Paper*, 2021

The Impact of Option Hedging on the Spot Market Volatility

Benjamin Anderegg[a], Florian Ulmann[b,*], Didier Sornette[b,c]

[a]*Swiss National Bank*[†]*, Boersenstrasse 15, 8001 Zuerich*

[b]*ETH Zurich, Department of Management, Technology, and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland*

[c]*Swiss Finance Institute, c/o University of Geneva, 40 blvd. Du Pont d'Arve, 1211 Geneva 4, Switzerland*

Abstract

We theoretically model and empirically quantify the feedback effect of delta hedging for the spot market volatility of the forex market. We start from an economy with two types of traders, an aggregated option market maker (OMM) and an aggregated option market taker (OMT), whose exposures reflect the total outstanding positions of all option traders in the market. A different hedge ratio of the OMM and OMT leads to a net delta hedge activity, which introduces market friction. We represent this friction by a simple linear permanent impact model. This approach allows us to derive the dependence of the spot market volatility on the gamma exposure of the traders. Our model shows that the spot market volatility is increased (decreased) by a negative (positive) gamma exposure of the OMM, whereby the amount of the increase depends on the net delta hedge amount executed in the spot market. To empirically test this model, we first reconstruct the aggregated OMM's gamma exposure using trade repository data and find that it is negative. The empirical analysis performed over the period from 21st October 2017 to 30th June 2018 using our reconstructed OMM data then strongly supports our theoretical model: The gamma exposure of the OMM turns out to be highly significant for the spot market volatility and, as expected, the spot market volatility is increased with the OMM's short gamma exposure. Quantitatively, a negative gamma exposure of the OMM of -1000 billion USD (which is what we observe from our reconstructed OMM data) leads to an absolute increase in volatility of 0.7% in EURUSD and 0.9% in USDJPY.

JEL: G12 G15 G21 G23 G32 C32 C51 C55 C58 C55 C80

Keywords: DTCC FX options delta hedging price impact trade repository data volatility modelling

---

[*] corresponding author

Email addresses: `Benjamin.Anderegg@snb.ch` (Benjamin Anderegg), `fulmann@ethz.ch` (Florian Ulmann), `dsornette@ethz.ch` (Didier Sornette)

[†] The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank. The Swiss National Bank takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

Revise and resubmit at the Journal of International Money and Finance

1. Introduction

This paper addresses the theoretical description and empirical quantification of the effect of delta hedging strategies on spot market volatility. By accounting for a linear, permanent market impact of spot market trades, we show that, in theory, spot market volatility is influenced by the net delta hedge demand arising from the options market. This theory is subsequently empirically validated with the use of trade repository data.

1.1. Literature review

Option pricing theory has been studied extensively since the seminal works of Black and Scholes (1973) and Merton (1973). Their theory is fundamentally based on an explicit trading strategy for the underlying asset and riskless bonds whose final payoff is equal to the payoff of the derivative security at maturity. As this trading strategy provides an insurance against the risk that comes with buying and selling the derivative, it is called a dynamic hedging strategy (dynamic, since the replication should ideally be continuous) Sircar and Papanicolaou (1998).

A crucial assumption of the Black-Scholes model is a frictionless and elastic market, i.e., a market in which trades have no price impact, neither temporary nor permanent Gueant and Pu (2015). However, a vast body of work concerned with market microstructure clearly shows that one can reject the assumption of a frictionless market Farmer et al. (2006). In the context of dynamic hedging strategies, this implies a feedback effect on market dynamic parameters such as the spot price and volatility structure: the dynamic hedging strategy requires the investor who sells (buys) an option to sell (buy) the asset in the spot market when its price declines and to buy (sell) it when its price goes up. From the nature of such a trading pattern, one can expect an increase (decrease) in spot market volatility Frey and Stremme (1997) Sircar and Papanicolaou (1998). Indeed, spot market volatility has been reported to have changed due to the introduction of futures and options Miller (1997) MacKenzie (2008). Note that this feedback effect can even result in two extreme price dynamics, market crashes and pinning. For example, Almgren (2015) attributes the Oct. 15, 2014, US Treasury price action, in which the US Treasury market underwent the largest intraday move since 2009, to the hedging of short option positions held by the dealer community. Avellaneda and Lipkin (2003), on the other hand, discuss the case where hedge funds, which have a long option position, pin the spot price to the strike at expiration due to their hedging behaviour. Furthermore, since changes in market dynamic parameters (e.g., spot market volatility) also change the derivative's price and therefore its dynamic hedging strategy, market dynamic parameters cannot be regarded as exogenous. In this context, Lions and Lasry (2006) cite practitioners who observe that the future values of the parameters they use for the pricing of options might be modified by their own activities (which is called a feedback effect).

Different work streams have emerged from the discussion of market frictions in the context of dynamic hedging strategies of options. Some are more focused on accounting for market frictions in the pricing of options (and in deriving an optimal hedging strategy), while others are more concerned by describing the actual feedback effect that dynamic hedging has on market dynamics. Our paper contributes to the latter stream. More specifically, we focus on the effect that delta hedging has on the spot market volatility. Nevertheless, we first present a brief literature review for both streams.

Accounting for feedback effects in the Black-Scholes pricing model has been the objective of a large body of academic literature. First, Brennan and Schwartz (1989) and Frey and Stremme (1997) focus on the derivation of the price dynamics through a clearing condition. Sircar and Papanicolaou (1998) even goes a bit further and introduces an option pricing model that accounts self-consistently for the effect of dynamic hedging on the volatility of the underlying asset. They find that spot market volatility increases by approximately 10%-18% for traders who write options and hedge their delta in the spot market.

The other approach is to account for the price impact of hedging trades in the spot market directly. In this approach, the price impact is often decomposed into temporary and permanent price impact components Almgren and Chriss (2000).

Temporary price impact considerations are mostly relevant in the context of optimal hedging strategies, as presented in Rogers and Singh (2010) and Naujokat and Westray (2011). The basic problem is that continuous hedging is prohibitive due to execution costs stemming from temporary market impact, whereas low-frequency hedging leads to large tracking error of the frictionless Black-Scholes delta hedge Gueant and Pu (2015).

The consideration of a permanent market impact is dedicated to the hedging of derivatives in situations where the notional of the option hedged is such that the delta-hedging is non-negligible compared to the average daily volume traded on the spot market Bouchard et al. (2017). Note that this large position may be the position of a single large trader or may be the aggregate position of a collection of traders, for example, the entire sell-side community, who all hold similar positions and who all hedge while their counter-parties do not Li and Almgren (2016). Permanent price impact is e.g. considered by Abergel and Loeper (2016), Loeper (2013) and Bouchard et al. (2016). All of their theoretical works reveal that the market dynamics of the spot market are changed by dynamic hedging of derivatives. Thereby Lions and Lasry (2006), also using a permanent market impact, find that the impact upon the volatility of the asset depends on the gamma of the option of the dominant single large investor they model. Further, Li and Almgren (2016) then take both temporary (with quadratic execution costs) and permanent price impacts into account (following Almgren and Chriss (2000)) and find that permanent market impact causes an increase or decrease in the realised volatility of the market price, depending on whether the large investor or the entire community of hedging traders is net long or short options.

1.2. Our contribution

We will focus on the theoretical and empirical discussion of the feedback effect of delta hedging strategies in the FX spot market for the FX sport price volatility.

First, similar to the approach presented in Frey and Stremme (1997), we work in a model economy with only two agents. Our agents are represented by two types of traders, namely, an aggregated option market maker (OMM), who provides liquidity to the options market, and an aggregated option market taker (OMT), who consumes liquidity. As a consequence of this setup, the OMM and the OMT always have exact opposite positions from one another, and each of their books represent the total outstanding position on the options market.

We then assume that the two traders' hedging trades in the spot market have a linear, permanent market impact and no temporary market impact. This framework results in a simple volatility model that shows that the spot market volatility increases above its fundamental value (which is the volatility in case of no market impact of delta hedging trades) if the hedge ratio of the trader, who is net short options (and therefore has a negative gamma exposure) is larger than the hedge ratio of the trader who is net long options. Note that this model economy is motivated by the idea that typical OMMs, e.g., large banks, dynamically hedge their positions while OMTs, e.g., investors, do not. This would result in an asymmetric net delta hedge demand, which will lead to an increase (if the OMM is short options) or decrease (if the OMM is long options) in the spot market volatility. Thus, our model framework and our theoretical results are similar to those presented by Lions and Lasry (2006) and Li and Almgren (2016). Nevertheless, nowhere in the paper do we a priori assume that the OMM hedges a larger share of her exposure compared to the OMT.

Second, we want to validate this model empirically for the FX market. While much work has been done on the theoretical side, empirical work on feedback effects arising from dynamic hedging has been limited. This gap exists because most models complicated setup does not allow for straightforward empirical testing and because an empirical study of such feedback effects often requires proprietary data, e.g., data that allow for the identification of the hedging demand. As our model only incorporates measurable or observable quantities, empirical validation is straightforward, as we later show. And we bypasses the need for proprietary data by reconstructing it from trade-repository data (see below).

So to validate our model, we first need to obtain the data of the aggregated OMM and OMT. The BIS triennial survey 2016 states that the FX market turnover is composed of 33% spot transactions and 67% derivatives such as FX swaps, outright forwards and options BIS (2016). Since 2016, those 67% are subject to mandatory reporting in the US, for which our main data source in this analysis, the Depository Trust and Clearing Corporation (DTCC) trade repository, is an authorised trade repository covering approximately 20% of the worldwide options market transactions Shtauber and Marone (2014). These transaction data allow for profound insights on the market positioning in terms of new and outstanding positions. DTCC positioning data have been used previously to analyze spillover effects from the FX options market on the FX spot market but mostly with a focus on the spillover effect on the spot price rather than its volatility Weng and Grover (2017) Shtauber and Marone (2014) Winkler (2017). The downside of the DTCC data is that we do not know the side behind each trade, and thus we do not know whether the transaction is buy or sell initiated. However, this information is necessary to reconstruct the aggregated exposure of the OMT and OMM, as we assume that the OMM is always the liquidity provider and the OMT is always the liquidity taker. We solve this problem by using a simple trade classification algorithm (the quote rule of Hasbrouck (1988)) to classify the aggressor side behind each option trade stored in the DTCC repository. [2] Thus, we can identify the OMM (liquidity provider) and OMT (liquidity taker) behind each trade and can therefore also calculate their aggregated position at each point in time. This allows us to estimate our volatility model. [3]

The remainder of this paper is structured as follows. The second section introduces our volatility model. The third section briefly explains how we reconstruct the OMM's position behind each transaction of the DTCC trade repository data. In the fourth section, we show our empirical results, supporting our volatility model structure. In the fifth section, we discuss our results and draw conclusions from our findings.

---

[2] See Frömmel et al. (2020) for an overview on trade classification algorithms and their quality.

[3] Note that we only examine vanilla options, as the details of exotic options in the DTCC trade repository are not sufficient.

2. A model for the feedback effects of delta hedging

2.1. Derivation of the volatility model

In this section, we derive an expression for the spot market volatility that includes the price impact of the delta hedging activity. First, we assume the observable exchange rate, denoted by S_t, is given by a linear stochastic differential equation.

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t \tag{2.1}$$

Further we first assume that the non-observable fundamental value of the exchange rate F_t, which has the interpretation of the market price unaffected by delta hedging (the friction in our model), follows a geometric Brownian motion.

$$dF_t = \alpha F_t dt + \nu F_t dW \tag{2.2}$$

Applying Ito's lemma to the log-price changes $df_t = d(ln(F_t))$ results in $df_t = \frac{dF_t}{F_t} - \frac{1}{2}\nu^2 dt$. Using dF_t from equation 2.2 we can then write df_t as

$$df_t = \underbrace{(\alpha - \frac{1}{2}\nu^2)}_{\alpha'} dt + \nu dW = \alpha^{'} dt + \nu dW \tag{2.3}$$

Now we assume that there are only two market participants, the OMM and the OMT. Their net delta exposures are $N_t^{OMM}\Delta_t^{OMM}$ and $N_t^{OMT}\Delta_t^{OMT}$, respectively, where N_t is the notional (always positive) and Δ_t is the option delta (the delta is positive for a long call and a short put and negative for a short call and a long put). Both hedge their overall delta exposure to a certain degree in the spot market, given by the hedge ratios h^{OMM} and h^{OMT}, which we assume to stay constant in time. If the price changes by a certain percentage between time $t - \Delta t$ and t, the net traded delta hedging volume, executed in the spot market at time t, is given by the negative of the change in the delta exposure between $t - \Delta t$ and t, multiplied by the hedge ratio, so $V_t^{OMM} = h^{OMM} N_t^{OMM}(-(\Delta_t^{OMM} - \Delta_{t-\Delta t}^{OMM}))$ for the OMM and $V_t^{OMT} = h^{OMT} N_t^{OMT}(-(\Delta_t^{OMT} - \Delta_{t-\Delta t}^{OMT}))$ for the OMT. Note that in a market with two participants, the notional is the same for both participants ($N_t^{OMM} = N_t^{OMT} = N_t$), and the delta of the OMM is always equal to the negative delta of the OMT ($N_t\Delta_t^{OMM} = -N_t\Delta_t^{OMT}$). Furthermore, the positions of each trader represent the total outstanding position in the market (which is why our two agents are the OMM and the OMT, as by definition the aggregated positions from all OMMs sum up to the total outstanding positions in the market). Defining the asymmetry in the hedge ratio of the OMM and the OMT from the OMM's point of view as $h_{net}^{OMM} = h^{OMM} - h^{OMT}$, the total net traded delta hedge volume between $t - \Delta t$

and t is given by

$$V_t = V_t^{OMM} + V_t^{OMT} \tag{2.4a}$$

$$= h^{OMM} N_t^{OMM}(-(\Delta_t^{OMM} - \Delta_{t-\Delta t}^{OMM})) + h^{OMT} N_t^{OMT}(-(\Delta_t^{OMT} - \Delta_{t-\Delta t}^{OMT})) \tag{2.4b}$$

$$= (h^{OMM} - h^{OMT}) N_t(-(\Delta_t^{OMM} - \Delta_{t-\Delta t}^{OMM})) \tag{2.4c}$$

$$= h_{net}^{OMM} N_t(-(\Delta_t^{OMM} - \Delta_{t-\Delta t}^{OMM})) \tag{2.4d}$$

In the continuous limit, when $\Delta t \longrightarrow 0$, (2.4d) becomes

$$dV_t = -h_{net}^{OMM} N_t d\Delta_t^{OMM} \tag{2.5}$$

Now we assume that the traded delta hedge volume executed in the spot market, dV_t, impacts the spot market price in a linear and permanent fashion, with β defined as the log-price change (market impact) per traded volume. This is the friction that we introduce and which will, as we see later, result in the feedback effect discussed in this paper. Thus, the observed log price change at time t, ds_t, is given by the log price change of the fundamental price at time t, df_t, as well as the price impact of the delta hedge volume executed in the spot market, βdV_t. Equation 2.6 reflects this ansatz.

$$ds_t = df_t + \beta dV_t \tag{2.6a}$$

$$= df_t - \beta h_{net}^{OMM} N_t d\Delta_t^{OMM} \tag{2.6b}$$

We thereby assume that the price impact β is a constant. This assumption basically reflects Kyle's price impact model (the price impact is permanent, linear and depends on the volume of the transaction only), presented in Kyle (1985). The assumption that the permanent price impact scales linearly with the transaction volume is also in line with current price impact theory Bouchaud (2010).

Applying Ito's lemma to the log-price changes $ds_t = d(ln(S_t))$ results in $ds_t = \frac{dS_t}{S_t} - \frac{1}{2}\sigma_t^2 dt$. Using this, as well as ds_t from equation 2.6b, we can then write dS_t as

$$dS_t = S_t df_t + \frac{1}{2}\sigma_t^2 S_t dt - \beta h_{net}^{OMM} N_t S_t d\Delta_t^{OMM} \tag{2.7}$$

Applying Ito's lemma to $d\Delta_t^{OMM}$ with respect to dS_t leads to

$$d\Delta_t^{OMM} = \underbrace{(\underbrace{\frac{\partial \Delta_t^{OMM}}{\partial t}}_{-Charm_t^{OMM}} + \mu_t S_t \underbrace{\frac{\partial \Delta_t^{OMM}}{\partial S}}_{\gamma_t^{OMM}} + \frac{1}{2}\sigma_t^2 S_t^2 \underbrace{\frac{\partial^2 \Delta_t^{H}}{\partial S^2}}_{Speed^{OMM}})}_{\gamma'_t} dt + \sigma_t S_t \underbrace{\frac{\partial \Delta_t^{OMM}}{\partial S}}_{\gamma_t^{OMM}} dW_t \tag{2.8}$$

The option greeks are labeled with brackets. γ_t^{OMM}, which is important in our theory, is the option gamma of OMM at time t. Inserting 2.3 and 2.8 into 2.7 results in

$$dS_t = (\alpha' S_t + \frac{1}{2}\sigma_t^2 S_t - \beta h_{net}^{OMM} N_t S_t \gamma_t')dt + (\nu S_t - \beta h_{net}^{OMM} N_t S_t^2 \sigma_t \gamma_t^{OMM})dW_t \tag{2.9}$$

Comparing equation 2.9 with 2.1 results in

$$\sigma_t = \frac{\nu}{(1 + \beta h_{net}^{OMM} N_t S_t \gamma_t^{OMM})} \tag{2.10}$$

2.2. *Discussion of the volatility model*

We quickly discuss the intuition behind our presented model now. $h_{net}^{OMM} S_t N_t \gamma_t^{OMM}$ is the change of the total market's hedged delta exposure in base currency in response to a relative change in S_t. It is equal to $h^{OMM} S_t N_t \gamma_t^{OMM} + h^{OMT} S_t N_t \gamma_t^{OMT} = h_{net}^{OMM} S_t N_t \gamma_t^{OMM}$ by using the asymmetry in the hedge ratio of the OMM and the OMT from the OMM's point of view ($h_{net}^{OMM} = h^{OMM} - h^{OMT}$) as well as the fact that $\gamma_t^{OMM} = -\gamma_t^{OMT}$. The negative of $h_{net}^{OMM} S_t N_t \gamma_t^{OMM}$ is then exactly the amount which has to be executed in the spot market in response to a relative change in S_t in order to keep the hedged delta exposure constant. And if there is price impact β, this execution influences the volatility ν by the factor $(1 + \beta h_{net}^{OMM} S_t N_t \gamma_t^{OMM})^{-1}$.

In order to simplify the notation, we now define the **market's delta-hedged gamma exposure** Γ_t^M

$$\Gamma_t^M = h_{net}^{OMM} S_t N_t \gamma_t^{OMM} \tag{2.11}$$

Using the market's delta-hedged gamma exposure, given by equation 2.11, we can then rewrite equation 2.10 as

$$\sigma_t = \frac{\nu}{(1 + \beta \Gamma_t^M)} \tag{2.12}$$

We see that the spot market volatility of a market unaffected by delta hedging, ν, is rescaled by a factor $(1 + \beta\Gamma_t^M)^{-1}$. Thus it is the Γ_t^M that determines whether the volatility is scaled up or down. If Γ_t^M is positive, the total market's hedged delta exposure increases in response to an increase in S_t. The negative of this increase in the delta exposure is then executed in the spot market, i.e. the market is selling the underlying if S_t increases, which counteracts the increase in S_t (if there is a price impact) and thus lowers the fundamental volatility ν. But if Γ_t^M is negative, the total market's hedged delta exposure decreases in response to an increase in S_t. The negative of this decrease in the delta exposure is then executed in the spot market, i.e. the market is buying the underlying if S_t increases, which supports the increase in S_t (if there is a price impact) and thus increases the fundamental volatility ν.

So the intuition behind the above described dynamic stems from the fact that a short delta-hedged gamma exposure of the market is associated with a short position in the net hedged option contracts of the market, which then again implies a net delta hedging strategy of the market in which she has to buy the asset if the spot price increases and sell the asset if the spot price decreases. This trading behavior leads to an increase in the spot market volatility given that the hedging trades are assumed to have a price impact (vice versa for the long delta-hedged gamma exposure of the market).

Whether Γ_t^M is positive or negative thereby depends on the gamma of the OMM, γ_t^{OMM} (or the gamma of the OMT as together they represent the full market), as well as on the asymmetry in the hedge ratio of the OMM and the OMT, $h_{net}^{OMM} = h^{OMM} - h^{OMT}$. In other words, to gauge the feedback of the option market on the spot market volatility, the gamma exposure of the trader who hedges a larger part of her delta exposure compared to the other trader is relevant. As described in equation 2.11, we call this the market's delta-hedged gamma exposure. And while the direction of the feedback effects depends on the sign of Γ_t^M, the strength of the feedback effect depends on the absolute value of the market's delta-hedged gamma exposure $|\Gamma_t^M|$ and on the price impact β. Note that as the volatility cannot be below zero by definition, the factor $\beta\Gamma_t^M$ must be in the open interval $(-1, +\infty)$.

Figure 1 displays the above discussed model dynamics and figure 2 presents a case study for the case where the OMM holds a short call option and $h_{net}^{OMM} = 1$.

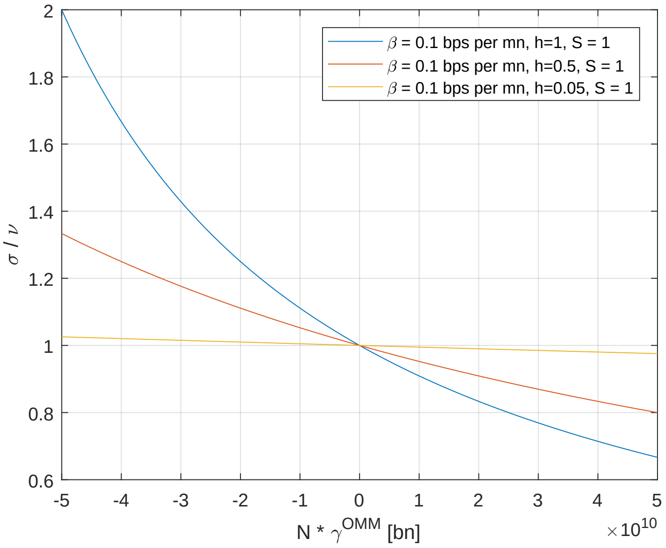


Figure 1: Discussion of our volatility model: This figure shows the dependence of $\frac{\sigma_t}{\nu}$ as a function of $N_t\gamma_t^{OMM}$ as predicted by equation 2.10. On the vertical axis, we display the observable spot market volatility normalized by the volatility of a market unaffected by delta hedging, so $\frac{\sigma_t}{\nu}$. On the horizontal axis, we display the factor Γ_t^M, where we fix the market impact to $\beta = 0.1\ [bps/\textit{mn base currency}]$, the exchange rate to $S_t = 1$ and h_{net}^{OMM} to three different values. We thereby always assume that the OMM hedges a larger part of its delta exposure, so $h_{net}^{OMM} > 0$, and thus the sign of her gamma exposure is the relevant one for the discussed feedback effect. Thus Γ_t^M reduces to $N_t\gamma_t^{OMM}$, as all other variables are fixed. We see that the volatility is scaled up or down depending on the sign of the delta-hedged gamma exposure of the market Γ_t^M (or rather by the sign of the gamma exposure of the OMM, as we assumed $h_{net}^{OMM} > 0$). The strength of this effect thereby depends non-linearly on the absolute value of the gamma exposure $N_t\gamma_t^{OMM}$.

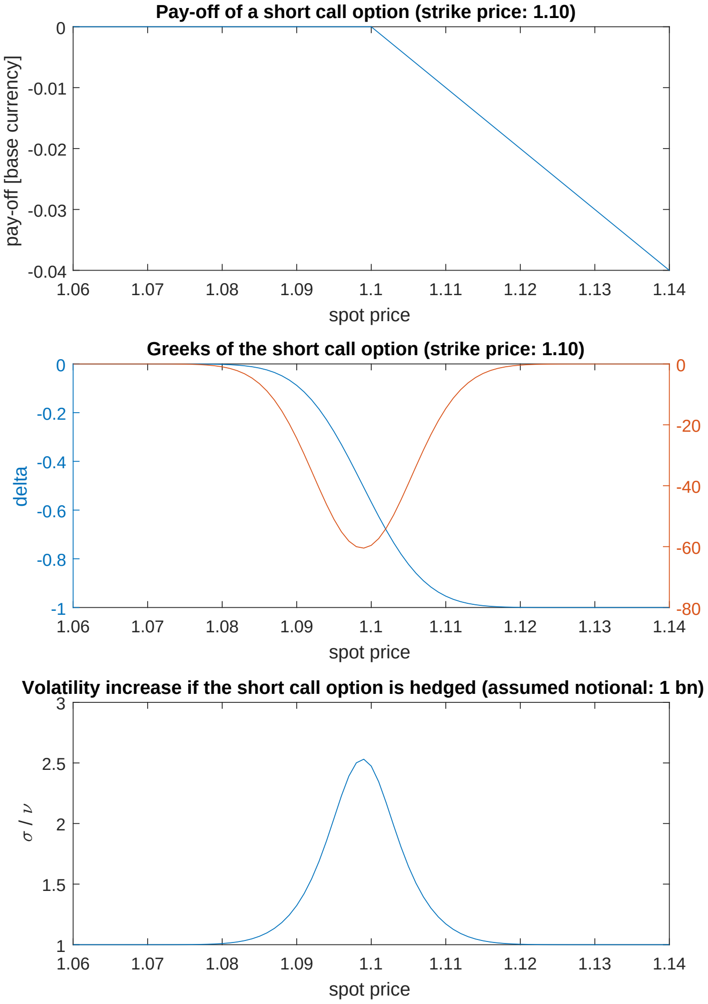


Figure 2: Case study of our volatility model: We assume that the OMM has a short call position. The upper panel shows the pay-off at maturity of the OMM's short call position, with a strike price at $K = 1.10$. The middle panel displays the corresponding delta and gamma profile of the short call option with a volatility of 6% and a time-to-expiry of three months. We thereby see that the delta is decreasing in the spot price, and thus the corresponding delta hedge profile (which is the negative of the delta profile) is increasing in the spot price. This implies that the OMM needs to buy the underlying if the spot price goes up and sell the underlying if the spot price goes down. The corresponding gamma profile is negative. Now we assume that the notional behind the short call option is one billion ($N_t = -1$ bn) and that the OMM hedges her total position in the spot market, while the OMT does not hedge her position at all (hedge ratio of $h_{net}^{OMM} = 1$). Note that thus the OMM's gamma position is equal to the market's delta-hedged gamma position. The resulting increase in the volatility ($\frac{\sigma}{\nu}$, based on equation 2.10 with a permanent market impact of $\beta = 0.1[bps/mn]$, a hedge ratio of $h_{net}^{OMM} = 1$ and a notional of $N_t = -1$ bn) is displayed in the lower panel. We see that the volatility locally increases at spot prices for which the gamma profile of the OMM is negative.

3. Reconstruction of the OMM's gamma exposure from trade repository data

In order to empirically validate equation 2.10, we need data on the gamma exposure of the OMM, $S_t N_t \gamma_t^{OMM}$ (or, equivalently, as we only have two traders in our model, the gamma exposure of the OMT, $S_t N_t \gamma_t^{OMT}$). In this section we discuss how we construct data on the gamma exposure of the OMM, $S_t N_t \gamma_t^{OMM}$. We thereby focus on the FX options market.

The Dodd-Frank Act, which was signed into law in July 2010, requires the reporting of all over-the-counter FX option transactions that were traded in the US, or involve US persons, to a data repository. Note that there are several data repositories that market participants can report to. We use the Depository Trust and Clearing Corporations repository (DTCC), which covers approximately 20% of the worldwide options market transactions Shtauber and Marone (2014).

Option trades stored in the DTCC trade repository are the result of OTC transactions. Large banks mostly act as liquidity providers for their clients, e.g., hedge funds, corporations, institutional investors, asset managers or private clients, which then act as liquidity takers. The banks provide liquidity by constantly offering a bid and ask price (which is quoted in terms of implied volatility) to their clients, and the clients then take the liquidity by either buying on the ask price, this deal is called a paid, or selling on the bid price, this deal is called a given. The trade is then reported to the DTCC trade repository, but the information on whether the deal happened on the bid or the ask price is not included.

As presented before, our model economy is motivated by this structure of the OTC option market, and models only two traders, the option market maker (OMM) and the option market taker (OMT). The OMM represents the aggregate of the liquidity providers (mostly banks) and the OMT represents the aggregate of the liquidity takers (hedge funds, corporations, institutional investors, asset managers and private clients). All deals are attributable to the aggregated position of the OMT, whereas the OMM then just has the exact opposite position.

The main problem is that the public DTCC data does not provide information on the type of deal that happened, so whether a deal was a paid or a given. Thus, we need to reconstruct this information in order to quantify the OMT's positioning in the options market. We thus follow a trade classification procedure. In the first step, the implied volatility of each option transaction stored in the DTCC trade repository is reconstructed by numerically minimizing the difference between the premium that was paid (which is reported to the DTCC trade repository) and the premium that Black's model suggests for the trade (based on the specifications that are reported to the DTCC trade repository). [4] Next, this reconstructed implied volatility, which was traded at a certain time (that is also reported to the DTCC trade repository), is matched onto the simultaneous reference bid-ask price stream for that time, that we take from Bloomberg. Following the quote rule of Hasbrouck (1988), we label the transaction a paid (buy order on the ask price) if the reconstructed implied volatility is closer to the ask price compared to the bid price or a given (sell order on the bid price) if the reconstructed implied volatility is closer to the bid price compared to the ask price. Those deals then reflect the OMT's position. And as the OMM is always on the other side of each deal, it follows that she has a short position if the transaction is labeled a paid and a long position if the transaction is labeled a given. A discussion of the quality of our reconstruction proceedure is presented in Appendix B.3.

---

[4] For other asset classes, other option pricing models might be appropriate.

As we now know the OMM's position behind each transaction, we can calculate the delta and gamma exposure of the OMM at any given point in time. We simply calculate the delta and gamma of the OMM's open positions at any given point in time, multiply it by the signed notional of the corresponding position (i.e., the negative notional if the OMM is short the option) and take the corresponding sum over all open positions. This results in the aggregated delta and gamma exposure of the OMM's portfolio at any given point in time ($S_t N_t \gamma_t^{OMM}$), which we now can use to validate our volatility model.

A more detailed discussion on the underlying DTCC data quality and on how the OMM's positioning is derived from the DTCC data can be found in the appendix.

3.1. Discussion of the OMM's gamma exposure reconstructed from DTCC data

For the reconstruction of the OMM's gamma exposure from DTCC data we focused on the FX pairs EURUSD and USDJPY, as for those two FX pairs the outstanding volume in the FX options market is large enough to expect significant feedback effects on the FX spot market. As described before, we reconstructed the delta and gamma exposure of the OMM at all points in time. The corresponding results are displayed in figure 3, where each trading day is split into 12 two-hour intervals. [5]

As we see in figure 3, the OMM is short gamma at all times. This implies that the OMM is shorting put and call options (on a net basis). This is in line with our expectation. First, most investors are purchasers of options rather than sellers. They use options as a hedge or with consideration of the spot price and, therefore, mostly purchase options. Furthermore, they refrain from going short an option contract (put or call), as the risk involved in doing so is too high. Those traders most certainly trade as market takers in the market, e.g., as clients of banks. On the other side of each transaction of a market taker sits a market maker, e.g., the banks; consequently, the market maker is expected to have a short option position and therefore a short gamma position. Second, a short position in an option implies a short vega position, and as the implied volatility usually trades with a premium to the realized volatility, this is one way for option market makers to make money. For those two reasons, the OMM is expected to have a net short gamma position, which is exactly what we observe. As stated above, we see from our data that the OMM is short 75% of all traded option contracts in EURUSD and 78% of all traded options in USDJPY.

For the delta exposure, we have no a priori expectation of how it should look, since the delta exposure is not directly related to the positioning (if the OMM is long or short options) but rather is a function of where the spot price trades relative to the notional weighted strike price of the aggregated OMM's positions.

If the OMM would hedge her total position in the spot market, her delta hedge strategy would be to trade the negative difference of her delta exposure between two points in time in the spot market. The OMM would then be delta neutral by the end of each time period. Note that, as gamma is the change in delta per unit change in the spot price (derivative of the delta with respect to the spot price), gamma indicates how much delta would have to be hedged if the spot price moved one unit.

---

[5] UTC 00:00-02:00, 02:00-04:00, 04:00-06:00, 06:00-08:00, 08:00-10:00, 10:00-12:00, 12:00-14:00, 14:00-16:00, 16:00-18:00, 18:00-20:00, 20:00-22:00, 22:00-24:00

In that sense, gamma is a proxy for the intensity of the delta hedge activity.

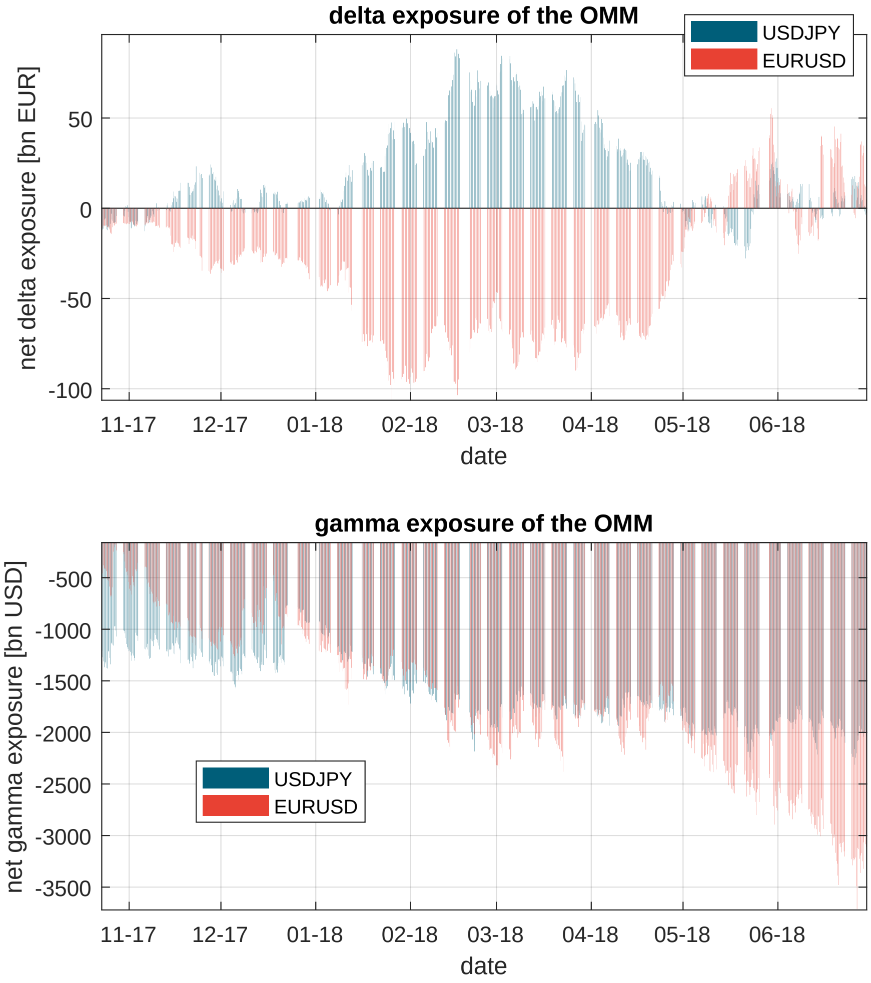


Figure 3: Reconstructed delta and gamma exposure of the OMM: The figure shows the evolution of the accumulated net delta and gamma exposure of the OMM over the analysed time frame for EURUSD and USDJPY.

4. Validation of the volatility model

As outlined in the previous sections, the market's delta hedged gamma exposure is assumed to generate feedback effects: a positive gamma exposure implies that the market sells the underlying asset in a rising market to hedge its delta exposure, and as the executed hedging trades in the spot market have a market impact, this activity should have a negative effect on the realised spot market volatility. Following the same line of argument, a negative gamma exposure should have a positive effect on the realised volatility. Our volatility model reflects this logic. In this section, we will validate and quantify this feedback effect empirically with the data on the gamma exposure of the OMM that we reconstructed from the DTCC data.

$$\sigma_t = \frac{\nu}{(1 + \beta h_{net}^{OMM} N_t S_t \gamma_t^{OMM})} \tag{4.1}$$

We validate the volatility model presented in equation 4.1 with the use of MATLAB's nonlinear regression model Matlab (2019). Thereby the data input is σ_t, the realised spot market volatility (annualised) for the time period between t and $t+1$, and $N_t S_t \gamma_t^M$, the reconstructed OMM's gamma exposure at time t. The parameters that we estimate in the regression model are ν and $\beta * h_{net}^{OMM}$ [6].. As the volatility strongly depends on the intraday liquidity cycle, we validate the model with daily data. [7]

Table 1 and figure (4) present the results of the nonlinear regression model.

	EURUSD	USDJPY
ν (%)	7.2***	5***
βh_{net}^{OMM} ($\frac{\%}{\text{bn}}$)	0.00009***	0.00015***
RMSE (min. σ, max σ)	2.1 (3.8, 17.5)	1.7 (3.0, 14.6)
F-test vs. const. model	1430***	1280***

Table 1: This table shows the least-square fit results of the nonlinear model fit for EURUSD and USDJPY. t-statistics are presented in brackets. The stars *** denote significance at the 1% level.

The high values for the F-test strongly support our proposed model and thus the existence of the described feedback effect. And we see highly significant parameter estimates for the factor βh_{net}^{OMM} as well as for ν, which tells us that the gamma exposure of the OMM has high relevance for the spot market volatility.

The root mean squared error (RMSE)of the estimation is fairly low, given the high noise of daily volatility estimates and the fact that we analyse the observed volatility only with respect to one isolated effect (the feedback effect) in our model.

---

[6] Starting values for βh_{net}^{OMM} and ν are chosen as 0 and the convergence criteria is set at a tolerance of $1e-8$.

[7] For EURUSD, we look at the European trading session (UTC 09:00 to UTC 16:00), and for USDJPY, we look at the Asian trading session (UTC 2:00 to UTC 9:00). So we e.g. regress the gamma position of the OMM in EURUSD at 09:00 UTC onto the realized spot volatility between 09:00 and 16:00.

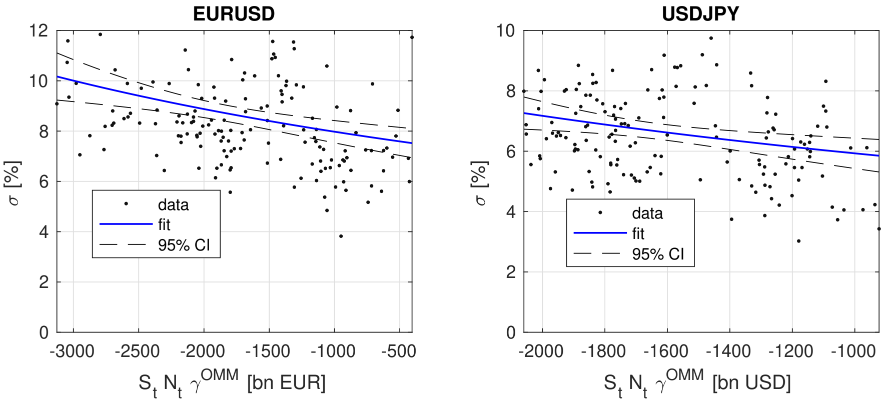


Figure 4: In these panels, we show the nonlinear model fit for EURUSD and USDJPY (solid blue line) and the respective 95% confidence intervals (dashed black lines).

Appendix E.1 presents the analysis on intraday data in a random intercept linear mixed-effect model, for which we realise very similar parameter estimates and a high goodness of fit in terms of R^2 (a measure not valid for the here presented nonlinear model fit).

The results are very robust against changes in the estimation period, i.e. the significance or parameter estimate change with negligible amplitude upon a change in the estimation period.

The estimate of ν reflects the fundamental volatility, which can be interpreted as the spot market volatility in the absence of any delta hedging activity. It is higher for EURUSD than for USDJPY. This difference stems from the fact that the volatility data for the daily estimate for USDJPY originate from the Asian trading hours, in which the volatility is usually lower than during the European or American trading hours.

The value of βh_{net}^{OMM} is not further disentangled here, as we would need an estimate of one of them, which is empirically difficult. But as the price impact β for sure is positive, we can conclude from the parameter estimates that h_{net}^{OMM} is also positive. This confirms that the OMM hedges a larger share of her delta exposure compared to the OMT, which is in line with our expectation that the OMM hedges his positions as he provides liquidity to the market as a service and thus needs to hedge his exposures.

The estimate of the parameter βh_{net}^{OMM} can now be used to quantify the feedback effect of the OMM's reconstructed gamma exposure on the observed spot market volatility. This discussion is presented in figure 5, using the fitted parameter values. In the upper panel, we present the fitted volatility model. The middle panel presents the absolute volatility increase due to the feedback effect, and the lower figure presents the relative volatility increase. We see that a negative gamma exposure of the OMM of approximately -1000 billion in the base currency (which is around what we observe in figure 3) leads to an increase of the EURUSD volatility above the fundamental volatility level of 0.7% [8] and to an increase of the USDJPY volatility above the fundamental volatility level of 0.9%. In the lower panel, we subsequently see that this gamma exposure of approximately -1000 billion in the base currency amounts to a relative increase above the fundamental volatility level of approximately 10% for EURUSD [9] and 18% for USDJPY. [10] This feedback effect stems from the fact that a short gamma exposure of the OMM (who is found to hedge a larger share of her delta exposure than the OMT) is associated with a short option position, which implies a delta hedging strategy in which the OMM has to buy the base currency if the spot price increases and sell the base currency if the spot price decreases. This leads to an increase in the spot market volatility given that the hedging trades have a price impact (vice versa for the long gamma exposure). [11]

---

[8] $\sigma - \nu = \frac{\nu}{1-\beta * h_{net}^{OMM} * N_t S_t \gamma_t^{OMM}} - \nu = \frac{7.2\%}{1-0.00009\frac{\%}{bn} * 1000bn} - 7.2\% = 0.7\%$

Apparently, the feedback effect is stronger (per billion gamma exposure of the OMM) in USDJPY than in EURUSD. As the net hedge ratio h_{net}^{OMM} is probably of similar size in both currency markets, we can conclude that this higher effect can be attributed to a higher β in the USDJPY market (as then the higher estimate of βh_{net}^{OMM} in USDJPY can only be explained by a higher β). This can be explained by the higher liquidity of the EURUSD spot market and thus the lower market share of the delta hedging strategy in EURUSD and the lower price impact of the traded volume BIS (2016).

Lastly, note that very similar results were found with other empirical approaches, as e.g. with a random intercept linear mixed-effect model in which we used intraday data but accounted for the corresponding intraday time with a grouping variable. The intraday results are are presented in the appendix Appendix E.

---

[9] $\frac{\sigma-\nu}{\nu} = \frac{0.7\%}{7.2\%} = 10\%$

[10] Note that the true effect per billion is probably a bit smaller when we adjust for the fact that our OMM positioning is only based on the DTCC data, which represent only approximately 20% of the total options market Weng and Grover (2017). But as most relevant OMM report to DTCC, we regard it as a good representation of the total options market.

[11] Even though in theory one could conclude that our model is tradable, this is not the case, as no risk-less profit could be made based on our model if one takes transaction costs into account. A short discussion of that is presented in appendix Appendix D.

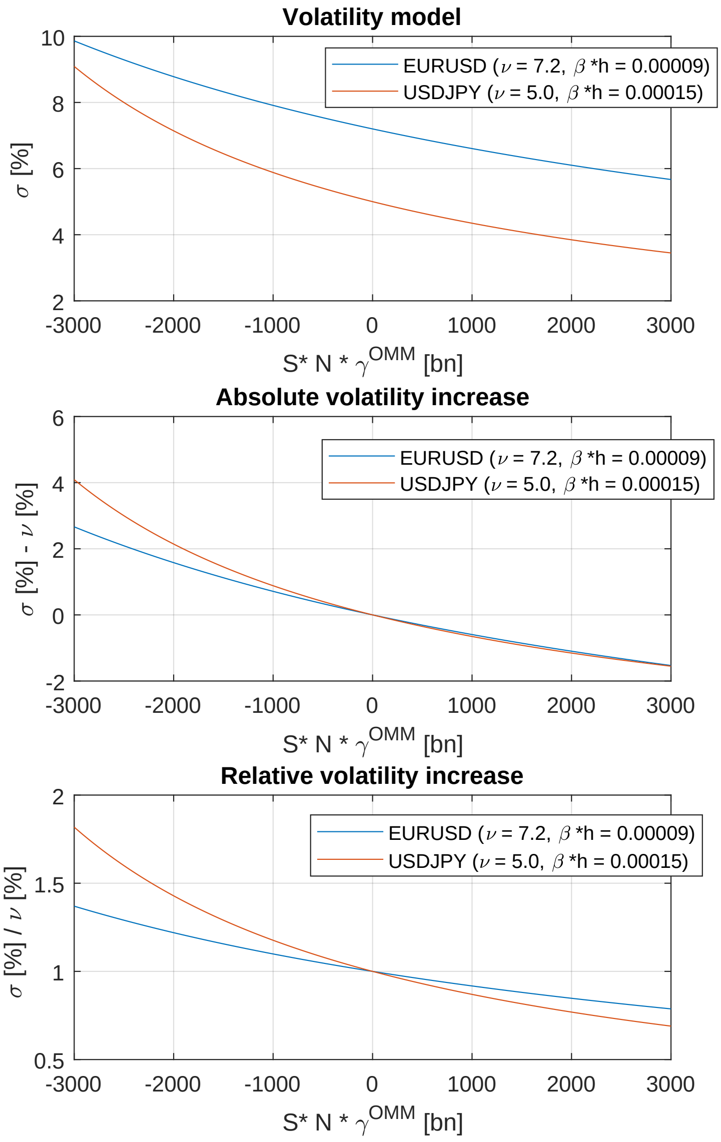


Figure 5: In these panels, we discuss our volatility model with the fitted parameters for EURUSD and USDJPY. We present the fitted volatility model in the upper panel, the absolute volatility increase due to delta hedging activity in the middle panel and the relative volatility increase in the lower panel.

5. Discussion and Conclusion

To model the feedback effect of delta hedging for spot market volatility, we have proposed a model economy of two types of traders, an OMM and an OMT, whose exposures each reflect the total outstanding positions in the market. If the OMM and the OMT have different hedge ratios, this will result in a net delta hedge activity that introduces market friction and feedback effects. We conceptualise this friction by assuming a simple linear, permanent impact model for the net delta hedge volumes that are executed in the spot market. From this framework we subsequently derive a spot market volatility model, given by equation 2.12, that depends on the markets delta hedged gamma exposure as well as on the magnitude of the price friction (i.e. the price impact of delta hedge transactions in the spot market). For example, if the OMM has a negative gamma exposure and she hedges a higher degree of her delta exposure compared to the OMT (two of our empirical findings), the spot market volatility is structurally high, as the net delta hedge amounts, executed in the sport market, would be positive if the spot price moves up and negative if the spot price moves down (which results in an increase in the spot market volatility if the transactions come with a price impact). The implications of this model are in line with the theoretical results of Lions and Lasry (2006) and Li and Almgren (2016).

Feedback effects of this nature have not been empirically discussed, as prior models did not allow for straightforward empirical testing, mainly due to their complicated setup or since data to do so is often not observable. In the case of our volatility model, we only needed the aggregated OMM's gamma exposure in order to empirically validate the model and therefore quantify the feedback effect. We reconstructed this data by using the publicly available DTCC trade repository data. This reconstructed net gamma exposure of the OMM was found to be negative, as expected: investors usually buy options with either consideration of the spot price or with the desire to hedge other positions. As those traders usually act as OMTs in the markets, we expect the OMT to be net long on options. As the OMM provides liquidity as a service to the market, his position will be the reverse of the OMT. Therefore, a net short position makes sense from our point of view.

We then use this reconstructed gamma exposure of the OMM to validate our volatility model. A least-square fit thereby shows a high goodness of fit, which strongly supports our model and the discussed feedback effect. The two main empirical findings are first, that the OMM hedges a larger share of her delta exposure compared to the OMT and second, the validation and quantification of the feedback effect. We will elaborate on the to findings here:

(1) We find that the OMM hedges a larger share of her delta exposure compared to the OMM. This, again, is in line with our expectation and in accordance with the line of reasoning presented above: As OMTs are usually investors who buy options with consideration of the spot price or to hedge other positions in their portfolio, they are expected to not hedge their option's delta. As the OMMs provide liquidity to the market as a service but usually without taking a view on the spot, they are assumed to hedge their large portfolio in order to reduce the risk involved in their business model. Therefore, we expected the hedge ratio of the OMM to be higher than that of the OMT, which is exactly what we found. Consequently, we also observe that the volatility is increased by the OMM's short gamma exposure.

(2) We find very strong statistical support of our volatility model structure and are able to use the model fit to quantify the feedback effect of the OMM's gamma exposure on the spot market volatility. A negative gamma exposure of the OMM of

approximately -1000 billion in the base currency (which is around what we observe from our reconstructed OMM data) leads to an absolute increase in volatility of 0.7% in EURUSD and 0.9% in USDJPY. If we assume that the hedge ratios in both markets are the same, this difference can be directly explained by the higher market impact of a transaction in the USDJPY spot market compared to the EURUSD spot market, which makes sense, since the liquidity of the EURUSD spot market is higher than that of the USDJPY spot market. This absolute increase amounts to a relative increase in the volatility above its fundamental level (volatility level in the absence of any delta hedging) of approximately 10% for EURUSD ($\frac{0.7\%}{7.2}$) and 18% for USDJPY ($\frac{0.9\%}{5.0}$). Therefore, our results are in line with previous theoretical work on the feedback effect that delta hedging strategies have on the spot market volatility, as e.g., Sircar and Papanicolaou (1998) suggest a relative increase in the spot market volatility of approximately 18%.

Note that it would be wrong to conclude now that the options market introduces a destabilizing or stabilizing effect on the spot market via the delta hedging strategies per se. Whether the effect of the option market on the spot market is stabilizing or not only depends on the net gamma position of the OMM (as the OMM is found to hedge a higher share of her delta exposure compared to the OMT). However, since the OMM is usually short gamma, as other market participants refrain from selling options, the option market is found to lead to an increase in the spot market volatility. Moreover, while these effects appear to exist, implied volatility is still trading at a premium to the realized volatility! One would assume that the OMM would thus only stop selling options if her hedging strategy would drive the realized volatility above the implied volatility.

In future work, we intend to discuss how the persistence of the gamma exposure of the OMM could explain part of the autocorrelation that is observed in the spot market volatility; we also intend to discuss the local volatility phenomena and to examine the connection of an extreme gamma position in the context of market crashes.

Acknowledgements

The authors thank all of the Foreign Exchange and Gold unit members of the Swiss National Bank for various stimulating discussions. We further thank the anonymous referee of the Swiss National Bank working paper series, who reviewed a preprint of this paper before it was accepted for publication in the SNB working paper series. And we thank the anonymous referee of the Journal of International Money and Finance for the very helpful comments and suggestions.

References

F. Abergel and G. Loeper. *Option Pricing and Hedging with Liquidity Costs and Market Impact, Proceedings of the International Workshop on Econophysics and Sociophysics*. Springer, 2016. doi: 9783319477053.

R. Almgren. Treasury price swings on october 15, 2014. Technical report, Quantitative Brokers, 2015. Technical report.

R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Mathematics and Financial Economics*, 3(1):5–39, 2000.

M. Avellaneda and M. D. Lipkin. A market-induced mechanism for stock pinning. *Quant. Finance*, 3(6):417–425, 2003.

BIS. Triennial central bank survey of foreign exchange and otc derivatives markets in 2016. Technical report, Bank for International Settlements, 2016. Triennial Central Bank Survey.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654, 1973.

M. E. Blume, A. C. Mackinlay, and B. Terker. Order imbalances and stock price movements on october 19 and 20, 1987. *The Journal of Finance*, 44(4):827–848, 1989. doi: https://doi.org/10.1111/j.1540-6261.1989.tb02626.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1989.tb02626.x`.

B. Bouchard, G. Loeper, and Y. Zou. Almost-sure hedging with permanent price impact. *Finance and Stochastics*, 20(3):741–771, 2016.

B. Bouchard, G. Loeper, and Y. Zous. Hedging of covered options with linear market impact and gamma constraint. *SIAM J. Control Optim.*, 55(5):3319–3348, 2017.

J. P. Bouchaud. Price impact. In *Encyclopedia of Quantitative Finance*. Wiley, 2010.

M. Brennan and E. Schwartz. Portfolio insurance and financial market equilibrium. *Journal of Business*, 62(4):455–476, 1989.

B. Chakrabarty, B. Li, V. Nguyen, and R. A. Van Ness. Trade classification algorithms for electronic communications network trades. *Journal of Banking & Finance*, 31(12):3806–3821, 2007. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2007.03.003. URL `https://www.sciencedirect.com/science/article/pii/S0378426607001021`.

DTCC. Explanation of trade information warehouse data. Technical report, The Depository Trust & Clearing Coorporation, 2011. Technical Document.

J. Farmer, A. Gerig, F. Lillo, and S. Mike. Market efficiency and the long memory of supply and demand: Is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6(2):107–112, 2006.

R. Frey and A. Stremme. Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4):351–374, 1997.

M. Frömmel, D. D'Hoore, and K. Lampaert. The accuracy of trade classification systems on the foreign exchange market: Evidence from the rub/usd market. *Finance Research Letters*, page 101892, 2020. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2020.101892. URL `https://www.sciencedirect.com/science/article/pii/S1544612320317062`.

O. Gueant and J. Pu. Option pricing and hedging with execution costs and market impact. *Mathematical Finance*, 27(3):803–831, 2015.

J. Hasbrouck. Trades, quotes, inventories, and information. *Journal of Financial Economics*, 22(2):229–252, 1988. ISSN 0304-405X. doi: https://doi.org/10.1016/0304-405X(88)90070-0. URL https://www.sciencedirect.com/science/article/pii/0304405X88900700.

A. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

T. M. Li and R. Almgren. Option hedging with smooth market impact. *Market Microstructure and Liquidity*, 2(1):1650002, 2016.

P.-L. Lions and J.-M. Lasry. Towards a self-consistent theory of volatility. *J. Math. Pures Appl.*, 86(6):541–551, 2006.

G. Loeper. Option pricing with market impact and non-linear Black and Scholes pdes. *SSRN Electronic Journal*, 2013. doi: 10.2139/ssrn.2240199.

D. MacKenzie. *An Engine, Not a Camera: How Financial Models Shape Markets*. MIT Press, 1 edition, 2008. ISBN 0262134608.

Matlab. Matlab fit nonlinear regression model. https://www.mathworks.com/help/stats/fitnlm.html, 2019. Accessed: 2019-06-01.

R. C. Merton. The theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

M. Miller. *Financial Innovations and Market Volatility*. Basil Blackwell Ldt., 1997.

F. Naujokat and N. Westray. Curve following in illiquid markets. *Mathematics and Financial Economics*, 4(4):299–335, 2011.

L. C. G. Rogers and S. Singh. The cost of illiquidity and its effects on hedging. *Mathematical Finance*, 20(4):597–615, 2010.

A. Shtauber and G. Marone. Can DTCC positioning data predict EMFX. Technical report, Deutsche Bank, 2014. Deutsche Bank Markets Research.

K. R. Sircar and G. Papanicolaou. General Black - Scholes models accounting for increased market volatility from hedging strategies. *Applied Mathematical Finance*, 5(1):45–82, 1998.

N. Weng and R. Grover. Options flows matter: introducing the deutsche bank skew volume indicators. Technical report, Deutsche Bank, 2017. Deutsche Bank Markets Research.

R. Winkler. Foreign exchange coffee report. Technical report, Deutsche Bank, 2017. Deutsche Bank Markets Research.

Appendix A. Data Specification

Appendix A.1. Depository Trust and Clearing Corporations repository (DTCC) data

The Dodd-Frank Act, which was signed into law in 2010, requires the reporting of all over-the-counter FX option transactions that were traded in the US or involve US persons to a data repository. There are several data repositories to which market participants can report. We use the largest one of those, the Depository Trust and Clearing Corporations repository (DTCC) DTCC (2011).

The DTCC data are a representative source of the total options market. The share of the DTCC transactions among the total turnover in the options market, as published by the BIS, is approximately 20%. Due to the large volumes traded in the EURUSD and USDJPY options market, we focus on those two instruments.

Most relevant information of the options trade is included in the repository, such as the type of options traded, the notional amount (volume), the strike price, the expiry date, the premium paid, the currency for quoting the notional and the premium as well as the option trade timestamp. The missing information is the investor type (hedge fund, corporation, asset manager, private client, etc.) and the side of the trade (if it was a paid or given); in addition, trades done outside the US are not represented. Our reconstruction logic fills one of the gaps, namely, the missing information about the side behind a transaction, by reconstructing the implied volatility of each trade and subsequently matching it with the reference bid/ask price at the time of the trade to check if the trade happened on the bid or the ask side (for more detail, see appendix Appendix B.1 and Appendix B.2).

The key in this reconstruction and matching procedure therefore is to have accurate time stamps from the DTCC repository in a sense that they need to be reported immediately to the repository. To get an indication if this immediate reporting is given, we looked at the intraday turnover distribution of the DTCC data time stamps in order to check whether the distribution of the time stamps follows the usual intraday turnover pattern observed in all financial markets, or if there is an unusual aggregation of timestamps in the end of the day (due to delayed reporting). Figure reffig:intradayTurnover shows this analysis.

All of the above factors make the DTCC repository an ideal source for our analysis purposes.

Appendix A.2. General remarks on the data used in this paper

- The options price stream onto which the reconstructed implied volatilities are matched, comes from Bloomberg. Those options price streams are quoted in terms of implied volatility for standardised contracts (standard tenor and delta).
- The forward points and interest rates which we further used in the estimation of the implied volatility behind each option trade, as well as for the calculation of the delta and gamma of the OMM at each point in time, are also from Bloomberg.
- Only vanilla option contracts with a time to expiry below 600 days are taken into account.

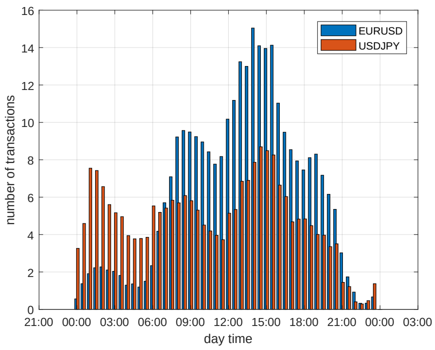


Figure A.6: This figure shows the intraday turnover of the data that underlie our analysis. The intraday turnover, measured as the number of transactions per half hour, seems to follow a typical financial market turnover pattern and therefore serves as a strong indication that the trades are reported in a timely manner to the DTCC trade repository. This is a crucial condition for the analysis, as it means that the timestamps are reliable, which is a necessary precondition for the reconstruction of the side (if it was traded on the bid or the ask price). Furthermore, the volumes traded for EURUSD and USDJPY are comparable in size.

- The observed annualised spot market volatility is calculated based on mid-price data from the corresponding primary FX market of the currency pairs (EBS).
- Analysis date range: 20.9.2017 to 30.6.2018.
- Intraday Bloomberg price streams are only available 60 days into the past. We thus cannot estimate the OMM's positions (from the DTCC data) for the period before the analysis period. Therefore, the estimated gamma exposure of the OMM is inaccurate in the beginning of the analysis period, as only the "new" transactions contribute to our estimate of the OMM exposure. However, the farther we go into the analysis period, the more accurately we can know the OMM exposure. As the median time to expiry of the analysed option contracts turned out to be 30 days, we leave out the first 30 days in our analysis to obtain a representative data set of the OMM positioning. All figures are displayed accordingly.

Appendix B. Derivation of the OMM's positioning behind each transaction

The DTCC repository provides useful information on each transaction but lacks one key point for our analysis: whether the deal was a paid (buyer initiated) or given (seller initiated). We solve this question by using a trade classification algorithm. First, we reconstruct the implied volatility of each trade by using the data from the DTCC trade repository. Second, using a trade classification algorithm, we match the reconstructed implied volatility onto a reference bid/ask price stream from Bloomberg and estimate whether it was traded on the bid or the ask price. By doing so, we estimate the side behind each transaction, that is, if it was a paid (active buy transaction) or a given (active sell transaction). Third, we derive the OMM's position, as she is always liquidity provider by definition (i.e., if the transaction was identified as paid, we assume the OMM is short the contract, as she provided liquidity on the ask).

Appendix B.1. Reconstruct the implied volatility from the transaction data

Here, we outline Blacks model. These formulas are used in the determination of the implied volatility.

$$C(F,\tau) = D \cdot (\Phi(d_+)F - \Phi(d_-)K) \tag{B.1}$$

$$P(F,\tau) = D \cdot (\Phi(-d_-)K - \Phi(-d_+)F) \tag{B.2}$$

$$d_\pm = \frac{1}{\sigma\sqrt{\tau}} \cdot (ln(\frac{F}{K}) \pm \frac{1}{2} \cdot \sigma^2\tau) \tag{B.3}$$

$$d_\pm = d_\mp \pm \sigma\sqrt{\tau} \tag{B.4}$$

$$\tau = T - t \tag{B.5}$$

$$D = e^{-r_{price\ ccy} \cdot \tau} \tag{B.6}$$

$$F = S \cdot e^{(r_{price\ ccy} - r_{base\ ccy}) \cdot \tau} \tag{B.7}$$

Now we briefly explain the variables and comment on where the information for each variable comes from.

- Φ: Cumulative distribution function of the standard normal distribution
- T: Expiry date; from DTCC data
- t: Trade time stamp; from DTCC data
- τ: Time-to-expiry (also called tenor); can be calculated with equation (B.5)
- K: Strike price; from DTCC data
- $r_{price\ ccy}$: Interest rate of the price currency for the tenor τ; from Bloomberg; obtained by either looking at the bond yield of the corresponding country with tenor τ or, if τ is in between two standard bond contracts, calculating an interpolated yield from the two corresponding bond contracts (e.g., for a tenor of $\tau = 1.3$ years, the bond yield of the $\tau = 1$ year and $\tau = 2$ years is interpolated)

- $r_{\textit{base ccy}}$: Interest rate of the base currency for the tenor τ; from Bloomberg; obtained by either looking at the bond yield of the corresponding country with tenor τ or, if τ is in between two standard bond contracts, calculating an interpolated yield from the two corresponding bond contracts.
- $F(\tau)$: Forward price (amount of price currency per unit of base currency in time τ); comes from Bloomberg; obtained by either looking at the forward points of the corresponding currency with tenor τ or, if τ is in between two standard contracts, calculating an interpolated forward point from the two corresponding contracts (Note: We can get the data for the forward point from Bloomberg directly or calculate the forward with the interest rates $r_{\textit{price ccy}}$ and $r_{\textit{base ccy}}$. We prefer the first method, as it probably is a more realistic measure because it includes the cross-currency basis.).
- S: Spot price (amount of price currency per unit of base currency); from EBS
- C, P: Premium of the call (C) and put (P) option; from DTCC data

Using the above variables and formulas, we see that the only unknown quantity in equation (B.1), for a call option, or in equation (B.2) for a put option, is the implied volatility σ (note that this is not the same sigma as in the main part of the paper; here it is the implied volatility), which is our objective. Thus, we only need to numerically minimise the following equation and estimate the implied volatility of each transaction.

$$\min_{\sigma} ||C_{\textit{Black formula}}(\sigma) - C_{\textit{observed in DTCC}}|| \rightarrow \sigma_{reconstructed} \tag{B.8}$$

Appendix B.2. Determine the OMM's side behind each transaction

Once we estimate the implied volatility behind a transaction, we can match the transaction time stamp (which is given by the DTCC data) to a corresponding reference bid/ask price stream for the implied volatility (from Bloomberg). Note that it is important that the reference price stream is adjusted to match the tenor and delta of the reconstructed implied volatility of the transaction. Thus, the reference price stream onto which we match the reconstructed implied volatility of the transaction results from spline interpolation of the volatility surface (from Bloomberg) at the time when the transaction occurred. If the difference between the estimated implied volatility and the ask price of Bloomberg is *smaller* than the corresponding difference between the estimated implied volatility and the bid price from Bloomberg, we label the transaction as a paid, as the transaction probably occurred on the ask price. Conversely, if the difference between the estimated implied volatility and the ask price from Bloomberg is *larger* than the corresponding difference between the estimated implied volatility and the bid price from Bloomberg, we label the transaction as a given.

$$|\sigma_{estimated} - \sigma^{ask}_{reference}| - |\sigma_{estimated} - \sigma^{bid}_{reference}| \rightarrow \begin{cases} < 0, & \text{paid} \\ > 0, & \text{given} \end{cases} \tag{B.9}$$

Then, as the OMM is always liquidity provider, she is short the position if the transaction is a paid and long if the transaction is a given.

Appendix B.3. Quality of the reconstruction

Frömmel et al. (2020) present a nice discussion of the quality of trade reconstruction algorithms. They conclude that the modified EMO rule suggested by Chakrabarty et al. (2007) performs best, which works in the following manner: First, for trades at the quote and up to 30% below the ask or above the bid, the quote rule of Hasbrouck (1988) should be applied (quote rule: if price above (below) midpoint: buy (sell); Prices at midpoint unclassified; This method uses quote data.). Then, for all trades in the inner 40% of the spread, the tick rule of Blume et al. (1989) should be applied (tick rule: if price higher (lower) than previous: buy (sell); If no price change, same as previous. This method uses transaction data.). In our reconstruction, we only use the quote rule. The rationale for this decision is discussed in this subsection.

While the modified EMO rule is straightforward to apply for a centralized market, the application to the FX option market is a bit challenging, as the FX options market is a fragmented OTC market, so there is no one central exchange that records everything. There are rather several repositories to which the transactions that occur on exchanges or between two counterparties (OTC) are reported to. This means there is not one trade data set to use for the tick rule and not one reference price data set to use for the quote rule.

We take the Bloomberg best bid and ask price as reference quote price, which is an aggregated price from several market places. And as it is an aggregated best bid and ask price stream, the corresponding spread is small. This helps for the reconstruction, as most transactions have prices that lie outside of this tight spread (as most market participant cannot access those best bid and ask prices) and thus can be clearly identified as paid or given, according to the quote rule. With this reference price we apply the quote rule.

To discuss the quality of the reconstruction based on the quote rule, we calculate the difference of each trade price, $p_{i,t}$, from the contemporaneous far touch reference price, $p_t^{fartouch}$ and normalize it by the contemporaneous spread, $p_t^{ask} - p_t^{bid}$. If the trade is a paid, the far touch is the offer and if the trade is a given, the far touch is the bid. D_i is a +1 if the transaction is a given and -1 if the transaction is a paid. Thus m_i tells us how many spreads away the transaction occurred from the far touch.

$$m_i = D_i \frac{(p_{i,t} - p_t^{fartouch})}{p_t^{ask} - p_t^{bid}} \tag{B.10}$$

This is what we see in figure B.7. A positive value indicates that the trade occurred in the spread, while a negative value indicates that the trade occurred outside of the spread. As expected, most trades occur outside of the tight Bloomberg spread.

The modified EMO rule assigns all trades at and up to 30% below the ask or above the bid to the corresponding far touch according to the quote rule (so it assigns a paid if the far touch is the ask and it assigns a given if the far touch is the bid). Only the transactions that lie in the inner 40% of the spread need to be reassigned using the tick rule. In our case, this is only amounts to 16% of all EURUSD transactions and 19% of all USDJPY transactions. Thus, for only a minor share of our transactions the modified EMO rule actually requires further analysis.

And as discussed above, using the tick rule is now a bit more tricky, as we do not have a complete set of transactions, but only the transactions reported to DTCC. And since applying the tick rule on an incomplete set of transactions is not ideal,

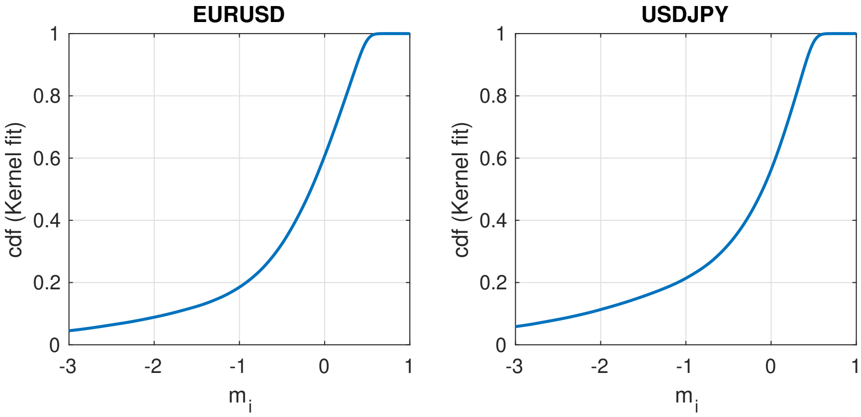


Figure B.7: Cumulative distribution function (cdf) of the variable as defined equation B.10. m_i tells us how many spreads away the transaction occurred from the far touch. A positive value indicates that the trade occurred in the spread, while a negative value indicates that the trade occurred outside of the spread.

we stick to the pure quote rule for our classification.

Nevertheless, for the sake of completeness we applied the tick rule (on the DTCC data set) to analyze how many trades would be classified differently when applying the tick rule (instead of the quote rule) to the trades that occurred in the inner 40% of the spread, as suggested by the modified EMO rule. We find that for EURUSD only 6% and for USDJPY only 7% of the trades in the inner 40% would actually be classified differently. In total, this means using the modified EMO rule vs. the quote rule impacts less than 1% of all transaction of EURUSD and 1.3% of all transaction of USDJPY. In other words, the quote rule and the modified EMO rule is identical for around 99% of all transactions in our data set.

Thus, as applying the tick rule on an incomplete set of transactions is not ideal and as it further does not even impact the outcome, we use the quote rule for the reconstruction.

Appendix C. Determine the OMM's delta and gamma exposure at each time step

Now that we have derived the OMM's position behind each transaction, we can calculate her potential delta and gamma exposure at each point in time using equations (C.1) and (C.2) for the delta and equation (C.3) for the gamma. K is the strike price, S is the current spot price, F is the forward price for the remaining time, σ is the implied volatility for the remaining time, and τ is the remaining time.

$$\Delta_{call} = \frac{\partial C(F,K,S,\sigma,\tau)}{\partial S} = \Phi(d_+) \tag{C.1}$$

$$\Delta_{put} = \frac{\partial P(F,K,S,\sigma,\tau)}{\partial S} = -\Phi(-d_+) \tag{C.2}$$

$$\Gamma = \frac{\phi(d_1)}{S\sigma\sqrt{\tau}} \tag{C.3}$$

Note that ϕ is the standard normal probability density function and Φ is the cumulative distribution function of the standard normal distribution. Both measures, gamma and delta, are subsequently multiplied by the signed notional of each position (e.g., -1 million base currency if the OMM is short an option with a notional of 1 million and +1 million base currency if the OMM is long an option with a notional of 1 million), resulting in the gamma and delta exposure of the OMM. Equation (C.4) describes the delta exposure and equation (C.5) the gamma exposure, whereby N is the notional of the option contract. If we sum up the gamma exposures of all active (not yet expired) positions for a certain point in time, we end up with the net gamma exposure of the OMM for this time.

$$\delta_{exposure} = \Delta * N \tag{C.4}$$

$$\gamma_{exposure} = \Gamma * N \tag{C.5}$$

$$sign(D) = \begin{cases} 1 & \text{if market maker is long} \\ -1 & \text{if market maker is short} \end{cases} \tag{C.6}$$

Appendix D. Is our model tradable?

Note that our non-linear regression used for the model estimation relates the OMM's gamma exposure in the beginning of the time period to the realised volatility during the time period. Thus, our regression suggests that based on the publicly available trade repository data we use in this report (and the subsequent estimation of the OMM's gamma exposure), a good forecast of the realised spot market volatility is possible. As the realised volatility is strongly correlated with the implied volatility, this could suggest a potential trading strategy. If we see that the OMM increased her short gamma exposure by i.e., $10bn$ EURUSD from time $t-2$ to time $t-1$ (so the net gamma exposure changes by $-10bn$), we know from our model parameter estimates that the realised spot market volatility of EURUSD in the time period $t-1$ to t will go up by $(-10bn) \cdot (-0.0007\frac{\%}{bn}) = 0.007\%$. Furthermore, we know that the realised spot market volatility is correlated with the implied volatility. If we therefore buy an option at $t-1$ and sell the option again at t, we sell 0.007% higher (option prices are usually quoted in terms of volatility). The converse is true if we see that the OMM decreases her short gamma exposure by $10bn$. Of course, this trading strategy is based on an isolated view that only takes the observed feedback effect that the OMM gamma exposure has on the spot market volatility into account. We empirically tested this trading strategy with the data behind the results presented in table 1 and found that the observed effect has a Sharpe ratio close to zero, which indicates efficient market behaviour. Additionally, the bid ask spreads in the option markets are usually wider than 0.007% anyway, so taking transaction costs into account, the strategy would even be unprofitable.

Appendix E. Estimation with intraday data

In this section we discuss the feedback effect on the spot market volatility independent of our model framework on an intraday basis.

Appendix E.1. Estimation of a random intercept linear mixed-effect model

As we can reconstruct the gamma exposure of the OMM on an intraday frequency, we can also quantify the effect with this intraday data. One difficulty of this intraday estimation is that the fundamental volatility (the constant of the linear regression model) depends on the intraday time, as it is influenced by the intraday liquidity cycle (trades influence prices and therefore also the volatility). We therefore cannot just mix the data from different intraday times and estimate the linear regression model. Thus, we use a mixed-effect model, namely, a random intercept linear mixed-effect model, as presented in equation E.1. As when we presented the gamma exposure, each trading day is split into 12 two-hour intervals. [12] We define a dummy variable for the corresponding intraday time stamp, I_m (where $m = \{1, ..., 12\}$), that indicates the corresponding intraday interval. I_m then is our grouping variable for the constant $\alpha_1[\%]$, as the fundamental volatility depends on the intraday time stamp. This construct, the constant conditional on the grouping variable, $(\alpha_1[\%]|I_m)$, is called the random effect, while the fixed effect is given by $\alpha_1[\%]$ and $\alpha_2[\frac{\%}{\text{bn}}]$. As the volatility can have extreme values on an intraday level, due to e.g., data releases, we remove outliers from the intraday data set. [13]

$$\sigma_{t,m}[\%] = \alpha_1[\%] + (\alpha_1[\%]|I_m) + \alpha_2[\frac{\%}{\text{bn}}] \cdot N_t S_{t,m} \gamma^M_{t,m}[\text{bn}] + \epsilon_{t,m} \tag{E.1}$$

Table E.2 displays the results of the estimation of equation E.1. Looking at the fixed effect, we see that the results confirm those obtained with the daily data. Note that now the fundamental volatility is nearly equal between EURUSD and USDJPY, as we account for the different intraday volatility regimes.

Appendix E.2. Relation to the volatility model

Note that the regression model presented in the previous section is just the linear approximation of equation 2.10. Equation E.2 shows the derivation of the linear regression model, where we first write the series expansion and then a linear approximation of it (the last step in the derivation is justified by $|\beta h^{OMM}_{net} N_t S_t \gamma^M_t| <= 1$).

---

[12] UTC 00:00-02:00, 02:00-04:00, 04:00-06:00, 06:00-08:00, 08:00-10:00, 10:00-12:00, 12:00-14:00, 14:00-16:00, 16:00-18:00, 18:00-20:00, 20:00-22:00, 22:00-24:00

[13] We remove the highest and lowest 2.5 percentiles.

	EURUSD	USDJPY
α_1 (%)	5.7***	6.2***
	(10.8)	(11.1)
α_2 ($\frac{\%}{\text{bn}}$)	-0.0007***	-0.0008***
	(-8.6)	(-3.3)
R-squared	0.34	0.17
F-test vs. const. model	74.5***	10.8***

Table E.2: This table shows the regression results of the random intercept linear mixed-effect model presented in equation E.1 for EURUSD and USDJPY. t-statistics are presented in brackets. The stars *** denote significance at the 1% level.

$$\sigma_t = \frac{\nu}{(1+\beta h_{net}^{OMM} N_t S_t \gamma_t^M)} \tag{E.2a}$$

$$= \nu \sum_{k=0}^{\infty} (-\beta h_{net}^{OMM} N_t S_t \gamma_t^M)^k \tag{E.2b}$$

$$= \nu(1 - \beta h_{net}^{OMM} N_t S_t \gamma_t^M + O(2)) \tag{E.2c}$$

$$\approx \underbrace{\nu}_{=\alpha_1} + \underbrace{(-\nu\beta h_{net}^{OMM})}_{=\alpha_2} N_t S_t \gamma_t^M \tag{E.2d}$$

Therefore, the estimations presented above provide support for our volatility model or, rather, its linear approximation.

$N_t S_t \gamma_t^M$ is the gamma exposure of the OMM. As $\alpha_2 = -\nu\beta h_{net}^{OMM}$, where h_{net}^{OMM} can be positive or negative (depending on whether the OMM or the OMT hedges a larger share of their exposure), $\nu > 0$ (fundamental volatility) and $\beta > 0$ (market impact), the sign of α_2 tells us whether the OMM hedges a larger share of her delta exposure compared to the OMT (as explained before). We estimate α_2 to be negative, so we can conclude that $h_{net}^{OMM} > 0$ and therefore that the OMM hedges a higher share of her delta exposure compared to the OMT. The value of α_2 quantifies the percentage increase in spot market volatility per gamma exposure of the OMM.

As we estimate $\alpha_1 = \nu$ separately, we could calculate βh_{net}^{OMM} from $\alpha_2 = -\nu\beta h_{net}^{OMM}$. Nevertheless, we cannot further dissect βh_{net}^{OMM} in its parts, as β and h_{net}^{OMM} are not further specified.

Appendix E.3. What can we learn from the intraday estimation

First, we want to highlight that the results presented above are very much in line withe the model validation results presented in section 4. This is further strong evidence that the quantification of the feedback effect works well with our reconstructed OMM data set.

So the quantification of the feedback effect of the OMM's gamma exposure onto the spot market volatility is given by the parameter α_2 of the intraday estimation (equation E.1).

The significance of α_2 tells us that the gamma exposure of the OMM actually has relevance for the spot market volatility, and its value tells us what the corresponding linear effect is per billion gamma exposure of the OMM. Concretely, we see that the overall feedback effect stemming from the OMM's gamma exposure, which is given by the value of α_2, suggests that a short gamma exposure of the OMM of -1 billion in the base currency results in an increase of the realised spot market volatility of approximately 0.001% for both EURUSD and USDJPY. This does not look like a large effect, but as we see from the data displayed in figure 3, the gamma exposure of the OMM is of order -10^3 billion in both EURUSD and USDJPY. Consequently, the resulting effect is actually an increase in the spot market volatility of approximately 1% $(-0.001[\frac{\%}{\text{bn}}] \cdot -10^3[bn] = 1[\%])$. Note that the feedback effect of the gamma exposure on the spot market volatility is (most probably) not linear, as we know from our volatility model. Therefore, the quantification of the feedback effect presented in this section is only a linear approximation.

The constant α_1 is also found to be highly significant. α_1 can be interpreted as the spot market volatility in the absence of any delta hedging activity, the fundamental volatility. Note that for USDJPY, we find a strong difference in the estimate of ν from the model validation section 4, in which we used daily data, compared to the intraday USDJPY estimate presented above. The spot market volatility in the absence of any delta hedging activity is found to be lower for the daily estimate (5.0%) compared to the intraday estimate (6.2%). This difference stems from the fact that the volatility data for the daily estimate for USDJPY from the model validation section 4 originate from the Asian trading hours, in which the volatility is usually lower than during the European or American trading hours. As the intraday model takes the intraday time explicitly into account (and therefore also the intraday volatility regime), and corrects for it with a random effect, we obtain a more meaningful estimate of α_1 for USDJPY in the intraday estimation.

On the Directional Destabilizing Feedback Effects of Option Hedging

Didier Sornette[a,b,d,e], Florian Ulmann[a], Alexander Wehrli[a,c]

[a] ETH Zurich, Department of Management, Technology, and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland

[b] Swiss Finance Institute, c/o University of Geneva, 40 blvd. Du Pont d'Arve, CH–1211 Geneva 4, Switzerland

[c] Swiss National Bank†, Boersenstrasse 15, 8001 Zuerich, Switzerland

[d] Tokyo Tech World Research Hub Initiative, Institute of Innovative Research
Tokyo Institute of Technology, Tokyo, Japan

[e] Institute of Risk Analysis, Prediction and Management (Risks-X)
Academy for Advanced Interdisciplinary Studies
Southern University of Science and Technology (SUSTech), Shenzhen, 518055, China

Abstract

We investigate the feedback effect of option delta hedging activity on the stability of the price of the underlying. While previous literature has focused on the effect of hedging activity on the volatility of the underlying, this paper focuses on directional instabilities arising from feedback effects. We propose a model, in which the drift of the underlying is affected by delta hedging and couple the predictions of this model with an approach to identify short-lived, locally explosive trends in the expected price evolution. Using simulations, we show that such drift bursts, as defined in Christensen et al. (2020), indeed occur with higher or lower intensity (depending on the option parameters and its delta) compared to the unaffected asset price, where our model predicts such effects. Our analytical results and synthetic experiments furthermore suggest that the effect of hedging on price stability is strongly asymmetric for the case where the option market maker is short vs. long the option. When the market maker is short, we predict a significantly more pronounced effect on the spot price than when she is long. Since the regime where the market maker is short is the predominant regime empirically, this suggests that option hedging indeed can be expected to impact prices. Using the example of the GameStop stock price in early 2021, we investigate the predictions of our model on empirical option positions and document that there are indeed instances where explosive trends due to hedging were more likely.

JEL: G12 C51 C53 C58

Keyword: option hedging feedback effects instability drift burst

---

Email addresses: `dsornette@ethz.ch` (Didier Sornette), `fulmann@ethz.ch` (Florian Ulmann), `alexander.wehrli@snb.ch` (Alexander Wehrli)

[†] The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank. The Swiss National Bank takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

1. Introduction

At intraday time scales, financial price properties have been documented to deviate significantly from what a random walk hypothesis would predict. This disaccord suggests the presence of frictions, and naturally raises the question whether such frictions are owed to particular structural aspects of financial markets. Three main categories of explanations have been proposed to answer this question. The first category explains such effects with economic frictions, stemming from capital constraints and margin requirements (Brunnermeier and Pedersen, 2008). The second category argues with the specifics of how information gets incorporated into asset prices (Easley et al., 1998). The third one finally draws from the pool of discovered behavioural biases of investors (Ben-David and Hirshleifer, 2012).

Complementary to these explanations, this paper focuses on a liquidity friction that is introduced in the market of the underlying (the spot market) by a feedback mechanism from the corresponding derivatives market. Derivatives markets, and in particular options markets, have obtained significant importance due to risk management practices, as well as regulatory constraints. As a consequence, potential feedback effects stemming from options markets also present natural candidates to explain the type of frictions we are interested in. As a recent example, hedging demand from options markets was a frequently cited explanation for the erratic price fluctuations observed in the GameStock (GME) stock price at the beginning of 2021. The Financial Times (FT) for example explained the GME price moves with:

“Small investors can influence share prices by buying large amounts of call options, [...]. Typically, this means that wholesale brokers such as banks need to purchase shares on the open market to hedge themselves [...]. This hedging can have an outsized impact on prices [...].” (FT, 2021)

Motivated by these recent events, we formalise this discussion in an attempt to answer the question whether the options market indeed has a measurable and relevant effect on the stability of the spot price.

Our work builds on the idea first discussed in Sircar and Papanicolaou (1998) and on the respective formalism presented by Anderegg et al. (2019) and Ni et al. (2020), which all investigate the relation between option hedging and the volatility of the spot price. We go beyond these contributions and also take into account the effect of option hedging on the *drift* of the spot price. From there, we discuss the potential for directional price instabilities to arise due to feedback from option markets. Our work is also related to research into the feedback effects of derivatives markets on the price of the underlying in the context of exchange traded funds (ETF) (Shum Nolan et al., 2015; Ben-David et al., 2018).

The two main questions addressed in this paper are first, whether option positioning can lead to price instabilities, which are quantified using *drift bursts* as introduced by Christensen et al. (2020). Both our theoretical discussion as well as our simulation studies clearly show that the hedging of options with high notional and a low time to expiry can lead to either unusually stable or unusually explosive prices, depending on the delta of the option(s). A key insight of our findings is that the effect of option hedging is strongly *asymmetric* for when the option market maker is short the option, compared to when she is long. Since the regime where the market maker is short is the predominant one empirically, this suggests that option hedging indeed can be expected to have a strong impact on prices.

The second question we address is whether option positioning was a potential contributor to the observed price distortion in the GME stock price during the first quarter of 2021. Calibrating our model to price and position data for GME, we indeed find that there were instances where our model predicts an increased option-induced instability in the GME price during the period in question. However, with our limited resolution price data, only a circumstantial link between the predicted drift burst intensity and the realized dynamics can be established.

The rest of the paper is structured as follows. Section 2 introduces the model of feedback effects from the derivative markets onto the price dynamics of the underlying. Section 3 discusses how option positioning can lead to price instabilities in the underlying price and section 4 applies our framework to the GME stock price. Section 5 concludes.

2. A model for the feedback effect of option hedging

Note that up to equation 2.10, the following derivation is identical to the one presented in Anderegg et al. (2019), but still repeated here for coherence. Everything after equation 2.10 are novel contributions to the literature.

2.1. Derivation of the spot price dynamics in the presence of option hedging

Our goal is to derive the dynamics of the spot market price that includes the price impact of option hedging activity – in particular delta hedging. To this end, we assume the observable price, denoted by S_t, to follow a linear stochastic differential equation

$$\mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sigma_t S_t \mathrm{d}W_t. \tag{2.1}$$

Furthermore, we assume that the unobservable fundamental value of the asset F_t, which has the interpretation of the market price unaffected by delta hedging (the friction in our model), follows the dynamics

$$\mathrm{d}F_t = \alpha F_t \mathrm{d}t + \nu F_t \mathrm{d}W_t. \tag{2.2}$$

Applying Itô's lemma to the log-price changes $\mathrm{d}f_t = \mathrm{d}(\log(F_t))$ results in $\mathrm{d}f_t = \frac{\mathrm{d}F_t}{F_t} - \frac{1}{2}\nu^2 \mathrm{d}t$. Using $\mathrm{d}F_t$ from (2.2) we can then write $\mathrm{d}f_t$ as

$$\mathrm{d}f_t = \underbrace{(\alpha - \frac{1}{2}\nu^2)}_{\alpha'} \mathrm{d}t + \nu \mathrm{d}W_t = \alpha' \mathrm{d}t + \nu \mathrm{d}W_t. \tag{2.3}$$

We assume that there are only two market participants, an option market maker (OMM) and an option market taker (OMT). Their net delta exposures are respectively $N_t^M \Delta_t^M$ and $N_t^T \Delta_t^T$, respectively, where N is the notional (always positive) and Δ is the option delta (being the derivative of the option value with respect to S, it is positive for a long call and a short put, and negative for a short call and a long put). Both market participants hedge their overall delta exposure to a certain degree in the spot market, given by the hedge ratios h^M and h^T, which we assume to stay constant over time. Assume that the market participants hedge their risk at discrete intervals of length Δt. If the price changes by a certain percentage between time $t - \Delta t$ and t, the net traded delta hedging volume, executed in the spot market at time t, is given by the negative of the change in the delta exposure between $t - \Delta t$ and t, multiplied by the hedge ratio, so $V_t^M = h^M N_t^M (-(\Delta_t^M - \Delta_{t-\Delta t}^M))$ for the OMM and $V_t^T = h^T N_t^T(-(\Delta_t^T - \Delta_{t-\Delta t}^T))$ for the OMT. Note that, in a market with two participants, the notional is the same for both participants ($N_t^M = N_t^T = N_t$), and the delta of the OMM is always equal to the negative of the delta of the OMT ($N_t \Delta_t^M = -N_t \Delta_t^T$). Furthermore, the position of each trader represents the total outstanding position in the market (which is why our two agents are the OMM and the OMT, as by definition the aggregated positions from all OMMs sum up to the total outstanding positions in the market). Defining the difference in the hedge ratio between the OMM and the OMT from the OMM's point of view as $h = h^M - h^T$, the total net traded delta hedging volume between $t - \Delta t$ and t is given by

$$\begin{aligned} V_t &= V_t^M + V_t^T && \text{(2.4a)} \\ &= h^M N_t^M(-(\Delta_t^M - \Delta_{t-\Delta t}^M)) + h^T N_t^T(-(\Delta_t^T - \Delta_{t-\Delta t}^T)) && \text{(2.4b)} \\ &= (h^M - h^T) N_t(-(\Delta_t^M - \Delta_{t-\Delta t}^M)) && \text{(2.4c)} \\ &= h N_t(-(\Delta_t^M - \Delta_{t-\Delta t}^M)). && \text{(2.4d)} \end{aligned}$$

In the continuous limit, when $\Delta t \longrightarrow 0$, (2.4d) becomes

$$V_t = -h N_t \mathrm{d}\Delta_t^M. \tag{2.5}$$

Next, we assume that the delta hedging volume executed in the spot market, V_t, impacts the spot market price in a linear and permanent fashion, with β defined as the log-price change (market impact) per traded volume. This is the friction that we introduce and which will, as we see later, result in the feedback effect discussed in this paper. Thus, the observed log-price change, $\mathrm{d}s_t$, is given by the log-price change of the fundamental price, $\mathrm{d}f_t$, as well as the price impact of the delta hedging volume, βV_t, as

$$\begin{aligned} \mathrm{d}s_t &= \mathrm{d}f_t + \beta V_t && \text{(2.6a)} \\ &= \mathrm{d}f_t - \beta h N_t \mathrm{d}\Delta_t^M, && \text{(2.6b)} \end{aligned}$$

assuming that the price impact coefficient β is constant. This specification reflects a price impact model in the spirit of Kyle (1985), where the price impact is permanent, linear and depends only on the volume of the transaction. The assumption that permanent price impact is proportional to the transaction volume is however also in line with modern dynamical theory of liquidity and price impact (Donier et al., 2015).

Applying Itô's lemma to the log-price changes $\mathrm{d}s_t = \mathrm{d}(\log(S_t))$ results in $\mathrm{d}s_t = \frac{\mathrm{d}S_t}{S_t} - \frac{1}{2}\sigma_t^2 \mathrm{d}t$. In combination with $\mathrm{d}s_t$ from (2.6b), we can write $\mathrm{d}S_t$ as

$$\mathrm{d}S_t = S_t \mathrm{d}f_t + \frac{1}{2}\sigma_t^2 S_t \mathrm{d}t - \beta h N_t S_t \mathrm{d}\Delta_t^M. \tag{2.7}$$

Applying Itô's lemma to $\mathrm{d}\Delta_t^M$ with respect to $\mathrm{d}S_t$ leads to

$$\mathrm{d}\Delta_t^M = \underbrace{\left(\underbrace{\frac{\partial \Delta_t^M}{\partial t}}_{-\mathrm{Charm}_t^M := \chi_t} + \mu_t S_t \underbrace{\frac{\partial \Delta_t^M}{\partial S}}_{\gamma_t} + \frac{1}{2}\sigma_t^2 S_t^2 \underbrace{\frac{\partial^2 \Delta_t^M}{\partial S^2}}_{\mathrm{Speed}_t^M := \psi_t} \right)}_{:=\gamma_t^{'}} \mathrm{d}t + \sigma_t S_t \underbrace{\frac{\partial \Delta_t^M}{\partial S}}_{\gamma_t} dW_t. \tag{2.8}$$

The option greeks are labeled with brackets and we suppress the superscript M in the sequel, as all the greeks express the sensitivities regarding the OMM's exposure. γ_t, which is central in our theory, is the option gamma of the OMM at time t. Inserting (2.3) and (2.8) into (2.7) results in

$$\mathrm{d}S_t = (\alpha^{'} S_t + \frac{1}{2}\sigma_t^2 S_t - \beta h N_t S_t \gamma_t^{'})\mathrm{d}t + (\nu S_t - \beta h N_t S_t^2 \sigma_t \gamma_t)\mathrm{d}W_t \tag{2.9}$$

Comparing (2.9) with (2.1) results in

$$\sigma_t = \frac{\nu}{1 + \beta h N_t S_t \gamma_t} \tag{2.10}$$

and

$$\mu_t = \frac{\alpha^{'} + \frac{1}{2}\sigma_t^2(1 - \beta h N_t S_t^2 \psi_t) - \beta h N_t \chi_t}{1 + \beta h N_t S_t \gamma_t}. \tag{2.11}$$

Using (2.10) and the definition of $\alpha^{'}$ from (2.3) finally leads to

$$\mu_t = \frac{\alpha - \frac{1}{2}\nu^2 \left(1 - \frac{1 - \beta h N_t S_t^2 \psi_t}{(1 + \beta h N_t S_t \gamma_t)^2}\right) - \beta h N_t \chi_t}{1 + \beta h N_t S_t \gamma_t}. \tag{2.12}$$

An illustration of the dynamics of $\sigma(S_t)$ and $\mu(S_t)$ from (2.10) and (2.12) is given in Figure 1 for the case of a call option position of the OMM (assuming the OMM hedges a larger share of her delta compared to the OMT). As documented in Anderegg et al. (2019), the gamma exposure of the trader who hedges a larger part of the delta exposure, in this example the OMM, determines how the volatility of the affected price S is scaled away from its fundamental level ν. Additionally however, we see that the drift of the asset price is also affected by the option hedging activity. If the OMM for example has a short call option position (left column), the drift is negative for prices larger than the strike price ($S_t > K$). Intuitively, the reason for this behaviour can most easily be inferred from the delta exposure of the OMM, which follows a sigmoid shape. The negative of this delta exposure is the OMM's hedging demand, i.e. the amount of the underlying she needs to hold given a spot price level in order to be delta neutral. The corresponding change of these holdings determines the amount of the underlying she needs to buy/sell in order to maintain delta neutrality. Thus in the region $S_t > K$, a negative unit change in the price, $\mathrm{d}S_t < 0$, forces the OMM to sell a larger quantity of the underlying to stay delta neutral compared to what the OMM has to buy in case of a positive unit change in the price, $\mathrm{d}S_t > 0$. In other words, the sigmoid shape of Δ^M leads to an asymmetry in the hedging demand, and if those traded quantities have a price impact, then $\mu_t < \alpha$ in the region $S_t > K$, and the effect is reversed for $S_t < K$. These dynamics are reproduced with opposite signs in case of a long option position (right column), but the effect is less pronounced for both the volatility as well as the drift.

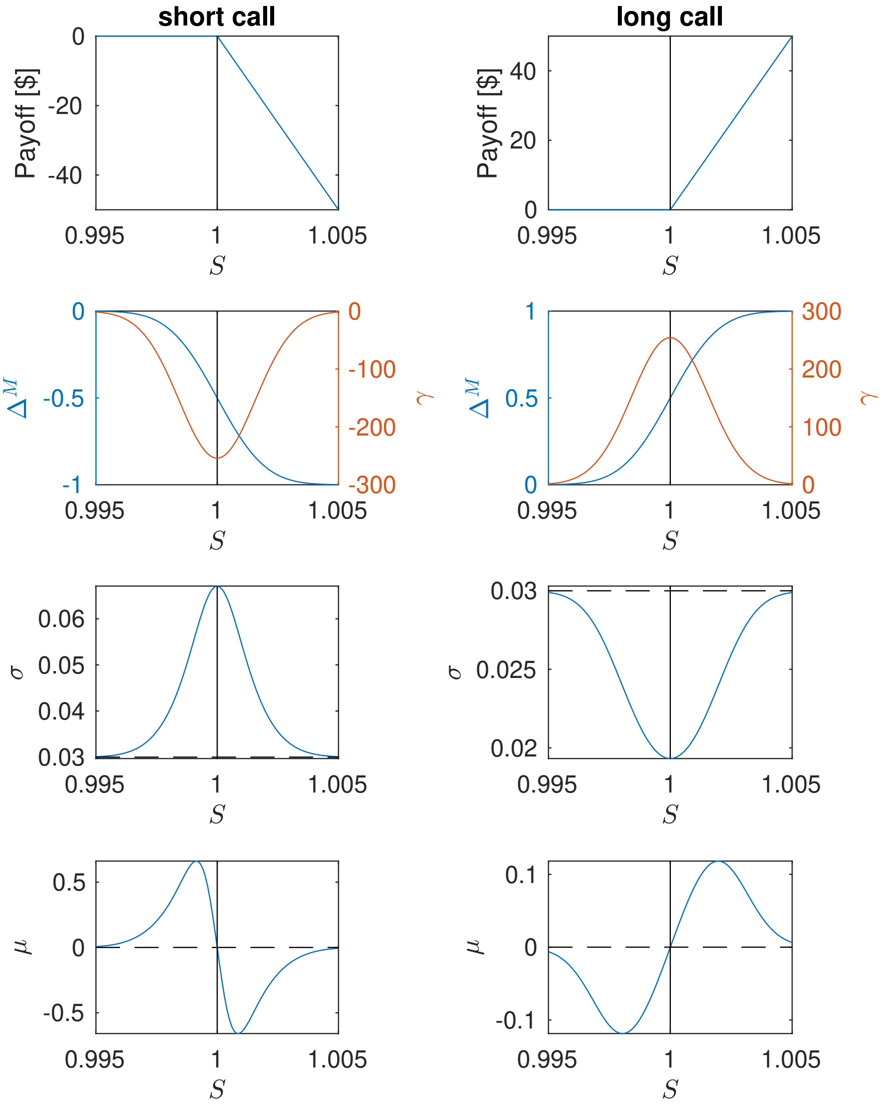


Figure 1: Model dynamics when the OMM has a short call position (left column) or a long call position (right column). The top row shows the payoff at maturity, with a strike price at $K = 1$. The second row displays the corresponding delta and gamma profile with a volatility of 3% and a time-to-expiry of 1 day. We assume that the notional behind the option is $N_t = 10^4$ shares and βh are motivated from empirical results for GME (where we find $h = 0.63$). The third row then shows the resulting volatility $\sigma(S)$, based on (2.10). The corresponding drift $\mu(S)$, of (2.12) is displayed in the bottom row. Horizontal dashed lines indicate the fundamental values of the respective quantities, $\nu = 0.03$ and $\alpha = 0$. Note that for a put option position rows two to four are identical, only the payoff in row one would change accordingly.

2.2. Stochastic processes of μ

Due to the dependence on S_t, both μ_t and σ_t^2 are stochastic processes themselves. Using equations (2.12) and (2.10), we can solve for the dynamics of $\mathrm{d}\mu_t$ using Itô's Lemma to obtain

$$\mathrm{d}\mu_t = a_{\mu,t}\mathrm{d}t + b_{\mu,t}\mathrm{d}W_t. \tag{2.13}$$

The coefficient $a_{\mu,t}$ is illustrated in Figure 2. Several observations can be made for the case of a short call option:

- $\mathrm{d}\mu_t$ is mean-reverting around the strike ($K = 1$). This can be seen by noticing that the deterministic part of the differential equation has the opposite sign of μ_t around K. This indicates mild behaviour of μ_t close to the strike, or a certain "stabilisation" of the price through option hedging.
- Going further away from the strike, $a_{\mu,t}$ has the same sign as μ_t at those prices, hinting towards a clear directional behaviour of μ_t, or a certain "explosiveness" of the price.
- Lastly, we note that the "stabilisation" effect very close to the strike is in fact stronger than the "explosiveness" induced further away from it, which gets obvious if one compares the respective magnitude of the amplitudes.

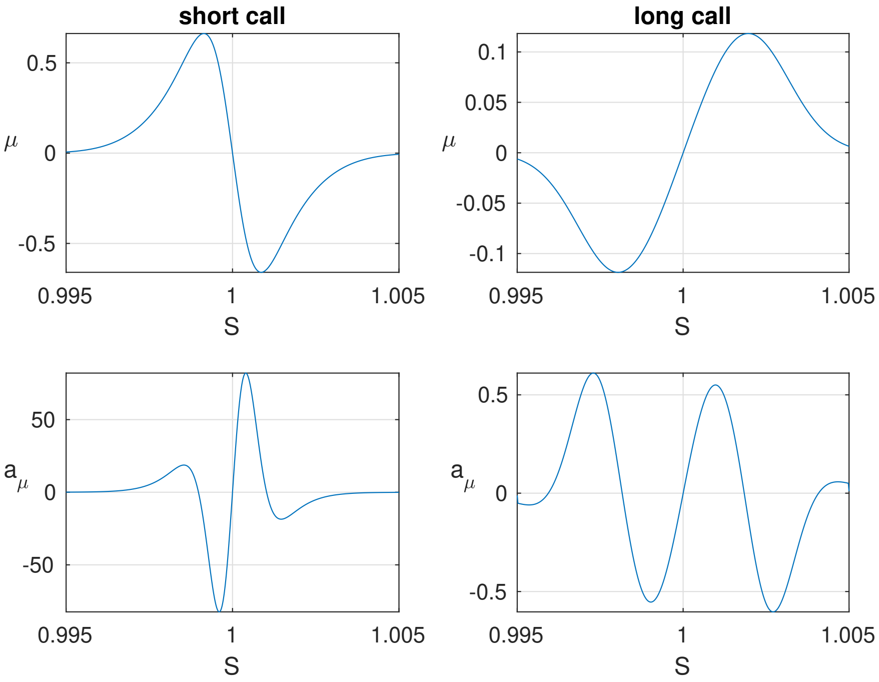


Figure 2: Illustration of the deterministic part of the drift μ for the two cases discussed in Figure 1. The dynamics of the process is given by (2.13) and we solve numerically for the relevant coefficient displayed here. Note that for a put option position the picture is identical.

The observations we can make from the case of a net long positioning of the OMM are clearly very different from the case of a short position:

- In contrast to the short case, $\mathrm{d}\mu_t$ is not mean-reverting around the strike price ($K = 1$), but it is further away from the strike. This can be seen by noticing that the deterministic part $a_{\mu,t}$ of the differential equation $\mathrm{d}\mu_t$ has the same sign like μ_t around K, while it is opposite further away from the strike. This indicates "explosiveness" of μ_t close to the strike, and a "stabilisation" effect further away from the strike. This is opposite from what we observed for the short case.
- More importantly however, the magnitude of the effects are much less pronounced than when the OMM is short. Comparing the left and right column on Figure 2 reveals that the effects of option hedging are something like three orders of magnitude larger when the OMM is short. Already μ itself is less disturbed if the OMM has a long call position, as gets obvious from the first row. A comparison of the deterministic part reveals the same. The deterministic part of μ changes in a range of [-0.5, +0.5] when the OMM is long. When she is short on the other hand, the deterministic part of μ ranges from approx -100 to 100 for the same option parameters. This observation has particular relevance since the OMM being short is the predominant regime in practice. This significant asymmetry between the two regimes can already be seen when comparing the left and right column of Figure 1 and analytically can be explained with the structure of the denominator in equations (2.12) and (2.10). A short OMM position increases the effect of option hedging on the spot price by forcing the denominator (quickly) to zero, whereas a long position (slowly) forces the denominator to infinity, decreasing the effect of option hedging.

In summary, our model predicts that, whenever the OMM is net short in her option positioning, the price is unusually stable very close to the strike and unusually *un*stable further away from the it. And whenever the OMM is net long in her option positioning, the price is not affected too much, but get slightly more unstable close to the strike and slightly more stable further away from it. We test these predictions with a specific measure of instability in the following section. But before going there, we want to shortly discuss the intuition behind our model.

2.3. Intuition for the asymmetry in the effect

The net delta hedge volume of the option market acts linearly on prices. As shown in the previous section, according to our model, we expect an asymmetric effect on the spot price when the OMM is net short, compared to when she is long. What is the intuition behind this asymmetry? If the net delta hedge volume is dominated by a short (long) option positioning, option hedging implies buying (selling) in a rising market and selling (buying) in a falling market. Buying in a rising market thus leads to a positive feedback due to price impact: price rise, the OMM needs to buy the asset to stay delta neutral, and by doing so pushes the price even higher, requiring more buying[2]. In contrast, selling in a rising market does not result in such a positive feedback effect. This fundamental difference between the hedging strategies in the two cases, i.e. the absence of a second order feedback effect in the case of a long position, explains the asymmetric effect that option hedging has on the directional stability of the underlying.

---

[2] This positive feedback loop is in fact not too different from the portfolio insurance strategy that was at least in part responsible for the 1987 stock market crash. Portfolio insurance strategies would sell futures contracts when prices declined, the buyers of these futures would in turn sell the underlying to hedge their derivatives position, driving down prices even further, triggering portfolio insurance strategies to sell even more futures contracts.

3. Option hedging and price (in)stability

3.1. Motivations

In our setup, both the observable volatility and drift, σ_t and μ_t, are functions of the delta exposure of the OMM (or derivatives of it). In particular, the drift and the volatility of the observed price are *nonlinear* functions of S_t and the feedback from option hedging to expected returns can be both positive or negative, depending on the position of the OMM/OMT and their hedge ratio. Models of nonlinear feedback $\mu(S_t)$ have been widely considered as models for financial bubbles and crashes (e.g. Johansen et al. (1999, 2000)). In the same spirit, previous contributions have used the insight that hedging strategies of general Black-Scholes option models lead to a positive feedback on the volatility, to propose nonlinear price models with multiplicative noise, see e.g. Sornette and Andersen (2002); Li and Sornette (2013); Li et al. (2014); Schatz and Sornette (2020). In these constructions, the particular form of the dependencies $\mu(S_t)$ and $\sigma(S_t)$ are *postulated* to be a nonlinear increasing power law of S_t. Crucially, such power law feedback to expected return and volatility can lead to finite time (stochastic) singularities in the price process, which represent times of system instability or regime change (Corsi and Sornette, 2014). The general mathematical conditions for a stochastic finite time singularity to occur in nonlinear stochastic dynamics are given in Cherny and Engelbert (2005) and summarized in Li et al. (2019).

In our setup here, the feedback mechanisms are endogenously derived from a specific generating mechanism (option hedging), but their resulting dynamics are potentially more intricate than power laws, because they additionally also depend on the option positioning and hedging behavior of market participants. These quantities are a priori exogenous to the model, but can naturally be envisaged to themselves depend on the asset price in practice, leading to complex interactions. Our goal is to determine whether the presence of these option-induced feedback mechanisms increases the susceptibility of the asset price to instabilities. One way to nonparametrically test for the presence of directional instabilities – so called *drift bursts* – has recently been proposed by Christensen et al. (2020). Specifically, this procedure allows one to test for short-lived explosive drift dynamics, while being robust to the simultaneous presence of volatility bursts, which makes it particularly attractive for our purposes.

Our strategy will be to use the drift burst test as a measure for the instability of the price process and analyze how changes in the option positioning affect this measure according to our model. For this purpose, we first give a high-level introduction to the drift burst methodology and then apply the procedure to synthetic realizations from our model.

3.2. Drift bursts

Standard formulations for continuous-time arbitrage-free price processes are typically cast in the form of Itô semimartingales, such as (2.1). From the additional typical assumption of local boundedness of drift and volatility coefficients of these processes, it follows that, on vanishingly small time intervals, contributions of volatility dominate contributions of the drift. In this case, Christensen et al. (2020) show that suitably scaled ratios of nonparametric, kernel-based estimates of the drift ($\hat{\mu}_t$) and volatility ($\hat{\sigma}_t$) coefficients, are asymptotically normally distributed. The proposal is thus to consider a test statistic

$$T_t \propto \frac{\hat{\mu}_t}{\hat{\sigma}_t}, \tag{3.1}$$

which under the null of no drift burst has a limiting normal distribution. If however the drift term locally explodes fast enough relative to the volatility, then this statistic can grow arbitrarily large, providing the grounds to perform a statistical test for the presence of a bursting drift coefficient. The detection strategy is thus to compute the test statistic progressively over time and reject the null of a non-explosive drift coefficient when $|T_t|$ gets significantly large.

3.3. *Types of singularities covered by drift burst tests*

Before applying the drift burst test to our specific problem, it is useful to illustrate in more detail what type of divergences the test will identify. To be precise, the test identifies times t as drift bursts, where over some time interval $\delta \longrightarrow 0$ we have

$$\int_{t-\delta}^{t+\delta} \mu_s \mathrm{d}s = O_p(\delta^{\gamma_\mu}) \quad \text{for} \quad 1/2 > \gamma_\mu > 0 \tag{3.2}$$

and

$$\int_{t-\delta}^{t+\delta} \sigma_s \mathrm{d}W_s = O_p(\delta^{\gamma_\sigma}) \quad \text{for} \quad \gamma_\sigma > \gamma_\mu. \tag{3.3}$$

Thus the assumption is that, for any ϵ, there exists a $C > 0$, so that the probability of $\left|\int_{t-\delta}^{t+\delta} \mu_s \mathrm{d}s\right|$ exceeding $C \cdot |\delta^{\gamma_\mu}|$ is small, i.e. the probability of exceeding is stochastically bounded by ϵ, as $\delta \longrightarrow 0$. It turns out that this condition covers a wide range of (very) different types of singularities. To motivate our thinking, consider first the simple case where μ is locally deterministic and differentiable. In this case, condition (3.2) implies that μ is of the form

$$\mu_t(s) \sim \frac{1}{|t-s|^h} \tag{3.4}$$

with $0 < h < 1$, so that μ diverges locally (at t) with exponent h, while remaining integrable. By integration in (3.2), we furthermore see that the exponent $\gamma_\mu = 1 - h$, so in fact we must have $1/2 < h < 1$ to satisfy (3.2).

A second case to consider are functions that are locally deterministic, but non-differentiable. Let us assume for example that μ is given by the measure defined as the local derivative of the Weierstrass function $f_\alpha(t)$ with parameters a and b. By construction, this measure is almost everywhere singular, but obeys the Lipschitz condition $\int_{t-\delta}^{t+\delta} f_\alpha(s)\mathrm{d}s = |f_\alpha(t+\delta) - f_\alpha(t-\delta)| \sim |\delta|^\alpha$ for the exponent $\alpha = -\ln(a)/\ln(b)$ (Gluzman and Sornette, 2002). Thus in this case, we have the correspondence $\gamma_\mu = \alpha$.

Finally, we can also envisage the singularity in μ to be stochastic. For example, if the measure defining μ was derived from a geometric Brownian motion, we would find $\gamma_\mu = 1/2$, which is however excluded by the condition $\gamma_\mu < 1/2$ from (3.2).

While these differences between the three types of singularities might appear to be mathematical niceties, it is important to stress that they also differ notably in both their amplitude and financial interpretation. For locally deterministic and differentiable bursts, there are local, super-exponential dynamics, similar to bubble regimes at longer time scales. For non-differentiable dynamics, as illustrated with the Weierstrass construction, there is a complex and subtle hierarchy of trend-changes at all scales. Finally, for stochastic singularities, there is again a complex but random set of trend-changes at all scales. Commensurate with these differences, one should also assume that the underlying generating mechanisms of the various singularity types are not the same. Since all of the singularity types satisfy the condition (3.2) posed by the drift burst statistic, the resulting test is also incapable of distinguishing the different types of singularities and their origin.

For the present study, we are mainly interested in having a general measure of the instability of the price process, as characterized by divergences in μ. As such, the drift burst statistic is a useful tool despite its general condition of divergence. To make progress on distinguishing different types of divergences, the so-obtained identification of singularities could be complemented with estimates of the (local) Hölder exponent, e.g. using tools from signal processing, which can be used to tell apart differentiable from non-differentiable singularities. To also distinguish deterministic from stochastic singularities, additional diagnostics are required. We leave this study for future work.

3.4. Drift bursts induced by option hedging in synthetic data

In order to evaluate how the presence of option hedging affects the probability of drift bursts being realized in our model, we implement the following simulation study. We assume that there is only one call option contract outstanding in the market, with strike $K = 1$, and fix $\beta = 3.4 \times 10^{-7}$, $h = 0.64$, $\nu = 0.03$ and $\alpha = 0$ (these parameters are motivated by empirical results presented later). On this baseline, we vary the remaining parameters of the option contract (again, motivated by empirical numbers) as

- $N \in \{8000, 10'000, 12'000, 14'000\}$
- $\tau \in \{1, 2, 3, 5\}$
- sign$(\Delta) \in \{-1, +1\}$

where τ is the time-to-maturity of the option in days. The parameters of the asset price process, i.e. S_0, are varied so that we cover the full range of the option delta

- $|\Delta| \in [0, 1]$,

which results in P=672 different parameterizations for the simulation. For each possible parametrization, we draw L=20’000 independent realizations of our model. We then use the Euler discretization scheme to approximate the continuous-time stochastic processes on intervals of 100 milliseconds.

As with any statistical test, the drift burst test will be subject to some type I error. Since we are interested in drift bursts that are induced into S through option hedging, we want to correct for the false positive rate of the test on the null model, i.e. the fundamental price F. For this purpose, for every parameter combination $k = 1, ..., P$ and realization $n = 1, ..., L$, we record the values of S_t and F_t. On each of these time series separately, we compute the drift burst test statistic on a regular one-second grid[3]. Additionally, we compute and record critical values for the test statistic that are adjusted for the autocorrelation of the test statistic on a given realization (see Appendix B of Christensen et al. (2020)). To account for the rolling calculation of the test statistic and avoid double counting of events, we allow for at most one drift burst to be established over any 30 second window (i.e. the bandwidth of μ) at which the test statistic attains a local extremum and exceeds a critical value corresponding to the 90% quantile of the reference distribution. This procedure will yield, for a given parameter combination k and realization n, sets of drift burst times and their associated test statistic for the affected price S, $\{(t_i^{n,k,S}, T_{t_i^{n,k,S}})\}_{i=1,\ldots,C^{n,k,S}}$, and the fundamental price F, $\{(t_i^{n,k,F}, T_{t_i^{n,k,F}})\}_{i=1,\ldots,C^{n,k,F}}$. Thus, $C^{n,k,X}$ is the total number of drift bursts registered for parameter combination k, realization n and price X.

As an estimator for the count of drift bursts that are the result of option hedging dynamics in our model and not due to false positives of the test procedure, we compute the difference in the mean number of drift bursts (across the L realizations) between S and F, i.e

$$\bar{B}_k = L^{-1} \sum_{n=1}^{L} \left(C^{n,k,S} - C^{n,k,F} \right). \tag{3.5}$$

Each $C^{n,k,F}$ hereby corrects for the type I error the test incurs on the same realization n, without the presence of option hedging[4]. To make estimates for $\bar{B}$ comparable, here and subsequently in the paper, we always express it as an intensity per trading day of 6.5 hours.

If $\bar{B}$ is positive, this means that there is a measurable increase in the intensity of instabilities arising from option hedging in our model. Thus for $\bar{B} > 0$ we speak of an explosive price process. Conversely, if $\bar{B}$ is negative, the number of drift bursts in the affected price is below the null level of the unaffected price. Thus in this case, option hedging has a stabilizing effect on the asset price according to our model. It is important to note that the sign of $\bar{B}$ does not make a statement about the sign of the price divergences, they can be both positive or negative, depending on option positioning, but their intensity is increased significantly.

Figure 3 presents estimates of $\bar{B}_k$ for the various simulation parameterizations. We first notice that 27% of the parameter combinations result in a statistically significant (see caption of Figure 3 for a description of the significance test applied) non-zero $\bar{B}_k$. So option positioning has a measurable effect on the stability of the asset price. Moreover, if $\bar{B}_k$ is statistically significantly different from zero, it is positive in around 48% of the parameter combinations, indicating that, as suggested by the elaborations in section 2.2, option hedging can result in both unusually explosive, as well as unusually stable price dynamics.

The drift burst statistic furthermore identifies instabilities in line with what the dynamics of our model would suggest: First, the effect of option hedging is more strongly visible for short positions. Second, for short positions, the stabilization ($\bar{B}_k < 0$)

---

[3] We use a 30 second bandwidth for the drift estimate, so that the kernel estimate for μ contains 300 observations, and base the volatility estimates on a 2.5-minute bandwidth as suggested by Christensen et al. (2020).

[4] We provide some sanity checks for the drift burst test in Appendix C (robustness to vanishing volatility) and Appendix D (size of the test and sensitivity to level of volatility).

is more pronounced than instability ($\bar{B}_k > 0$), as can be inferred from the left skewed nature of the distribution of $\bar{B}_k$. The dependence of $\bar{B}_k$ on the options parameters and the price are discussed in the following subsection.

3.5. Dependence on option characteristics

In order to assess the dependence of $\bar{B}_k$ to the characteristics of the option being hedged, we bin the simulations according to their parameter values, and look at differences in the corresponding values of $\bar{B}_k$, see Figure 4. For parameter combinations, where $\bar{B}_k$ is not statistically significantly different from zero, we set $\bar{B}_k = 0$. The conclusion from this analysis is that $\bar{B}_k$ is different from zero in situations where

1. the notional is high
2. the time to expiry of the option is short
3. the OMM holds a short position in the option

Looking at the dependence of $\bar{B}_k$ to the delta of the option (to aggregate different option positions we have to use their delta, not the price S, because we want to aggregate along the dimension where hedging demand materializes, irrespective of the particular set of parameters), we see that our model prediction is confirmed. For short positions, $\bar{B}_k$ is negative around the strike ($|\Delta| = 0.5$), indicating that the price is stable, while further away from the strike $\bar{B}_k$ is positive, indicating that the price is unstable. For long positions, $\bar{B}_k$ is rarely statistically different from zero, and if it is, it is very small in absolute terms compared to the case of short positions. Only if the notional gets very high and the time to expiry very low, we see that the price get slightly more unstable close to the strike, as expected.

3.6. Dependence on time scale of the drift estimate

A key parameter in computing the drift burst test statistic is the time scale of the kernel used to estimate the drift $\hat{\mu}_t$ in (3.1), because it also controls the time scale over which an identified singularity/divergence developed.

As Figure 5 shows, $\bar{B}_k$ has non-zero entries on all time scales, indicating that option positioning has an impact on the price stability across various time scales. Obviously, as we present $\bar{B}_k$ on a per hour basis, the level of $\bar{B}_k$ in absolute terms gets smaller for large scales. In addition we see that the dependence on the option delta is very similar on all time scales. Also, the dependence on the delta is weaker for larger time scales, which is a consequence of the simulation setup, as the price first has to diffuse over a time corresponding to the bandwidth over which σ^2 is estimated, which is 5 times as long as for μ (so e.g. around 4 hours for the largest scale), before a first bursts can even be registered.

In summary, option positioning can lead to price instability on various time scales in our setting, and specific points of instability are identified by the test statistic where our model predicts them. For empirical applications, it is important to note that any estimate of $\bar{B}$ must always be interpreted in conjunction with the time scale over which the test statistic was computed (i.e. the time scale of the relevant kernel estimates of μ and σ^2).

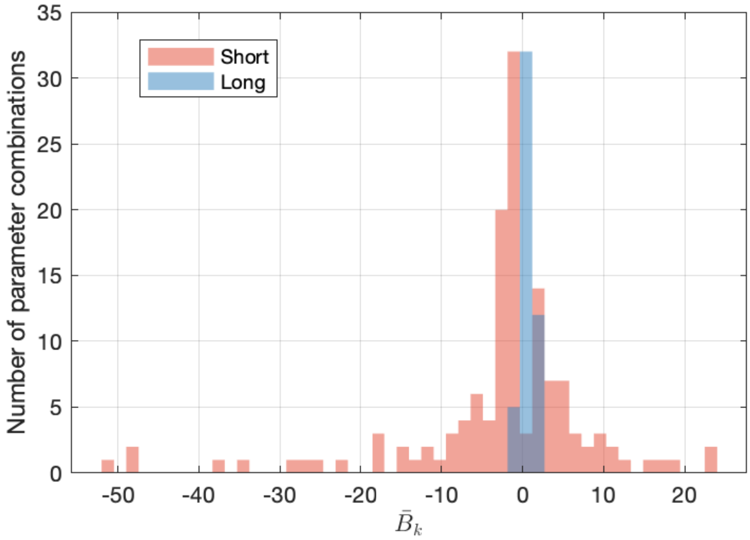


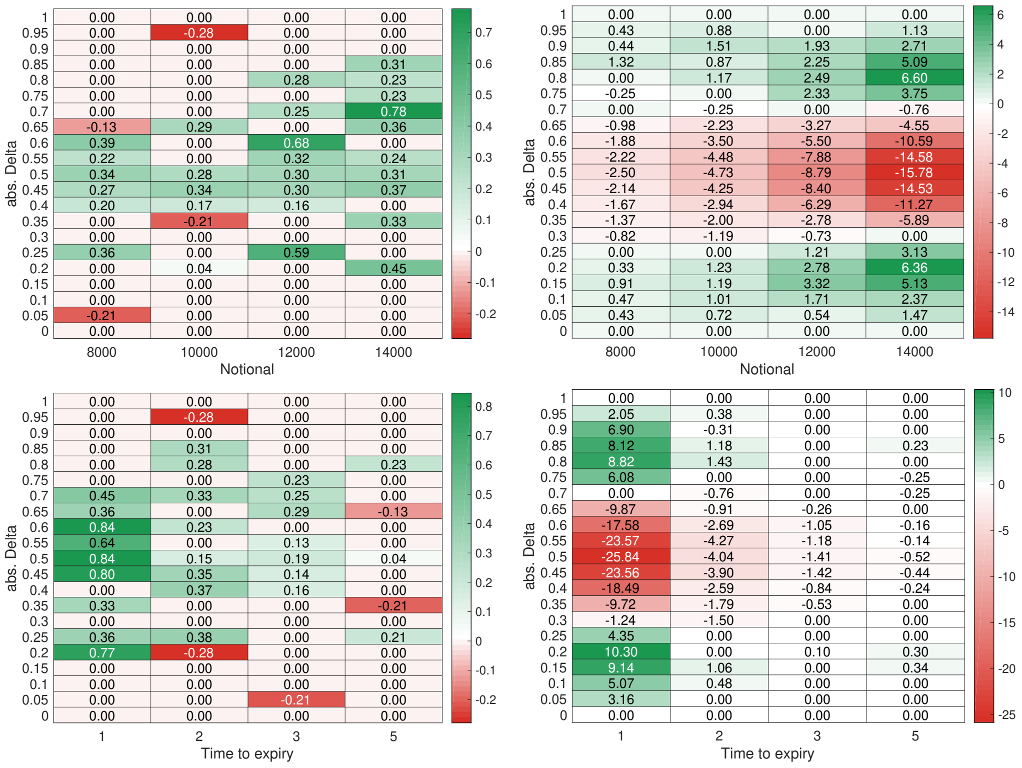
Figure 3: Shows a histogram of $\bar{B}_k$, the difference in the average number of drift bursts per trading day of 6.5 hours across 20'000 realizations of the model, cf. (3.5). Each observation corresponds to one parameter combination (see main text). The histogram only counts parameter combinations where the intensity of drift bursts in the affected price (across the $L = 20'000$ realizations) is statistically significantly different from the intensity of drift bursts in the unaffected price, i.e. when $\bar{B}_k$ is statistically significantly different from zero. This is determined by bootstrapping the 95% confidence intervals for $\bar{B}_k$. If this interval includes the value of zero, we exclude the parameter combination from the histogram. The red bars only count parameterizations where the OMM is short (138/672 significant parameter combinations), the blue ones when she is long (49/672 significant parameter combinations). All simulations were discretized on intervals of 0.1 seconds and the drift burst test applied with a kernel size of 30 seconds for μ.

3.7. Direction and magnitude of drift bursts

Last, we discuss the *direction* of the option induced drift bursts in Figure 6. We take a closer look at the simulation results presented in the second column of Figure 5 and for each delta compare the probability density functions (pdf) of the returns of the bursts stemming from S and the returns of the bursts stemming from F. The returns of the bursts are called μ_{kernel}, which are annualised and have units of %. We only display the deltas for which we registered a $\bar{B}_k$ that is statistically significantly different from zero.

The following observations can be made:

1. The direction of the bursts are in line with the dynamics of μ for a short call position displayed in Figure 1. For delta values below the strike (i.e. -0.25, -0.3, -0.35, -0.4, -0.45), the pdf mass is shifted towards large positive returns. This is in line with the positive deviation of μ for prices below the strike in Figure 1. For delta values above the strike (i.e. -0.55, -0.6, -0.65, -0.7, -0.75, -0.8), the pdf is shifted towards large negative returns. Again, this is in line with the negative deviation of μ for prices above the strike in Figure 1, confirming the predictions of our model.
2. Looking at the delta values for which the shift in probability mass is most extreme (-0.55 and -0.45), we see that



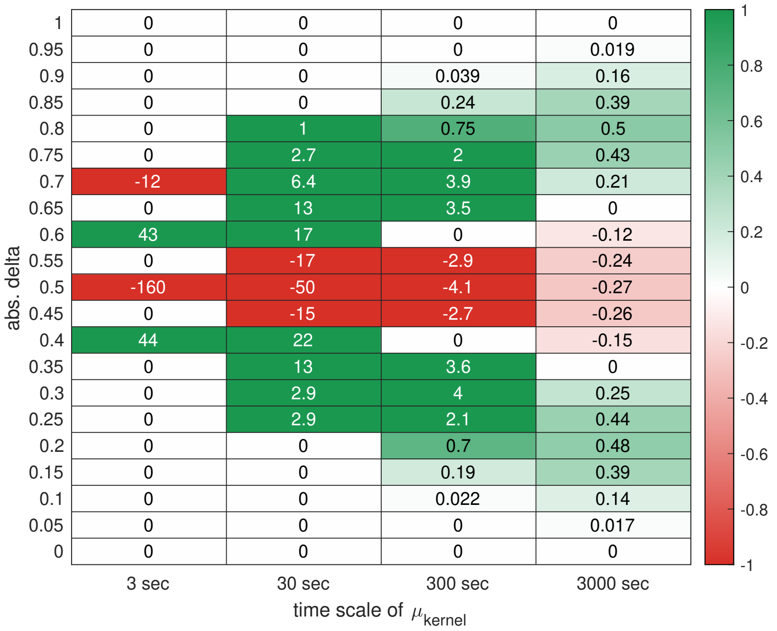
abs. Delta	8000	10000	12000	14000
1	0.00	0.00	0.00	0.00
0.95	0.00	-0.28	0.00	0.00
0.9	0.00	0.00	0.00	0.00
0.85	0.00	0.00	0.00	0.31
0.8	0.00	0.00	0.28	0.23
0.75	0.00	0.00	0.00	0.23
0.7	0.00	0.00	0.25	0.78
0.65	-0.13	0.29	0.00	0.36
0.6	0.39	0.00	0.68	0.00
0.55	0.22	0.00	0.32	0.24
0.5	0.34	0.28	0.30	0.31
0.45	0.27	0.34	0.30	0.37
0.4	0.20	0.17	0.16	0.00
0.35	0.00	-0.21	0.00	0.33
0.3	0.00	0.00	0.00	0.00
0.25	0.36	0.00	0.59	0.00
0.2	0.00	0.04	0.00	0.45
0.15	0.00	0.00	0.00	0.00
0.1	0.00	0.00	0.00	0.00
0.05	-0.21	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00

abs. Delta	8000	10000	12000	14000
1	0.00	0.00	0.00	0.00
0.95	0.43	0.88	0.00	1.13
0.9	0.44	1.51	1.93	2.71
0.85	1.32	0.87	2.25	5.09
0.8	0.00	1.17	2.49	6.60
0.75	-0.25	0.00	2.33	3.75
0.7	0.00	-0.25	0.00	-0.76
0.65	-0.98	-2.23	-3.27	-4.55
0.6	-1.88	-3.50	-5.50	-10.59
0.55	-2.22	-4.48	-7.88	-14.58
0.5	-2.50	-4.73	-8.79	-15.78
0.45	-2.14	-4.25	-8.40	-14.53
0.4	-1.67	-2.94	-6.29	-11.27
0.35	-1.37	-2.00	-2.78	-5.89
0.3	-0.82	-1.19	-0.73	0.00
0.25	0.00	0.00	1.21	3.13
0.2	0.33	1.23	2.78	6.36
0.15	0.91	1.19	3.32	5.13
0.1	0.47	1.01	1.71	2.37
0.05	0.43	0.72	0.54	1.47
0	0.00	0.00	0.00	0.00

abs. Delta	1	2	3	5
1	0.00	0.00	0.00	0.00
0.95	0.00	-0.28	0.00	0.00
0.9	0.00	0.00	0.00	0.00
0.85	0.00	0.31	0.00	0.00
0.8	0.00	0.28	0.00	0.23
0.75	0.00	0.00	0.23	0.00
0.7	0.45	0.33	0.25	0.00
0.65	0.36	0.00	0.29	-0.13
0.6	0.84	0.23	0.00	0.00
0.55	0.64	0.00	0.13	0.00
0.5	0.84	0.15	0.19	0.04
0.45	0.80	0.35	0.14	0.00
0.4	0.00	0.37	0.16	0.00
0.35	0.33	0.00	0.00	-0.21
0.3	0.00	0.00	0.00	0.00
0.25	0.36	0.38	0.00	0.21
0.2	0.77	-0.28	0.00	0.00
0.15	0.00	0.00	0.00	0.00
0.1	0.00	0.00	0.00	0.00
0.05	0.00	0.00	-0.21	0.00
0	0.00	0.00	0.00	0.00

abs. Delta	1	2	3	5
1	0.00	0.00	0.00	0.00
0.95	2.05	0.38	0.00	0.00
0.9	6.90	-0.31	0.00	0.00
0.85	8.12	1.18	0.00	0.23
0.8	8.82	1.43	0.00	0.00
0.75	6.08	0.00	0.00	-0.25
0.7	0.00	-0.76	0.00	-0.25
0.65	-9.87	-0.91	-0.26	0.00
0.6	-17.58	-2.69	-1.05	-0.16
0.55	-23.57	-4.27	-1.18	-0.14
0.5	-25.84	-4.04	-1.41	-0.52
0.45	-23.56	-3.90	-1.42	-0.44
0.4	-18.49	-2.59	-0.84	-0.24
0.35	-9.72	-1.79	-0.53	0.00
0.3	-1.24	-1.50	0.00	0.00
0.25	4.35	0.00	0.00	0.00
0.2	10.30	0.00	0.10	0.30
0.15	9.14	1.06	0.00	0.34
0.1	5.07	0.48	0.00	0.00
0.05	3.16	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00

Figure 4: Shows heatmaps of $\bar{B}_k$ across simulation parameterizations. All simulations were discretized on intervals of 0.1 seconds and the drift burst test applied with a kernel size of 30 seconds for μ. The parameterizations are binned according to three characteristics – the notional N, the time to expiry τ and the absolute value of the delta $|\Delta|$. The heatmaps then show the average within each bin of the corresponding $\bar{B}_k$. The left columns is for when the OMM is long the contract, the right column for when she is short.



abs. delta	3 sec	30 sec	300 sec	3000 sec
1	0	0	0	0
0.95	0	0	0	0.019
0.9	0	0	0.039	0.16
0.85	0	0	0.24	0.39
0.8	0	1	0.75	0.5
0.75	0	2.7	2	0.43
0.7	-12	6.4	3.9	0.21
0.65	0	13	3.5	0
0.6	43	17	0	-0.12
0.55	0	-17	-2.9	-0.24
0.5	-160	-50	-4.1	-0.27
0.45	0	-15	-2.7	-0.26
0.4	44	22	0	-0.15
0.35	0	13	3.6	0
0.3	0	2.9	4	0.25
0.25	0	2.9	2.1	0.44
0.2	0	0	0.7	0.48
0.15	0	0	0.19	0.39
0.1	0	0	0.022	0.14
0.05	0	0	0	0.017
0	0	0	0	0

Figure 5: Shows $\bar{B}_k$ per trading day of 6.5 hours, for one particular short option parameterization $\{N = 14'000, \tau = 1, \text{sign}(\Delta) = -1\}$, depending on the time scale of the kernel estimate for μ that enters the drift burst statistic, and the absolute value of the delta ($|\Delta|$) of the option. The simulations are performed 20'000 times for each combination. The time scale of the kernel estimate for σ^2 is chosen to be five times the one for μ and we scale the time intervals for the discretization in the simulations, so that one kernel estimate for μ always contains 300 observations of the process (both points are suggested by Christensen et al. (2020)). The drift burst test is applied every ten time steps of the simulation. Note that the colorbar is cut at minus one and plus one, in order to better display the stable and unstable areas.

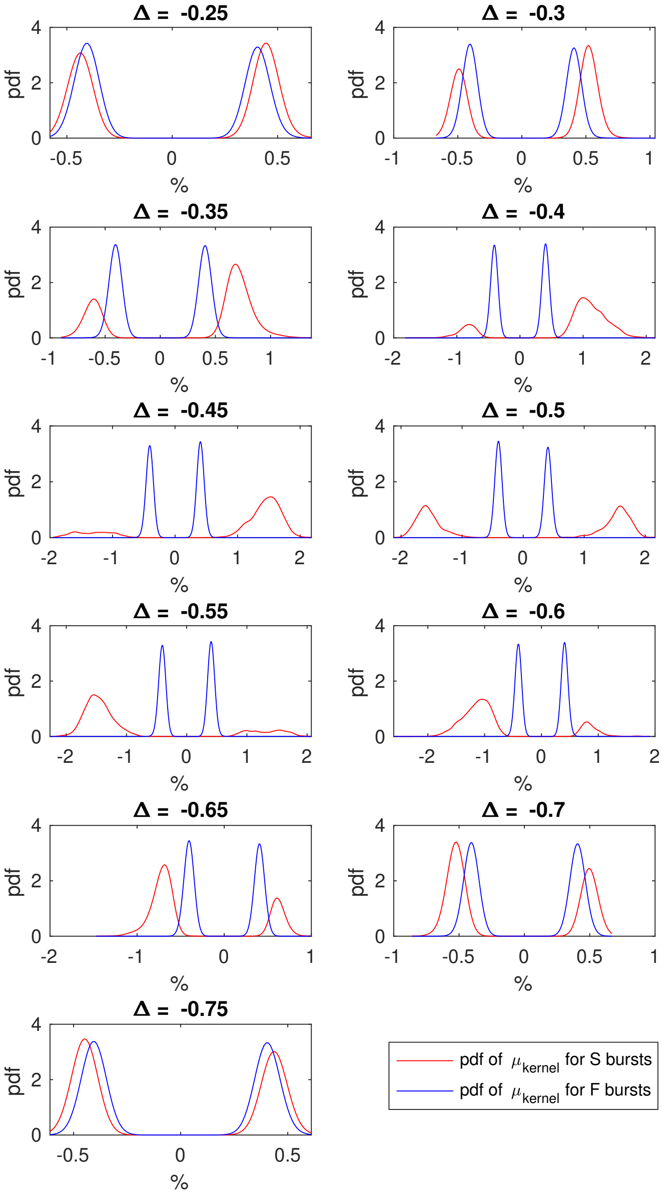


Figure 6: Shows the probability density function of μ_{kernel} (which is annualised and has units of %) for the bursts stemming from S (red lines) and the bursts stemming from F (blue lines), for a μ bandwidth of 30 seconds (second column of Figure 5) and for all delta which have a $\bar{B}_k$ that is statistically significantly different from zero. The probability density functions are calculated using a kernel fit with a bandwidth of 0.05.

the mean of the absolute value of the bursts in S is shifted to around 1.5% (annualised), compared to the mean for the bursts in F of 0%. Thus, the directional effect is in fact rather large and given that, e.g. the average annual return on the S&P 500 is of the order of 10%, economically significant.

4. Application to the GameStop stock price

As motivated in the introduction, we turn to the GameStop (GME) stock price for an empirical application of the proposed approach. GME experienced highly elevated volatility and erratic price moves in the first quarter of 2021, as displayed in Figure 7. As option hedging activity was a frequently cited explanation for the unusual price dynamics, we investigate this hypothesis through the lens of the approach from section 3.

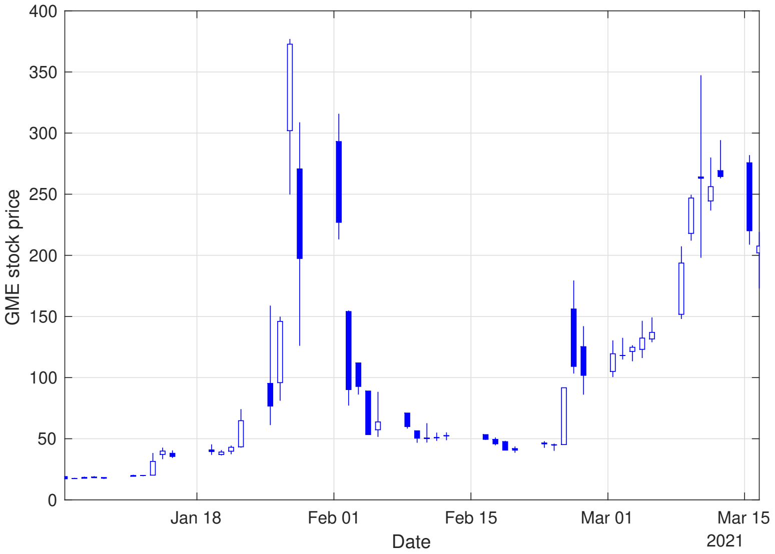


Figure 7: Candle daily price chart of the GME stock traded on the NYSE. If the closing price is greater than the opening price, the body of the candle (indicating the region between the open and close price) is unfilled, otherwise it is filled.

4.1. Model calibration for GME

In a first step, we calibrate our model to price and option position data for GME. A first quantity to estimate is the price impact coefficient β. As this is not central to the ideas presented in this paper, we defer details of the calibration of β to Appendix A. We estimate the price impact to be $\beta \approx 0.0034$ bps per share.

4.1.1. Estimation of fundamental volatility ν and net hedge ratio h

The next step is to obtain estimates for the fundamental volatility ν and the net hedge ratio h from relation (2.10). To do this, we estimate the nonlinear regression model

$$\sigma_t = \frac{\nu}{1 + \beta h S_t \sum_j N_{t,j} \gamma_{t,j}} \tag{4.1}$$

using daily volatilities σ_t, the underlying price S_t and the total gamma exposure $\Gamma_t := \sum_j N_{t,j}\gamma_{t,j}$. The total gamma exposure is the composite of all outstanding positions j, each of which is specified by a combination of $\{K_j, \tau_j, c_j\}$, where K_j is the strike price, τ_j the time-to-maturity and c_j indicating whether the contract is a call or a put. $N_{t,j}$ is the signed[5] open interest of the OMM behind position j. Open interest hereby represents the total number of contracts, either long or short, that have been entered into and not yet offset by delivery for one position j, see CBOE (2021) for details. Each open position has a buyer and seller, but for calculation of open interest, only one side of the contract is counted. $\gamma_{t,j}$ is the gamma of the option behind an individual position j.

We use open interest data of Trade Alert[6] from January 4, 2021 to March 15, 2021 to calculate $N_{t,j}$. This data set covers the total market for GME options and includes the information on which side of the market the open interest was traded (bid or offer price). Using this information, we compute the signed open interest of the OMM behind position j as $N_{t,j} = OI^b_{t,j} - OI^o_{t,j}$, where $OI^b_{t,j}$ is the open interest traded on the bid price and $OI^o_{t,j}$ is the open interest traded on the offer price. The assumption here is that the OMM is always liquidity provider, i.e. she is willing to buy the stock at the bid price, and willing to sell the stock at a slightly higher offer price, earning the spread between the two as a compensation for the inventory risk arising from liquidity takers matching with these prices. Under this assumption, the OMM is long if the trade occurred on the bid and short if the trade occurred on the offer.

The price data for S_t is taken from Bloomberg[7] and since we use option positioning at the start of the trading day, we use the opening price for S_t when computing the gamma of each option.

Finally, to compute $\gamma_{t,j}$, we rely on the Black-Scholes model, i.e.

$$\begin{aligned} d_1 &= \frac{1}{\sigma_{t,j}\sqrt{\tau_{t,j}}} \cdot \left[\log(\frac{S_t}{K_j}) + (r_{t,j} - q_t + \frac{\sigma^2_{t,j}}{2})\tau_{t,j})\right] \\ d_2 &= d_1 - \sigma_{t,j}\sqrt{\tau_{t,j}} \\ \gamma_{t,j} &= e^{-q_t\tau_{t,j}}\frac{\phi(d_1)}{S_t\sigma_{t,j}\sqrt{\tau_{t,j}}}, \end{aligned} \tag{4.2}$$

where ϕ is the standard normal probability density function and we set the dividend yield $q_t \equiv 0$ and also fix the risk-free interest rate to $r_{t,j} \equiv 0$. Calculation of the $\gamma_{t,j}$ requires calculation of the implied volatility of each position j. To calculate the implied volatility $\sigma_{t,j}$ behind each position j, we use daily volatility surface data from the CBOE data shop[8] and linearly interpolate it to the corresponding strike K_j and time-to-expiry $\tau_{t,j}$ of position j. Having computed $\gamma_{t,j}$ for each position, we can calculate the gamma exposure of the OMM as $\Gamma_t = \sum_j N_{t,j}\gamma_{t,j}$.

Using Γ_t, the spot price S_t and the at-the-money implied volatility as a proxy for σ_t, finally allows estimation of ν and h by performing a nonlinear regression according to (4.1). The results of this procedure are presented in Figure 8 and Table 1.

---

[5] The sign is negative if the position is a short position and positive if the position is a long position.

[6] Top Open Interest product from Trade Alert LLC: https://www.trade-alert.com

[7] We obtain minute-by-minute sampled prices.

[8] CBOE data shop: https://datashop.cboe.com/end-of-day-volatility-skew

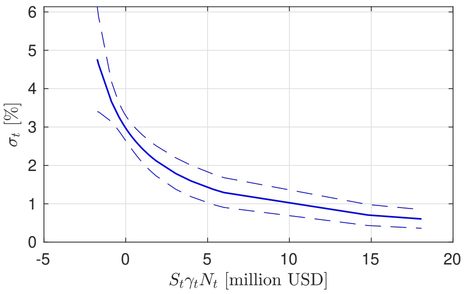


Figure 8: Estimation of equation (2.10) from the nonlinear regression (4.1). The dashed lines indicate nonsimultaneous 95% confidence bounds for the fitted mean values.

$\sigma_t = \nu/(1 + h \cdot \beta S_t \Gamma_t) + e_t$	
ν (%)	2.97 ***
(t-statistic)	(18.3)
h (%)	63.6 ***
(t-statistic)	(4.3)
F-test vs. const. model	172 ***

Table 1: Estimates for ν and h. The *** denote significance at the 1% level.

Our estimates suggest that the model is highly significant vs. a constant model and can explain a large part of the variance in the data ($R^2 = 27\%$). Also, the estimates for ν and h are highly significant and in an economically sensible range. We estimate ν to be around 3% and h around 64%. The fact that $h < 1$ makes sense given that we can expect the OMT to also hedge her position to some extent and that the OMM also has alternatives to hedging than directly buying the underlying.

4.1.2. Estimation of the fundamental drift α

Using the estimated ν and h now allows us to estimate α from the relation (2.12). To this end, we again translate this into a nonlinear regression model

$$\mu_t = \frac{\alpha - \frac{1}{2}\nu^2\left(1 - \frac{1-\beta h S_t^2 \sum_j \psi_{t,j} N_{t,j}}{\left(1+\beta h S_t \sum_j \gamma_{t,j} N_{t,j}\right)^2}\right) - \beta h \sum_j \chi_{t,j} N_{t,j}}{1 + \beta h S_t \sum_j \gamma_{t,j} N_{t,j}}, \tag{4.3}$$

where similar to before, $\gamma_{t,j}$, $\chi_{t,j}$ and $\psi_{t,j}$ are calculated for each position j on day t using equations (4.4) and (4.5) respectively as

$$\chi_{t,j} = q_t e^{-q_t \tau_{t,j}} \Phi(d_1) - e^{-q_t \tau_{t,j}} \phi(d_1) \frac{2(r_{t,j} - q_t)\tau_{t,j} - d_2 \sigma_{t,j}\sqrt{\tau_{t,j}}}{2\tau_{t,j}\sigma_{t,j}\sqrt{\tau_{t,j}}}. \tag{4.4}$$

$$\psi_{t,j} = -\frac{\gamma_{t,j}}{S_t}\left(\frac{d_1}{\sigma_{t,j}\sqrt{\tau_{t,j}}} + 1\right) \tag{4.5}$$

Φ denotes the standard normal cumulative distribution function. The corresponding overall charm (speed) exposure is then given by the sum over the product of $N_{t,j}$ and $\chi_{t,j}$ ($\psi_{t,j}$), i.e. $\mathrm{X}_t := \sum_j N_{t,j}\chi_{t,j}$ ($\Psi_t := \sum_j N_{t,j}\psi_{t,j}$).

Estimating this model using realized log-price returns for the first four hours of each trading day t as μ_t results in no significance for the model (F-statistc vs. constant model of 0.37) and its parameters (p-value for α of 0.85)[9]. Thus we conclude that $\alpha = 0$, i.e. there is no fundamental drift.

---

[9] As a robustness check we have also estimated a model where the fundamental drift is allowed to have a monthly seasonality. Also in this extended specification, neither the constant, nor the respective monthly dummies are significant.

In summary, the calibrated model parameters are

$$\beta = 0.0034 \text{ bps per share}, \qquad \nu = 2.97\%, \qquad h = 63.6\%, \qquad \alpha = 0\%. \tag{4.6}$$

Statistics of the main variables used in the calibration are additionally presented in Table B.3.

4.2. Option-induced price instability for GME

As a first sanity check, we evaluate whether the model-implied drift from option positions at the start of a trading day is correlated with the subsequent daily returns. The analysis deferred to Appendix E shows that there is a positive and significant correlation between the two, confirming the basic adequacy of our model's predictions. Turning to price instabilities, a principled way to assess the potential for drift bursts to occur in the GME stock price based on the approach so far in this paper, would be to take the open option positions at the start of each trading day and perform simulations of S_t, conditional on these positions and the day's opening price S_0 and then infer the option-induced drift burst intensity similar to how this was done in section 3. In order to alleviate the computational burden associated with this procedure, we instead resort to an approximation scheme based on the synthetic data already generated in section 3. Since these simulations were based on parameter values that correspond to the ones we have calibrated empirically, they can be used to construct a function[10]

$$\mathbb{E}[\bar{B}_j |\, S_t] = F(N_j, \tau_j, \Delta_j, \text{sign}(\Delta_j)), \tag{4.7}$$

which for a given option position j with parameters $\{N_j,\, \tau_j,\, \Delta_j,\, \text{sign}(\Delta_j)\}$ and a prevalent spot price S_t approximates the expected number of drift bursts (on a time scale of 30 seconds, over a time horizon of 6.5 hours) induced by hedging this position[11].

Our specific empirical procedure consists in taking the open option positions and the opening price on a given day, and then fitting this function for a range of S_t around the opening price of the particular day. The overall predicted option-induced drift burst intensity at each spot price S_t is then just the sum of the intensities predicted for each individual position

$$\mathbb{E}[\bar{B}|S_t] = \sum_j \mathbb{E}[\bar{B}_j|S_t] = \sum_j F(N_j, \tau_j, \Delta_j, \text{sign}(\Delta_j)). \tag{4.8}$$

In Figure 9, we evaluate this prediction using the open option positions of a given day, and minute-by-minute prices S_t. The predictions are made for each minute of each trading day and summarized as violin plots per trading day in the upper panel of the Figure.

According to this analysis, the predicted option-induced drift burst intensity is on average very close to zero for most trading days. We however see that there are a number of days where the predicted drift burst intensity is clearly positively and/or

---

[10] We use a simple nonparametric interpolation on a grid of the parameter values.

[11] To validate the approximation, we have performed the full daily simulation mentioned in the beginning for some exemplary days and checked that the approximation and full simulation yield comparable results.

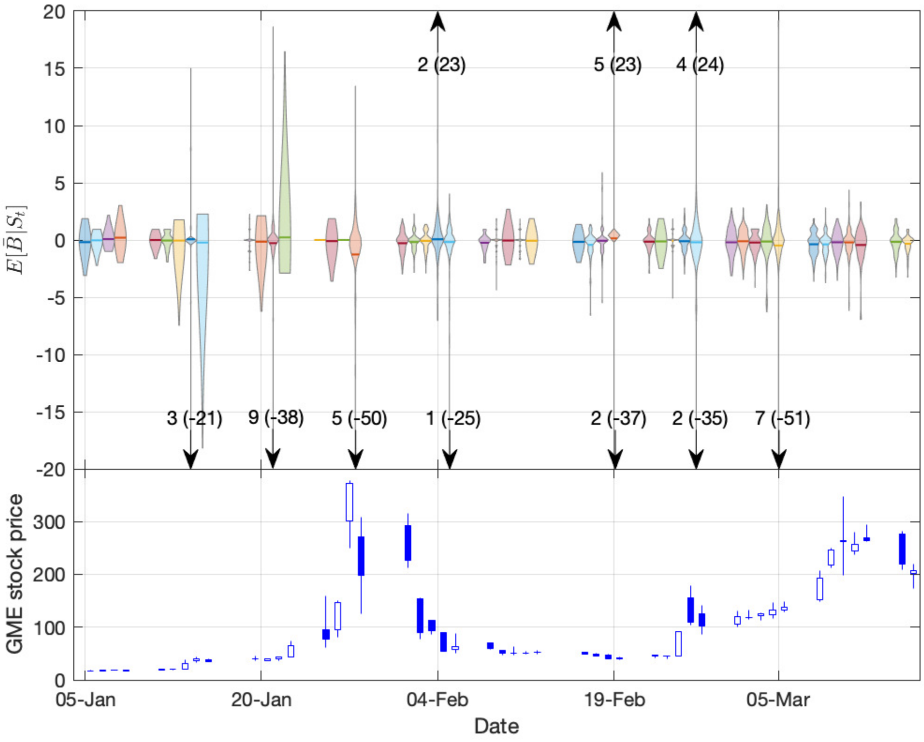


Figure 9: Shows the total option-induced drift burst intensity $\mathbb{E}[\bar{B}|S_t]$ (upper panel) for the GME stock price (lower panel, in the form of a candle chart). $\mathbb{E}[\bar{B}|S_t]$ is defined by expression (4.8) and $\bar{B}_j$ is given by expression (3.5). Positive values indicate instability in the price due to hedging, negative values suggest a stabilizing effect of option hedging. The predictions are made for each minute of each trading day and summarized as violin plots per trading day. The kernel density estimates used to construct the violin plots span the entire range of the predicted values. The y-axis is limited to the range $[-20, 20]$. For days with observations outside this range, the arrows indicate the number of observations truncated, with the value of the maximum observation in parentheses.

negatively elevated. These days appear to coincide with periods of erratic price moves, whereas e.g. during the first week of January, or the week of February 8th, where prices evolved smoothly, the predicted drift burst intensity also appears rather stable.

Nevertheless, linking the predictions of our model to the observed massive price moves is difficult. It is important to realise that $\mathbb{E}[\bar{B}|\ S_t]$ predicts the probability/intensity of drift bursts at time scales of 30 seconds occurring over the subsequent 6.5 hours. Our approach makes no prediction of the exact timing of the burst. Furthermore, assessing this prediction by comparing it to drift burst estimates on empirical prices with our granular minute-by-minute price data is not possible, as higher resolution data would be needed to determine empirical drift bursts at 30 second time scales[12].

Instead, we present a simple correlation analysis and focus on whether we actually see stable or unstable price behaviour after points with significant (positive or negative) spikes in $\bar{B}_k$. To do this, we select times where the predicted drift burst intensity is above an absolute level of 10 and then evaluate how the maximum absolute return of the subsequent five minutes $r_{max,t} := \max\{|r_{t+1}|, |r_{t+2}|, |r_{t+3}|, |r_{t+4}|, |r_{t+5}|\}$, ranks among the absolute (minute) returns of the first two hours of the day $r_{day} := \{|r_0|, ..., |r_{2h}|\}$. So every time we find a value $\mathbb{E}[\bar{B}_j | \, S_t] > 10$ or $\mathbb{E}[\bar{B}_j | \, S_t] < -10$ we look at the next five returns, take their maximum absolute value, $r_{max,t}$, and calculate the percentile rank $PR_{r_{max,t}} := PR(r_{max,t}, r_{day})$ against all absolute minute returns of the first two hours of the trading day, r_{day} (we choose the first two hours to have a conservative benchmark, as the returns in the morning are structurally more volatile than the returns of the afternoon).

The hypothesis is that a strong negative option-induced drift burst intensity $\mathbb{E}[\bar{B}_j | \, S_t] << -1$ is indicative of a stabilizing price effect, so $PR_{r_{max,t}}$ should be low. In contrast, a positive option-induced drift burst intensity $\mathbb{E}[\bar{B}_j | \, S_t] >> +1$ is indicative of an explosive price, so $PR_{r_{max,t}}$ should be high. We then calculate the probability density function (pdf) of the percentile ranks of $PR_{r_{max,t}}$ given $\mathbb{E}[\bar{B}_j | \, S_t] > 10$ as well as the pdf of the percentile ranks of $PR_{r_{max,t}}$ given $\mathbb{E}[\bar{B}_j | \, S_t] < -10$ and compare them to a benchmark pdf which consist of the percentile rank from 10'000 random draws of $PR_{r_{max,random}}$ (the percentile ranks of randomly selected times within the first two hours of the day). This is displayed in Figure 10. We clearly see that the pdf for the positive option-induced drift burst intensities is shifted towards high percentile ranks, compared to the pdf of the random benchmark, which we take as indication that the option positioning at those points indeed leads to explosive price moves. On the other hand, the expected negative shift in the pdf of negative predicted drift bursts is absent. One potential explanation for this could be due to the stabilizing effect of option hedging being most pronounced very close to the strike (which is where the predicted drift burst intensity will be most negative). With closer distance to the strike, also the probability of open option positions to exercise is elevated. If options are exercised, the OMM will not need to hedge the related exposure anymore and thus the stabilizing effect from hedging might not materialize.

5. Conclusion

We have provided a framework to study the impact of option hedging on the stability of the underlying asset price. Our model suggests that the friction introduced by option hedging not only affects the volatility of the asset price, but also its expected return. In particular, we found evidence that the presence of feedback effects from option hedging can increase or decrease the intensity of so-called drift bursts – local divergences in the drift – depending on the option's parameters and its delta. Our model predicts that the effect of option hedging is asymmetric in the sense that a short position of the OMM has a significant impact on the price stability, whereas a long position only impacts the prices slightly. Since the OMM being short is the predominant regime in practice, this suggests that empirically, we can expect the effect of option hedging to be strong. Investigating the option positioning during the erratic developments in the GME stock price during the first quarter of 2021

---

[12] Note that the drift burst method recommends to use a rolling window of 1500 data points for the estimate of σ, which would correspond to 25 hours when using minute-by-minute data. This significantly exceeds the duration of a 6.5 hour trading day, over which we make our predictions.

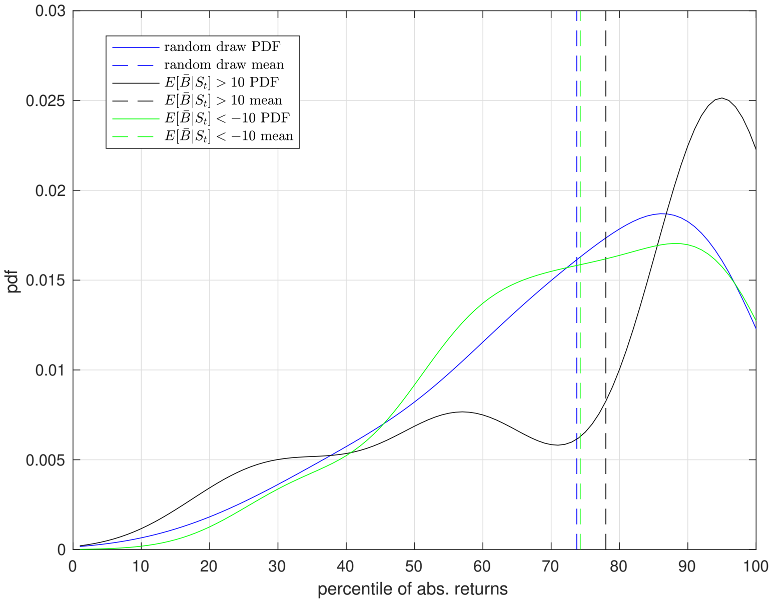


Figure 10: PDF estimates for the percentile rank (against the distribution of returns in the first two hours of the trading day) of the maximum absolute returns within 5 minutes after a predicted drift burst intensity above a level of +10 (black) and below -10 (green). Also shown as a reference is a distribution obtained when using randomly sampled times within the first two hours of the day (blue).

using our framework, we found that option positioning on some days indeed predicts an increased intensity of option-induced drift bursts, but due to limitations in our data, we can only establish a circumstantial link between our model predictions and the actually observed price dynamics.

Acknowledgements

The authors thank Kim Christensen for very helpful comments on the application of his paper Christensen et al. (2020).

References

B. Anderegg, D. Sornette, and F. Ulmann. Quantification of feedback effects in FX options markets. *SNB Working Paper Series*, 2019-03:1–27, 2019.

I. Ben-David and D. Hirshleifer. Are Investors Really Reluctant to Realize Their Losses? Trading Responses to Past Returns and the Disposition Effect. *The Review of Financial Studies*, 25(8):2485–2532, 08 2012. ISSN 0893-9454. doi: 10.1093/rfs/hhs077. URL https://doi.org/10.1093/rfs/hhs077.

I. Ben-David, F. Franzoni, and R. Moussawi. Do ETFs Increase Volatility? *The Journal of Finance*, 73(6):2471–2535, 2018. doi: https://doi.org/10.1111/jofi.12727. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12727.

J. P. Bouchaud. Price Impact. In *Encyclopedia of Quantitative Finance*. Wiley, 2010.

M. K. Brunnermeier and L. H. Pedersen. Market Liquidity and Funding Liquidity. *The Review of Financial Studies*, 22(6):2201–2238, 11 2008. ISSN 0893-9454. doi: 10.1093/rfs/hhn098. URL https://doi.org/10.1093/rfs/hhn098.

CBOE. CBOE glossary. https://www.cmegroup.com/trading/about-volume.html, 2021. Accessed: 2021-04-22.

A. Cherny and H.-J. Engelbert. *Singular Stochastic Differential Equations*, volume 1858 of *Lecture Notes in Mathematics*. Springer Verlag, Heidelberg, 2005.

K. Christensen, R. Oomen, and R. Reno. The drift burst hypothesis. *Journal of Econometrics, in press*, 2020. ISSN 0304–4076. doi: https://doi.org/10.1016/j.jeconom.2020.11.004. URL https://www.sciencedirect.com/science/article/pii/S0304407620303912.

F. Corsi and D. Sornette. Follow the money: The monetary roots of bubbles and crashes. *International Review of Financial Analysis*, 32:47–59, 2014.

J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud. A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15(7):1109–1121, 2015.

D. Easley, M. O'Hara, and P. Srinivas. Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2):431–465, 1998. doi: https://doi.org/10.1111/0022-1082.194060. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.194060.

FT. GameStop and BlackBerry shares soar on amateur traders' fervour. https://on.ft.com/3oipqrG, 2021. Accessed: 2021-04-23.

S. Gluzman and D. Sornette. Log-periodic route to fractal functions. *Phys. Rev. E*, 65:036142, Mar 2002. doi: 10.1103/PhysRevE.65.036142. URL https://link.aps.org/doi/10.1103/PhysRevE.65.036142.

A. Johansen, O. Ledoit, and D. Sornette. Predicting financial crashes using discrete scale invariance. *Journal of Risk*, 1(4):5–32, 1999.

A. Johansen, O. Ledoit, and D. Sornette. Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(2):219–255, 2000.

A. S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913210.

L. Li and D. Sornette. Diagnostics of rational expectation financial bubbles with stochastic mean-reverting termination times. *The European Journal of Finance*, 19(5-6):344–365, 2013.

L. Li, R. Ren, and D. Sornette. The volatility-confined lppl model: A consistent model of 'explosive' financial bubbles with mean-reversing residuals. *International Review of Financial Analysis*, 33:210–225, 2014.

L. Li, M. Schatz, and D. Sornette. A simple mechanism for financial bubbles: time-varying momentum horizon. *Quantitative Finance*, 19(6):937–959, 2019.

S. X. Ni, N. D. Pearson, A. M. Poteshman, and J. S. White. Does Option Trading Have a Pervasive Impact on Underlying Stock Prices? SSRN, 2020.

M. Schatz and D. Sornette. Inefficient bubbles and efficient drawdowns in financial markets. *International Journal of Theoretical and Applied Finance*, 23(7):2050047, 2020.

P. Shum Nolan, W. Hejazi, E. Haryanto, and A. Rodier. Intraday Share Price Volatility and Leveraged ETF Rebalancing. SSRN, 2015.

K. R. Sircar and G. Papanicolaou. General black-scholes models accounting for increased market volatility from hedging strategies. *Applied Mathematical Finance*, 5(1):45–82, 1998. doi: 10.1080/135048698334727. URL `https://doi.org/10.1080/135048698334727`.

D. Sornette. Critical Phenomena in Natural Sciences, Chaos, Fractals, Self-organization and Disorder: Concepts and Tools. Springer Series in Synergetics. Springer Verlag, Heidelberg, 2nd edition, 2004.

D. Sornette and J. V. Andersen. A nonlinear super-exponential rational model of speculative financial bubbles. *International Journal of Modern Physics C*, 13(02):171–187, 2002. doi: 10.1142/S0129183102003085.

Appendix A. Estimation of the market impact coefficient β

In order to have an estimate for the linear, permanent market impact per share, β, which is introduced in equation (2.6), we assume following Bouchaud (2010) that spot market liquidity providers quote bid and offer prices, and adopt the basic strategy of requiring a compensation for the adverse impact of the option-hedging trader hitting one of their quotes. Because the liquidity providers do not know which of their quotes are taken at any given point in time, they will post bid and offer prices, which ensure no ex-post regret for the liquidity provider. Thus the bid and offer quotes are posted around the mid-price m_t as

$$b_t = m_t - \beta v^{\phi}(1 + x_t) \quad \text{and} \quad o_t = m_t + \beta v^{\phi}(1 - x_t), \tag{A.1}$$

where x_t is the conditional forecast of the sign (buy or sell) of the next trade, v its random volume and $\phi = 1$ since we postulate linear volume-dependence. This implies that the bid-offer spread s_t is given as

$$s_t = o_t - b_t = 2\beta v. \tag{A.2}$$

Relation (A.2) suggests that β can be estimated from a simple linear regression of the spread on the volume of liquidity taking trades. For this purpose we use GME trade and price data from the NYSE exchange[13] of January 2021. Summary statistics of the data are presented in Table A.2.

Using this approach, we find

$$\beta \approx 0.0034 \text{ bps per share.} \tag{A.3}$$

We have validated the magnitude of β using alternative approaches and data sets.

	mean	std	skewness	kurtosis	min	5th quantile	25th quantile	median	75th quantile	95th quantile	max	N
spreads [bps]	30.5	34.3	0.2	91.1	-1060.4	2.6	10.3	21.1	39.7	90.9	839.2	178001
shares per trade	249.5	961.2	43.7	3188.9	100	100	100	100	200	634.4	102500	178001

Table A.2: Summary statistics of the spread and trade size (shares per trade) data used to estimate the price impact coefficient β.

Appendix B. Summary statistics of the variables used in the model calibration for GME

Appendix C. Robustness against zero-volatility estimates

One concern could be that the drift burst test statistic is spuriously identified to be large at times where the kernel estimator of the volatility is very small[14]. As we can see from Figure C.11, the times where the test statistic identifies a burst is characterized by drift estimates that are in the tail of the distribution, whereas the corresponding volatility estimates are

---

[13] CBOE data shop: https://datashop.cboe.com/equity-trades

[14] This effect is in fact at the origin of the Student-t distribution, see Sornette (2004), section 14.2.1.

	mean	std	skewness	kurtosis	min	5th quantile	25th quantile	median	75th quantile	95th quantile	max	N
gamma exposure	46000	140000	4.8	28	-14000	-6400	-55	1400	11000	300000	910000	48
charm exposure	-21000	240000	-2.6	17	-1200000	-280000	-38000	-1.4e-33	26000	370000	590000	48
speed exposure	67000	1000000	2.2	13	-2700000	-1500000	-8900	1.4e-34	13000	1700000	4300000	48
ATM vol [%]	2.7	1.2	0.15	2.3	0	1.2	1.8	2.6	3.6	4.9	5.3	48
daily mu realized [%]	-0.46	15	0.038	4.7	-40	-31	-6.8	-0.33	6.6	22	46	48
option implied mu (ann.) [%]	-1.4	7	-3.9	20	-39	-12	-1.6	-3.1e-38	1.1	3.7	5.9	48

Table B.3: Summary statistics for the variables used in the model calibration for GME.

concentrated in the center of the distribution. There is thus no evidence for false positives due to locally vanishing volatility estimates.

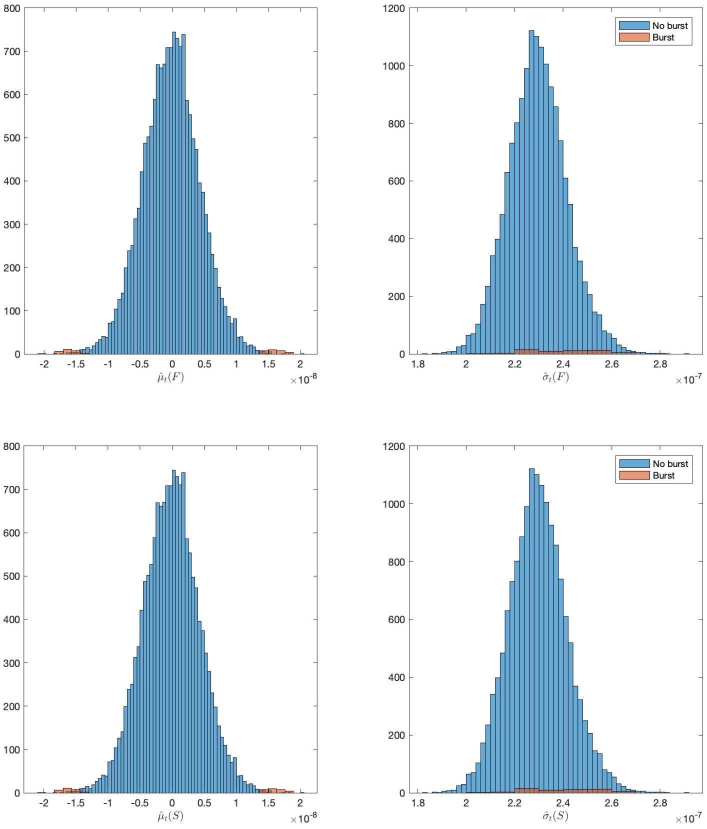


Figure C.11: Histogram of test times, for when no burst was identified (blue) and for times where a drift burst was identified at the 90% confidence level. Top: for the fundamental price F, bottom: for the affected price S.

Appendix D. Size of the drift burst test

In Table D.4, we present how often the drift burst test leads to the rejection of the null of no drift burst on a simple Brownian motion (our fundamental price F), for different values of the fundamental volatility ν and different confidence levels. We can see that the test is conservative compared to the nominal level[15] and exhibits no sensitivity to the level of volatility.

ν	90%	95%	99%
0.03	0.055	0.029	0.005
0.06	0.052	0.017	0.001
0.09	0.048	0.025	0.002
0.12	0.062	0.032	0.002
0.15	0.059	0.025	0.003
0.18	0.054	0.022	0.003
0.21	0.06	0.023	0.007
0.24	0.05	0.021	0.004
0.27	0.062	0.024	0.004
0.3	0.056	0.025	0.003

Table D.4: Size of the drift burst test on 1000 realizations per value of the fundamental volatility ν. The simulated process is driftless ($\alpha = 0$). The rejection rate is computed across Monte Carlo realizations.

Appendix E. Correlation between hedging-implied drift and daily realized returns

Using the estimated parameters from (4.6) also allows us to compute the drift implied by hedging $\mu_{impl,t}$ for day t as

$$\mu_{impl,t} = \frac{\alpha - \frac{1}{2}\nu^2 \left(1 - \frac{1-\beta h S_t^2 \Psi_t}{(1+\beta h S_t \Gamma_t)^2}\right) - \beta h \mathrm{X}_t}{1 + \beta h S_t \Gamma_t}. \tag{E.1}$$

In particular, we are interested in whether the sign of the realized log-price return for day t is actually correlated with the drift implied by hedging $\mu_{impl,t}$ for day t. Estimating this using the realized log-price returns for the first four hours of trading day t as μ_t, we find a statistically significant (p-value of 0.01) positive correlation of 0.66, which we take as additional support for the utility of our model[16].

---

[15] Christensen et al. (2020) shows that the test statistic becomes less conservative with larger bandwidths of the kernels used to compute the drift and volatility estimates. Our levels of the rejection rates are comparable to what they report on a more complicated process with jumps.

[16] We only consider absolute $\mu_{impl,t}$ values above 2.5%, as lower $\mu_{impl,t}$ values indicate no directional impact on the price from the option hedging, and thus are not relevant for a correlation analysis of the return sign.

3 Price Dynamics under a Circuit Breaker

In early 2016, a new circuit breaker (CB) rule was implemented on the CSI 300 financial index which replicates the top 300 stocks traded in the Shanghai and Shenzhen stock exchanges. The exchange specified that, if the price drops below (or rises above) a pre-specified threshold within a trading session, then trading is halted for a pre-specified amount of time, until regular trading resumes. The rational behind a CB is to give traders time to calm down, interact with their counterparts and rethink their trading decision. On January 4th 2016, the first trading day of the year, the price movement in the CSI 300 was already sufficient to trigger the CB twice. Then, in the very same week on January 7th, the mechanism was triggered twice again. The Chinese authorities subsequently abandoned the CB mechanism from their index.

Thereby, such measures to intervene with the normal trading process during extreme times, with the hope to stabilize prices and maintain proper functioning, have been around for years, and come in various forms. They range from market-wide trading halts, price limits on the whole market or individual assets, to limits on order flow, positions or margins and even transaction taxes. These measures have grown to be an important part of the broad market architecture. Yet, their effect still is rather unclear (**Grossman**, 1990), as can e.g. be inferred from China's recent struggles discussed above. This is a problem, especially as in today's highly automated markets, in which turbulences stemming from erroneous algorithms or fake news seem to occur with growing frequency, such stabilising measures become more and more important.

The most prominent of these measures is the price-based market-wide CB (like the one introduced in the CSI 300 in China) which was first advocated by the Brady Commission (**Presidential Task Force on Market Mechanisms**, 1988) after the Black Monday of 1987 and subsequently implemented in 1988. All stocks and related derivatives are temporarily halted when a market index drops by a certain amount. Following this lead, price-based market-wide CBs have been widely adopted by exchanges around the globe.

And naturally, this adoption lead to a vast body of academic research on CBs by both theorists and empiricists in the past 30 years. However, the regulatory nightmare of China in 2016, described above, revived the discussion around the effects of CBs and gave it new relevance. And thereby especially the question of how to *design* a CB came to the center of the attention, as apparently the implementation of the price-based CB on the CSI 300 lead to strong adverse effects. The paper presented in this chapter contributes to this CB design question. While the paper discusses relevant literature at the corresponding places in detail, the following paragraphs give a brief overview of the most important academic work done on the topic of CBs and embeds our contribution on a high level.

Early proponents of CBs are **Greenwald** and **Stein** (1991) and **Lee** et al. (1994). **Greenwald** and **Stein** (1991) argues that, in a market with limited participation and thus execution risk, CBs can help to better synchronize trading for market participants and improve the efficiency of allocations. Similar, **Lee** et al. (1994) argue that by lowering informational asymmetries between traders, halts could permit the orderly emergence of a new consensus price. And in line with that, **Hautsch** and **Horvath** (2019) recently showed that the mere presence of CBs limits adverse selection risk for liquidity providers and that CBs enhance price discovery during the break.

On the other side stands the seminal work of **Subrahmanyam** (1994). Focused on price-based CBs, he presents an inter-temporal equilibrium model with three agents, two types of traders with asymmetric information and a risk-averse market maker. He shows that price-based CBs can increase price volatility as well as the probability that the price crosses the CB bound (which is called the magnet effect) by causing investors to shift their trades to earlier periods with lower liquidity supply (see also **Subrahmanyam** (1995)). **Chen** et al. (2018) slightly extend this work by developing a similar intertemporal equilibrium model in which investors trade for risk sharing. They also find that a CB tends to lower the stock price and increase its ex ante price volatility. The work of **Arak** and **Cook** (1997) and **Kim** and **Rhee** (1997) further contribute to a better understanding of the magnet effect.

While there is a lot of theoretical work around for the effect of price-based CBs, empirical work is rather scarce due to the fact that the likelihood of CBs to be approached, or even triggered, is very small by design. **Harris** (1997) even argues that it is perhaps impossible to reliably estimate the net effect of a CB on markets, because of a contrafactual problem: without making strong assumptions about what the price dynamics would have been had the CB not been imposed, it is impossible to make reliable inferences about the specific effects of the CB. Our paper focuses on theory only.

Nevertheless, there is some empirical work around. **Goldstein** and **Kavajecz** (2004) provide a detailed analysis on the behaviour of market participants in the period around October 27, 1997 and find that leading up to the trading halt, market participants accelerated their trades, which is consistent with the mechanism described by **Subrahmanyam** (1994). In the same spirit, a recent experimental study of **Magani** and **Munro** (2020) observes that in some, particularly noisy, environments, market participants tend to liquidate their assets earlier than in markets where there are no CBs. However, there are also several empirical papers that do not find evidence of such trading patterns and thus of a magnet effect, like for example **Huang** et al. (2001) and **Abad** and **Pascual** (2007).

The paper presented in this chapter proposes a new theoretical point of view, as our model is fundamentally different from the ones currently existing in the CB literature. Our model does not incorporate any microstructure assumptions like in **Subrahmanyam** (1994), where asymmetric information between traders and a risk-averse market maker is assumed. Neither does our model make any assumptions about the fundamental price dynamics like in **Subrahmanyam** (1994), where the informed trader can anticipate the price drift. Instead, our phenomenological model is a discontinuous generalisation of **Krugman** (1991)'s seminal work on target zones, which couples the probably of a future price change, in our case during the trading halt if a CB mechanism is triggered, back into the price itself. We show that this framework in essence is a linear factor model, where the risk factor is the expected price change occurring over the trading halt.

More specifically, in the paper presented in this chapter, **Lera** et al. (2021), we present a general CB equation that describes the price dynamics in presence of a CB. Since the CB trigger mechanism is often pre-specified by the exchange, market participants are aware of its existence. Anticipation of a potential trading halt is additional information that is fed back into the current price via tactical trading (usually by advancing trade in time). This gives rise to a non-linear coupling between the observed price and CB trigger point. And while we introduce our theory on a price-based circuit breaker, our framework is readily generalisable to other trigger mechanisms. This framework is our main contribution.

In the context of price-level based CBs, our theory confirms the two main effects that have been reported, i.e. an increased price volatility prior to the lower price level, and that the sole existence of a CB leads to the negative effect of attracting the price to the CB level. This gives our model a certain validity.

But more importantly, we believe that our model provides an ideal starting point to answer CB design questions. An ill-constructed design renders the CB either ineffective or even counter-productive, as discussed above. Thus, modelling of CBs should ideally be done with the aim to solve the question of how to optimally design a CB. Our model serves this exact purpose. In our models phenomenological structure only the investors' anticipation of the probability of a CB trigger feedbacks on the price process. Thus, the framework is a priori independent of the trigger mechanism. Therefore different trigger mechanism can be specified and studied, which makes our model a powerful tool for the discussion of CB design questions. Our main contribution is thus to suggest a framework that is flexible enough to study the effects of different CB design parameters and trigger mechanisms in a general way. And to the best of our knowledge, there is no other framework that allows for such general CB design discussions.

This is especially relevant as various CB design questions are unanswered. One example of such a design question is whether the trading halt should be rule-based (e.g. that the market will stop when a certain pre announced trigger point is reached) or if it should occur at the discretion of exchanges. The general debate over rule-based versus discretionary policies thereby goes back to **Kydland** and **Prescott** (1977), who show that in the context of macroeconomic planning, rule-based policies generally dominate discretionary policies. Applied to CBs, **Telser** (1981) then suggests that rule-based price limits in futures markets are superior to the discretionary trading halts, simply because of their predictability. Similar, **Mann** and **Sofianos** (1990) also advocate rule-based trigger mechanisms. **Subrahmanyam** (1995) presents counterarguments to these opinions. He shows that discretionary, or even randomised, trading halts may be less susceptible than rule-based halts to magnet effects. Another example of a design question is the impact of the duration of the trading halt. **Clapham** et al. (2017) points towards the relevance of this question and shows that shorter interruptions have less negative impact and lead to more effective CBs. In contrast to that, **Hautsch** and **Horvath** (2019) document that it is the mere presence of trading pauses, rather than their actual length, which makes them effective. And in the context of price-level CBs, the design question could e.g. concern the width to the pre announced price level. If the level is chosen too low, the CB is de facto ineffective, and if it is chosen too high, it is overly sensitive to ordinary intraday price movements (**Berkman** and **Lee**, 2002).

In a more general sense, we just ask which CB trigger mechanism is most effective. And while we introduce our model on a price-level based CB mechanism, we also show how our main CB formula can be generalised to different trigger mechanisms and how it thus can be used to study different CB designs. Specifically, we hint at a more robust CB design, i.e. a volatility-based CB trigger mechanism, which does not lead to a magnet effect.

In summary, we hope to provide exchanges and policy makers with a framework on how to think about CBs, and hope it will serve as starting point for the design of more resilient CB mechanisms.

References

D. **Abad** and R. **Pascual**. On the magnet effect of price limits. *European Financial Management*, 13(5):833–852, 2007

M. **Arak** and R. **Cook**. Do daily price limits act as magnets? the case of treasury bond futures. *Journal of Financial Services Research*, 121(1):5–20, 1997

H. **Berkman** and J. B. T. **Lee**. The effectiveness of price limits in an emerging market: Evidence from the korean stock exchange. *Pacific-Basin Finance Journal*, 10(5):517–530, 2002

H. **Chen**, A. **Petukhov**, and J. **Wang**. The Dark Side of Circuit Breakers. *Working Paper*, 2018

B. **Clapham**, P. **Gomber**, M. **Haferkorn**, and S. **Panz**. Managing excess volatility: Design and effectiveness of circuit breakers. SAFE Working Paper Series 195, Leibniz Institute for Financial Research SAFE, 2017

M. A. **Goldstein** and K. A. **Kavajecz**. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 7(3):301–333, 2004

B. **Greenwald** and J. **Stein**. Transactional risk, market crashes, and the role of circuit breakers. *Journal of Business*, 64(4):443–462, 1991

S. **Grossman**. Introduction to NBER symposium on the October 1987 crash. *The Review of Financial Studies*, 3:1–3, 1990

L. **Harris**. Circuit Breakers and Program Trading Limits: What Have We Learned? 1997

N. **Hautsch** and A. **Horvath**. How effective are trading pauses? *Journal of Financial Economics*, 131(2): 378–403, 2019. doi: 10.1016/j.jfineco.2017.12

Y. **Huang**, T. **Fu**, and M. **Ke**. Daily price limits and stock price behavior: evidence from the taiwan stock exchange. *International Review of Economics & Finance*, 10(3):263–288, 2001

K. A. **Kim** and S. G. **Rhee**. Price limit performance: Evidence from the tokyo stock exchange. *The Journal of Finance*, 52(2):885–901, 1997

P. R. **Krugman**. Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, 106(3): 669–682, 1991

F. E. **Kydland** and E. C. **Prescott**. Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3):473–491, 1977

C. M. **Lee**, M. J. **Ready**, and P. J. **Seguin**. Volume, volatility, and new york stock exchange trading halts. *The Journal of Finance*, 49(1):183–214, 1994

S. **Lera**, D. **Sornette**, and F. **Ulmann**. Price Dynamics with Circuit Breakers. *Revise and resubmit at the Journal of Banking and Finance*, 2021

J. **Magani** and D. **Munro**. Dynamic runs and circuit breakers: an experiment. *Experimental Economics*, 23(1): 127–153, 2020

R. **Mann** and G. **Sofianos**. Circuit breakers for equity markets. *Market Volatility and Investor Confidence*, pages 225–253, 1990

Presidential Task Force on Market Mechanisms. The report of the presidential task force on market mechanisms, 1988

A. **Subrahmanyam**. Circuit breakers and market volatility: A theoretical perspective. *The Journal of Finance*, 49(1):237–254, 1994

A. **Subrahmanyam**. On rules versus discretion in procedures to halt trade. *Journal of Economics and Business*, 47(1):1–16, 1995

L. G. **Telser**. Margins and futures contracts. *Journal of Futures Markets*, 1(2):225–253, 1981

Price Dynamics with Circuit Breakers

Sandro Claudio Lera[a,b,c,*], Didier Sornette[a,d,e], Florian Ulmann[d]

[a] *Institute of Risk Analysis, Prediction and Management, Southern University of Science and Technology, Shenzhen, China*

[b] *Business School, Southern University of Science and Technology, Shenzhen, China*

[c] *Connection Science, Massachusetts Institute of Technology, Cambridge, USA*

[d] *Department of Management, Technology, and Economics, ETH Zurich, Switzerland*

[e] *Tokyo Tech World Research Hub Initiative, Tokyo, Japan*

Abstract

We model the dynamics of an asset price in the presence of a circuit breaker. A price-based circuit breaker for example is a trading halt imposed by the exchange after the price drops below (or rises above) a pre-specified level within a trading session. The investors' anticipation of the probability of a halt feedbacks on the price process. In a general stochastic financial framework, this leads to coupled integral and stochastic differential equations. In the context of price-based circuit breakers, our theory predicts an increased price volatility prior to the trigger point and that the sole existence of a circuit breaker leads to the adverse effect of attracting the price to the circuit breaker level (magnet effect), as has been reported in the literature. As our model is not a priori dependent on a specific circuit breaker trigger mechanism, we show how it can be used to study different circuit breaker designs. We propose first steps towards a more robust design of circuit breakers, i.e. that a trading halt that is conditioned on a pre-specified volatility level, instead of price level, mitigates the negative side effects of increased volatility. The methodology proposed in this article serves as starting point for the design of more resilient circuit breaker mechanisms.

JEL: G15, G23, C51, C54

Keywords: circuit breaker, magnet effect, circuit breaker design, exchange

---

[*] corresponding author

Email addresses: `leras@sustech.edu.cn` (Sandro Claudio Lera), `dsornette@ethz.ch` (Didier Sornette), `fulmann@ethz.ch` (Florian Ulmann)

Revise and resubmit at the Journal of Banking and Finance

1. Introduction

1.1. Motivation

In a perfectly efficient market, prices are thought to reflect the fundamental value of their underlying assets at all times. While this may be a reasonable approximation of real markets over horizons of weeks or months for most of the time, except during financial bubbles and crashes, it is well known that, on shorter time-scales, prices can deviate as a result of herding behaviours created either by social interactions and/or machine trading. For instance, overly-optimistic trend-following, when adopted by a sufficiently large minority, may lead to the formation of financial bubbles (Lin et al., 2018). Panic sell-offs based on over-reaction to negative news can trigger further sell-offs and self-accelerate into an endogenous market crash. In addition, market turbulences stemming from erroneous algorithms, fake news or macroeconomic events seem to occur with growing frequency in today's highly automated markets. Thus, it is common that exchanges employ so-called market circuit breakers (from hereon CBs). The exchange specifies that if a trigger mechanism is activated, for price-based CBs e.g. if the price drops below (or rises above) a pre-specified threshold value within a trading session, then trading is halted for a pre-specified amount of time, until regular trading resumes. The rational behind this CB mechanism is that it gives traders time to calm down, interact with their counterparts, rethink their trading decision, and/or adjust their algorithms. Note that also other form of CBs exist. Yong and Yang (2004) provide a detailed overview. While we introduce our theory with a focus on the most prominent CB, which is the price-based trading halt just described, it is generalisable to other trigger mechanisms as we show.

In a recent survey, Gomber et al. (2017) report that the popularity of circuit breakers on trading venues has risen from 60% in 2008 to 86% in 2016. The implementation of a CB leaves exchanges with the question of how to best *design* them. In the context of price-barriers, this e.g. means how to choose the pre-defined threshold and halting time. But the design questions can also be more general in nature. For example if the trigger mechanism should be based on a price level or a volatility level, or if the trigger level should be completely rule-based or discretionary. Given that investors are continuously forming anticipations, their tactical positioning in the presence of a CB may lead to a feedback of the CB on the price itself. We start with carefully accounting for such anticipations and feedbacks in a rigorous formalism. Our motivation is that a lack of understanding of the intricate price dynamics resulting from the interaction with the CB mechanism may lead to an ill-constructed design and thereby renders it either ineffective or counter-productive.

More specifically, we present a circuit breaker equation that describes the price dynamics in presence of a price-based circuit breaker. In a first simplifying view, the price follows a geometrical Brownian motion, and the CB acts as an absorbing boundary. But as we describe below, this simplifying assumption is inaccurate. Since the CB trigger point is often pre-specified by the exchange, market participants are aware of its existence. Anticipation of a potential trading halt is additional information that is fed back into the current price via tactical trading (usually by advancing trade in time). This gives rise to a non-linear coupling between the observed price and CB trigger point. And while we introduce our theory on a price-based circuit breaker, our framework is generalisable to other trigger mechanisms, as we show in section 5. This general framework is our main contribution. Our theory is flexible enough to study the effects of different CB trigger mechanisms as well as CB design parameters. Since in our model only the investors' anticipation of the probability of a CB trigger feedbacks on the price process, the framework is a priori independent of the specific trigger mechanism. Thus, different trigger mechanisms can be specified and studied, which makes the discussion of the effect of different CB designs straightforward. To the best of our knowledge, there is no other framework that allows for such a general CB design discussion.

1.2. Circuit Breaker Events

Figure 1 shows three examples of price dynamics that triggered a price-based CB. In early 2016, a price-based CB was implemented on the CSI300 financial index which replicates the top 300 stocks traded in the Shanghai and Shenzhen stock exchanges. On January 4th 2016, the first trading day of the year, the price movement was already sufficient to trigger the CB twice. In the very same week, on January 7th, the mechanism was triggered twice again. The Chinese authorities subsequently abandoned the CB mechanism from their index. Another CB event happened on Chicago's Cboe Futures Exchange, offering bitcoin futures since December 10, 2017. On that first day of bitcoin trading, two CB got triggered on their upper bounds at 10% and 20%, respectively (upper plot in Figure 1). More recently, amid uncertainty around the COVID-19 crisis, the S&P 500 Index triggered a market wide price-based CB at the -7% level four times in March 2020. These was the first time a CB was triggered on the S&P 500 since 1997. Unlike the CSI and BTC halts, here the opening values were already close to the CB level, indicating that a lot of information must have been priced in from the oversight gap

In the following section we summarize past and related work.

1.3. Related Literature

Market-wide price-based CBs rose to prominence in the aftermath of the October 1987 stock market crash, first advocated by the Brady Commission (Presidential Task Force on Market Mechanisms (1988)). Ever since, several academic investigations have been conducted. We refer to the work of Yong and Yang (2004), Abad and Pascual (2013) and Sifat and Mohamad (2019) for in depths overviews of earlier work on price-based CBs, but nevertheless provide a summary of the most relevant literature in the following paragraph.

As an early proponent of price-based CBs, Greenwald and Stein (1991) argues that, in a market with limited participation and thus execution risk, CBs can help to better synchronize trading for market participants and improve the efficiency of allocations. Similar, Lee et al. (1994) argue that by lowering informational asymmetries between traders, halts could permit the orderly emergence of a new consensus price. And in line with that, Hautsch and Horvath (2019) recently showed that the mere presence of CBs limits adverse selection risk for liquidity providers and that CBs enhance price discovery during the break.

But there are also contrasting views to the ones above on the effect of price-based CBs. Corwin and Lipson (2000) empirically find trading halts increase price volatility and report that the order book depth near the best bid and ask prices is unusually low before, during, and after trading halts. However, the associated diverging bid-ask spread prior to the halt explains, at best, a small portion of the increased volatility. They surmise that the remaining volatility may be the result of unmeasured information effects or the market closure itself. On top of that Goldstein and Kavajecz (2004) additionally provide empirical

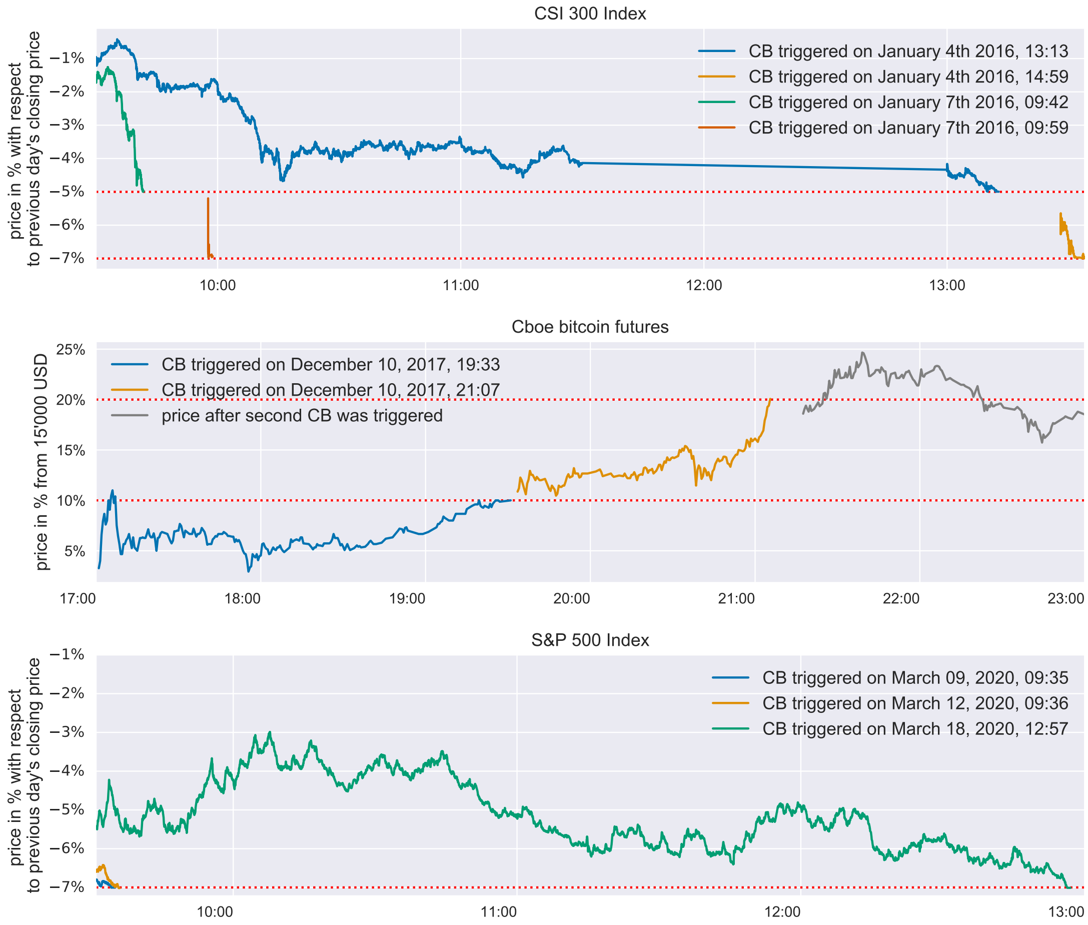


Figure 1: The top plot shows the CSI300 index on January 4th, 2016 and January 7th, 2016. On both days, two price-based CB were triggered. The price-based CB was defined in a symmetric way: If the Index moves by $\pm 5\%$ with respect to the last day's closing price, trading is stopped for 15 minutes and continued afterwards. If, subsequently, the index reaches changes by a total of $\pm 7\%$, trading is stopped for the remainder of the day. The middle plot shows the price trajectory of the bitcoin future on the first trading day in December 10, 2017. The Cboe rule book specifies that a price-based CB is first triggered for 2 minutes if the price falls below or rises above 10% with respect to $15'000$ USD (or with respect to the previous day's closing price, if it is not the first day of trading). If it rises above 20%, it is halted for yet another 5 minutes. The first CB would have been triggered already around 5 minutes after the opening at 17:00 UTC. However, the CFE Help Desk exercised discretion under CFE rules not to halt trading in XBT futures at the 10% up level due to a fair and orderly market. At 19:31 UTC, the CFE halted XBT trading, in accordance with CFE Rule 1302(i)(ii). XBT re-opened for trading two minutes after the time of the halt. At 21:05 UCT, the CFE halted XBT trading again in accordance with CFE Rule 1302(i)(ii), and re-opened five minutes later. The bottom plot displays three out of four market wide halts triggered on the S&P 500 at the -7% level. Already the open prices lie close to the CB boundary, indicating that a lot of (COVID-19 related) overnight information must have been priced. Accordingly, the fourth trigger event on March 16 is not shown since it was triggered within the first minute of trading.

evidence of a ‘magnet effect’, stating that the price level of the CB is reached faster than it would in absence thereof, because traders want to make up for the looming halt ahead of time. This attraction towards the circuit break is confirmed by Cho et al. (2003). And even in a recent experimental study, Magani and Munro (2020) observe that in some, particularly noisy environments, market participants tend to liquidate their assets earlier than in markets where there are no circuit breakers. Nevertheless, there is also empirical work that does not find any evidence for the existence of the magnet effect Abad and Pascual (2007).

An early theoretical model on price-based CBs thereby stems from Subrahmanyam (1994), who presents an intertemporal equilibrium model with three agents, two types of traders with asymmetric information and a risk-averse market maker. He shows that price-based CBs can increase price volatility as well as the probability that the price crosses the CB bound (which is called the magnet effect) by causing investors to shift their trades to earlier periods with lower liquidity supply (see also Subrahmanyam (1995)). Chen et al. (2018) slightly extend this work by developing a similar intertemporal equilibrium model in which investors trade for risk sharing. They also find that a price-based CB tends to lower the stock price and increase its ex ante price volatility.

Our approach is different from the ones above in that we do not rely on any specifications of different types of traders and their risk preferences. Instead, our phenomenological model can be seen as discontinuous generalization of Krugman (1991)’s seminal work on target zones, which couples an expected (log) price change back into the (log) price itself. Lera and Sornette (2018) have shown that Krugman’s phenomenological model is an approximation of the ‘exact’ representation in terms of a portfolio of fundamental value and option prices. Depending on the question under scrutiny, one of these two representations may be more convenient than the other. Similarly, here we prose a point of view that serves as alternative to the ones rooted in microscopic assumptions cited above. In particular, we show in section 5 that this phenomenological representation proofs useful when studying the design of circuit breakers.

As discussed above, one example of such a design question is whether the trading halt should be rule-based or if it should occur at the discretion of exchanges. Telser (1981) suggests that rule-based price limits in futures markets are superior to the discretionary trading halts, simply because of their predictability. Similar, Mann and Sofianos (1990) also advocate price-triggered halts. Subrahmanyam (1995) presents counterarguments to these opinions. He shows that discretionary, or even randomised, trading halts may be less susceptible than rule-based halts to magnet effects. Further, the impact of the duration of the trading halt is also a point of discussion. Benjamin et al. (2017) e.g. shows that shorter interruptions have less negative impact and lead to more effective CBs. Contrary, Hautsch and Horvath (2019) document that it is the mere presence of trading pauses, rather than their actual length, which makes them effective. And in the context of price-based CBs, design questions e.g. concern the width to a potential pre announced price level. If the level is chosen too low, the CB is de facto ineffective, and if it is chosen too high, it is overly sensitive to ordinary intraday price movements Berkman and Lee (2002). And besides these examples, one could also think about new CB designs in a more general way, e.g. to discuss which CB trigger mechanism is most effective.

We believe that our model provides an ideal starting point to answer such CB design questions. Its phenomenological structure simply links the CB trigger anticipation with todays's price, so that the effect of different CB trigger mechanisms and CB design parameters on the price dynamics can be discussed in a general way. This makes our model a powerful tool to discuss CB design questions.

The paper is organised as followed. In section 2 we thereby start by modelling a pre-defined price-based CB mechanism. We present the equation embodying the effective attractive force exerted by the CB, which we then solve perturbatively in section Appendix A. Examining the solution as a function of different parameters, we can explain the increased volatility towards the CB boundary under a broad range of assumptions. Section 4 shows how the empirically observed magnet effect can be accounted for naturally within our framework. We then show in section 5 how our main CB formula can be generalised to different trigger mechanisms and how it can be used to study CB designs. And we also hint at more robust circuit breaker designs, i.e. a volatility-based CB trigger mechanism. Last, we conclude in section 6.

2. The Circuit Breaker Equation

2.1. General Formulation

We denote by $s = s_t$ the (logarithm of the) exchange quoted price of an asset or index at time t. Our model is focused on the intraday dynamics of any given, specific day, and hence we denote by s_0 the market opening price. The market closing price is denoted by s_{t_c}, i.e. the total trading time is equal to t_c. Unless stated otherwise, we set $t_c = 1$ from hereon. We assume that the market has a circuit breaker implemented, which is triggered if s drops below a predefined value $\underline{s}$. [1] If the price reaches $\underline{s}$ at some time t_s (the stopping time), i.e. if $s_{t_s} = \underline{s}$, then trading is halted for a predetermined time Δt_h during which no trading can occur. The market reopens at time $t_r = t_s + \Delta t_h$ and trading continues normally until the end of the day, or until another CB is triggered.

We denote by f_t the (logarithm of the) the asset price that would prevail on a perfectly efficient and rational market and in absence of a CB. From hereon, we also refer to this quantity as the 'fundamental' value. [2] It is customary to assume that the log-fundamental price dynamics obeys

$$\mathrm{d}f_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t \tag{2.1}$$

with drift μ and volatility σ. Since there is a CB in place, the rational agent has to price in the possibility of a trading halt. Similar to the rational expectation models, we write s_t as a probabilistic linear combination of two outcomes: the CB is triggered, or it is not triggered. Denoting by $p_{\text{trig}} = p_{\text{trig}}(s_t, t; t_c)$ the probability at time t that the CB is triggered at any time between t and t_c given that the price is s_t, we can write

$$s_t = \left(1 - p_{\text{trig}}\right) \cdot s_t^{\text{NT}} + p_{\text{trig}} \cdot s_t^{\text{T}} \tag{2.2a}$$

---

[1] In practice, the value of $\underline{s}$ is usually specified in form of a percentage drop with respect to the closing price of the previous day. Hence, it is a well-known, fixed quantity already before the market opening. The formulation in terms of an upper-boundary $\overline{s}$ is analogous.

[2] This must not be understood as the fair, rational price predicted by a fully efficient market. Our theory does however not rely on this assumption and f can itself be inefficient, or show irrational herding behaviour. The deviation of s from f just measures the additional effect of the CB.

where the superscript ‘T’ and ‘NT’ stand for ‘trigger’ and ‘not trigger’, respectively. This linear probability weighting is the canonical, simplest starting point to model rational agents[3]. It is for instance used by Krugman (1991) in his forex target zone model to account for the possibility that the target zone is not defended perfectly by the central bank with a probability p. This leads to considering a superposition of a free floating exchange rate and a bounded exchange rate, like the above equation (2.2a) (cf. equation (14) in (Krugman, 1991)).

For $p_{\text{trig}} = 0$ (or when there is no CB), it holds trivially that $s_t^{\text{NT}} = f_t$, as alluded above.

For $p_{\text{trig}} = 1$, the asset price s_t^{T} departs from f_t as a result of the market impact of investment decisions of traders who anticipate the occurrence of the trigger. The main concern of investors is the risk associated with the impossibility of trading between the stopping time t_s and the reopening time t_r. During this time period, the underlying fundamental value f_t is likely to change but investors are unable to adjust continuously their allocation to the risky asset correlatively as would be the case in absence of the circuit breaker. In order to remain invested in this risky asset, the investors thus require a risk premium to compensate for the risks incurred by the possible occurrence of the trading halt. In a risk neutral world, this risk premium should be proportional to the expected price change $\mathbb{E}_t\left[s_{t_r}\right] - \mathbb{E}_t\left[s_{t_s}\right]$ occurring over the trading halt[4], as this price changes sets the size of the possible loss incurred by the investors during the trading halt. In other words, the expected price change $\mathbb{E}_t\left[s_{t_r}\right] - \mathbb{E}_t\left[s_{t_s}\right]$ acts as a market factor impacting all traders’ portfolios. We then capture the factor loading or exposition to this factor in the usual way via a coefficient γ. Since, by definition, $\mathbb{E}_t\left[s_{t_s}\right] = \underline{s}$, since the circuit breaker is triggered once the price s_t reaches the level $\underline{s}$ which happens at the stochastic stopping time t_s, we obtain the following asset price equation

$$s_t^{\text{T}} = f_t + \gamma\left(\mathbb{E}_t\left[s_{t_r}\right] - \underline{s}\right), \tag{2.2b}$$

where $\mathbb{E}_t$ denotes the expectation value with information up to time t and we refer to $\mathbb{E}_t\left[s_{t_r}\right] - \underline{s}$ as the trading risk premium. Extensions of model (2.2b) include going beyond risk-neutrality, which implies that the risk premium should include volatility terms in addition to the expected price change $\mathbb{E}_t\left[s_{t_r}\right] - \underline{s}$.

The price equation (2.2b) is reminiscent of the forex target zone model of Krugman (1991), $s_t = f_t + \gamma\mathbb{E}_t\left[\mathrm{d}s/\mathrm{d}t\right]$, where the trading risk premium term is replaced by the expected momentum $\mathbb{E}_t\left[\mathrm{d}s/\mathrm{d}t\right]$. In Krugman’s model, investors anticipate the influence on the price of possible central bank interventions. These interventions can be also modelled as the provision of an effective put option guaranteeing that the price will not fall below the lower target value (Lera and Sornette, 2018). Another difference between equation (2.2b) and Krugman’s model is found in the different boundary conditions. Whereas the forex target zone can be considered as a repelling boundary, we are dealing here with an absorbing boundary condition at $\underline{s}$. In other words, in a target zone regime, the market expects the central bank to step in and push the price back. In contrast, the circuit breaker just halts trading.

---

[3] Within the framework of cumulative prospect theory and behavioral finance, one could adjust these weights, to take into account non-linear or non-rational effects. This might complicate some of the equations shown below, but does not otherwise change the presentation of our results.

[4] Strictly speaking, we should write $\mathbb{E}_t\left[s_{t_r} \mid \text{trig}\right]$ instead of just $\mathbb{E}_t\left[s_{t_r}\right]$, i.e. we should make clear that the expectation value is conditional on the fact that the CB is triggered. Indeed, this conditioning is important in our calculations below. For notational simplicity, we consider this conditioning as implicit in the following.

In summary, while they are not micro-founded, expressions (2.2a) and (2.2b) constitute reduced form equations that capture in a concise and elegant way the impact of a circuit breaker. They should be considered at a similar level of reduced form abstraction as Krugman (1991)'s forex target zone model. And similarly to Krugman (1991)'s model[5], the price dynamics solution of (2.2a) and (2.2b) is not exactly a martingale, which seems in contradiction with textbook asset pricing theory that arbitrage opportunities cannot exist in a competitive markets because they would be instantly exploited, and thereby eliminated, by arbitrageurs. But with constrained arbitrageurs with limited capital and subjected to solvency requirements and in the presence of transaction costs and other frictions, the arbitrage opportunities become risky and, in practice, often cannot be exploited. In Krugman (1991)'s model, we observe that the arbitrage opportunity vanishes linearly in time and shrinks exponentially fast as f_t moves away from the target zone. In our model (2.2a) and (2.2b), the price dynamics does not have a simple analytical form as for Krugman (1991)'s model, but similar considerations are expected to hold.

2.2. The Assumption of One Circuit Breaker

By definition, it holds that $\mathbb{E}_t\left[s_{t_s}\right] = s_{t_s} = \underline{s}$, as stated in equation (2.2b). On the other hand, the price at the reopening s_{t_r} is not known in advance and depends on the stochastic stopping time t_s via $t_r = t_s + \Delta t_h$. This is why the exact price s_{t_r} has to be replaced by its expectation value. Once the circuit breaker mechanism was triggered, and after the market reopens at time t_r, the observable price is back at its fundamental value, i.e. $s_{t_r} = f_{t_r}$. This assumption is reasonable whenever there is either only one CB a day, or when the triggering of a second CB is relatively unlikely within the same day $(p_{\text{trig}}(t_r) \ll 1)$. Recall that f_t is the price in absence of a circuit breaker which can, but must not necessarily be identified with the fundamental value of the asset in a market efficiency type of sense. Further, a no-arbitrage argument is embodied in the fact that there is no aftermath effect of the circuit breaker. A distortion of the price process only arises due to the trading risk premium, which by construction is zero upon market reopening. From $s_{t_r} = f_{t_r}$ follows trivially $\mathbb{E}_t\left[s_{t_r}\right] = \mathbb{E}_t\left[f_{t_r}\right]$. Plugging this into (2.2b) and then into (2.2a) yields

$$s_t = f_t + \gamma p_{\text{trig}} \left(\mathbb{E}_t\left[f_{t_r}\right] - \underline{s}\right). \tag{2.3}$$

In the remainder of this article, we are concerned with finding a solution $s = \phi(f,t)$ that satisfies (2.3). Note that this is a deterministic relationship that holds for any value f. Stochasticity arises by plugging the realized fundamental trajectory f_t into this deterministic relationship, such that $s_t = \phi(f_t,t)$. [6] In practice, only the process s_t can be observed. We can make inferences about the price trajectory in absence of the circuit breaker by considering the functional inverse, i.e.

---

[5] The Martingale condition reads $\mathrm{E}_t[s_{t'}] = s_t,\ \forall t' \geq t$. In Krugman (1991)'s model, the log-exchange rate is given by $s_t = f_t + \frac{1}{\rho}e^{\rho(\underline{f}-f_t)}$ where $\rho = \sqrt{\frac{2}{\gamma\sigma^2}}$, with σ being the volatility of the fundamental term f_t. We thus obtain $\mathrm{E}_t[s_{t'}] = s_t + (s_t - f_t)\left(e^{\frac{t'-t}{\gamma}} - 1\right)$, using standard Gaussian integration. In the limit $\gamma = 0$, $s_t = f_t$ and the Martingale condition holds. But when $\gamma > 0$, there is an apparent arbitrage opportunity $\mathrm{E}_t[s_{t'}] - s_t = (s_t - f_t)\left(e^{\frac{t'-t}{\gamma}} - 1\right) \approx \frac{1}{\gamma}(s_t - f_t)(t'-t)$ for small $t'-t$.

[6] We will use the notation $s = \phi(f,t)$ and $s_t = \phi(f_t,t)$ interchangeably, as they are synonymous. The function ϕ maps a given fundamental value x at time t onto an observable price s, $s = \phi(x,t)$. Usually, the value of x we are most interest about is f_t, i.e. the value of the fundamental at time t.

$f_t = \phi^{-1}(s_t, t)$. [7] [8]

In the limit $p_{\text{trig}} \longrightarrow 0$, we recover $s_t = f_t$, as it should. The probability p_{trig} is itself a function of s_t and t. The closer s_t is to the barrier $\underline{s}$, and the more time until the closing time t_c there is, the larger p_{trig} is. Similarly, the stopping time t_s depends non-linearly on the position of s_t, as is explained in more detail below. We can thus see that s_t is coupled to its expected future trajectory, making equation (2.3) non-trivial. In the limit case where $p_{\text{trig}} \longrightarrow 1$, we arrive at $s_t = f_t + \gamma\left(\mathbb{E}_t\left[f_{t_r}\right] - \underline{s}\right)$, which enjoys an evident interpretation: If the CB is triggered with certainty, then the market immediately prices in the expected difference between the value at the trigger point $\underline{s}$ and the reopening price f_{t_r}.

2.3. *Illustrative Numerical Solutions of the CB Equation*

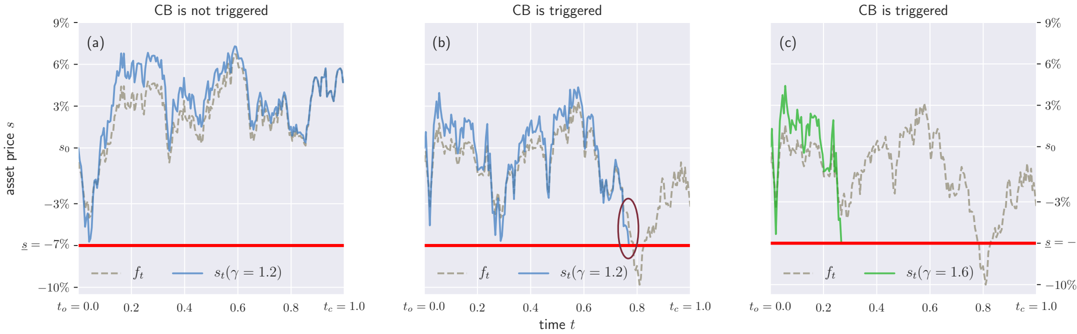


Figure 2: The grey dashed lines in all three plots show two independent stochastic realizations of the fundamental process (2.1). The parameters of the fundamental price are $\mu = -5\%$ and $\sigma = 10\%$. We have used two random number seeds, one to generate the realized trajectory in plot (a) and one to generate two identical realizations in (b) and (c). The blue lines in (a) and (b) are the associated observed, tradable asset price s_t related to f_t via (2.3) with $\gamma = 1.2$ and $\Delta t_h = 0.3$. While for the trajectory f_t in (a) the CB at -7% level is not triggered, for the one in (b) it is. We also observe a magnet effect (inside red ellipsis), meaning that s_t reaches that level earlier than f_t. This is accentuated in the green line of plot (c), which depicts the same f_t trajectory as in (b), but s_t is linked to f_t with a larger coupling strength ($\gamma = 1.6$). This leads to a price dynamics that reaches the -7% level about half a day earlier (around $t = 0.3$) than it would in absence thereof (around $t = 0.8$).

To gain some intuition for the subsequent quantitative discussion, we present two illustrative solutions of equation (2.3). We assume that the underlying fundamental value f_t governed by (2.1) has a drift of $\mu = -5\%$ and volatility of $\sigma = 10\%$, in units of days. In annual terms, this amounts to a whopping drift of more than $-1'000\%$ and an annual volatility around 150%. While this seems a lot, such high drifts and volatility are actually very common during tumultuous trading days. In the remainder of this article, we work in units of days. Time runs continuously from $t = 0$ (market opening) to $t = 1$ (market closing). As the total duration of one day is normalized to 1, this means that the fundamental price f_t is expected to drop by -5% over the full day of trading, on average. The large value of the standard deviation σ means f_t can be found anywhere between +5% and -15% at a $\pm\sigma$ probability level. Thus, the two price trajectories presented here are typical.

---

[7] Here, ϕ^{-1} denotes the functional inverse in the first argument. This function is well-defined since $\mathrm{d}s/\mathrm{d}f > 0$ is required for stability and consistency.

[8] A similar method has been applied to estimate the fundamental value of the Swiss Franc/Euro exchange rate (Hanke et al., 2019) and the USD/HKD exchange rate (Lera et al., 2019) in presence of target zones.

The dashed, grey lines in all three plots of Figure 2 show two stochastic realizations of such a process (2.1). Both are generated with $\mu = -5\%$ and $\sigma = 10\%$, but with different random number seeds. We have used two random number seeds, one to generate the realized trajectory in plot (a) and one to generate two identical realizations in (b) and (c). After a strong initial down-trend, the realization in plot (a) moves mostly side-ways for the remainder of the day. The realized path for f_t in plot (b) and (c) is punctuated by three intraday drops, of which the last one falls almost below 10% of its opening level.

We now assume that a CB is employed, to be triggered once the log difference between opening price s_0 and current price s_t falls below -7%. Here, we have picked $\underline{s} = s_0 - s_{t_{\text{trig}}} = -7\%$ because it is the CB level that is imposed on the S&P500 index by the New York Stock Exchange (NYSE). As a consequence of this installation, the observed price s_t is now distinct from the unobservable fundamental value f_t that would prevail in absence of the CB. However, their relationship is deterministic, given by $s_t = \phi(f_t, t)$ where $\phi(f, t)$ is a solution of equation (2.3). The associated observable price trajectories s_t are shown as the blue lines in (a) and (b) with $\gamma = 1.2$ and as green line in (c) with $\gamma = 1.6$.

In (a), no CB is triggered, although s_t approaches the lower boundary closely around $t = 0.05$ (i.e. about 20 minutes after opening if we assume trading hours from 9:30am to 4pm). The price s_t is observed to both under- and overshoot its underlying fundamental value. However, when f_t approaches $\underline{s}$, the observable price is pushed below the fundamental one. As the fundamental price f_t moves away from $\underline{s}$, we encounter a regime where s_t is above f_t. As the day goes on, s_t converges to f_t, since the closer to the end of the day $t = 1$, the less likely a CB will be triggered when the price is sufficiently far from it; hence $p_{\text{trig}} \to 0$, and thus $s_t \to f_t$ at $t \to 1$.

The process in plot (b) is similar to the the left one. However, here the circuit breaker is actually triggered, and then stopped out for the rest of the day (as we have set $\Delta t_h = 0.3$ in this example). As in the left plot, one can observe a “magnet effect” where s_t is pushed below f_t towards $\underline{s}$, thus advancing the stopping time by the circuit breaker. This is even more apparent in plot (c), which is exactly the same as (b) just with a larger γ parameter. This leads to a price dynamics that reaches the -7% level about half a day earlier (around $t = 0.3$) than it would in absence thereof (around $t = 0.8$). We refrain from showing the continued process upon reopening at time $t_r = t_s + \Delta t_h$, in line with our assumption of considering only one CB. We will expand on these qualitative observations in the following.

3. Discussion of the Circuit Breaker Equation

The solution of the CB equation (2.3) is presented in Appendix A. The presence of a circuit breaker turns out to modify price trajectories in a very non-trivial manner, as shown in Figure 3. One can observe that the non-linear deviation of s from f is governed by the non-linear decay of p_{trig}. As anticipated, $s \to f$ is recovered far away from $\underline{s}$, where the presence of the CB is not felt, and correspondingly $\alpha \to \mu$ and $\nu \to \sigma$. The other end of the curves is more interesting. Depending on the parameter settings, the CB pushes observed prices s either above or below its fundamental value f. This is rationalized by plugging (A.4) into (2.3) which yields

$$s_t = f_t + p_{\text{trig}}\gamma(f_t - \underline{s}) + p_{\text{trig}}\gamma\mu\left(\mathbb{E}_t\left[t_s - t\right] + \Delta t_h\right) \tag{3.1}$$

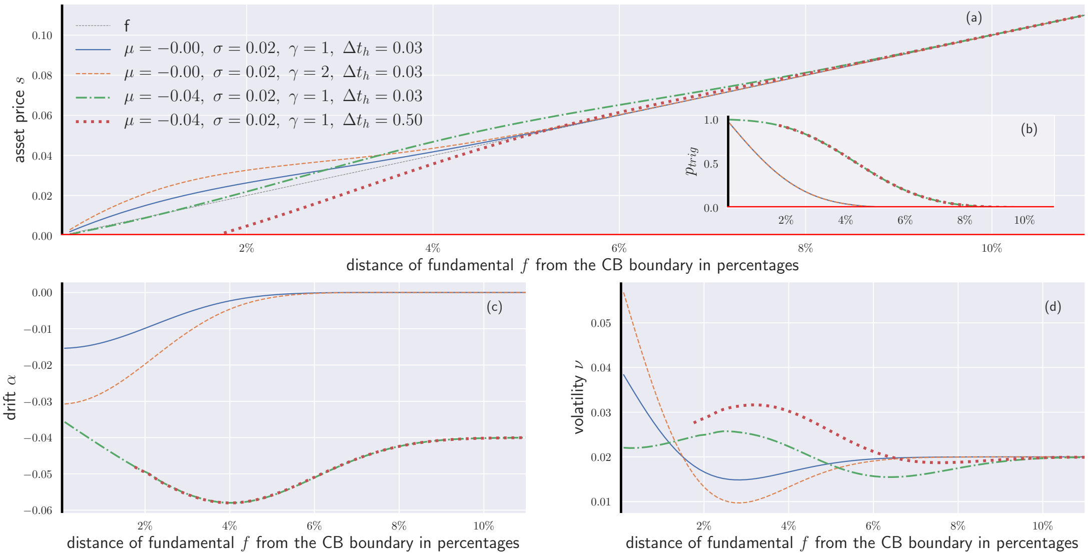


Figure 3: The top panel shows the first order solution to equation (2.3) for different sets of parameters, with CB location at $\underline{s} = 0$, as a function of the distance of the fundamental price f_t from the CB boundary (0.1 means a distance of 10%). The volatility σ of the underlying fundamental (2.1) is held fixed, whereas two values of μ are shown: no drift ($\mu = 0$), and strong negative drift ($\mu = -4\%$). Note that we are working in daily units, such that a volatility of $\sigma = 2\%$ amounts to an annualized volatility of roughly 30%. Correspondingly, a drift of $\mu = -4\%$ would correspond to an extremely high annualized drift of $-1'000\%$. While this seems extreme, we shall see in the empirical section below that, during market crashes, such a range of values is realistic. The inset panel is the hitting probability, which is independent of γ and Δt_h by construction. Drift α and volatility ν are shown in the bottom two panels. See text for an interpretation of these functional shapes.

and reveals a competition between the second and third term in the right hand side. In the case where $|\mu|$ is small, the second term dominates and pushes s above f, as the potential downshift in f between halting- and reopening-time is negligible. The larger the halting time Δt_h and the more negative μ, the more important is the downward correction induced by the third term. In turn, this pushes s below its value in absence of a CB, contrary to its desired effect. As the CB trigger point $\underline{s}$ is approached, the drift α first enters a regime of increased negative downwards acceleration with α much more negative than μ, which is then later dampened out, or even reverted. Interestingly, the dependence of the stopping time Δt_h factors out, and does not have an influence on the drift. As the distance from the CB boundary decreases, the volatility first also decreases and then steeply increases compared to the constant volatility σ in absence of a CB. This explains the increased price volatility observed by Lee et al. (1994), and confirms the hypothesis of Corwin and Lipson (2000) that this volatility increase is not (only) the result of a decreasing liquidity, but derives from the anticipation of the price disturbance induced by the CB itself. For negative drifts of the fundamental price, volatility is then again decreased closed to the trigger point. A discussion of the implications regarding the effectiveness of CBs is postponed to section 5, where we will examine the prevalence of the ‘magnet effect’.

3.1. The Case of an Unknown Drift

Throughout this article, we work under the assumption that the underlying drift μ and volatility σ are known. But since we consider intraday dynamics, and in particular tumultuous trading days, one may cast reasonable doubt on this assumption. More generally, we can assume that there is a prevailing belief $\hat{\mu}$ of what the value of μ is (and analogously for σ). This is similar in spirit to the approach of Chen et al. (2018) who assume that there are two types of traders with different perceptions of the underlying drift μ. Starting from $\hat{\mu} \neq \mu$, we consider the arrival of fundamental news at a rate η, that ultimately makes $\hat{\mu}$ converge exponentially to the true μ over a time-scale $\tau \sim \eta^{-1}$. We have run numerical calculations over a reasonable large range of values for η and have found no qualitative change in our results. A more detailed account of unknown fundamental parameters under various assumptions is part of future research.

4. The Magnet Effect

In this section, we pick up a question previously discussed in the literature: Is there a magnet effect? The magnet effect, as observed empirically by Goldstein and Kavajecz (2004) and subsequently confirmed by others (see (Yong and Yang, 2004; Abad and Pascual, 2013) for reviews), claims that the very existence of a CB pushes the prices to lower levels than what would be observed if there was no CB in place. To assess the existence of such a magnet effect, we consider the difference in hitting probabilities. Starting from some opening price level $s_0 > 0$ at time $t = 0$, we ask what is the probability that some lower threshold $\underline{s} = 0$, the (hypothetical) location of the circuit breaker, is reached within a day (up to $t = 1$). In absence of a circuit breaker, the market follows the process (2.1) with constant drift μ and volatility σ. Then, we denote by $p_{\text{trig}}(\mu, \sigma)$ the probability that the lower threshold $\underline{s} = 0$ is reached in absence of circuit breaker. This expression is known analytically and equal to expression (A.9). If a circuit breaker is introduced, with halting time Δt_h, the drift and volatility are modified into non-linear functions α and ν, respectively (Figure 3). We denote by $p_{\text{trig}}(\alpha, \nu)$ the associated probability to hit this boundary within a day. We then calculate the magnet effect (ME) as

$$\text{ME}(\mu, \sigma, s_0, t_h, \gamma) \quad = \quad p_{\text{trig}}(\alpha, \nu) - p_{\text{trig}}(\mu, \sigma), \tag{4.1}$$

i.e. as the difference of hitting probabilities with and without a circuit breaker. [9] A detailed description of how these quantities are determined numerically is found in Appendix C. If ME is positive, then the hitting probability is larger with than without a circuit breaker, i.e. the magnet effect occurs. Conversely, if the quantity is negative, the magnet effect does not exist, and the CB has a repulsive effect. As is visible from Figure 4, the ME holds for all type of realistic market configurations. This shows that the very existence of a CB tends to push the price to the CB level, when the drift μ is sufficiently negative and the price approaches the circuit breaker. The intuition behind this effect is clear and have already been hinted in previous sections: if the market drops as the result of a negative drift μ, the market prices the additional movement of $\mu \Delta t_h$ during

---

[9] There are other quantifications of the magnet effect that appear reasonable. For instance, we have applied the same analysis by comparing expected hitting times under (μ, σ) and (α, ν). The conclusions are essentially the same as with measure (4.1). We have decided to show here on the hitting probability because p_{trig} is naturally bounded between 0 and 1 and can be calculated by considering time just from $t = 0$ up to $t = 1$.

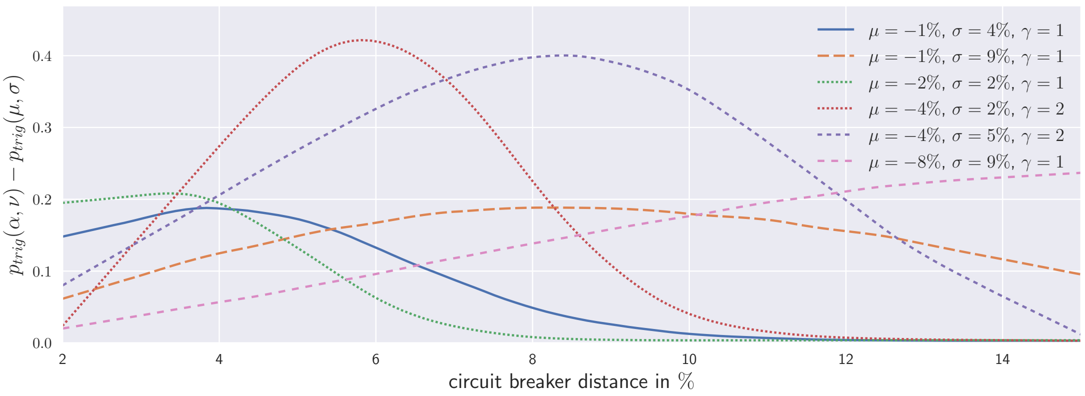


Figure 4: Dependence of the magnet effect (4.1) as a function of the distance of the opening price to the circuit breaker level $\underline{s}$. The magnet effect, as quantified by expression (4.1), is seen to hold for all parameter configurations. Since p_{trig} is independent of Δt_h, only $\Delta t_h = 0.1$ is shown.

market closure as well as the drift until the halting time, already prior to the halt.

5. Towards a More Robust Circuit Breaker Design

The previous section confirms the existence of a magnet effect for a price-based CB. This can be attributed to the fact that the market knows exactly about the trigger point $\underline{s}$ and the halting time Δt_h and account for a looming trading halt ex-ante. So is there a better CB design (i.e. a better trigger mechanism)? To answer this question, we consider a generalisation of our initial equation (2.3) in the form

$$s_t = f_t + \gamma p_{\text{trig}} \left(\mathbb{E}_t \left[f_{t_r} \right] - \mathbb{E}_t \left[s_{t_s} \right] \right) \tag{5.1}$$

where we still assume that $\mathbb{E}_t \left[s_{t_r} \right] = \mathbb{E}_t \left[f_{t_r} \right]$, i.e. that there is no aftermath effect or second circuit breaker. Different from equation (2.3), here we refrain from replacing $\mathbb{E}_t \left[s_{t_s} \right]$ by the fixed trigger threshold $\underline{s}$. After all, exchanges have the flexibility to define the trigger condition differently. In this section, we outline how this degree of freedom may be used to implement circuit breaker structures that may alleviate negative side effects such as increased volatility or magnet effect. Equation (5.1) in combination with the perturbative solution approach (A.7) provides a good starting point to examine different CB design. This section illustrates the generality and value of our proposed framework. An in-depth examination of different CB mechanisms is for future research.

A relatively straightforward way to avoid the negative side effects of increased volatility and magnet effect is by imposing a trading halt not based on a price level but based on a volatility level $\overline{\nu}$. Beyond its technical benefits, we propose that a volatility-based trading halt is also conceptually more relevant insofar as volatility is a measure of risks (and to some extent also of uncertainty), whereas a rapid price increases or decay may also be the result of a well-informed reaction to news. Moreover, as reported by Figlewski and Wang (2000), a large price drop usually triggers a sharp volatility increase, a phenomenon known as the leverage affect. Therefore, a volatility based CB is a priori more adapted to address price drops that are significant and abrupt.

To see further the technical benefits of a volatility based circuit breaker as can be examined with our framework, we first express the expectation values in (5.1) explicitly:

$$\mathbb{E}_t\left[f_{t_r}\right] = f_t + \mu\mathbb{E}_t\left[t_s + \Delta t_h - t\right] \tag{5.2a}$$

$$\mathbb{E}_t\left[s_{t_s}\right] = s_t + \mathbb{E}_t\left[\int_t^{t_s} \mathrm{d}\tau\ \alpha(s_\tau)\right] + \mathbb{E}_t\left[\int_t^{t_s} \mathrm{d}W_\tau\ \nu(s_\tau)\right] = s_t + \mathbb{E}_t\left[\int_t^{t_s} \mathrm{d}\tau\ \alpha(s_\tau)\right] \tag{5.2b}$$

Plugging (5.2a) and (5.2b) into (5.1) and grouping some terms yields

$$s_t = f_t + \gamma p_{\mathsf{trig}}\left(s_t - f_t\right) + \gamma p_{\mathsf{trig}}\left(\mu\mathbb{E}_f\left[t_s - t\right] - \mathbb{E}_t\left[\int_t^{t_s} \mathrm{d}\tau\ \alpha(s_\tau)\right]\right) + \gamma p_{\mathsf{trig}}\mu\mathbb{E}_t\left[\Delta t_h\right]. \tag{5.3}$$

A general treatment of expression (5.3) will require a perturbative analysis along the lines of (A.7). Here, we show the effects of a first order treatment. Assuming on the right-hand-side of (5.3) that $s = f$ (minimal perturbation of the price by the presence of the CB) and hence $\mathbb{E}_t\left[\int_t^{t_s} \mathrm{d}\tau\ \alpha(s_\tau)\right] = \mu\mathbb{E}_f\left[t_s - t\right]$, we find that, to first order

$$s_t^{(1)} = f_t + \gamma p_{\mathsf{trig}}\mu\mathbb{E}_t\left[\Delta t_h\right]. \tag{5.4}$$

The volatility ν is related to σ via Itô's lemma, $\nu = \sigma\partial s_t/\partial f_t$.[6] Within the price-based circuit breaker discussed so far, p_{trig} has an explicit dependence on the level of the fundamental f. The lower f, the more likely a CB is triggered. Therefore, $\partial p_{\mathsf{trig}}/\partial f < 0$ (cf. Figure 3 (b)) and

$$\nu^{(1)} = \sigma\frac{\partial s^{(1)}}{\partial f} = \sigma\left(1 + \underbrace{\gamma\mathbb{E}_t\left[\Delta t_h\right]}_{>0}\mu\frac{\partial p_{\mathsf{trig}}}{\partial f}\right) > \sigma \tag{5.5}$$

for $\mu < 0$, which is the relevant case for a circuit breaker at a lower boundary. This explains the increased volatility prior to a trading halt. By contrast, a volatility based CB does not depend explicitly on the level of f. This is intuitive: the level of the fundamental does not directly influence its volatility and is hence not predictive of the trigger probability or expected stopping time. Therefore, $\partial p_{\mathsf{trig}}/\partial f = 0$ to a first-order approximation and it follows from (5.5) that $\nu^{(1)} = \sigma$. At least to first order, the volatility-based circuit breaker thus avoids increased volatility levels, in stark contrast to price level based ones. Similarly, one can show that $\alpha^{(1)} \approx \mu$ up to an additional explicit time-dependence of p_{hit} (suggesting a decreased magnet effect). There may be, however, higher-order effects via $\frac{\partial p_{\mathsf{trig}}}{\partial f} = \frac{\partial p_{\mathsf{trig}}}{\partial s}\frac{\partial s}{\partial f}$ and the leverage effect. The iterative framework (A.7) allows one to take such dependencies into consideration.

In addition to stopping mechanisms, the construction of CBs has another degree of freedom: the stopping time. In this article, we have kept Δt_h as a fixed constant variable, as is the case for the vanilla CB mechanism. But this is not a requirement, and the exchange might also impose stochastic, event-dependent criteria. Studying the implications of such rules is relatively straightforward within the framework presented in this article. We leave this for future research.

6. Conclusions

We have studied the dynamics of an asset price in the presence of a circuit breaker (imposed trading halt). We have considered the customary form of a price-based circuit breaker (CB) whereby trading is halted for a time Δt_h after the price drops below an upper threshold $\overline{s}$ or lower threshold $\underline{s}$. At its starting point is a stochastic process specifying the dynamics of the (log) 'fundamental' value that would prevail if no circuit breaker was installed. Throughout this article, we have assumed that this stochastic process is a continuous random walk with constant drift μ and constant volatility σ. This assumption could be relaxed, and would generally not alter the main conclusions, however at the cost of analytical intractability.

The effect of the CB was then imposed on top of the dynamics of the fundamental f. We have proposed to describe the difference between the observed price s and the fundamental value f by a term that embodies the pricing of the expected evolution of f during the halting time when no trading occurs. This correction term was weighted by a term proportional to the probability that the CB is triggered. Since that probability depends itself on the realized price trajectory, a coupled McKean-Vlaslov equation was obtained in Section 2. Its solution was provided in section Appendix A, both numerically and analytically to first order. We found a modified, non-constant drift $\alpha(s)$ and volatility of $\nu(s)$ that differ from their constant counterparts μ and σ of the fundamental value. Most notably, for the price-based CB, we found an increased volatility ν as the price approaches the CB boundary, in agreement with previous empirical observations for this trigger mechanism.

In section 4, we then showed how the magnet effect can be accounted for naturally within our framework. The magnet effect expresses the fact that the sole presence of a price-based CB tends to push the price towards the trigger level faster than in its absence. We have shown that, within our theory, this magnet effect occurs for all parameter combinations that are realistic, and hence confirms previous studies for price-based CBs.

In section 5 we then showed how our framework can be easily generalised to other trigger mechanisms. And based on that, we then suggested an alternative circuit breaker design that aim at avoiding perturbing too much the price dynamics due to the presence of the circuit breaker. We have shown that a volatility-based CB e.g. mitigates negative side effects of price-based CBs, at least to first order, and argued that it is also more suitable conceptually, as volatility represents risk perceptions and some level of market uncertainty. The framework presented in this article provides a starting point to develop better CB designs.

Acknowledgements

The authors acknowledge stimulating discussions with Matthias Leiss and Alexander Lipton.

References

D. Abad and R. Pascual. On the magnet effect of price limits. *European Financial Management*, 13(5):833–852, 2007. doi: https://doi.org/10.1111/j.1468-036X.2007.00399.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-036X.2007.00399.x`.

D. Abad and R. Pascual. Holding Back Volatility: Circuit Breakers, Price Limits, and Trading Halts. In *Market Microstructure in Emerging and Developed Markets: Price Discovery, Information Flows, and Transaction Costs*. John Wiley & Sons, 2013.

C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory, 1999th edition*. Springer, 1999.

C. Benjamin, P. G., M. H., and S. P. Managing excess volatility: Design and effectiveness of circuit breakers. SAFE Working Paper 195, 2017. URL `http://hdl.handle.net/10419/174902`. urn:nbn:de:hebis:30:3-456406.

H. Berkman and J. B. T. Lee. The effectiveness of price limits in an emerging market: Evidence from the korean stock exchange. *Pacific-Basin Finance Journal*, 10(5):517–530, 2002. URL `https://EconPapers.repec.org/RePEc:eee:pacfin:v:10:y:2002:i:5:p:517-530`.

H. Chen, P. A., and W. J. The Dark Side of Circuit Breakers. *Working Paper*, 2018.

H. Chen, A. Petukhov, and J. Wang. The Dark Side of Circuit Breakers. *Working Paper*, 2018.

D. D. Cho, J. Russell, G. C. Tiao, and R. Tsay. The magnet effect of price limits: evidence from high-frequency data on taiwan stock exchange. *Journal of Empirical Finance*, 10(1-2):133–168, 2003.

S. A. Corwin and M. L. Lipson. Order Flow and Liquidity around NYSE Trading Halts. *The Journal of Finance*, 55(4): 1771–1805, 2000.

S. Figlewski and X. Wang. Is the 'leverage effect' a leverage effect? *Available at SSRN 256109*, 2000.

M. A. Goldstein and K. A. Kavajecz. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 7(3):301–333, 2004. ISSN 13864181. doi: 10.1016/j.finmar.2003.11.003.

P. Gomber, B. Clapham, M. Haferkorn, S. Panz, and P. Jentsch. Circuit Breakers - A Survey among International Trading Venues. *The Journal of Trading*, 12(1):42–54, 2017.

B. Greenwald and J. Stein. Transactional risk, market crashes, and the role of circuit breakers. *Journal of Business*, 64(4): 443–462, 1991.

M. Hanke, R. Poulsen, and A. Weissensteiner. The chf/eur exchange rate during the swiss national bank's minimum exchange rate policy: a latent likelihood approach. *Quantitative Finance*, 19(1):1–11, 2019.

N. Hautsch and A. Horvath. How effective are trading pauses? *Journal of Financial Economics*, 131(2):378–403, 2019. doi: 10.1016/j.jfineco.2017.12.

P. R. Krugman. Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, 106(3):669–682, 1991.

C. M. Lee, M. J. Ready, and P. J. Seguin. Volume, volatility, and new york stock exchange trading halts. *The Journal of Finance*, 49(1):183–214, 1994.

S. C. Lera and D. Sornette. An Explicit Mapping of Currency Target Zone Models to Option Prices. *International Review of Finance: 10.1111/irfi.12196*, 2018.

S. C. Lera, M. Leiss, and D. Sornette. Currency target zones as mirrored options. *The Journal of Derivatives*, 26(3):53–67, 2019. ISSN 1074-1240. doi: 10.3905/jod.2019.26.3.053.

L. Lin, M. Schatz, and D. Sornette. A simple mechanism for financial bubbles: time-varying momentum horizon. *Quantitative Finance*, DOI:10.1080/14697688.2018.1540881, 2018.

J. Magani and D. Munro. Dynamic runs and circuit breakers: an experiment. *Experimental Economics*, 23(1):127–153, 2020.

R. Mann and G. Sofianos. Circuit breakers for equity markets. *Market Volatility and Investor Confidence*, pages 225–253, 1990.

H. P. McKean. A class of markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences*, 56(6):1907–1911, 1966.

A. Molini, P. Talkner, G. Katul, and A. Porporato. First passage time statistics of Brownian motion with purely time dependent drift and diffusion. *Physica A*, 390(11):1841–1852, 2011.

Presidential Task Force on Market Mechanisms. The report of the presidential task force on market mechanisms, 1988.

I. M. Sifat and A. Mohamad. Circuit breakers as market stability levers: A survey of research, praxis, and challenges. *International Journal of Finance & Economics*, 24(3):1130–1169, 2019.

A. Subrahmanyam. Circuit breakers and market volatility: A theoretical perspective. *The Journal of Finance*, 49(1):237–254, 1994.

A. Subrahmanyam. On rules versus discretion in procedures to halt trade. *Journal of Economics and Business*, 47(1):1–16, 1995.

L. G. Telser. Margins and futures contracts. *Journal of Futures Markets*, 1(2):225–253, 1981.

H. K. Yong and J. J. Yang. What Makes Circuit Breakers Attractive to Financial Markets? A Survey. *Financial Markets, Institutions & Instruments*, 13(3):109–146, 2004. ISSN 0963-8008. doi: 10.1111/j.0963-8008.2004.00074.x.

Appendix A. Solution of the Circuit Breaker Equation

In this section, we present the solution of equation (2.3), which requires the determination of p_{trig} and $\mathbb{E}_t\left[t_s\right]$ with respect to two distinct probability measures, as we will now elaborate. We will find that the constant drift μ and volatility σ of the fundamental f are altered into non-trivial functional shapes $\alpha(s)$ and $\nu(s)$ for the observed price s. Readers who are interested only in a qualitative discussion of our results may skip the following subsections and continue with subsection 3.

Appendix A.1. Calculation of the Trigger Probability

Denote by $\alpha(x,t)$ and $\nu(x,t)$ the drift and volatility of $s_t = s(f_t,t)$, respectively, i.e. $\mathrm{d}s = \alpha \mathrm{d}t + \nu \mathrm{d}W_t$. This drift and volatility functions are related to the constant drift μ and volatility σ of f_t through Itô's lemma:

$$\mathrm{d}s = \underbrace{\left(\frac{\partial\phi}{\partial t} + \mu\frac{\partial\phi}{\partial f} + \frac{1}{2}\sigma^2\frac{\partial^2\phi}{\partial f^2}\right)}_{\equiv\alpha}\mathrm{d}t + \underbrace{\sigma\frac{\partial\phi}{\partial f}}_{\equiv\nu}\mathrm{d}W_t. \tag{A.1}$$

Let us further denote by $\rho(x,\tau;s_t,\underline{s})$ the probability density of the price x at time $\tau > t$ given that the price was s_t at time t, and in presence of an absorbing boundary at $\underline{s}$. Generally, this probability density ρ is the solution to the Fokker-Planck equation

$$\frac{\partial\rho}{\partial\tau} = -\frac{\partial}{\partial x}\left[\alpha(x)\,\rho(x)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\nu(x)^2\,\rho(x)\right] \tag{A.2}$$

with the Dirac-Delta initial condition $\rho(x,t) = \delta(x-s_t)$. Given the ρ that solves (A.2), the survival probability $S(t)$ that the process has not yet hit the boundary from time 0 up to time $t > 0$ calculates as $S(t;s_0) = \int_{\underline{s}}^{\infty}\mathrm{d}s\;\rho(s,t;s_0,\underline{s})$. At time t, when the price is at s_t, we are interested in the probability that the CB is triggered before the end of the dayt_c. Hence,

$$p_{\text{trig}}(t) = 1 - S(t_c - t; s_t) = 1 - \int_{\underline{s}}^{\infty}\mathrm{d}s\;\rho(s,t_c-t;s_t,\underline{s})\,. \tag{A.3}$$

Appendix A.2. Calculation of the Expected Reopening Time

With f_t following (2.1), the expected value taken at time t of f_τ with $\tau > t$ is given by $\mathbb{E}_t\left[f_\tau\right] = f_t + \mu(\tau - t)$. Conditional on a CB being triggered and, if the time t_r of reopening was known, we would thus have $\mathbb{E}_t\left[f_{t_r}|t_r\right] = f_t + \mu(t_r - t)$. But t_r is itself a random time since $t_r = t_s + \Delta t_h$, where t_s is the random first-passage hit time of the barrier. Taking the additional expectation over all possible $0 < t_r < 1$ yields the expected value of the fundamental upon market reopening as a function of the expected stopping time

$$\mathbb{E}_t\left[f_{t_r}\right] = f_t + \mu\left(\mathbb{E}_t\left[t_s\right] + \Delta t_h - t\right). \tag{A.4}$$

In order to calculate the expected stopping time $\mathbb{E}_t\left[t_s\right]$, we should not work directly with the density ρ from (A.2). As mentioned in the previous section, we have to condition all events on a CB being triggered before the end of the trading day. By contrast, equation (A.2) is the solution for the unconditional density in the presence of the absorbing boundary. To arrive at the conditional density ρ^h, we have to replace ρ by [10]

$$\rho^h \equiv \rho(s, t; s_0, \underline{s} \mid \text{trig}) = \frac{p_{\text{trig}}(s, t_c - t)\ \rho(s, t; s_0, \underline{s})}{p_{\text{trig}}(s_0, t_c)} \tag{A.5}$$

where p_{trig} is derived from the unconditional density ρ as explained above. [11] In the remainder of this article, we will be calculating the expected hitting probability with respect to the unconditional measure ρ and the expected stopping with respect to the conditional measure ρ^h. As we show in Appendix B, this distinction is particularly relevant in the case where the impact of the fundamental volatility σ dominates over that of the fundamental drift μ. For notational convenience, we refrain from writing p^{ρ}_{trig} and $\mathbb{E}^{\rho^h}_t\ [t_s]$.

Given density ρ^h, the first passage probability density $\text{FPD}(\tau)$ of the process first hitting $\underline{s}$ at time τ is determined from $\text{FPD}(\tau) = -\partial S^h(\tau)/\partial\tau$, where S^h is just the survival probability associated with ρ^h. The expected hitting time then calculates via

$$\mathbb{E}_t\ [t_s - t] = \int_0^{t_c - t} \mathrm{d}\tau\ \tau\ \text{FPD}(\tau)\ . \tag{A.6}$$

Appendix A.3. General Scheme Using Recurrence Equations

The value of p_{trig} depends on the current position s_t in the form of an initial value of the Fokker-Planck equation (A.2). The solution $\rho(x; \tau)$ then characterizes probabilistically the future trajectories of s_t. This is intuitively reasonable: the realized market price s_t is observed in real-time, and its probabilistic properties are inferred in order to anticipate its future value, which is then fed back onto the price. Dynamics where drift or volatility are functions of the process distribution are known as McKean-Vlasov processes, first studied by McKean (1966). Analytical solutions to this type of problem are rare. We thus turn to an iterative solution instead using the framework of perturbation theory (Bender and Orszag, 1999). Specifically, we treat the deviation of s from f as a perturbation, [12] such that we can replace equation (2.3) by the recurrence equation

$$s_t^{(n+1)} = f_t + p^{(n)}_{\text{trig}}\ \gamma\ \left[f_t + \mu\left(\mathbb{E}^{(n)}_t\ [t_s - t] + \Delta t_h\right) - \underline{s}\right] \tag{A.7}$$

---

[10] A ‘back of the envelope’-derivation of this results follows from Bayes’ theorem, $\rho(A|B) = \rho(B|A)\ \rho(A)/\rho(B)$, where, ‘A = being at position s at time t’ and ‘B = hitting the barrier’.

[11] Conditional on s_t triggering a CB, we know that s will be at $\underline{s}$ at time t_s. Therefore, the fundamental value f_t will be at $f_{t_s} = \phi^{-1}(\underline{s}, t_s)$ with t_s unknown. Thus, we can also write $\mathbb{E}_t\left[f_{t_r}\right] = \mathbb{E}_t\left[\phi^{-1}(\underline{s}, t_s)\right] + \mu\Delta t_h$ with the expectation over the first-passage density of t_s. But ϕ needed to calculate $\mathbb{E}_t\left[f_{t_r}\right]$ is itself the solution of (2.3). Below, we will present a perturbative solution which has, to first order, no explicit dependence on t_s. Working with ϕ^{-1} thereby renders our perturbative approach infeasible. In line with our overall approach, we thus stick to the phenomenological equation (A.4).

[12] Within intraday dynamics for the log prices considered in this article, it is reasonable to assume that $|f - \underline{s}| \ll 1$ at all times. Naturally, p_{trig} is bounded between 0 and 1. We derive an approximate upper bound estimate for γ by realizing that $\mathrm{d}s/\mathrm{d}f > 0$ is required for consistency and stability. Replacing s by its first-order approximation (A.10), we can solve the inequality $\mathrm{d}s^{(2)}/\mathrm{d}f > 0$ analytically, to find that $\gamma < \left(\text{Erf}(1) + \frac{2}{\sqrt{\pi e}} - 1\right)^{-1} \approx 3.87$. This might change somewhat by the presence of a drift, but we can at least conclude that γ is of order 1.

where the hitting probability p_{trig} and expectation value $\mathbb{E}_t^{(n)}[\cdot]$ are calculated from the n-th order approximation of ρ and ρ^h, respectively. Applying Itô's lemma (A.1) to $s^{(n)}$, we obtain the drift $\alpha^{(n)}$ and volatility $\nu^{(n)}$ at n-th order. These are then plugged into a numerical solver of the Fokker-Planck equation (A.2) to obtain $\rho^{(n)}$ and from there the hitting probability and expected hitting time at order n. This is finally plugged into (A.7) for the next iteration. The fixed point of the recurrence equation (A.7) then provides the solution of equation (2.3).

Appendix A.4. First Order Solution

In this section, we show explicitly the first order solution $s_t^{(1)} = \phi^{(1)}(f_t, t)$. Drift $\alpha^{(1)}$ and volatility $\nu^{(1)}$ are then obtained by taking numerical derivatives of $\phi^{(1)}$. Starting from $s_t^{(0)} = f_t$ such hat $\phi^{(0)}$ is the identify function, we find from (2.1) that $\alpha^{(0)} = \mu$ and $\nu^{(0)} = \sigma$. For constant drift and volatility, there is an explicit solution of equation (A.2) with absorbing boundary condition. Setting for notational convenience, and without loss of generality $\underline{s} = 0$, the solution reads

$$\rho^{(0)}(s,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\left[\exp\left(-\frac{(s_0 - s + \mu t)^2}{2\sigma^2 t}\right) - \exp\left(-\frac{2 s_0 \mu}{\sigma^2}\right)\exp\left(-\frac{(s_0 + s + \mu t)^2}{2\sigma^2 t}\right)\right] \tag{A.8}$$

from which we obtain

$$p_{\text{trig}}^{(0)}(s,t) = \frac{1}{2}\exp\left(-2s\mu/\sigma^2\right)\text{Erfc}\left(\frac{s - \mu t}{\sqrt{2t}\sigma}\right) + \frac{1}{2}\text{Erfc}\left(\frac{s + \mu t}{\sqrt{2t}\sigma}\right) \tag{A.9}$$

where Erfc is the complementary error function. [13] For the special case where $\mu = 0$, the term $\mathbb{E}_t[t_s]$ drops out and we can write the first-order approximation of (A.7) explicitly as

$$s_t^{(1)} = \phi^{(1)}(f_t, t) = f_t + \gamma \cdot (f_t - \underline{s}) \cdot \text{Erfc}\left[\frac{f_t - \underline{s}}{\sqrt{2(t_c - t)}\sigma}\right], \tag{A.10}$$

from which we derive

$$\alpha^{(1)} = -\sqrt{\frac{2}{\pi}}\frac{\gamma\sigma}{\sqrt{t_c - t}}\exp\left(\frac{(f - \underline{s})^2}{2(t_c - t)\sigma^2}\right) \tag{A.11a}$$

$$\nu^{(1)} = \sigma + \sqrt{\frac{2}{\pi}}\frac{\gamma}{\sqrt{t_c - t}}\,(\underline{s} - f)\,\exp\left(-\frac{(f - \underline{s})^2}{2(t_c - t)\sigma^2}\right) + \sigma\,\gamma\,\text{Erfc}\left[\frac{f - \underline{s}}{\sqrt{2(t_c - t)}\sigma}\right], \tag{A.11b}$$

via Itô's lemma (A.1). We have compared analytical and numerical form of $\alpha^{(1)}$ and $\nu^{(1)}$ and found an excellent match. [14] Similarly, for $\mu \neq 0$, we can write $\phi^{(1)}$ analytically up to quadrature of the term $\mathbb{E}_t^{(1)}[t_s - t] = \int_0^{t_c - t} \mathrm{d}\tau\,\tau\,\text{FPD}^{(1)}(\tau)$. Drift $\alpha^{(1)}$ and volatility $\nu^{(1)}$ are then obtained by taking numerical derivatives of $\phi^{(1)}$. The results are shown in Figure 3.

Higher order iterations, $n = 2, 3, \ldots$ can only be performed numerically. This is evident already from the case $\mu = 0$, as no closed-form solution of the Fokker-Planck equation (A.2) can be found for drift and volatility (A.11). For the remainder of this article, we thus work with the first order approximation, which already presents a fairly good approximation of the limit $n \longrightarrow \infty$.[14] For notational simplicity, we shall denote $\alpha^{(1)}, \nu^{(1)}$ by α, ν, respectively.

---

[13] See for instance Molini et al. (2011) for a derivation of the probability distribution in presence of an absorbing boundary.

[14] Details on how to solve this iteration numerically, as well as numerical evidence that $s^{(1)} \approx s^{(\infty)}$ is available from the authors upon request.

Appendix B. Comparing Conditional with Unconditional Density

Given conditional density ρ^h, one then derives the first-passage-density and the expected stopping time as explained in section Appendix A. The necessity to replace the unconditional density ρ by its conditional counter-part ρ^h is visualized in Figure B.5. s If μ is strongly negative ($\mu < -6\%$ daily change), it holds that $p_{\text{trig}} \lesssim 1$ and hence $\rho \approx \rho^h$, such that one can work with the unconditional density. For less extreme cases, the difference between unconditional and conditional density leads to differences in expected hitting times reaching levels of up to 0.5, i.e. a miss-calculation of up to half a trading day. In this article, we have worked with ρ^h.

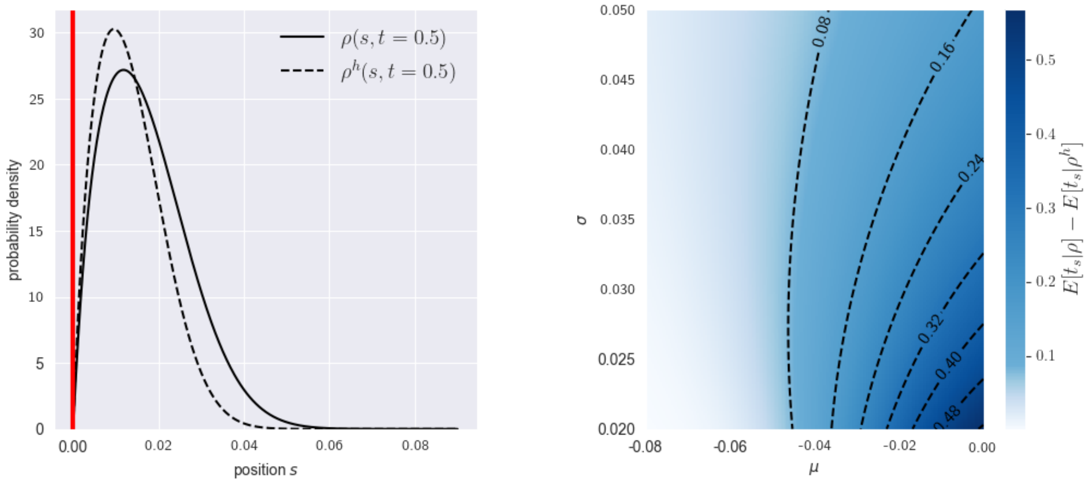


Figure B.5: In the left plot, we compare the conditional and unconditional densities, $\rho^h \equiv \rho(s, t; s_0, \underline{s} \mid \text{trig})$ and $\rho(s, t; s_0, \underline{s})$ respectively. The continuous line denotes the probability density $\rho(s, t)$ of a random walk with constant drift $\mu = -4\%$ and constant volatility 2%, in presence of an absorbing boundary at $\underline{s} = 0$ and a starting position at 0.03. The dashed line shows the associated conditional density $\rho^h(s, t)$ that corresponds to realizations such that $\underline{s} = 0$ is reached within a day, i.e. before $t = 1$. The plot on the right hand side depicts a systematic classification of the difference in expected hitting time based on ρ vs. ρ^h. For small drifts but still negative drifts, the difference is off by up to half a trading day.

Appendix C. Calculation of Trigger Probability and Expected Hitting Time

Consider a general, stationary stochastic process $\mathrm{d}s = \alpha \mathrm{d}t + \nu \mathrm{d}W_t$ for some drift function α, volatility function ν, initial position $s = s_0 > \underline{s}$ at time $t = 0$ and absorbing boundary condition at $s = \underline{s}$. The probability density $\rho(s, t)$ of the process at position s at time t is determined by the solution of the Fokker-Planck equation (A.2), which, in general, must be solved numerically. We are not aware of any readily usable numerical solver of the Fokker-Planck equation (for Python). While there exist numerical solvers of partial differential equations, it is often not trivial to adjust for absorbing boundary conditions, and Dirac-Delta initial distributions. We have thus written our own solver. [15] The hitting probability, and expected hitting time is then determined as follows:

1. The numerical solver takes as input two functions α and ν, a starting point s_0 and an end-time $T(=1)$, and returns the functional object $p(s,t) = p(s,t;s_0)$ which can be evaluated for $s \in [\underline{s}, \infty)$ and $t \in [0,T]$.
2. Using Simpson's numerical integration rule, this density is numerically integrated along s to yield the survival probability $S(t)$.
3. Thee probability to hit $\underline{s}$ from time $t = 0$ to time $t = T$ is given by $p_{\text{trig}}(T) = 1 - S(T)$.
4. Taking the negative derivative with respect to time, approximated through numerical first differences, we obtain the first-passage density FPD(t).
5. Numerical integration of $\int\limits_0^{t_c} \mathrm{d}t \; t \; \text{FPD}(t)$ up to $t_c = 1$, again with Simpson's rule, results in the expected hitting time.

---

[15] The code is available here: `https://github.com/slera90/`.

4 Dynamical Cost of Capital Driven by Dynamical Cash Flows

Valuing a firm with high expected cash flow growth is a difficult task. This can be seen in the high price volatility of firms for which the market has high cash flow growth expectations. This price volatility can stem from the hype surrounding these firms, but also is a consequence of the high sensitivity of standard valuation models to small changes in the expected growth rate path. Discounted cash flow (DCF) models are fundamentally driven by the assumed future path of cash flow growth and the discount rate used to value these uncertain future cashflows, with variations in either of these key assumptions producing enormous variability in the models' determination of fair value, thereby limiting their practical utility. With the NASDAQ Composite at its all-time high, the limited use of these standart DCF models when applied to high growth firms is a problem of great relevance, as large-scale mispricing can have sever economic consequences. This chapter is devoted to this problem by providing a better understanding of the often neglected relationship between the expected return (i.e. the discount rate) and cash flow growth and with it to propose a new valuation model.

While relevant literature is discussed at different places throughout the paper, I want to give a brief overview of the (vast) literature concerned with the estimation of the expected return of a stock in the next few paragraphs.

The most common approach to estimate the expected return is to employ a linear factor model, as suggested by the arbitrage pricing theory of **Ross** (1976). The most prominent examples thereby are the **Fama** and **French** (1992) three-factor model or the **Fama** and **French** (2015) five factor model. However, such linear factor models are not very accurate in estimating the cost of capital at the firm level (**Fama** and **French**, 1997). In addition, these linear factor models often completely lack theoretical motivation.

Common alternatives are implied cost of capital (ICC) models, pioneered by **Gebhardt** et al. (2001), which are more theoretically anchored, as a (valuation) model structure is necessary. Such ICC models often assume a cost of capital with a flat term structure. But besides various technical problems of such flat term structure ICC models, well documented by **Magni** (2010), the major drawback of these models is that they do not allow for the interpretation of an expected return process that can be realised by investing in the company today and holding it for a long time (**Fama** and **French**, 1996).

Lately, **Fama** and **French** (2015) started to redirect the "factor model community" onto more theoretical grounds. Using clean surplus accounting (**Miller,** M. and **Modigliani** (1961)) and some minor manipulation, they argue that the expected return of each stock is only determined by its price-to-book ratio and by expectations of its future profitability and investment. Variables not explicitly linked to this decomposition, such as size and momentum, that are found to help forecast returns, must only do so by implicitly improving forecasts of profitability and investment or by capturing horizon effects in the term structure of expected returns. Thus, only the so called "profitability", "investment" and "value" factors can be motivated from theory.

In the spirit of this argument, and based on solid theoretical developments from investment theory, **Hou** et al. (2020) then suggest an expected investment growth factor that is shown to forecast future investment and thus returns. While our paper has the same spirit, we suggest to go a slightly different route. Profitability, investment and profit growth form a triangle, i.e. profit growth is mechanically the result of invested capital and its corresponding profitability. And while **Hou** et al. (2020) focus on forecasting investment and, assuming constant profitability, thus implicitly capture the future profit growth dynamics, we suggest that it is more natural to focus on profit growth directly. Rather than forecasting investment, an investor would directly think about the future cash flow growth structure (also e.g. under the assumption of constant profitability). Thus, rather than concentrating on future investment, the paper presented in this chapter, **Solo** et al. (2021), brings future cash flow growth to the center of the attention. We briefly present the key ideas in the following paragraphs.

We start by analysing the risk exposure of a firm with regard to its cash flow growth rate level, using the consumption based asset pricing model, and find that those two variables are in good approximation linearly related to each other. We further show that this result can also be derived from basic economic intuition. Motivated by these insights, we then propose a time-varying discount rate which has the same structural form as the assumed cash flow growth rate process. This is our main contribution.

Using our time-varying discount rate in a DCF model, we present a valuation model that is less sensitive to the cash flow growth rate process assumptions. As our time-varying discount rate is proportional to the assumed cash flow growth rate term structure, these two naturally complementary variables interact to massively reduce the range of reasonable valuation outcomes of the model. This important feature greatly enhances the practical utility of a DCF model using our time-varying discount rate versus models that use either a constant discount rate or even a time-varying discount rate that is not explicitly proportional to the modelled path of future cash flow growth.

In addition, we elaborate a time-varying internal cost of capital (ICC) model that reflects the expected return implied by the assumed term-structure of the growth rate. This time-varying expected return process allows for a straightforward interpretation as investment return of a buy and hold strategy, compared to e.g. a constant ICC model, which by construction is a constant compound return. The predictive power of our time-varying ICC model for future returns is thereby confirmed in a Fama MacBeth regression model and the economic significance is confirmed in a trading strategy with yearly rebalancing.

In short, we show first that a time-varying discount rate, that borrows its structural form from the assumed cash flow growth rate process, aligns with economic intuition and empirical evidence. Second, we show this idea conceptually improves standard valuation models, as they get less sensitive to cash flow growth assumptions. And third, we show that it also corrects conceptual shortcoming of flat-term structure ICCs, as our proposed time-varying ICC does have the interpretation of an investment return of a buy and hold strategy, and thus is a valid proxy for the expected return (i.e. cost of capital). Our empirical results strongly support these statements.

In the end, I hope this chapter contributes towards a better understanding of the often neglected, but incredibly relevant relationship between the expected return and cash flow growth, and thus, helps investors make improved capital allocation decisions, especially when considering an investment in high growth firms. Further, I also hope to contribute to the literature stream that aims at refocusing the academic discussion around the expected return factor zoo. All kind of equity return factor are constantly suggested in the financial economics literature, without any theoretical justification. I suggest to go back to the essence of what an equity investors should think about, which only should be the expected return stemming from a discounted cash flow model, i.e. our proposed time-varying ICC.

References

E. **Fama** and K. **French**. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992

E. **Fama** and K. **French**. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84, 1996

E. **Fama** and K. **French**. Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193, 1997

E. **Fama** and K. **French**. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015

W. **Gebhardt**, C. **Lee**, and B. **Swaminathan**. Industry costs of equity. *Journal of Accounting Research*, 39(1): 135–176, 2001

K. **Hou**, H. **Mo**, C. **Xue**, and L. **Zhang**. An Augmented q-Factor Model with Expected Growth. *Review of Finance*, 25(1):1–41, 02 2020

C. **Magni**. Average internal rate of return and investment decisions: A new perspective. *The Engineering Economist*, 55(2):150–180, 2010

S. **Ross**. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976

D. **Solo**, D. **Sornette**, and F. **Ulmann**. Internal Cost of Capital Driven by Dynamical Cash Flows. *Reprint submitted to the Journal of Financial Economics*, 2021

Dynamical Internal Cost of Capital Driven by Cash Flow Growth

David Solo[a], Didier Sornette*[b,c], Florian Ulmann*[b]

[a] Diversified Credit Investments, LLC, San Francisco, California, USA

[b] ETH Zurich, Department of Management, Technology, and Economics, Switzerland

[c] Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies
Southern University of Science and Technology (SUSTech), Shenzhen, China

---

Abstract

Based on economic intuition and empirical evidence from the CCAPM, we argue that the term structure of expected returns should be identical to that of the cash flow growth rates. We introduce a discounted cash flow model with a time-varying expected return process matching the assumed cash flow growth rate process, which reduces the range of reasonable valuations outcomes to useful levels. In the same manner, we elaborate a time-varying internal cost of capital (ICC) model that reflects the expected return implied by growth rate process assumptions at each future point in time. In contrast to flat term structure ICCs, our time-varying ICC can be conceptually interpreted as an investment return process and thus can serve as valid proxy for the cost of capital. We show that our time-varying ICC model is indeed superior to the constant ICC models in a Fama-MacBeth regression setting to predict future realised returns. And using the expected return of the time-varying ICC model as control in a Fama-MacBeth regression of the profitability, the investment and the value factor, both the profitability and the value factor become insignificant in explaining future realised returns. The superiority and economic significance of the time-varying ICC model is further confirmed in a trading strategy with yearly rebalancing.

JEL: C20, C53, G11, G12, O16

Keywords: Internal Cost of Capital, consumption based asset pricing model, cash flow growth, equity risk premium, implied risk, Fama-MacBeth regression, time-varying risk premium, trading strategy, discounted cash flow model

---

*corresponding authors

Email addresses: `davidmsolo@outlook.com` (David Solo), `dsornette@ethz.ch` (Didier Sornette*), `fulmann@ethz.ch` (Florian Ulmann*)

Reprint submitted to the Journal of Financial Economics

1. Introduction

A solid theoretical foundation of the cost of capital is essential in accounting and finance research dealing with asset pricing and investigating market anomalies. And accurate estimates of the cost of capital are crucial for practitioners for optimal fund allocation. Nevertheless, academics as well as practitioners find it still challenging to estimate firms' costs of capital.

The most common approach used to estimate the cost of capital in the financial economics literature is to employ a linear factor model, such as e.g. the Fama and French (1992) three-factor model or the Fama and French (2015) five factor model. However, the very same authors have warned about the difficulties encountered in accurately estimating the cost of capital with linear factor models at the firm level (Fama and French, 1997).

Building on these concerns, Gebhardt et al. (2001) pioneered the implied cost of capital (ICC) approach, which is more theoretically anchored. The ICC approach begins by assuming a valuation model based on discounted cash flows, e.g. the dividend discount model. Next, some assumption about future cash flows are made, with which one can then solve for the ICC (also called the implied expected return or implied discount rate) that equates the present value of the expected future cash flows to the current stock price.

Such ICCs are often assumed to have a flat term structure (we refer to such a flat term structure ICC as "cICC", constant ICC, from now on; the model is discussed in Appendix A). cICCs are often used as a proxy for the next period expected return or the cost of capital. And the quality of those estimates are then examined with regard to their association to future realised returns. Botosan et al. (2011) discuss the empirical work done in this field, and they document mixed results. Some cICC measures show a positive association to future realised returns, but most of them show no significance.

Besides various technical problems of cICC models, well documented by Magni (2010), there is basically one major drawback in using cICC models as proxy for the next period expected return for stocks. Whenever one assumes a time-varying cash flow growth rate process in a valuation model, the expected return process should also be assumed to be time-varying, as growth and expected return are naturally proportional variables. So any constant compound expected return structure fails to appropriately reflect the expected return stemming from the assumed growth path at each future point in time. Thus, a cICC cannot have the interpretation of an expected return process that can be realised by investing in the company today and holding for a long time. And therefore cICCs are also not good proxies for the cost of capital (Botosan et al., 2011), whenever the growth rate process is assumed to be time-varying.

This argument is in the spirit of Fama and French (1996), who also emphasise that already by construction the constant ICC is a (constant) compound return, and not an expected one period simple return of the investment. In other words, the constant ICC is not the return that can be expected from investing a dollar in a firm at the beginning of the ICC estimation period (Fama and French, 1999). Hughes et al. (2009) formalise this and show that the constant ICC measure differs from the expected return on average if the expected return is stochastic in nature, based on which they call empirical findings from the ICC literature into question. In the fixed income literature for example, the insight that the constant ICC for bonds (called the bond yield) may differ from the bond's expected returns on average has been long recognised (Vasicek, 1977; Cox et al., 1985).

We take this point as motivation to introduce an ICC framework that conceptually has the interpretation of an investment return process since it properly reflects the term structure of the risk exposure implicitly assumed in the valuation model. Specifically, our main contribution is to introduce a time-varying expected return structure matching the term structure of the assumed cash flow growth rate process. One can motivate this term-structure either from an empirical argument or by intuition.

We then use the concept of a discount rate with the same term structure as the cash flow growth rate in a discounted cash flow (DCF) model as well as an ICC model. To test the quality of our idea, we execute standard factor regression analysis with the expected next period return suggested by our time-varying ICC measure. We find that it performs significantly better than the expected return obtained under a flat term structure ICC measure for explaining future realised returns. Using the expected return of the time-varying ICC model as control in a Fama-MacBeth regression of the profitability, investment and value factors of Fama and French (2015), both the profitability and the value factor become insignificant in explaining future realised returns.

A trading strategy using our time-varying measure shows a significant improvement of the information ratio with respect to the constant ICC model, especially if one focuses on the fastest growing firms for which the growth path becomes more relevant. This confirms the economic significance of our time-varying ICC framework.

Our work can find some underpinning in the discussion presented by Fama and French (2015), i.e. that only the profitability, investment and value factor can be motivated from theory. In their approach, the expected return of each stock is determined by its price-to-book ratio and by expectations of its future profitability and investment. Variables not explicitly linked to this decomposition, such as size and momentum, that are found to help forecast returns, must only do so by implicitly improving forecasts of profitability and investment or by capturing horizon effects in the term structure of expected returns. With this reasoning, Fama and French (2015) conclude that their three-factor model (Fama and French, 1992) should be extended to five factors (including profitability and investment). In the spirit of this argument, and based on solid theoretical developments from investment theory, Hou et al. (2020) then suggest an expected (investment) growth factor that is shown to forecast future investment and thus returns. Their basic intuition is that firms with high expected investment growth earn higher expected returns than firms with low expected investment growth, holding investment and expected profitability constant. They thus suggest to extend the q factor model of Hou et al. (2014) into an expected (investment) growth factor to form the q^5 factor model.

It is important to realise that profitability, investment and profit growth form a triangle: profit growth from t to $t+1$ is mechanically a function of (a) new invested capital at time t multiplied by the profitability from this new invested capital between t ant $t+1$ plus (b) previous (before t) invested capital multiplied by the change in profitability between t and $t+1$. Hou et al. (2020) focus on forecasting investment and assume constant profitability, and thus implicitly capture the future profit growth dynamics. We suggest that it is more natural to focus on profit growth directly. Inspiration from practice [1] suggests that, rather than forecasting investment, an investor would directly think about the future cash flow growth structure. Thus, rather than concentrating on future investment, our approach focuses on future cash flow growth, also holding expected profitability constant.

---

[1] This paper is motivated by the model developed by the first author for actual market investment purposes when he was working with portfolio investors in high growth companies while based in Palo Alto, California.

Further, instead of using a factor approach, we stay in the structural form of a classical discounted cash flow model and calculate the expected return as the time-varying internal cost of capital that equates our future cash flow projections to the current market price. Hou et al. (2020) interpret their q^5 model as a linear approximation to the firm-level cost of capital (given the nonlinear dynamics involved). We actually do capture the nonlinear dynamics, by using the structural form of a classical discounted cash flow model directly.

In summary, our contribution to this literature stream is to refocus the discussion around what could be loosely referred to as an expected return factor. We suggest to go back to the essence of what equity investors are obsessed with, which is the expected return coming from a discounted cash flow model given current prices, current profitability and their assumption of future cash flows.

The next section expands on the motivation for our model in terms of an expected return process with the same term structure as the cash flow growth rate process. Section 3 then presents our model in the context of stock valuation models. Section 4 presents our empirical results in terms of standard Fama-Macbeth factor regressions of future realised returns. We also show the strong economic significance of our framework when applied to trading high growth firms. Section 5 concludes.

2. Motivation for a Cost of Capital Model Driven by Cash Flow Growth

As mentioned before, we motivate our expected return term structure from economic intuition and from an empirical argument, which are successively presented in the next subsections.

2.1. Economic Intuition for a Cost of Capital Model Driven by Cash Flow Growth

Here, we present the economic intuition for our idea. In the long run, the return from any stock investment must be close to the performance of the underlying business, i.e. to its cash flow growth. And as cash flow growth rates usually are expected to decrease with firm age to the industry average, the expected return's term structure should have a similar structural form. Of course, changes in the valuation multiple can dilute or increase this price return.

Let us assume a firm has a high cash flow growth and a high valuation multiple. This firm's cash flow growth rate most probably will decrease to the industry average over the years, while the valuation multiple at the same time will be compressed. Thus, the price return will be a function of both the cash flow growth and the multiple compression. The multiple compression does not usually dilute all of the return stemming from the cash flow growth because, otherwise, there would be no incentive to invest in these inherent riskier high growth firms. This dynamic suggests that the expected return is always proportional to the cash flow growth rate, but not necessarily equal to it, due to dilution stemming from multiple compression. And this proportionality is maintained across the future time-horizon. In this picture, the expected return is assumed to have the same term-structure as the cash flow growth rate.

2.2. Empirical Motivation from the Consumption based Asset Pricing Model

The similarity of the term-structure of expected return and cash flow growth rate is additionally motivated with well established empirical measures of risk exposure. The empirical motivation stems from the insight that the risk exposure of a firm is found to be linearly proportional to its cash flow growth rate. Thus, the expected return, which itself is linearly proportional to the risk exposure, should be assumed to have the same term structure as the cash-flow growth rate. Such a time-varying expected return ensures that the risk exposure is appropriately reflected at each future point in time, which is why the expected return then also has a proper interpretation as the investment return that can be expected at each future point in time. This empirical argument is now presented in more details.

2.2.1. Revival of the CCAPM

In asset pricing models, expected returns are determined by the exposure of the return innovation to a source of risk. In the consumption based asset pricing model (CCAPM) of Rubinstein (1976), Lucas (1978) and Breeden (1979), risk exposure of a firm is measured by the consumption beta, the slope coefficient from regressing return innovations onto consumption innovations. The CCAPM states that differences in expected return across assets are determined by differences in the assets exposure to consumption risk (the systematic risk of the CCAPM). Similarly to the CAPM with the market beta, the CCAPM's consumption beta, a measure of the co-movement between asset return and aggregate consumption, determines the expected returns of the assets. More specifically, the covariance of the asset returns with the stochastic discount factor, the relevant measure of risk premium (also called price of risk), is proportional to the covariance of the asset returns with aggregate consumption growth in the CCAPM. The intuition is that the risk premium of an asset is determined by its ability to insure against consumption fluctuations.

This key financial economic insight has for long been discarded, as early empirical tests did not support the theory (Breeden et al., 1989; Hansen and Singleton, 1982). But recently the CCAPM was revived due to improved empirical treatment of the model [2]. We build on this revival of the CCAPM and now use two established measures of risk exposure and analyse them with respect to their relationship to cash flow growth rates. One is from Dittmar and Lundblad (2017), a measure of consumption risk exposure, i.e. the consumption beta. The other is from Bansal et al. (2005), a proxy for consumption risk exposure, i.e. the cash flow beta.

---

[2] First, empirical results improved by using longer time horizons. Bansal et al. (2005) first recognised that the systematic consumption risk, if measured over longer time horizons, is able to explain a large part of the cross-sectional variation in expected returns. This is because measuring earnings and consumption over the long run alleviates problems caused by short-term earnings management (Jones, 1991; Teoh et al., 1998), short-term consumption commitment, or delayed response (Parker and Julliard, 2005; Jagannathan and Wang, 2007). Similar to Bansal et al. (2005), Dittmar and Lundblad (2017) recently showed that, by measuring the consumption risk factor as the sum of lagged consumption innovations, the risk premia across assets are largely explained. This measure of Dittmar and Lundblad (2017) is the consumption risk exposure measure (consumption beta) that we use in this paper. An alternative proxy of the consumption risk exposure of a firm is the cash flow beta introduced by Bansal et al. (2005), which is also used in this paper. A second reason why the consumption beta failed to capture cross-sectional differences in risk premia could be due to the error in the measurement of consumption and of the consumption beta (Daniel and Marshall, 2005). In the presence of significant measurement error, estimated covariances between returns and consumption shocks would be contaminated and shifted downwards. This is called the error-in-variables problem. Under the classical measurement error model, the coefficient estimates are biased toward zero, which is known as the attenuation bias (Collot and Hemauer, 2020). However, besides the coefficient of the imprecisely measured variable, the coefficients of the other variables are also biased - known as contamination bias - and the direction of this bias is in general unknown (Collot and Hemauer, 2020). As the effect of measurement error may be less pronounced when one uses several lags of the consumption growth, the first point of using longer time horizons comes back into play.

2.2.2. *The Consumption Beta and Cash Flow Growth*

The consumption beta β suggested by Dittmar and Lundblad (2017) is a measure of the exposure of asset returns to consumption risk and is defined by

$$\prod_{j=0}^{K-1} R_{i,t-j} = a_i + \beta_i \sum_{j=0}^{K-1} \hat{\eta}_{t-j} + e_{i,t} \tag{2.1}$$

Dittmar and Lundblad (2017) estimate the consumption beta by regressing cumulated gross real excess returns R over K windows on cumulated innovations of log consumption growth η (the theoretical motivation for this regression model can be found in Appendix G). The innovation in log consumption growth $\hat{\eta}_{t-j}$ is the difference in log consumption growth in quarter $t-j$ (real, seasonally-adjusted, per capita) and its unconditional mean. An estimate of the univariate regression model presented in equation (2.1) for firm i then gives us the risk exposure or consumption beta $\hat{\beta}_i$. As both Bansal et al. (2005) and Daniel and Marshall (2005) find that the relation between the market return and consumption at eight quarters is best for interpreting the market risk premium, we use $K = 8$.[3]

Cash flows are calculated using data from the merged CRSP/Compustat database [4] and consumption data is taken from National Income and Product Account tables at the Bureau of Economic Analysis (see Appendix B for more details on the data).

We calculate the consumption β, our measure of risk exposure, by estimating the time-series regression of equation (2.1) on a firm level at each quarter. [5] We then fit and plot those estimated risk exposure β against the corresponding contemporaneous cash flow cumulative average growth rate (cagr over the total estimation period). [6] Figure 1 presents these results. We observe a statistically significant positive, linear dependence between the consumption beta and the cash flow growth rate level.

2.2.3. *Implications for the Expected Return Term Structure*

In the CCAPM, the expected excess return is given by

$$\mathbb{E}_t \left[r^i_\tau \right] - r^f_\tau = \beta^i_\tau \lambda_\tau . \tag{2.2}$$

where r^i_τ is the return of asset i between time t and time $\tau > t$, r^f_τ is the risk-free return between time t and time τ, β^i_τ is the exposure of asset i between time t and time τ, and λ_τ is the risk premium between time t and time τ.

---

[3] Note that Dittmar and Lundblad (2017) themselves find that $K = 4$ is optimal, but the difference to $K = 8$ is negligible.

[4] Using the indirect cash flow method: free cash flow = net income + accounting costs - changes in operating working capital - capital expenditures; Note that Bansal et al. (2005) find that risk exposure estimates are reasonably insensitive to how cash flows are measured.

[5] We follow the beta estimation regression approach suggested in chapter 8 of Bali et al. (2017). As we have quarterly data, we use a total time horizon for each estimation of 10 years, if available, and require a minimum of 20 valid data points (5 years).

[6] We make the fit in a quantile regression model to the median, as the consumption beta estimation is prone to generate positive outliers and as estimating the median relationship is more relevant for a per firm statement than an estimate based on the mean.

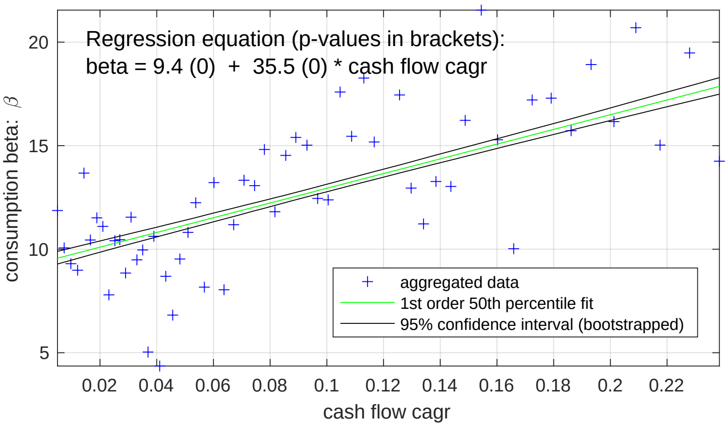


Figure 1: Consumption beta plotted against the cash flow growth rate. The risk exposure (consumption beta) increases with the absolute level of the cash flow growth rate in a linear fashion. The green line is the regression fit, the black lines are the 95% confidence intervals and the blue crosses display the aggregated data points (note that the regression is based on unaggregated data, but in order not to overload the figure, we aggregated the data points into quantile bins of the cash flow growth rate). The sample size behind the regression is large ($N \approx 274,000$), which naturally leads to high confidence. The results are focused on positive growth rates only, as for negative growth other dynamics come into play, which are not of interest here.

If we ignore the term structure of the risk premium λ_τ, as it is not firm specific, the linearity behind the risk exposure β^i_τ and the growth rate level suggests that the expected return process for firm i should just be given by a scaled version of the term structure of the growth rate process. As the risk exposure β^i_τ is found to have a positive linear relationship to the cash flow growth rate, $\beta^i_\tau \propto g^i_\tau$, it should be that β^i has the same term structure as g^i over all future times. And since the risk exposure β^i_τ is linearly proportional to the expected return $r^i_\tau \propto \beta^i_\tau$, r^i should be assumed to have the term same structure as g^i too, in order to represent the risk exposure appropriately at each future point in time. Thus, the expected return can and should be assumed to have the same term structure as the cash flow growth rate. [7] This is in line with Beaver et al. (1970), who show that abnormal earnings arising from growth opportunities are inherently more risky, leading to a positive association between the expected return and growth.

2.2.4. *Cash Flow Beta and Cash Flow Growth*

The cash flow beta provides an alternative solid proxy of the consumption risk exposure of a firm. Abel (1999) and Bansal and Yaron (2005) present a consumption-based model that shows that the differences in risk premia among assets mirror differences in the exposure of the asset's cash flows to consumption news[8]. Their theoretical and empirical work indicate that the covariance of the growth rates of firm cash flow with several lags of aggregate consumption growth rates identifies such a risk factor exposure of the firm. This risk factor exposure is called the cash flow risk exposure (or cash flow beta), and it measures differences in risk premia attributable to cash flow risk. They show that their measure of cash flow risk, the degree of co-movement of cash flow growth with consumption growth, is positively related to the risk premium of the corresponding firm (higher cash flow beta leads to a higher risk premium). They also show that it can explain a large part of the cross-sectional variation of risk premia across assets. Thus, we also present empirical results for the cash flow beta of Bansal et al. (2005).

---

[7] Note that we can neglect the intercept from figure 1, as it does not have a term structure by definition and also is not firm specific.

[8] Bansal et al. (2005) argue that, if asset returns have a multi-factor structure, then consumption betas, in the presence of cash flow and discount rate risks, may fail to account for the differences in the risk premia across assets, while both betas individually could explain it. This explains the failure of some tests of the CCAPM that measured and use the consumption beta directly.

The cash flow beta measures the cash flow risk exposure γ_i (cash flow beta) as the covariance of demeaned real cash flow growth rates with the sum of several lags of demeaned real aggregate consumption growth rates:

$$g_{i,t} = a_i + \gamma_i \left(\frac{1}{K} \sum_{k=1}^{K} g_{c,t-k} \right) + u_{i,t} \tag{2.3}$$

$g_{i,t}$ represents the demeaned log real cash flow growth rate and $g_{c,t}$ the demeaned log real growth rate in aggregate consumption (real, seasonally-adjusted, per capita). t is indicating years. We use $K = 2$ [9]. We calculate the cash flow γ_i by estimating the time-series regression of equation (2.3) on a firm level. We then plot these estimated risk exposure γ_i against the corresponding contemporaneous cumulative average cash flow growth rate (cagr over the total estimation period [10]). Figure 2 presents the respective regression results [11].

Cash flows are calculated using data from the merged CRSP/Compustat database [12] and consumption data is taken from National Income and Product Account tables at the Bureau of Economic Analysis (see Appendix B for more details on the data).

Also for the cash flow beta, another measure of risk exposure, we observe a statistically significant, positive linear relationship between the beta and the cash flow growth rate level.

Note that, for the definition of the cash flow beta γ_i, one already sees the linearity of the risk exposure to the cash flow growth rate level directly from equation (2.3), as the slope estimation in a univariate regression model is by definition linearly proportional to the dependent variable, which is the cash flow growth rate level in this case.

These results strengthen the motivation behind the framework of an expected return process that borrows its term structure from the assumed cash flow growth rate process. We build this logic into a DCF model in the next section.

---

[9] Same period as $K = 8$ in equation (2.1), when the regression was based on quarterly data

[10] Note that the cagr range is smaller than in figure 1. As we use yearly data in regression model (2.3), we estimate the time-series model over the total sample for each firm. Thus, the total estimation period is longer than for model (2.1), so naturally high contemporaneous cash flow cagr's are less frequently observed.

[11] We make the fit in a quantile regression model to the median, as the consumption beta estimation is prone to generate positive outliers and as estimating the median relationship is more relevant for a per firm statement than an estimate based on the mean.

[12] Using the indirect cash flow method: free cash flow = net income + accounting costs - changes in operating working capital - capital expenditures; Note that Bansal et al. (2005) find that risk exposure estimates is reasonably insensitive to how cash flows are measured.

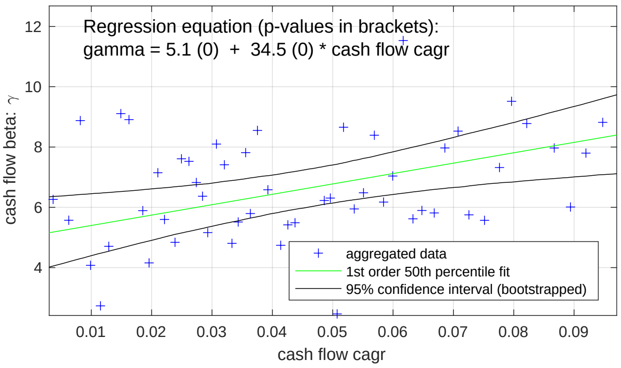


Figure 2: Cash flow beta plotted against the cash flow growth rate. The risk exposure increases with the absolute level of the cash flow growth rate in a linear fashion. The green line is the regression fit, the black lines are the 95% confidence intervals and the blue crosses display the aggregated data points (note that the regression is based on unaggregated data, but in order not to overload the figure, we aggregated the data points into quantile bins of the cash flow growth rate). The results are focused on positive growth rates only, as for negative growth, other dynamics come into play, which are not of interest here.

3. Cost of Capital Model Driven by Cash Flow Growth

As discussed before, using the constant internal cost of capital (ICC) model as proxy for the expected return for the next period has a major drawback. A constant ICC does not have the interpretation of an investment return, as its constant term structure fails to reflect the time-varying risk exposure associated with time-varying growth rates. We correct for this drawback by introducing an expected return process with the same term structure as the assumed growth rate process. This section now uses this time varying expected return process in the context of a DCF model as well as in an ICC model.

Our approach is in spirit most similar to what is presented by Da (2009). Using a model of equity strips, he shows that a high initial risk premium (stemming from consumption risk) decreases with time and the model choice of the cash flow process thereby determines the expected risk premium.

3.1. A DCF Model with a Time-Varying Discount Rate

This section uses the idea of a time-varying discount rate with the same term structure as the growth rate in the context of a DCF model. First we present the DCF equation in continuous form.

$$PV_t = \int_t^\infty \mathrm{d}\tau \; e^{-\int_t^\tau \mathrm{d}t' \; r(t')} \; \mathbb{E}_t \left[c(\tau) \right] \tag{3.1}$$

The index t represents the time, PV_t is the present value of all future expected cash flows: $\mathbb{E}_t \left[c_\tau \right]$ is the expected, at time t, cash flow c at future time $\tau > t$. The term $e^{-\int_t^\tau \mathrm{d}t' \; r(t')}$ is the time-varying discount factor using the time-varying expected return $r(t')$. We now impose a general structure for the evolution of the firm's cash flows:

$$\mathrm{d}c(\tau) = g(\tau)c(\tau)\mathrm{d}\tau + \nu(\tau)c(\tau)\mathrm{d}W_{\tau}^{'}, \tag{3.2}$$

where g is a growth rate that may change in time. Assuming that g is deterministic, we may write

$$\mathrm{d}\mathbb{E}_t\left[c(\tau)\right] = g(\tau)\mathbb{E}_t\left[c(\tau)\right]\mathrm{d}t. \tag{3.3}$$

We model cash flow growth rates as a deterministic mean reverting process (consistent with e.g. Nissim and Penman (2001), Menzly et al. (2004), Santos and Veronesi (2004) and Da (2009)). Specifically, we make the assumption that the convergence to the long-term growth rate g_∞ is described by an hyperbolic power law with exponent one

$$g(\tau) = \frac{g_t - g_\infty}{1+\alpha(\tau - t)} + g_\infty \tag{3.4}$$

where $\alpha > 0$. This expression (3.4) embodies the assumption that the growth rate of the cash flow converges to g_∞ as $\tau \longrightarrow \infty$.

As will become obvious later, the specific form of the cash flow growth rate process is not important for the resulting firm valuation in our model[13]. In principle, we can choose any other deterministic process in equation (3.4).

Solving equation (3.3) with (3.4) and initial condition $\mathbb{E}_t\left[c(\tau = t)\right] = c_t$ yields

$$\mathbb{E}_t\left[c(\tau)\right] = c_t\,\left(1-\alpha(\tau-t)\right)^{\frac{g_t - g_\infty}{\alpha}}\; e^{g_\infty(\tau - t)}. \tag{3.5}$$

Inserting equation (3.5) into (3.1), we get

$$PV_t = \int_t^\infty \mathrm{d}\tau\; e^{-\int_t^\tau \mathrm{d}t'\; r(t')}\left[c_t(1-\alpha(\tau-t))^{\frac{g_t-g_\infty}{\alpha}}\; e^{g_\infty(\tau-t)}\right] \tag{3.6}$$

Motivated by the findings presented in section 2, we now suggest to use a scaled version of the cash flow growth rate process as time-varying discount rate[14]

$$r(t') = r_{f,t'} + a\left(\frac{g_t - g_\infty}{1+\alpha(t'-t)} + g_\infty\right). \tag{3.7}$$

---

[13] In short, this is because of a compensation arising when the time-varying discount rate is proportional to the growth rate process, whatever its shape. In other words, our valuation model is not sensitive to the form of the cash flow growth rate process.

[14] In the CCAPM, the excess return $r^e_{i,t}$ of firm i is a function of the risk exposure ($\beta_{i,\tau}$, the consumption beta) scaled by the price of risk (λ_τ, the consumption growth rate), so $r^e_{i,t} = \beta_{i,\tau}\cdot\lambda_\tau$. From section 2, we know that $\beta_{i,\tau} \propto b\cdot g_{i,\tau}$, with b being the slope of figure 1 and $g_{i,\tau}$ being the cash flow growth rate for firm i. Thus, $r \propto b\cdot g_{i,\tau}\cdot\lambda_\tau$. Thus, $a = b\cdot\lambda_\tau$, the slope of figure 1 scaled by the price of consumption risk. Note that we neglect the intercept from figure 1 in this derivation, as it does not have a term structure by definition and is not firm specific.

The proposed time-varying discount rate properly reflects the assumed risk exposure at each future point in time stemming from the assumed growth rate path. Here, a is the scaled slope of figure 1, which we later estimate to be $a \approx 0.64$ (see lower panel of figure 4 or Appendix I). Inserting this into a DCF model leads to

$$PV_t = \int_t^{\infty} \mathrm{d}\tau \; e^{-\int_t^{\tau} \mathrm{d}t' \overbrace{\left[r_{f,t'} + a\left(\frac{g_t - g_\infty}{1+\alpha(t'-t)} + g_\infty\right)\right]}^{=r(t'),\text{ the time-varying discount rate}}} \left[c_t(1-\alpha(\tau-t))^{\frac{g_t-g_\infty}{\alpha}} \; e^{g_\infty(\tau-t)}\right] . \tag{3.8}$$

This is a first important result of our paper.

DCF models are fundamentally driven by the assumed future path of cash flow growth and the discount rate used to value these uncertain future cashflows, with variations in either of these key assumptions producing enormous variability in the models' determination of fair value, thereby limiting their practical utility. Our model solves this problem. By using a time-varying discount rate that is proportional to the *assumed* growth rate process, our valuation model is less sensitive to the cash flow growth rate process assumptions, as the discount rate is proportional to the assumed cash flow growth rate term structure, such that these two naturally complementary variables interact to massively reduce the range of reasonable valuation outcomes. This important feature greatly enhances the value of a DCF model using our time-varying discount rate versus models that use either a constant discount rate or even a time-varying discount rate that is not explicitly proportional to the modelled path of future cash flow growth.

The framework presented in this section is in contrast to most valuation models, which use constant discount rates, as e.g. the CAPM or the Fama-French three and five factor models. Guo (2004) attributes the failure of the CAPM from being empirically validated to precisely this static nature. Our idea is also supported by the work of Harvey (1989) and Jagannathan and Wang (1996), who introduce a conditional CAPM using time-varying beta's or risk premia, as well as Lyle and Wang (2015), who argue that a constant discount rate might be appropriate for the long term, but not suitable to value start-up or high growth firms. And even the (American Finance Association) presidential address of Cochrane (2011) highlighted the importance of time-varying discount rates in contemporaneous finance research.

3.2. A time-varying internal cost of capital (ICC) model

Equating equation (3.6) with the observable market value M_t, we get

$$M_t = \int_t^{\infty} \mathrm{d}\tau \; e^{-\int_t^{\tau} \mathrm{d}t' \; r(t')} \left[c_t(1-\alpha(\tau-t))^{\frac{g_t-g_\infty}{\alpha}} \; e^{g_\infty(\tau-t)}\right] . \tag{3.9}$$

The only unknown now is the expected return term structure $\int_t^{\tau} \mathrm{d}t' \; r(t')$. If we assume $r(t') = \bar{r}$ is constant over time, we get the cICC of this ICC model:

$$r^{cICC} \equiv \min_{\bar{r}} ||M_t - \int_t^{\infty} \mathrm{d}\tau \; e^{-\bar{r}(\tau-t)} \left[c_t(1-\alpha(\tau-t))^{\frac{g_t-g_\infty}{\alpha}} \; e^{g_\infty(\tau-t)}\right] || . \tag{3.10}$$

We have discussed above the drawbacks of using the r^{cICC} as proxy for the next period expected return. We now turn to a more appropriate model choice for the expected return process than to assume that it has a flat term structure.

As before, we build on the insights of section 2 that the expected return's term structure should be the same as the one of the assumed cash flow growth rate. But now, in an ICC model, this expected return process is additionally a function of the current price level M_t, or put different, of the risk exposure to a value factor (e.g. the price to cash-flow ratio[15]). Thus, the general form of the expected return process is not clear yet, as the process of the risk exposure to the value factor is not yet defined.

As discussed in section 2.1, intuition suggests that a valuation multiple of a high growth firm is compressed while the growth rate of the firm converges to the industry average. And while the valuation multiple is assumed to dilute some of the return from cash flow growth, it should not dilute all of it, because who would otherwise invest in the inherent riskier high growth firms. We thus assume that the price-to-cash flow ratio converges to an industry average with the same convergence process as the cash flow growth rate (consistent with e.g. Nissim and Penman (2001)). Or, in more academic terms, we assume that the risk exposure to a value factor follows a convergence process as given by equation (3.4) too, with the same α (i.e. it converges with the same time dynamics as the growth rate), but with other starting and end values.

Thus, the expected return process is a linear combination of two term structures, one stemming from the cash flow growth rate process and the other from the simultaneous valuation multiple compression[16]. Due to the simultaneity, both term structures are assumed to have the same α's, but different starting and end values, which then again results in a convergence process with the same α [17]. Thus, the general form of the expected return process is given by

$$r(\tau) = \frac{r_t - r_\infty}{1 + \alpha(\tau - t)} + r_\infty \,. \tag{3.11}$$

In this equation, the starting and end values of the total expected return process are denoted by r_t and r_∞ respectively.

The long-term value r_∞ is thereby the constant ICC after convergence of the cash flow growth rate and the cash flow-to-price ratio to the industry average. It is given by the ICC implied by the Gordon Growth model under the assumption of convergence of the cash flow growth rate and cash flow-to-price ratio to the industry average, so $g_\tau = g_\infty$ and $\frac{c_\tau}{M_\tau} = r_{cfy}$ (cfy for 'cash flow yield') respectively. If we define the time of convergence of the cash flow growth rate as t_c, we can write

---

[15] We call the price-to-cash flow ratio the risk exposure to a value factor. If one divides both sides of equation (A.2) by book value, one could also argue that the price-to-book ratio reflects the risk exposure to the value factor, and the cash flow-to-book ratio reflects the risk exposure to the profitability factor. Nevertheless, the price-to-cash flow ratio reflects both those dynamics, so we do not differentiate further here.

[16] This can easily be seen from the APT (presented in Appendix H), given that the price of risk does not have a firm specific term structure.

[17] Equation (3.11) is a linear combination of a scaled version of the cash flow growth rate term structure and the price-to-cash flow ratio term structure, both of which are given by equation (3.4), with different starting and end values, but the same α's. And since they have the same α's for the convergence processes, it is a linear combination of the respective risk exposure to growth and value in each point in time.

r_∞ with the Gordon Growth model as

$$r_\infty = r_{\tau \geq t_c} = \frac{c_\tau}{M_\tau}(1 + g_\tau) + g_\tau = r_{cfy}(1 + g_\infty) + g_\infty \, , \tag{3.12}$$

where r_{cfy} is the cash flow yield of the industry and g_∞ is the cash flow growth rate of the industry. In that sense, r_∞ is the industry ICC.

Inserting equation (3.11) into equation (3.6), we numerically solve for the only unknown r_t which is solution of

$$\min_{r_t} ||M_t - \int_t^\infty \mathrm{d}\tau \; e^{-\int_t^\tau \mathrm{d}t' \left(\frac{r_t - r_\infty}{1+\alpha(t'-t)} + r_\infty \right)} \left[c_t (1 - \alpha(\tau - t))^{\frac{g_t - g_\infty}{\alpha}} \; e^{g_\infty (\tau - t)} \right] || \, . \tag{3.13}$$

Last, we can obtain the next period expected return $r(t+1)$ using equation (3.11) with e.g. $\alpha = 1$, $\tau = t+1$, r_∞ as well as the estimated r_t. We denote the resulting next period expected return from the time-varying ICC model by r^{tvICC}.

$$r^{tvICC} \equiv r(t+1|\alpha, r_t, r_\infty) \, . \tag{3.14}$$

For the rest of the paper, we will use the notation r^{tvICC} to indicate next period expected returns predicted by the time-varying ICC model (tvICC for 'time-varying ICC'), given by equation (3.14), and r^{cICC} to indicate expected returns predicted by the constant ICC model (cICC for 'constant ICC'), given by equation (3.10).

Note that from a theoretical point of view, our r^{tvICC} is in the spirit of the loglinearized identity of Campbell and Shiller (1988). They generalize the Gordon growth model to the case in which the quantities are time-varying and relate the price-dividend ratio of an asset to its expected future log dividend growth and expected log returns. Our model is based on the same variables. But instead of using them in a two-period model based on loglinearization, which is shown to be problematic by Gao and Martin (2021), we solve for the ICC structure directly and thus capture more complex dynamics.

3.3. *Properties of the model*

Salient properties of the time-varying ICC process are presented in figure 3 and the corresponding next period expected return r^{tvICC} is studied in figure 4.

In figure 3, the parameters have been chosen such that the price-to-cash-flow ratio is already at the industry average. The time-varying ICC is thus just a scaled version of the cash flow growth rate process. And due to the empirically found linearity between the risk exposure and the cash flow growth rate, the time-varying ICC properly reflects the risk exposure at each future point in time. And since, in expectation, the future risk exposure is appropriately reflected, the time-varying ICC process, also at each future point in time, has the interpretation of an investment return that in expectation should be realised.

Let us give this point a bit more colour. The expected return process reflects the term structure of the risk exposure implied by the assumed cash flow growth rate path. Thus, at each future point in time, the expected return appropriately reflects the assumed risk exposure. Even in the long run this is true, as we assume that the total expected return process converges to the industry ICC given by the Gordon Growth model, so the expected return converges to a realistic, realisable long term return. Thus, at each future point in time, our process reflects the return an investor should expect given current information, which is in stark contrast to a flat term structure cICC model, as discussed before.

An investment today in the respective firm would result in the expected return process displayed by the time-varying ICC process of figure 3, and not in the constant compound return displayed by the constant ICC. Thus, the drawback of using a cICC model as proxy for the investment return is resolved by our time-varying ICC model, which can serve as accurate proxy for the expected return and the cost of capital.

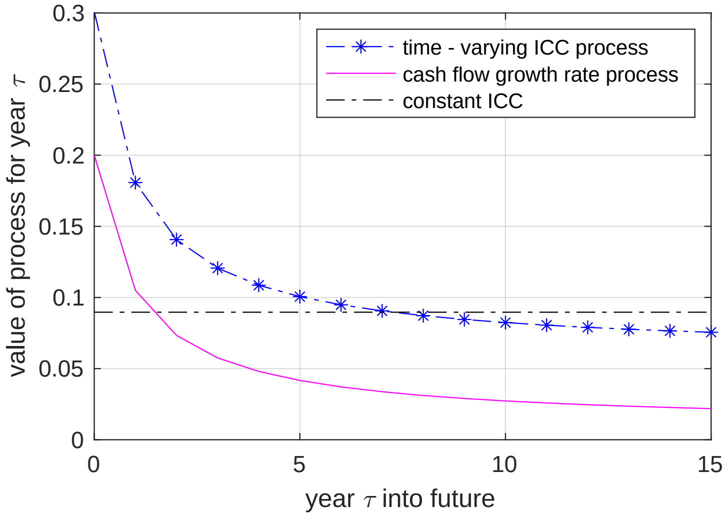


Figure 3: Salient properties of the time-varying ICC process given by equation (3.11), with r_∞ given by equation (3.12) and r_t solved with our time-varying ICC model given by equation (3.14) with the input parameters $\{M_t = 100mn,\ g_t = 0.2,\ g_\infty = 0.01,\ \alpha = 1,\ c(t) = 5mn,\ r_\infty = 0.061\}$ for $\tau = \{1 : 15\}$ years into the future. The horizontal dotted-dashed line indicates the value given by equation (3.10) of the constant ICC model.

The salient properties of the next period expected return of the time-varying ICC model, r^{tvICC}, are presented in figure 4. One can observe slopes as well as non-linear dependences with regard to all input parameters, which are in general stronger than for the cICC r^{cICC} given by equation (3.10). The expected return decreases with higher initial market prices (upper panel), increases with higher initial cash flow levels (middle panel) and increases with higher initial cash flow growth (lower panel). The r^{cICC} exhibits the same monotonic dependence as the r^{tvICC}, but with smaller amplitudes. The r^{tvICC} nicely recovers the empirically estimated linearity between the consumption beta (i.e. the scaled expected return) and the cash-flow growth rate level, inherent to our model construction[18].

---

[18] The value of the slopes between figure 1 and the lower panel of figure 4 are not comparable, since the former displays the return and the latter the beta. But using the proportionality of the return to the beta, we present a comparison of the theoretically predicted and the empirically estimated slope in Appendix I.

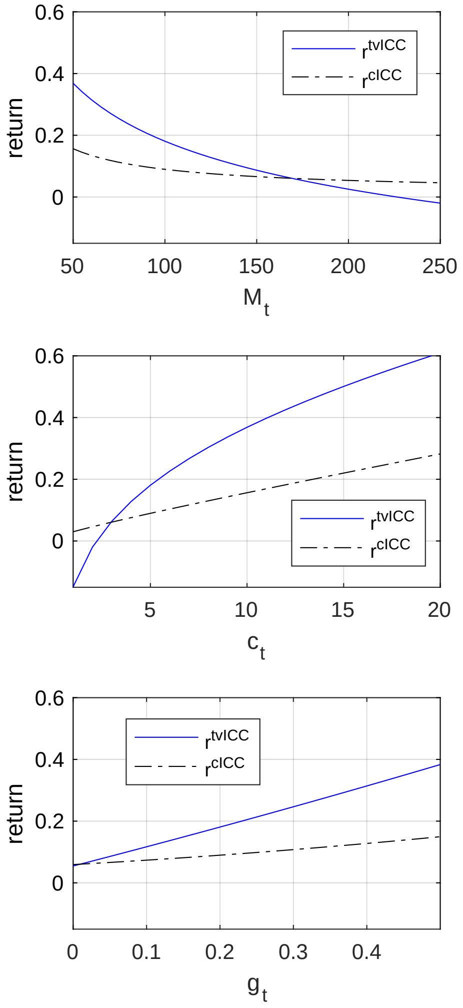


Figure 4: Properties of the next period expected return r^{tvICC} of the time-varying ICC model, given by equation (3.13), compared with the constant ICC r^{cICC} given by equation (3.10), as a function of a single parameter, while the others are fixed. The base values for the input parameters are: $\{M_t = 100mn, g_t = 0.2, g_\infty = 0.01, \alpha = 1, c(t) = 5mn, r_\infty = 0.061\}$. M_t is the market price of the asset; c_t is the cash flow per unit period; α controls the rate of decay of the cash flow growth to the long-term industry average according to equation (3.4); g_t is the growth rate of the cash flows defined in expression (3.2). The parameter set implicitly assumes that the price-to-cash flow ratios is already at its industry average, in order to isolate the return contribution stemming from the assumed cash flow growth rate path.

4. Statistical Significance for Future Realised Returns

4.1. Preliminaries

We now test the quality of the prediction of realised returns by the expected returns of the time-varying ICC model. We use the merged compustat/CRSP database (see Appendix B for details on the data) to numerically estimate the next period expected return r^{tvICC} (expected return as predicted by our time-varying ICC model) given by equation (3.13) for each firm i at each month t and use a Fama MacBeth (FM) regression approach to test if it has statistical relevance for future equity returns [19]. The realised return from month t to month $t+1$ is called $R_{i,t+1}$.

In order to benchmark our results, we further provide the empirical results for the next period expected return of the constant ICC model, r^{cICC}. In addition, we also analyse the significance of our next period expected return proxy r^{cICC} if we control for well known factors. We choose the value factor $BM_{i,t}$, the investment factor $INV_{i,t}$ and the profitability factor $OP_{i,t}$ as only these factors can be motivated by theory (e.g. by the model presented in equation (A.1), as discussed in Fama and French (2015) [20]), and thus present the ideal benchmark for our also theoretically well founded r^{tvICC}.

The summary statistics as well as correlations among all variables are presented in the Appendix C and Appendix D respectively.

Note that we do the empirical analysis on a firm level basis, using a FM regression approach, because it is our goal to provide an expected return proxy at the firm level. And since diversification into portfolios can mask cross-sectional phenomena in individual stocks and the choice of firm characteristics to sort portfolios has an impact on the significance and magnitude of risk premia, the portfolio approach is not appropriate for our purposes (Collot and Hemauer, 2020).

4.2. Fama MacBeth Regression Using the Expected Return

As described before, we want to analyse whether the calculated next period expected return $r(t+1)$ has predictive power for future equity returns, and how good its forecasts are relative to a constant ICC model and to those linear factor loading of Fama and French (2015), which can be motivated from theory. We thus run the following four FM regressions (see appendix Appendix E for more details on the FM regression approach):

$$R_{i,t+1} - r_{f,t} = a_{1,t} + \beta_t^{tvICC} r_{i,t}^{tvICC} + \epsilon_{1,i,t} \tag{4.1}$$

$$R_{i,t+1} - r_{f,t} = a_{2,t} + \beta_t^{cICC} r_{i,t}^{cICC} + \epsilon_{2,i,t} \tag{4.2}$$

---

[19] Appendix F discusses why empirical problems related to the FM regression approach are less pronounced for ICC expected return proxies.

[20] Using clean surplus accounting and some minor manipulation, we can write cash flows (i.e. dividends) as $c_t = (e_t - \Delta bv_t)$ (see Miller and Modigliani (1961) for details). So the cash flow is given by the earnings e and the change in book value of equity Δbv. Fama and French (2015) argue as followed. The decomposition of cash flows to $c_t = (e_t - \Delta bv_t)$ in equation (A.1) implies that the relevant expected return of each stock is determined by its price-to-book ratio and expectations of its future profitability and investment. If variables not explicitly linked to this decomposition, such as the size factor of the Fama and French (1992) three factor model or the momentum factor introduced in Carhart (1997), help forecast returns, they must do so by implicitly improving forecasts of profitability and investment or by capturing horizon effects in the term structure of expected returns.

$$R_{i,t+1} - r_{f,t} = a_{3,t} + \beta_t^{BM} BM_{i,t} + \beta_t^{INV} INV_{i,t} + \beta_t^{OP} OP_{i,t} + \epsilon_{3,i,t} \tag{4.3}$$

$$R_{i,t+1} - r_{f,t} = a_{4,t} + \beta_t^{tvICC} r_{i,t}^{tvICC} + \beta_t^{BM} BM_{i,t} + \beta_t^{INV} INV_{i,t} + \beta_t^{OP} OP_{i,t} + \epsilon_{4,i,t} \tag{4.4}$$

The idea of these four regression models is to first analyse r^{tvICC} in a univariate FM regressions (equation (4.1)), then compare it to r^{cICC} as the benchmark model (equation (4.2)), then estimate the standard $BM_{i,t}$ (book-to-market factor loading), $INV_{i,t}$ (investment factor loading), $OP_{i,t}$ (operating profitability factor loading) in the multivariate FM regression (equation (4.4)) to then see if those estimates change once we control with our $r_{i,t}^{tvICC}$ measure (equation (4.4)).

The independent variables are the factor loadings $r_{i,t}^{tvICC}$, $r_{i,t}^{cICC}$, $BM_{i,t}$, $INV_{i,t}$, $OP_{i,t}$, which have been calculated for each firm i at each month t. The dependent variable is the realised monthly excess return (in excess of the risk-free rate $r_{f,t}$) of stock i measured from month t to month $t+1$. In order to have equal units, we also work with monthly returns for $r_{i,t}^{tvICC}$ as well as for $r_{i,t}^{cICC}$.

Each month t, the cross-sectional linear regressions over all firms i are estimated. Thus, the resulting estimates of $\{\beta_t^{tvICC}$, β_t^{cICC}, β_t^{BM}, β_t^{INV}, $\beta_t^{OP}\}$ provide the estimated prices of risk or risk premia for all the factors at month t. The time averages of these estimations of the monthly price of risk are then taken as the final estimates for the prices of risk for all the factors. To examine whether these obtained prices of risks are statistically different from zero, we calculate the standard errors and the associated p-values to test the null hypothesis that the average coefficients are equal to zero. The standard errors are adjusted for heteroscedasticity and auto-correlation, following Newey and West (1987). Note that, in a FM regression, a statistically significant average slope coefficient indicates a cross-sectional relation between the given independent variable and the dependent variable in the average time period [21].

All results are presented in table 1.

---

[21] Appendix F discusses why empirical problems related to the FM regression approach are less pronounced for ICC expected return proxies.

	model (4.1)	model (4.2)	model (4.3)	model (4.4)
β^{tvICC}	0.072**	-	-	0.077***
β^{cICC}	-	0.1	-	-
β^{INV}	-	-	-0.01***	-0.0056***
β^{OP}	-	-	0.0037***	-0.0001
β^{BM}	-	-	0.0016**	0.0006

$*, **, ***$ indicates $p < 0.1, p < 0.05, p < 0.01$ respectively.

Table 1: Fama MacBeth regression results and their respective p-values.

Two points deriving from those results are worth highlighting.

(1) Comparing the estimates of the two univariate models (4.1) and (4.2), β^{tvICC} is found significant, while β^{cICC} is not. So we find a cross-sectional relation between our r^{tvICC} and the realised returns, while there is none for the r^{cICC}. And the parameter estimate of β^{tvICC} is positive, implying that a higher next period expected return is followed by a higher realised return, which makes sense economically [22]. The two models only differ in the assumptions with respect to the expected return process, namely that (4.1) stems from a time-varying ICC process that properly reflects the risk exposure at each future point in time, while (4.2) stems from the assumption that the expected return is constant over time. Thus, the model with the time-varying expected return process has some value. Note that the insignificance of β^{cICC} is consistent with the results of Botosan et al. (2011).

(2) Comparing the estimates of the two multivariate models (4.3) and (4.4), adding the next period expected return r^{tvICC} to the regression strongly changes the estimates of the well known OP and BM factor loading. While BM and OP are significant in model (4.3), they loose significance in (4.4) after controlling with the next period expected return r^{tvICC}. Remarkably, the estimate of β^{tvICC} has almost the same numerical value as in the univariate model β^{tvICC} and again is significant even after controlling for all other factor loadings (i.e. $BM_{i,t}$, $INV_{i,t}$ and $OP_{i,t}$). One can thus conclude that r^{tvICC} captures the dynamics intended to be reflected by OP and BM better than they do. As the model behind r^{tvICC} is able to reflect non-linear dynamics, this is not completely surprising. The estimate for INV on the other hand stays significant even after controlling with r^{tvICC}, while being reduced by almost half in numerical value.

4.3. Economic significance

Statistically significance is important but not sufficient and should be complemented by a test of the economic value of our time-varying ICC model. For this, we use the expected return models r^{tvICC} and r^{cICC} to implement a straightforward trading strategy. Each year by end of June, we perform our expected return forecasts for all firms, and group them in five quintiles, with the first (resp. last) quintile corresponding to the 20% lowest (resp. highest) return forecasts. We then form five equally weighted portfolios respectively invested in the firms in each of the five quintiles. For instance, portfolio 5 is invested in the firms making the fifth quintile of the distribution of expected return forecasts. Each portfolio is formed at the end of June of each year and is held for one year. Then, the five quintiles are recalculated for the next year. The corresponding portfolios sell their positions and then reinvest on the firms making the new quintiles. For instance, portfolio 5, which was invested in the firms making the fifth quintile of the distribution of expected return forecasts of the previous time period, is exiting all its positions and buying all the firms in the new calculated fifth quintile. The quality of this strategy is estimated by calculating the information ratio against the benchmark portfolio consisting of all firms. The left panel on figure 5 plots the information ratio of the five quintile portfolios for the r^{tvICC} and r^{cICC} models.

---

[22] A change in the expected return $r^{tvICC}_{i,t}$ of 1% on average (over all companies and times) results in a change in the observed return of that company of 0.072% for the following month ($0.01 * \beta^{tvICC}$). We show that this can be translated into a valuable trading strategy below.

Since we claim that our model ideally discounts future risk exposure, stemming from the growth path, we additionally want to gauge how good our model is in distinguishing between good and bad performing firms within a subset of firms with only high cash flow growth. This is displayed on the right panel on figure 5. We first remove a fraction of the slowest growing firms from the initial set of all firms. Second, we then construct the portfolio as the 5th quintile, i.e. the 20% highest return forecasts, of the remaining fraction of fast growing firms. The remaining fraction of fast growing firms is given by the x-axis in figure 5. The information ratio is thereby always calculated against the benchmark consisting of all firms in the corresponding fraction of fast growing firms. The idea behind this approach is to analyse how the quality of our model to distinguish between the good performing firms and the bad performing firms changes, if we apply it on a subset of firms characterized by high cash flow growth, as those firms are usually more noisy by nature, and thus more difficult to value. And since we claim that our model ideally discounts future risk exposure stemming from the growth path, we want to analyse how consistent our model is in distinguishing the good performing firms from the bad performing ones within a subset of firms especially exposed to this growth risk.

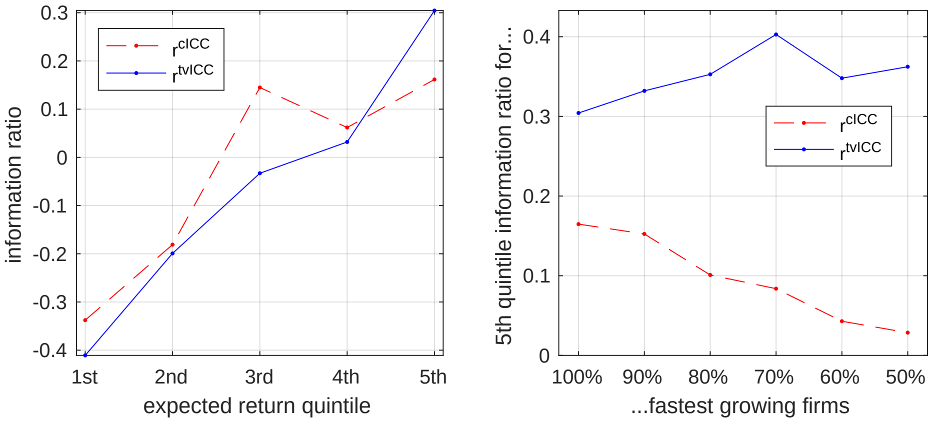


Figure 5: **Left panel**: information ratio against the benchmark portfolio consisting of all firms plotted as a function of the quintile order defined as follows. Using the expected return models r^{tvICC} and r^{cICC} in a trading setting, each year by end of June, we perform our expected return forecasts for all firms, and group them in five quintiles, with the first (resp. last) quintile corresponding to the 20% lowest (resp. highest) expected return forecasts. We then form five portfolios consisting of the firms in each of the five quintiles. For instance, portfolio 5th is always made with firms found in the 5th quintile. Each portfolio is formed at the end of June of each year and is held for one year. Then, each quintile portfolio is recalculated for the next year and so on. **Right panel**: Information ratio of the 5th quintile among the fraction of faster growing firms given along the x-axis. The information ratio is calculated against the benchmark consisting of all firms in the corresponding fraction of faster growing firms. For instance, 70% means that we have removed the 30% slowest growth firms among the whole set of firms and then constructed the 5th quintile (the 20% highest return forecasts) of the 70% faster growing firms. And the information ratio is calculated against the benchmark consisting of all firms in the 70% faster growing firm set.

The following points are worth highlighting.

(1) Left panel: The time-varying ICC model shows an approximately smooth monotonic increase in the information ratio, with the largest increase from the 4th to the 5th quintile. This suggests that r^{tvICC} can distinguish between good and bad performing firms. The constant ICC model on the other hand exhibits a noisy plateau for the three top quintiles, suggesting that the model r^{cICC} cannot distinguish between good and bad performing firms. This is in line with the results of table 1, in which r^{tvICC} is shown to have predictive power for future equity returns, while r^{cICC} has none.

(2) Left panel: The 5th quintile, which is made of the firms that are expected to be the top 20% performers in each one year investment period, is considerably better for the r^{tvICC} than for the r^{cICC}, again suggesting that r^{tvICC} has value in forming portfolios of good performing firms.

(3) Right panel: The more we focus on a subset of faster growing firms prior to constructing the 5th quintile portfolio with the r^{cICC} and r^{tvICC} models, the worse is the information ratio of the r^{cICC} portfolio, while the performance of the r^{tvICC} portfolio tends to increase even further. This clearly shows that our model is consistently good in distinguish the good performing firms from the bad performing firms, even within a subset of fast growing, and thus rather noisy firms. This again shows that it is key to reflect the future risk exposure stemming from the growth path with a time-varying expected return structure.

(4) Last, the absolute level of the information ratio of the time-varying ICC model is also reasonably good. And focusing only on the 50% fastest growing firms prior to choosing the 20% best expected performing firms (last data point on the right of the right panel indicated as ‘50%’ on the abscissa), the over-performance of the r^{tvICC} portfolio over the r^{cICC} portfolio is more than 0.3 as measured by their information ratios.

5. Conclusion

Based on economic intuition and empirical evidence from the CCAPM that risk exposure is linearly proportional to the growth rate, we argue that the term structure of the expected return should be the same as the one of the cash flow growth rate.

Based on this insight, we motivate an asset pricing model with a time-varying discount rate that reflects the implicitly assumed risk exposure stemming from the cash flow growth path of the valuation model at each future point in time. In contrast to DCFs with an expected return with a flat term structure, our valuation model is less sensitive to the assumptions underlying the cash flow growth rate process, as the discount rate is proportional to the assumed cash flow growth rate term structure, such that these two naturally complementary variables interact to massively reduce the range of reasonable valuation outcomes.

Similar, we introduce an internal cost of capital model, in which the expected return has a time-varying structure that reflects the term structure of the implicitly assumed risk exposure stemming from the assumed growth path, as well as the risk exposure to a value factor at each future point in time. Our time-varying ICC model resolves the major drawback of flat term structure ICC models when used as proxy for the next period expected return. Indeed, our time-varying ICC process can be interpreted as an investment return process that can be realised in expectation at each future point in time, as the term structure of the risk exposure is properly reflected over the future horizon.

We have shown that the time-varying ICC model is superior to the constant ICC model in a FM regression setting, as the time-varying ICC model is significant in predicting future realised returns, while the constant ICC model is not. As both models only differ with respect to the structural assumptions of the ICC process, we conclude that a time-varying ICC process, which reflects the implicitly assumed risk exposure at each future point in time, is key in valuing firms and in calculating their expected returns. Further, when we use the expected return of the time-varying ICC model as control in a FM regression with the profitability factor, the investment factor and the book-to-market factor loading, both the profitability and the value factor loading become insignificant in explaining future realised returns. This suggests that our time-varying ICC model expected return proxy reflects these dynamics better.

The superiority of the time-varying ICC model is also confirmed in a trading strategy with yearly rebalancing, especially when focusing on high growth firms for which the time-varying structure does a better job in reflecting the risk exposure stemming from cash flow growth.

Acknowledgements

We thank Emmanuel Munich and Tatyana Kovalenko for early explorations of how to best reflect high growth in a valuation model and thus laying the ground for the present work. We also thank Sandro Lera for insightful discussions and Yannick Malevergne for valuable feedback on prior versions of the paper.

References

Ken french's data library. `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. Accessed: 2020-07-28.

A. B. Abel. Risk premia and term premia in general equilibrium. *Journal of Monetary Economics*, 43(1):3–33, 1999.

T. Bali, R. Engle, and S. Murray. *Empirical Asset Pricing: The Cross Section of Stock Returns: An Overview*, pages 1–8. American Cancer Society, 2017. ISBN 9781118445112. doi: 10.1002/9781118445112.stat07954. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07954`.

R. Bansal and A. Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59 (4):1481–1509, 2005.

R. Bansal, R. F. Dittmar, and C. T. Lundblad. Consumption, dividends, and the cross section of equity returns. *The Journal of Finance*, 60(4):1639–1672, 2005.

W. Beaver, P. Kettler, and M. Scholes. The association between market determined and accounting determined risk measures. *The Accounting Review*, 45(4):654–682, 1970.

C. A. Botosan, M. A. Plumlee, and H. Wen. Industry costs of equity. *Contemporary Accounting Research*, 28(4):1085–1122, 2011.

D. Breeden. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(1):265–296, 1979.

D. T. Breeden, M. R. Gibbons, and R. H. Litzenberger. Empirical test of the consumption-oriented capm. *The Journal of Finance*, 44(2):231–262, 1989.

J. Y. Campbell. Two puzzles of asset pricing and their implications for investors. *The American Economist*, 47(1):48–74, 2003. ISSN 05694345. URL `http://www.jstor.org/stable/25604268`.

J. Y. Campbell and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1988.

M. Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82, 1997.

J. Cochrane. *Asset pricing*. Princeton Univ. Press, Princeton [u.a.], 2001. ISBN 0691074984. URL `http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+322224764&sourceid=fbw_bibsonomy`.

J. H. Cochrane. Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108, 2011. doi: https://doi.org/10.1111/j.1540-6261.2011.01671.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2011.01671.x`.

S. Collot and T. Hemauer. A literature review of new methods in empirical asset pricing: omitted-variable and errors-in-variable bias. *Financial Markets and Portfolio Management*, 1(1):2373–8529, 2020.

J. C. Cox, J. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(1):385–407, 1985).

Z. Da. Cash flow, consumption risk, and the cross-section of stock returns. *The Journal of Finance*, 64(2):923–956, 2009.

K. Daniel and D. Marshall. Equity-premium and risk-free-rate puzzles at long horizons. *Macroeconomic Dynamics*, 1(2): 452–484, 2005.

R. F. Dittmar and C. T. Lundblad. Firm characteristics, consumption risk, and firm-level risk exposures. *Journal of Financial Economics*, 125(2):326–343, 2017.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.

E. F. Fama and K. R. French. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84, 1996.

E. F. Fama and K. R. French. Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193, 1997.

E. F. Fama and K. R. French. The corporate cost of capital and the return on corporate investment. *Journal of Finance*, 54 (6):1939–1967, 1999.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81(3): 607–636, 1973.

C. Gao and I. W. R. Martin. Volatility, valuation ratios, and bubbles: An empirical measure of market sentiment. *The Journal of Finance*, n/a(n/a), 2021. doi: https://doi.org/10.1111/jofi.13068. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13068.

W. Gebhardt, C. Lee, and B. Swaminathan. Industry costs of equity. *Journal of Accounting Research*, 39(1):135–176, 2001.

H. Guo. A rational pricing explanation for the failure of capm. Working paper, Federal Reserve Bank of St. Louis, 2004.

L. P. Hansen and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5):1269–1286, 1982.

L. P. Hansen and K. J. Singleton. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, 91(1):249–265, 1983.

C. R. Harvey. Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, 45(5):38–45, 1989. doi: 10.2469/faj.v45.n5.38. URL https://doi.org/10.2469/faj.v45.n5.38.

K. Hou, C. Xue, and L. Zhang. Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, 28(3): 650–705, 09 2014. ISSN 0893-9454. doi: 10.1093/rfs/hhu068. URL https://doi.org/10.1093/rfs/hhu068.

K. Hou, H. Mo, C. Xue, and L. Zhang. An Augmented q-Factor Model with Expected Growth*. *Review of Finance*, 25(1): 1–41, 02 2020. ISSN 1572-3097. doi: 10.1093/rof/rfaa004. URL https://doi.org/10.1093/rof/rfaa004.

J. Hughes, J. Liu, and J. Liu. On the relation between expected returns and implied cost of capital. *Review of Accounting Studies volume*, 14(2):246–259, 2009).

R. Jagannathan and Y. Wang. Lazy investors, discretionary consumption, and the cross-section of stock returns. *The Journal of Finance*, 62(4):1623–1661, 2007.

R. Jagannathan and Z. Wang. The conditional capm and the cross-section of expected returns. *The Journal of Finance*, 51 (1):3–53, 1996. ISSN 00221082, 15406261. URL http://www.jstor.org/stable/2329301.

J. J. Jones. Earnings management during import relief investigations. *Journal of Accounting Research*, 29(2):193–228, 1991.

R. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(1):1429–1445, 1978.

M. R. Lyle and C. C. Wang. The cross section of expected holding period returns and their dynamics: A present value approach. *Journal of Financial Economics*, 116(3):505–525, 2015. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2015.03.001. URL https://www.sciencedirect.com/science/article/pii/S0304405X15000252.

C. A. Magni. Average internal rate of return and investment decisions: A new perspective. *The Engineering Economist*, 55 (2):150–180, 2010.

L. Menzly, T. Santos, and P. Veronesi. Understanding predictability. *Journal of Political Economy*, 112(1):1–47, 2004.

M. Miller and F. Modigliani. Dividend policy, growth, and the valuation of shares. *Journal of Business*, 34(4):411–433, 1961.

S. Nallareddy, E. Rouen, and J. C. Suarez Serrato. Do corporate tax cuts increase income inequality? Working Paper 24598, National Bureau of Economic Research, May 2018. URL `http://www.nber.org/papers/w24598`.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

D. Nissim and S. H. Penman. Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies*, 6(1):109–154, 2001.

J. A. Parker and C. Julliard. Consumption risk and the cross section of expected returns. *Journal of Political Economy*, 113 (1):185–222, 2005.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976. ISSN 0022-0531. doi: https://doi.org/10.1016/0022-0531(76)90046-6. URL `https://www.sciencedirect.com/science/article/pii/0022053176900466`.

M. Rubinstein. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics*, 7(1): 407–425, 1976.

T. Santos and P. Veronesi. Conditional betas. Working Paper 10413, National Bureau of Economic Research, April 2004. URL `http://www.nber.org/papers/w10413`.

J. Shanken. On the Estimation of Beta-Pricing Models. *The Review of Financial Studies*, 5(1):1–33, 05 1992. ISSN 0893-9454. doi: 10.1093/rfs/5.1.1. URL `https://doi.org/10.1093/rfs/5.1.1`.

S. H. Teoh, I. Welch, and T. J. Wong. Earnings management and the underperformance of seasoned equity offerings. *Journal of Financial Economics*, 50(1):63–99, 1998.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977).

Appendix A. Constant ICC Model

Discount cash flow (DCF) models are equivalent to dividend discount models (DDMs), but are defined in a broader sense as dividends are just one way of distributing earnings to shareholder.

The DCF has the following form (we use the letter c for free cash flows):

$$PV_t = \sum_{\tau=t+1}^{\infty} \frac{\mathbb{E}_t[c_\tau]}{\prod_{k=t+1}^{\tau}(1+r_{t,k})} \tag{A.1}$$

The time index t represents the time, PV_t is the present value of all future expected cash flows, $\mathbb{E}_t[c_\tau]$ is the at time t expected cash flow at a future time $\tau > t$ and $r_{t,k}$ is the discount rate used from time k to time $k+1$.

To calculate the expected return of a stock investment, one can now calculate the ICC by equating equation (A.1) to the currently observed market price M_t

$$M_t = PV_t = \sum_{\tau=t+1}^{\infty} \frac{\mathbb{E}_t[c_\tau]}{\prod_{k=t+1}^{\tau}(1+r_{t,k})} \,. \tag{A.2}$$

As the left hand side of the equation, M_t, is observable, an assumption about the future cash flow process leaves the ICC process $\prod_{\tau=t+1}^{T}(1+r_\tau)$ (or implied expected return process) as the only unknown.

If the expected return is assumed to have a flat term structure (i.e. constant over time), $r_\tau = \bar{r}$, the rate that equates all future expected cash flows to today's price is the cICC.

$$\min_{\bar{r}} ||M_t - \sum_{\tau=t+1}^{\infty} \frac{\mathbb{E}_t[c_\tau]}{(1+\bar{r})^\tau}|| \,. \tag{A.3}$$

Appendix B. Data Specification

Appendix B.1. Aggregate Consumption and Inflation Data

We use the per capita, seasonally adjusted, real personal consumption expenditures (PCE) on non-durable goods and services as aggregate consumption, as suggested by Hansen and Singleton (1983). Throughout the paper, conversion to real values is done by using the corresponding PCE deflator. All data is obtained from the National Income and Product Account tables at the Bureau of Economic Analysis at the quarterly frequency.

Appendix B.2. Fundamental and Price Data

The fundamental and price data of this paper comes from the merged Compustat/CRSP database (North America and annual for Compustat, monthly frequency for CRSP). Further we use Ken French's data library (ken) for the risk-free rate. The handling of the data, construction of factor loadings and the FM regression analysis is done as suggested by Bali et al. (2017). As is standard, we exclude financial services firms (SIC codes 6000-6999) (Nallareddy et al., 2018). As the cash flow is calculated using information from cash flow statements, the relevant period for those measures is 1989-2019, as SFAS 95 required firms to present a statement of cash flows for fiscal years ending after July 15 1988 only (Nallareddy et al., 2018). So for our expected return proxies (r^{tvICC} and r^{cICC}), which are calculated based on cash flow statement input, the relevant time period is 1989-2019. For the INV, OP and BM factor loading, the period is 1967-2019. Note that the backfill bias brought up by Fama and French (1999) cannot be a driver of any presented results, as the backfill was 1973 for Compustat, so the coverage is complete in the relevant time period of our study.

Appendix B.3. Specifics on the Estimation of the Next Period Expected Return

Our goal is to calculate for each company i and each month t (t is the time index, indicating the end of a month t) the next period expected return $r_{i,t}^{tvICC}$ as predicted by our time-varying ICC model (3.13). We fix the parameter α defined in expression (3.4) at 1 [23]. Thus, we need monthly data for each firm i on $\{M_t, g_t, g_\infty, c(t), r_\infty\}$.

M_t is observable at a monthly frequency, so the information available at the end of each month t is taken. g_t is estimated as the past 3 years cagr of the revenues for each month [24]. $c(t)$ is the past 3 year average of the cash flows for each month. Cash flows thereby are calculated using the indirect cash flow method [25]. g_∞ is calculated as the past 20 year cagr of the industry revenues for each month, $r_\infty = r^{cfy}(1+g_\infty)+g_\infty$ where r^{cfy} is the past 20 years average of the median industry cash-flow yield for each month.

In order to make sure that the fundamental parameters used in our analysis $\{g_t, g_\infty, c(t), r_\infty\}$ are publicly available at time t, we follow the procedure described in Bali et al. (2017) in chapter 10. The fundamental data used in the calculation of $r_{i,t}^{tvICC}$ of company i for the months from July of year $y+1$ to June $y+2$ is taken to be the fundamental data from calendar year y. The reason for this approach to the timing of the calculation of the $r_{i,t}^{tvICC}$ is that firms have 3 months from the end of their fiscal year-ends to report the required data, but many firms fail to meet this deadline. Thus, one generally assumes at least a six-month gap between the end of the fiscal year and the time at which the fiscal year-end data is publicly available.

For each firm i, we now have the parameter set $\{M_t, g_t, g_\infty, c(t), r_\infty\}$ and can thus numerically estimate the expected return $r_{i,t}^{tvICC}$ for firm i at time t using equation (3.13). The next period expected return calculated in this way, $r_{i,t}^{tvICC}$, then is used to explain the realised return of company i in the proceeding month $R_{i,t+1}$. Thus, the lag between the publication of the fundamental data used to calculate the expected return $r_{i,t}^{tvICC}$ and the observed return $R_{i,t+1}$ can potentially be larger than one full year.

Due to other dynamics involved in firms with negative growth rates, we only consider cash flow growth rates larger than 0% and smaller than 50%.

Note that $r_{i,t}^{tvICC}$ and $r_{i,t}^{cICC}$ have units of monthly returns in all empirical results presented in this paper.

Appendix C. Summary statistics

---

[23] Note that we find the dependence of the two ICC model's next period expected return on α to be very weak, suggesting that our empirical results are robust to mis-specifications of the assumed rate of convergence to the long-term asymptotic.

[24] As usual, revenues are used as proxy for cash flows when calculating the growth rate, since they are less volatile.

[25] free cash flow = net income + accounting costs - changes in operating working capital - capital expenditures

	mean	std	skewness	kurtosis	min	5th quantile	25th quantile	median	75th quantile	95th quantile	max	N
r^{tvICC}	0.01	0.021	0.28	14	-0.11	-0.022	0.0012	0.011	0.021	0.039	0.16	1300
r^{cICC}	0.013	0.011	9.6	160	0.00063	0.0053	0.0084	0.011	0.014	0.024	0.22	1600
INV	0.096	0.3	1.1	23	-2.4	-0.27	-0.021	0.068	0.18	0.57	3.2	3200
OP	-0.14	16	-14	1400	-600	-0.9	-0.0077	0.17	0.29	0.54	270	3200
BM	-0.54	0.91	-0.69	5.6	-6.6	-2.1	-1	-0.45	0.041	0.75	3.2	3200
R	0.0073	0.17	3.4	76	-0.8	-0.21	-0.072	-0.0038	0.069	0.25	2.8	3800

Table C.2: This table presents summary statistics for our sample. Each month, the mean, standart deviation, skewness, excess kurtosis, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, maximum and number of data points of the cross-sectional distribution of each variables (r^{tvICC}, r^{cICC}, BM, INV, OP and R) are calculated. The table presents the time-series means for each cross-sectional value.

Appendix D. Correlations

	r^{tvICC}	r^{cICC}	INV	OP	BM	R
r^{tvICC}	1	0.882	-0.018	0.001	0.239	0
r^{cICC}	0.755	1	0.047	0.016	0.178	-0.007
INV	-0.007	0.021	1	0.27	-0.279	0.001
OP	0.015	0.008	0.083	1	-0.162	0.051
BM	0.246	0.232	-0.195	0.05	1	0.023
R	0.009	0	-0.02	0.002	0.026	1

Table D.3: This table presents the time-series averages of the monthly cross-sectional Pearson product-moment and Spearman rank correlations between pairs of r^{tvICC}, r^{cICC}, BM, INV, OP and R. Below-diagonal entries present the average Pearson product-moment correlations. Above-diagonal entries present the average Spearman rank correlation.

Appendix E. Fama MacBeth (FM) regression approach

The regression analysis used to see if a certain factor loading is helpful in forecasting the realised return is the Fama MacBeth (FM) regression approach (Fama and MacBeth, 1973). In short, it stems from the powerful insight that lack of predictability, with constant expected returns over time, implies that stock returns are uncorrelated over time, even though they are correlated across stocks at a given time (Fama and MacBeth, 1973).

We follow the Fama MacBeth (FM) regression approach described in chapter 6 of Bali et al. (2017).

Note that the literature often speaks about different steps in the FM regression, but often there is confusion about how many steps in total there are, as some never mention the first step. The first step of the FM regression is the time-series regression to estimate the factor loading, e.g. to estimate the market beta of each firm at each point in time. But as the factor loading not always has to be estimated, as e.g. for the profitability factor loading ($OP_{i,t}$) of Fama and French (2015) or also our

$r_{i,t}^{tvICC}$ (they can both just be calculated and do not have to be estimated by using a first stage time-series regression), this step is often not mentioned.

The second step in the FM regression is then to use the factor loadings to run a cross-sectional regression of the dependent variable $Y_{i,t}$ (realised returns) on the independent variables $X1_{i,t}, X2_{i,t}, \ldots$ (factor loadings) and an intercept:

$$Y_{i,t} = a_{0,t} + \beta_{1,t} X1_{i,t} + \beta_{2,t} X2_{i,t} + \ldots + \epsilon_{i,t} \tag{E.1}$$

The cross-sectional regression used is a standard ordinary-least-squares (OLS) regression. The independent variables are winsorised at the 0.5 and 99.5 percentiles. So at each time t, a cross-sectional regression is run. This results in a time-series of estimated intercepts and slope coefficients $a_{0,t}, \beta_{1,t}, \beta_{2,t}$ for each regression model. The estimated slope coefficients are the price of risk, or risk premium, stemming from the respective risk factor at time t.

The third step is then to compute the time-series averages of the estimated cross-sectional regression coefficients. To examine whether the average coefficient is statistically different from zero, we then also calculate the standard errors and the associated t-statistics and p-values to test the null hypothesis that the average coefficient is equal to zero. The standard errors are adjusted for heteroscedasticity and auto-correlation, following Newey and West (1987).

Interpretation of the results of FM regressions is fairly straightforward. A statistically significant average slope coefficient indicates a cross-sectional relation between the given independent variable X and the dependent variable Y in the average time period (Bali et al., 2017). When the regression specification includes more than one independent variable, statistical significance indicates that a relation between X and Y exists after controlling for the effects of the other independent variables included in the regression specification (Bali et al., 2017). If the coefficient of interest is statistically significant in one specification but insignificant when additional controls are added to the specification, then the relation between X and Y appears to be explained by some linear combination of the added control variables (Bali et al., 2017).

Appendix F. ICCs and the Fama MacBeth regression approach

The Fama MacBeth regression approach is known to have three biases, which we discuss below. However, these biases are less pronounced for ICC models, which makes the Fama MacBeth regression approach the ideal method to test them.

(1) The first bias is the error-in-variables bias. Under the classical measurement error model, the coefficient estimates are biased toward zero, which is known as the attenuation bias (Collot and Hemauer, 2020). However, besides the coefficient of the imprecisely measured variable, the coefficients of the other variables are also biased - known as contamination bias - and the direction of this bias is in general unknown (Collot and Hemauer, 2020). The two-pass regression approach suffers from the error-in-variables problem, because it uses the estimated coefficients from the first stage - which are by nature measured imprecisely - as explanatory variables in the second stage, where they introduce an error-in-variables bias in the price of risk estimates. Since ICC models are not estimated in a first stage, but can directly be calculated, this problem is not of relevance here.

(2) The second bias is the omitted-variable bias, which refers to the bias in the coefficient estimates obtained from regressions when some important explanatory variables are omitted from the model (Collot and Hemauer, 2020). The two-pass regression approach suffers from the omitted-variable bias in both stage one and two, as one in principle would have to always include all relevant factors in each regression. The omitted-variable bias in stage one then leads to a measurement error of the factor loading and thus even introduces an error-in-variables bias in the price of risk estimates. As in point (1), since ICC measures are not estimated in a first-stage regression, there is no omitted-variable bias in the first stage. And the omitted-variable bias in the second stage is also less pronounced, as an ICC by design captures all relevant factors which can be motivated from a valuation model and thus summarises other factors in an efficient way.

(3) The third problem is that, in some cases, the standard error of the estimated average slope coefficient has been shown to be biased downward, resulting in inflated statistical significance. Specifically, Shanken (1992) shows that an error-in-variables problem (Point 1) can result in overestimation of the statistical significance of the price of risk, as standard errors are shifted downwards when standard statistical methodologies are used (Bali et al., 2017). As our proxy is not estimated in a first-stage time-series regression, but directly calculated, this problem is not of relevance here.

Points 1, 2 and 3 suggest that the three biases inherent to the FM approach are not a problem for ICC proxies.

In addition, the linearity assumption inherent to the FM regression approach is not a problem for ICCs, as one regress expected return proxies on realised returns, a relationship which can be assumed to be linear.

Appendix G. Consumption Based Capital Asset Pricing Model

In this appendix, we summarise the derivation of the consumption based asset pricing model as presented in Cochrane (2001), in order to show the motivation behind the regression model presented in (2.1).

Appendix G.1. Presentation of the model

The basic objective of asset pricing is to figure out the value of any stream of uncertain cash flows. We want to find the value at time t of a future payoff x_{t+1}. E.g. if you buy a stock at time t, what is the value of its payoff $x_{t+1} = p_{t+1} + d_{t+1}$, where p_{t+1} is the price at $t+1$ and d_{t+1} is the dividend at $t+1$. x_{t+1} thereby is a random variable, as the investor does not know the payoff (future price and future dividend) with certainty, but she can assess the probability of various outcomes.

What is this payoff worth for a typical investor? This depends on what the investor wants. A common approach is to model the investor's preferences with a utility function defined over current (c_t) and future values of consumption ($u(c_{t+1})$).

$$U(c_t, c_{t+1}) = u(c_t) + \beta \mathbb{E}_t \left[u(c_{t+1}) \right] \ , \tag{G.1}$$

where β is the subjective discount factor of utilities. Note that consumption c_{t+1} is random, as the investor does not know his wealth tomorrow and hence what he will consume.

The period utility function $u(.)$ is concave in the level of consumption, so it captures the fundamental desire for more consumption, but with a declining marginal value of additional consumption. This assumption for $u(.)$ makes investors impatient (β discounts the future and captures impatience) and averse to risk as well as to inter-temporal substitution (due to the curvature of the utility function, the upside in value by additional consumption is lower than the downside in value of decreasing consumption). Investors thus prefer a constant consumption stream over time and across states of nature.

Now we assume that the investor can freely buy and sell as much of the payoff x_{t+1} as she wishes, at price p_t. This allows the investor to “transfer" consumption over time. Denote by e_t the original consumption, i.e. the consumption if the investor buys none of the assets, and by ξ the amount of the asset she chooses to buy. How much of the asset would the investor buy? This problem can be formalised as the following maximisation problem:

$$\begin{aligned} &\max_{\{\xi\}} \left(u(c_t) + \mathbb{E}_t\left[\beta u(c_{t+1})\right]\right), \text{ s.t.} \\ &c_t = e_t - p_t \xi \\ &c_{t+1} = e_{t+1} + x_{t+1}\xi \end{aligned} \tag{G.2}$$

Substituting the constraints into the objective, taking the derivative with respect to ξ and equating it to zero, we get the first-order condition for an optimal consumption and portfolio choice:

$$p_t = \mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)} x_{t+1}\right] \tag{G.3}$$

The investor buys more or less from the asset until this first-order condition holds [26]. Equation (G.3) is *the* central asset pricing equation. Given the payoff x_{t+1} and given the investors consumption choice c_t, c_{t+1}, it tells you what market price p_t to expect.

By defining the stochastic discount factor as $M_{t+1} \equiv \frac{\beta u'(c_{t+1})}{u'(c_t)}$, we can rewrite equation (G.3) as

$$p_t = \mathbb{E}_t\left[M_{t+1} x_{t+1}\right] \tag{G.4}$$

M_{t+1} discounts dollars, while β discounts utilities. M_{t+1} is stochastic in nature, because it is not known with certainty at time t.

Using the definition of the arithmetic average return of asset i, denoted by r^i_{t+1}, we can write the payoff as $1 + r^i_{t+1} \equiv \frac{x^i_{t+1}}{p^i_t}$ and thus rewrite the asset pricing equation (G.3) as

$$1 = \mathbb{E}_t\left[M_{t+1}(1 + r^i_{t+1})\right] \tag{G.5}$$

For a riskless asset, we know the payoff x^f_{t+1} already at time t and thus also know the assets return at time t $(1 + r^f_t \equiv \frac{x^f_{t+1}}{p^f_t})$.

---

[26] Note that $p_t u'(c_t) = \mathbb{E}_t\left[\beta u'(c_{t+1}) x_{t+1}\right]$ shows the trade-off faced by the investor nicely. The investor looses the utility $p_t u'(c_t)$ at time t if she buys another unit of the asset, while she gains the discounted, expected utility $\mathbb{E}_t\left[\beta u'(c_{t+1}) x_{t+1}\right]$ from the extra payoff of the asset at time $t + 1$.

Thus we can take $(1 + r_t^f)$ out of the expectation in equation (G.5) to get the asset pricing equation for a risk-less asset

$$1 + r_t^f = \mathbb{E}_t \left[M_{t+1}\right]^{-1} \tag{G.6}$$

With the definition of the covariance [27] and equation (G.6), we can rewrite equation (G.5) as

$$\begin{aligned} 1 &= \mathbb{E}_t \left[M_{t+1}(1 + r_{t+1}^i)\right] \\ &= cov_t \left[M_{t+1}, (1 + r_{t+1}^i)\right] + \mathbb{E}_t \left[M_{t+1}\right] \mathbb{E}_t \left[(1 + r_{t+1}^i)\right] \\ &= cov_t \left[M_{t+1}, (1 + r_{t+1}^i)\right] + \frac{\mathbb{E}_t \left[(1 + r_{t+1}^i)\right]}{1 + r_t^f} . \end{aligned} \tag{G.7}$$

The risk premium of the risky asset i then reads

$$\mathbb{E}_t \left[r_{t+1}^i\right] - r_t^f = -(1 + r_t^f) cov_t \left[M_{t+1}, (1 + r_{t+1}^i)\right] \tag{G.8}$$

With (G.6), this can be rewritten as a beta pricing model

$$\mathbb{E}_t \left[r_{t+1}^i\right] - r_t^f = \underbrace{\frac{cov_t \left[M_{t+1}, (1 + r_{t+1}^i)\right]}{Var(M_{t+1})}}_{=\beta^i} \underbrace{\left(-\frac{Var(M_{t+1})}{\mathbb{E}_t \left[M_{t+1}\right]}\right)}_{=\lambda_M} \tag{G.9}$$

This is the beta pricing model, which says that each expected return is proportional to the risk exposure β^i, i.e. the regression coefficient in a regression of the gross return $(1 + r_{t+1}^i)$ on the discount factor M_{t+1} (as β^i has the form of a regression coefficient in a regression of the gross return $(1 + r_{t+1}^i)$ on the discount factor M_{t+1}), and to the price of risk λ_M.

In order to express β^i in terms of observables, one can e.g. use the CRRA utility function (constant relative risk aversion, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$), which leads to the stochastic discount factor $M_{t+1} = \beta(\frac{c_{t+1}}{c_t})^{-\gamma}$. To estimate β^i, one would then regress the gross return $(1 + r_{t+1}^i)$ as dependent variable on $M_{t+1} = \beta(\frac{c_{t+1}}{c_t})^{-\gamma}$ as independent variable in a regression model with a constant discount rate. This is the theory behind the empirical model presented in (2.1) and (2.3).

---

[27] cov(m,x) = E(mx) - E(m)E(x)

Appendix G.2. The Log Normal Case

Let us now assume that the vector $\{M_{t+1}, (1+r^i_{t+1})\}$ is log-normally distributed and conditionally homoskedastic. Taking logs of equation (G.5) and then using properties of the log-normal distribution [28] gives

$$\begin{aligned} 0 &= \ln(\mathbb{E}_t\left[M_{t+1}(1+r^i_{t+1})\right]) \\ &= \mathbb{E}_t\left[\ln(M_{t+1}(1+r^i_{t+1}))\right] + \frac{1}{2}Var_t(\ln(M_{t+1}(1+r^i_{t+1}))) \\ &= \mathbb{E}_t\left[m_{t+1}\right] + \mathbb{E}_t\left[\ln(1+r^i_{t+1})\right] + \frac{1}{2}(\sigma^2_t(\ln(1+r^i_{t+1})) + \sigma^2_t(m) + 2\sigma_t(\ln(1+r^i_{t+1}), m)) \end{aligned} \tag{G.10}$$

where $\ln(M_{t+1}) = m_{t+1}$, $\sigma^2_t(\ln(1+r^i_{t+1})) = Var_t(\ln(1+r^i_{t+1}))$ and $\sigma_t(r_i, m) = cov_t(m_{t+1}, \ln(1+r^i_{t+1}))$. Assuming for equation (G.6) that the vector $\{M_{t+1}, (1+r^f_t)\}$ is also log-normally distributed, gives

$$\begin{aligned} 0 &= \ln(\mathbb{E}_t\left[M_{t+1}(1+r^f_t)\right]) \\ &= \mathbb{E}_t\left[\ln(M_{t+1}(1+r^f_t))\right] + \frac{1}{2}Var_t(\ln(M_{t+1}(1+r^f_t))) \\ &= \mathbb{E}_t\left[m_{t+1}\right] + \mathbb{E}_t\left[\ln(1+r^f_t)\right] + \frac{1}{2}(\sigma^2_t(m) + \sigma^2_t(\ln(1+r^f_t)) + 2\sigma_t(m_{t+1}, \ln(1+r^f_t))) \\ &= \mathbb{E}_t\left[m_{t+1}\right] + \ln(1+r^f_t) + \frac{1}{2}\sigma^2_t(m) \end{aligned} \tag{G.11}$$

Note that $\sigma_t(m_{t+1}, \ln(1+r^f_t))$ and $\sigma^2_t(\ln(1+r^f_t))$ are zero, as r^f_t is known at time t. Combining (G.10) with (G.11) leads to

$$\mathbb{E}_t\left[\ln(1+r^i_{t+1})\right] + \frac{1}{2}\sigma^2_t(\ln(1+r^i_{t+1})) - \ln(1+r^f_t) = -\sigma_t(\ln(1+r^i_{t+1}), m_{t+1}) \tag{G.12}$$

and if we assume that $r^i_{t+1} \approx \ln(1+r^i_{t+1})$ for $r^i_{t+1} \ll 1$ and that $r^f_t \approx \ln(1+r^f_t)$ for $r^f_t \ll 1$ we get

$$\mathbb{E}_t\left[r^e_{t+1}\right] = -\sigma_t(r^i_{t+1}, m_{t+1}) \tag{G.13}$$

where $\mathbb{E}_t\left[r^e_{t+1}\right] \equiv \mathbb{E}_t\left[r^i_{t+1}\right] - r^f_t + \frac{1}{2}\sigma^2_t(r^i_{t+1})$ is the expected geometric average excess return [29].

This can now be expressed in terms of observable variables again by using the log of the CRRA utility function $m_{t+1} = \ln(\beta) - \gamma\Delta\ln(c_{t+1})$ with which $\sigma_t(r^i_{t+1}, m_{t+1}) = cov_t(r^i_{t+1}, \ln(\beta) - \gamma\Delta\ln(c_{t+1})) = -\gamma \cdot cov_t(r^i_{t+1}, \Delta\ln(c_{t+1}))$, which we can insert into equation (G.12) to get

$$\mathbb{E}_t\left[r^e_{t+1}\right] = \gamma \cdot cov_t(r^i_{t+1}, \Delta\ln(c_{t+1}))\ . \tag{G.14}$$

---

[28] If $Z \sim N(\mu, \sigma^2)$, then $e^Z \sim logN(e^{\mu+0.5\sigma^2}, .)$; In our case $M_{t+1}(1+r^i_{t+1}) \sim logN(e^{\mu+0.5\sigma^2}, .)$, then $\ln(M_{t+1}(1+r^i_{t+1})) \sim N(\mu, \sigma^2)$ with $\mu = \mathbb{E}_t\left[\ln(M_{t+1}(1+r^i_{t+1}))\right]$ and $\sigma^2 = Var_t(\ln(M_{t+1}(1+r^i_{t+1})))$. Thus $\mathbb{E}_t\left[M_{t+1}(1+r^i_{t+1})\right] = e^{\mu+0.5\sigma^2} = e^{\mathbb{E}_t\left[\ln(M_{t+1}(1+r^i_{t+1}))\right]+0.5Var_t(\ln(M_{t+1}(1+r^i_{t+1})))}$ and $\ln(\mathbb{E}_t\left[M_{t+1}(1+r^i_{t+1})\right]) = \mathbb{E}_t\left[\ln(M_{t+1}(1+r^i_{t+1}))\right] + 0.5Var_t(\ln(M_{t+1}(1+r^i_{t+1})))$

[29] r^i_{t+1} and r^f_t are arithmetic average returns (the drift of a price process which is log normally distributed), while $r^i_{t+1} + \frac{1}{2}\sigma^2_t(r^i_{t+1})$ and $r^f_t + \frac{1}{2}\sigma^2_t(\ln(1+r^f_t))$ are the respective geometric average returns (the drift of the respective log price process, which is normally distributed). And with $\sigma^2_t(\ln(1+r^f_t)) = 0$, $\mathbb{E}_t\left[r^e_{t+1}\right] = \mathbb{E}_t\left[r^i_{t+1}\right] + \frac{1}{2}\sigma^2_t(r^i_{t+1}) - r^f_t$.

We can again write in terms of a beta pricing model as

$$\mathbb{E}_t\left[r^e_{t+1}\right] = \underbrace{\frac{cov_t(r^i_{t+1}, \Delta\ln(c_{t+1}))}{Var_t(\Delta\ln(c_{t+1}))}}_{=\beta^i} \underbrace{Var_t(\Delta\ln(c_{t+1}))\gamma}_{=\lambda_M} \tag{G.15}$$

Now β^i has the form of a regression coefficient in a regression of the arithmetic average returns on the log growth of consumption ($\Delta\ln(c_{t+1})$), which can be estimated. This β^i is exactly what is estimated by equation (2.1).

Appendix H. Arbitrage Pricing Theory

Multifactor asset pricing models like the arbitrage pricing theory (APT) of Ross (1976) are relating a stock's expected excess return to linear risk factors as follows

$$\mathbb{E}_t\left[R_{i,T}\right] - R_{f,T} = \sum_{j=1}^{k} \beta^j_{i,T}\lambda_{j,T} \tag{H.1}$$

where $R_{i,T}$ is the return of asset i at time T, $R_{f,T}$ is the risk-free rate at time T, $\beta^j_{i,T}$ is the risk exposure of asset i on risk factor j at time T, and $\lambda_{j,T}$ is the risk premium (price of risk) associated with factor j at time T.

The theory behind equation (H.1) does not specify the identity of the k factors though.

Appendix I. Slope comparison

Note that there is a difference between the slope of the lower panel in figure 4 and the slope of figure 1.

In our time-varying ICC model, r is a function of the risk exposure to the growth risk and of the risk exposure to the value factor. If we condition for the return contribution from the value factor (i.e. assume the price-to-cash flow ratio is already at its industry level), our model predicts $r \propto a \cdot g$, with a as the slope of the lower panel in figure 4 and g as the cash flow growth rate. a is found to be around $a = 0.64$ in our theoretical framework.

In the CCAPM, r is a function of the risk exposure β (the consumption beta) scaled by the price of risk λ (the consumption growth rate), so $r \propto \beta \cdot \lambda$. Further, from section 3, we know that $\beta \propto b \cdot g$, with b as the slope of figure 1 and g again the cash flow growth rate. Thus, $r \propto b \cdot g \cdot \lambda$.

A comparison of $r \propto b \cdot g \cdot \lambda$ and $r \propto a \cdot g$ implies $a = b \cdot \lambda$, which we can test by measuring $b \cdot \lambda$.

With $\lambda = 0.02$ (using $\lambda = Var(\Delta ln(c_{t+1})) \cdot \gamma$ from equation (G.15) with $Var(\Delta ln(c_{t+1})) = 0.001$ and $\gamma = 20$ from Campbell (2003)) and $b = 35.5$ in figure 4, this empirically implies the value $a = 35.5 \cdot 0.02 = 0.7$, which is close to the theoretical value $a = 0.64$ quoted above.

The same calculation can be done for the slope between the cash flow beta and the cash flow growth rate level.

5 Conclusion

The following paragraphs highlight the main contributions and some of the blank and blind spots of the conducted research and proposes direction for future research.

The extreme price moves of the GameStop shares in Q1 2021 gives the topic of feedback effects from options market onto the underlying's price dynamics a high relevance. Chapter 2 of this thesis contributes to this discussion by proposing a general theoretical model that describes the isolated effect of option positioning onto the dynamics of an underlying's price, namely its volatility and drift. Our model shows how the option positioning of market makers, or more precisely the price impact of the net delta hedge demand, can distort the underlying's price volatility and drift. With respect to the volatility, our model predicts that if the option market maker has a short (long) option positioning, the implied delta hedging strategy leads to an increase (decrease) of the underlying's price volatility. Further, depending on the option specifications and delta, the price can get unusually stable or explosive. This extensive theoretical discussion of the feedback effect from options markets onto the the underlying's price dynamics, especially with respect to the price (in)stability, is a first important novel contribution of chapter 2. Our model is validated with high confidence with data from the two largest foreign exchange markets, the EURUSD and USDJPY market, as well as the GameStop stock market. In both markets we find that the volatility is increased due to short option positioning of the option market maker. And we further show that there indeed were times where the option positioning suggested explosive behaviour of the GameStop stock price. This empirical discussion on the basis of our model framework is a second important contribution of chapter 2.

I see two limitations of the presented work. The first is on the theoretical side: We assume the price impact of the delta hedge transactions, our friction, to be linear. While this is true on longer time horizons (i.e. hours), on short time scales the price impact is known to scale with the square root of the traded volume. Thus, on short time scales, the model dynamics might slightly deviate from what our model predicts. Extending our friction to a square root form would thus be an interesting next step. The second limitation is on the empirical side: The net hedge demand of the market maker is not directly observable in most transaction data sets. We thus used transaction data in which we know (or reconstructed) the side of each transaction and then worked with the assumption that the market maker is always liquidity provider behind each transaction. While this seems reasonable, it is probably not always true. But on average the assumption should hold, and the fact that our model fit based on data from three different markets always results in very high confidence, serves as strong support. Nevertheless, future research in this area could be based on data in which the entity behind each trade as well as their hedging behaviour is known. This would allow for an even better estimation of the model parameters.

The extreme price moves that could be observed in the aftermath of the introduction of a new price-based circuit breaker mechanism in the CSI 300 in the beginning of 2016 motivated the work of chapter 3. We theoretically model the isolated effect of circuit breakers onto price dynamics and show that investors' anticipation of the probability of a halt feedbacks on the price process. We show that for price-based circuit breakers this results in an increased price volatility and a magnet effect and thus explains the observations in the CSI 300. But more importantly, as our model is a priori independent of the circuit breaker trigger mechanism, it can be used as starting point to study different circuit breaker designs. This framework is the main contribution of chapter 3. We thereby take first step towards a more resilient circuit breaker mechanism, e.g. a volatility based trigger mechanism that minimises the side-effects discussed for price-based circuit breakers.

The limitations of the research of extreme price moves such as circuit breaker events is that there are naturally not too many events which one can use for empirical studies. End even for the events that happened, the data set is very noisy, since various effects come together in such extreme events. Future research could try to focus more on the empirical side, i.e. find high frequency data around such events and try to validate our proposed model with it. In addition, future research should continue the work started by us, namely to use our model framework as starting point to answer the question of how to ideally design circuit breakers.

Last, the difficulty to value high growth firms and the resulting high volatility of such growth stocks is the motivation behind chapter 4 of this thesis. We present a model which allows investors to accurately estimate the expected return from an investment in a high growth firm. Specifically, using economic intuition and empirical evidence from the consumption based capital asset pricing model we argue that the expected return of a stock is linearly proportional to the firm's cash flow growth rate. Based on that, we suggest a time-varying discount rate that reflects this linearity between the expected return and the cash flow growth in its term structure. This idea is the main contribution of chapter 4. When used in a discounted cash flow model, the discount rate and growth rate interact to reduce the range of possible valuation outcomes to useful levels. This makes firm valuation less sensitive to the cash flow growth rate assumptions and thus increases its practical utility. And the time-varying internal rate of return of the corresponding internal cost of capital model (ICC) can conceptually be interpreted as an investment return process, compared to e.g. a constant ICC model, and thus is a valid proxy for the expected return (i.e. the cost of capital). We discuss the empirical quality of our cost of capital proxy in a Fama MacBeth regression setting as well as in a trading strategy, and show its superiority, especially if one focuses on high growth firms.

The limitation of the work presented in chapter 4 is mostly on the empirical side. We show that the discount rate of a valuation model should have the same structural form as the assumed cash flow growth rate process. As a consequence of this, the focus of an investor should solely be on the future cash flow growth rate structure, which then naturally dictates the discount rate structure. Thus, the focus of future research could be on empirically exploring cash flow growth rate structures. While we, motivated by prior research, used a generic cash flow convergence process for all firms, where only the convergence levels are industry specific, this could be refined by exploring structural similarities or differences in the cash flow growth rate process for different industries and firm types.

In summary, I hope this thesis provides three solid frameworks that allow any investor and policy maker to study the influence of option positioning, circuit breakers as well as cash flow growth onto an asset's price dynamics. Understanding asset price dynamics is complex, and probably a never ending quest, but I hope I could bring light into three corners of this enormous space.

Florian Ulmann

Curriculum Vitae

Education

PhD

2017 – 2021 **PhD Candidate in Finance**, *ETH Zurich*
Chair of Entrepreneurial Risk (MTEC)

2018 – 2019 **Swiss Program for Beginning Doctoral Students in Economics**, *Study Center Gerzensee*
Econometrics and Macroeconomics sequence

2017 – 2018 **PhD Exchange**, *INSEAD*, Singapore
Visiting Phd Student

Master

2012 – 2014 **Master of Science in Physics**, *University of Basel*
Specialised in Theoretical Physics: Stochastic Processes & Quantum Field Theory

2014 – 2014 **Master's Thesis Exchange**, *Yale University*, USA
Visiting student in computational biophysics: Machine Learning for Disease Prediction & Control

Bachelor

2013 – 2015 **Bachelor of Arts in Business and Economics**, *University of Basel*
Major in Business, Thesis: Predictive Force of Raw Accounting Data on Equity Returns

2009 – 2012 **Bachelor of Science in Nanoscience**, *University of Basel*
Biology, Chemistry and Physics with extra classes in Nanoscience

High School

2002 – 2008 **Matura**, *Sport High School of the Gymnasium Baeumlihof*, Basel
Specialised in Sport, Biology & Chemistry

Professional Experience

2021 – today **Junior Consultant**, Boston Consulting Group, Zurich
Generalist management consultant.

2019 – 2021 **Co-Founder**, Quantamental AG, Zurich
Quantamental is a Swiss based value investing boutique that pursues an active value investment strategy and thereby combines modern, quantitative methods with classical fundamental analysis to create a focused, global public equity portfolio.

2015 – 2019 **Quantitative Analyst, FX Unit**, Swiss National Bank, Zurich and Singapore
Involved in various quantitative analysis and development projects relevant for monetary policy decision-making and implementation.

2015 – 2015 **Intern, Monetary Policy Analysis Unit**, Swiss National Bank, Zurich
Worked on research projects with a focus on exchange rate equilibrium models.

Student Jobs

2015 – 2015 **Teacher**, academia Basel, Basel
Teacher for Business, Economics, Politics and Mathematics

2012 – 2014 **Student Trainee**, IBM Global Technology Services, Basel
Executive Management Support within the IBM Strategic Outsourcing Organisation.

2010 – 2012	**Communication Assistant**, Swiss Nanoscience Institute (SNI), Basel Assistant to public relation & event management department.

Extracurricular Activities

2014 – 2015	**Swiss Students and Affiliates at Yale (SSAY): Co-Founder**, *Yale University* Organisation of various social and academic events
2009 – 2012	**Students of Nanoscience Association: board member**, *University of Basel* Organisation of various social and academic events
2006 – 2010	**Jungwacht & Blauring St. Michael: board member** Organisation of numerous short- and longterm events for children and their families.

Research Projects

2021 – today	**On the Directional Destabilizing Feedback Effects of Option Hedging**, *ongoing*, Didier Sornette (ETH Zurich), Florian Ulmann (ETH Zurich), Alexander Wehrli (SNB)
2019 – today	**Dynamical Internal Cost of Capital Driven by Cash Flow Growth**, *Reprint submitted to JFE*, David Solo (DCI), Didier Sornette (ETH Zurich), Florian Ulmann
2017 – today	**The Impact of Option Hedging on the Spot Market Volatility**, *Revise and resubmit at JIMF*, Benjamin Anderegg (SNB), Didier Sornette (ETH Zurich), Florian Ulmann
2017 – today	**Price Dynamics with Circuit Breakers**, *Revise and resubmit at JBF*, Sandro Claudio Lera (MIT Media Lab), Didier Sornette (ETH Zurich), Florian Ulmann

Languages

German	Mother tongue
English	Fluent
French	Conversational

Awards

2002 – 2006	Several top three placements at Swiss Championships in Track and Field

Hobbies

Sport	Crossfit and mountaineering
Social	Family and friends.
Travelling	Discovering new cultures and countries.
Interests	Various topics ranging from business and economics, to science and technology.

© 2021 Florian Ulmann

All rights reserved.