Robin Valenta

The Long Run Return on Investment

- Introduction of a VEC model to business bank accounts

Masteroppgave i Samfunnsøkonomi
Veileder: Joakim Prestmo
Medveileder: Fabian Schlobach
Juni 2023

Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

NTNU
Kunnskap for en bedre verden

Robin Valenta

The Long Run Return on Investment

- Introduction of a VEC model to business bank accounts

Masteroppgave i Samfunnsøkonomi
Veileder: Joakim Prestmo
Medveileder: Fabian Schlobach
Juni 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

NTNU
Kunnskap for en bedre verden

Preface

This thesis concludes my master's course in economics at the Department of Economics at the Norwegian University of Science and Technology in Trondheim, where I have delved deep into the concepts of dynamic macroeconomic forces and astronomical vector models. I am grateful to NTNU for making this knowledge available, and for letting me teach others what I had just been thought myself. In the last couple of years, I have been given a shot at research from Leiv Opstad and his colleagues, and I am grateful for the opportunity, but even more so for the collaboration we have had along the way. Thanks also to Joakim Prestmo, who has guided me through the writing of this thesis.

In the fall on 2021, I met my soulmate Laura, whom I moved to Copenhagen for. That has been the best choice I have ever made, and I am infinitely grateful for the unconditional support and love you have given me, and for the home we have made together. Now we are taking our home back to Trondheim, where new adventures await. To my family I share the same infinite gratitude for always supporting and loving me, for helping me in my studies from the first grade to the writing of this thesis, and in general, for making something out of me and always being there.

In my stint in Copenhagen, I have been given the chance to work in, and write my thesis in collaboration with, the exciting fintech company MXNEY. I am grateful for the colleagues met, memories made, and the guidance given in the writing my thesis from Fabian Schlobach. I wish the company all the best in their future business.

To Laura: I love you.

Abstract

This master thesis formulates a strategy to estimate the return on investment using business bank accounts. The thesis is methodological at heart, but the data I use is part of the bigger picture of financial development. I construct an econometric model that can, I argue, estimate the return on investment, both within and between firms. The model consistently yields higher estimates for investments with higher return and higher estimates for firms with higher return on investment. To test the model's robustness, I estimate different variations of the model, showing that it can be applied to lower frequency data too. The model can be applied to rank the return on investments between firms, and in an extension of the model I show that the model can be used to rank the return on investment between investment types within firms. The data I use has been made available to fintech lenders through the open banking initiative, and I have been provided the data from one such fintech lender. Open banking aims to increase capital market efficiency by decreasing the informational barriers to entry in finance. The initiative's strategy is to enable firms to share the data withheld in their bank accounts with other finance companies, thereby incentivising innovation in the processing of that data. Fintech lenders use open banking data in the screening of potential borrowers, applying data-science strategies to identify credit risk and investment opportunities. But, competition in lending is different from other sectors, as lenders can always increase their supply of credit by increasing their credit risk, meaning that there is a fine balance between output and risk. Innovations in the processing of information directly reduces credit risk and increases credit supply, though the indirect effects of competition may be to increase risk in the financial sector. The general model I present in this thesis can be implemented by any bank, financial institution, or fintech lender, in the evaluation of potential borrowers' creditworthiness.

Sammendrag

Denne masteroppgaven formulerer en strategi for å estimere avkastningen på investeringer ved hjelp av forretningsbankkontodata. Masteroppgaven er metodologisk basert, men dataene jeg bruker er en del av det større bildet av finansiell utvikling. Jeg utvikler en økonometrisk modell som kan, etter min mening, estimere avkastningen på investeringer, både innenfor og mellom bedrifter. Modellen gir høyere estimater for investeringer med høyere avkastning, og høyere estimater for bedrifter med høyere avkastning på investeringer. For å teste modellens robusthet, estimerer jeg ulike variasjoner av modellen, og jeg viser at den også kan brukes på data med lavere frekvens. Modellen kan brukes til å rangere bedrifter etter avkastning på investering, og i en utvidelse av modellen viser jeg at den også kan brukes til å rangere investeringer etter avkastning innenfor bedrifter. Dataene jeg bruker er gjort tilgjengelig for fintech-utlånere gjennom åpen banking, og jeg har fått tilgang til dataene fra en slik fintech-utlåner. Åpen banking har som mål å øke effektiviteten i kapitalmarkedet ved å redusere informasjonsbarrierene for finansforetak. Initiativets strategi er å tillate bedrifter å dele dataene som holdes i bankkontoene deres med andre finansselskaper, for derved å stimulere innovasjon i bearbeidingen av disse dataene. Fintech-utlånere bruker forretningsbankkontodata i vurderingen av potensielle låntakere, og anvender datavitenskapsstrategier for å identifisere kredittrisiko og investeringsmuligheter. Konkurranse innen utlån er ulik andre sektorer, da långivere alltid kan øke kredittilbudet ved å øke kredittrisikoen, noe som betyr at det er en fin balanse mellom produksjon og risiko. Innovasjon i prosesseringen av informasjon reduserer kredittrisiko og øker kredittilbudet direkte, selv om den indirekte effekten av konkurranse kan øke risiko i finanssektoren. Den generelle modellen jeg presenter i denne oppgaven kan bli implementert av enhver bank, finansforetak, eller fintech-utlåner, i evalueringen av potensielle låntakeres kredittverdighet.

Contents

1 Introduction

The investment literature builds, in large part, on Keynes' (1936) marginal efficiency of capital. Each firm has one possible investment, which it invests in until the marginal return on that investment is zero (Jorgenson, 1963). Keynes built this on the intuition of Fisher's (1930) theory of investment opportunities, but Fisher's theory is crucially different in one aspect (Alchian, 1955): It is built to understand the investment *alternatives* of firms, giving a framework by which to rank investment opportunities both within and between firms. I apply Fisher's investment model in ranking the return on investment, not only between firms, but also within firms between investment types. By identifying firms that have profitable investment opportunities, and the ability to carry them out, banks and financial institutions drive economic growth (Schumpeter, 1939). If banks were able to *correctly* rank all firms according to their return on investment, external capital would cost as much as internal capital, as there would be no risk (Miller and Mogdiliani, 1958). The average return on investment (Tobin, 1969), the marginal return on investment (Mueller and Reardon, 1993), and the long run return on investment I propose, all aim to identify firms that have profitable investment opportunities but not the funds to carry them out.

The analytical point of departure for my analysis is the concept of cointegration (Engle and Granger, 1987). I show that, in a special case, the stationary residual is equal to profits, which enhances the insight provided by the equilibrium correction (EC) framework. Johansen (1988) combined the EC framework with the VAR framework (Sims, 1980), terming the result the "Vector Equilibrium (or Error) Correction" (VEC) model. Johansen's cointegration coefficient β estimates the long run cumulative effect of a variable on another, using restrictions on the cointegrating matrix space to enable the "partial" effect interpretation of regression coefficients (Frisch, 1933). To my knowledge, no previous literature has applied the concept of cointegration to open banking data, nor, to my knowledge, has the VEC model been applied to estimate the return on investment.

I explore several interconnected research questions related to the return on investment, but the goal of this thesis is singular: to validate my interpretation of Johansen's cointegrating coefficient β as the long run return on investment. I construct several hypotheses to this goal. First, I estimate $\hat{\beta}$ for investments and expenses, to validate that the estimate is higher for investment than for expenses. Second, I split the sample to estimate $\hat{\beta}_I$ for firms pre-classified as either high or low return on investment firms, to test whether the firm can rank the return on investment between firms. Third, I estimate which investments have the highest return within firms and between the firms in second hypothesis, to find *why* some firms are successful while others are not.

Financial technology ("fintech") lenders primarily lend to small business (Beaumont et al., 2022) that are financially constrained due to their short credit history (Maggio, et al., 2022). Regional banks have traditionally lent to small firms (Berger and Udell, 1995), but large banks have increasingly taken market shares from small banks through better information access-and processing

(Berger and Black, 2011). Fintech lenders compete with large banks in information processing (Cornelli et al., 2022), but they may also complement the banks through signalling creditworthy firms (Beaumont et al., 2022). In general, banks that implement fintech in their screening of borrowers decrease their credit risk (Lu et al., 2022). The entry of fintech lenders has put upward pressure on credit supply, and downward pressure on the cost of capital (Balyuk et al., 2022). It is argued that open banking has spurred innovation in the financial sector (Babina et al., 2022), and that this is driving a business cycle within a financial sector that has been in stagnation since the financial crisis. (Cortes et al., 2020). The financial sector accelerates economic-wide businesses cycles through credit (Bernanke et al., 1999), which leads to boom-and-bust cycles (Rajan, 1994), but credit expansion is necessary for economic growth (King and Levine, 1993).

This thesis contributes to financial development through the processing of information, through the formulation of a strategy to estimate the return on investment using business bank accounts. The data I use has been made available to fintech lenders through the open banking initiative, and I have been provided the data from one such fintech lender. MXNEY (”money”) is a Danish firm that finances ecommerce business through revenue-based finance. During the writing of this master thesis, I have been a part-time credit-and investment analyst at MXNEY, and the data I use have been made available for use in this thesis through MXNEY's graduate program.

The thesis is structure into four main parts: Theory, Model, Method, and Results. In Theory, I begin with the role of financial technology and credit in the economy, to which this thesis contributes. I am not aware of previous econometric applications with business bank accounts, so I review potential theories of investment that may suite my hypothesised model, discussing the average, marginal, and long run return on investment, along with asymmetric information, financial constraints, and dynamic business analysis. The VEC model I propose includes several concepts from different theories, which can be applied to answer questions posed in the first theory section, relating to the role of banks. In Model, I describe the components of my theoretical model, which bases itself on the existence of a long run relationship between revenue and costs within firms. A special case is confirmed, in which the residual of regression of revenue and costs equals profits. I implement several extensions to the standard bivariate VEC model, but there are far more extensions that I do not implement. In Method, I present the specific econometric strategy I use to validate my interpretation of the cointegration coefficient β as the long run return on investment, based on three main hypotheses and several secondary hypotheses. The formulation, specification, and presentation of the model, and the cointegration test used as selection criteria are formalised. In the last chapter, Results, I present the estimated VEC parameters in four tables: the estimated cointegration coefficients across all variables (table 5), the full VEC model (table 6), the simplified VEC model (table 7), and separate VEC models for high-and low profitability firms. I summarise, discuss, and conclude the results found in this introduction of a VEC model to estimate the long run return on investment in the final pages of the thesis.

2 Theory

In the first section of this chapter, I contextualise fintech financing and the banking sector. Here it is the economic effects of competition in the financial industry that are most relevant, where it is suggested that the current business cycle in the financial industry is driven by financial information technology, specifically through the application of open banking. It is maintained here that the main competitive edge of fintech lenders over banks is in the processing of hard information, such as business bank accounts. In the second section, I give an overview of the investment literature, which I draw heavily upon in the remainder of the thesis. In this chapter, I explain how credit and investment drive the economy, and how improvements in the processing of information compliment financial and economic efficiency.

2.1 Financial Technology and Economic Development

2.1.1 Fintech Competition

Borrowers with few tangible assets to pose as collateral, but stable streams of revenue for loans are often issued unsecured loans from fintechs (Felix, 2018). The unsecured debt can increase the ability of the firm to gain further funding by investing the original loan in assets, thereby signalling to other financial institutions that the firm is creditworthy (Beaumont, Tang, and Vansteenberghe, 2021). Borrowing from fintechs may thus act as an entry into the capital markets. The three authors find that fintech lenders provide quicker and larger funds to SMEs, and that fintech borrowers grow faster, invest more, default less, and rely less on trade credit than firms not receiving fintech loans. The greater informational efficiency of fintech lenders allow them to require less collateral. Some authors describe the customer base of fintech lender as "invisible prime"-borrowers with short credit history and little collateral, but with low risks of default (Maggio, Ratnadiwakara, and Carmichael, 2022), though there has been observed that fintech borrowers are more prone to bankruptcies (Maggio and Yao, 2019).

Studies show that fintech lenders lend more to geographical areas with higher levels of bankruptcy and unemployment, but that they are able to predict future loan performance more efficiently than the traditional approach to credit scoring (Cornelli, Frost, Gambacorta, and Jagtiani, 2022). Their advantage over banks is in the processing of hard information (Bayluk, Berger, and Hackney, 2022). There are fintech companies that are not lenders, but that offer other financial services, such as payment processing, accounting, or consulting. In general, companies implementing financial technology excel in performing traditional banking operations more efficiently than incumbent banks and financial institutions, often specialising in a single function of banking, such as lending or payment services (Navaretti, Calzolari, and Pozzolo, 2017).

Babina, Buchak, and Gornall (2022), compile a comprehensive overview over open banking initiatives across the world, presenting that 49 governments have a form of open banking policy implemented (as of Oct.21). Most of these governments are in Europe, were almost every country has either fully or partially implemented open banking. In the 49 governments where a form of open banking policy has been implemented, venture capital investments in fintech companies using open banking technology has doubled (Babina, Buchak, and Gornall, 2022).

Large banks have embraced fintech in China, increasing the supply of credit to SMEs, and taking market shares from small banks (Sheng, 2021). Large banks have an advantage in implementing financial technology compared to small banks, due to their existing advantage in hard information processing, but large banks are also most challenged by fintech lenders, who specialise in the same form of creditworthiness assessments (Cheng and Qu, 2020). Digital financial inclusion is most apt at increasing the credit supply to firms that rely less on relationships with their bank and to firms that have hard information to share with their bank. Small banks do not only provide relationship lending, but large banks implementing fintech are taking the non-relationship lending SME market share from small banks (Lu, Wu, Li, and Nguyen, 2022).

Small business lending has long relied on relationships between the firm and bank, through which the bank generates information about the firm's creditworthiness (Berger and Udell, 1995; 1998; 2002). Therefore, it is the small regional banks that have traditionally been the largest source of capital for small business, and their close ties benefit both parties, but with developments in financial information technology, small business lending no longer has to be close, first evident with the advent of widespread credit scoring (Petersen and Rajan, 1994; 2000).

Large banks have traditionally had an advantage in the processing of hard information over small banks (Liberti and Petersen, 2016). Credit scores and financial statements became available for small firms in the late nighties. Large banks then entered the small business lending market, where the large banks gained an advantage over the small banks through the processing of hard information (Berger and Black, 2011). When banks are large, they earn monopoly rents through interest rate premiums, as the lack of competition allows them to charge more for their capital (Boyd and Nicolo, 2005). The market power hypothesis states that less competition increases the cost of capital and decreases the supply of credit, which is confirmed in a large cross-county panel data analysis (Love and Peria, 2014).

The model I propose in this master thesis is a contribution to the processing of financial information. Furthermore, I argue that the model can applied in transaction-and relationship lending, as both large and small banks can use the model to gain insight into the business. In fact, I show that the model can be applied by the firm itself, to rank its investment choices. The application of the model is predicated on business bank account data, and it is an innovative approach based on the information sharing granted by the open banking initiative.

2.1.2 The Supply of Credit

The financial industry distinguishes itself from other industries by virtue of its primary output being credit. To be profitable, the interest made on issuing credit must offset the credit risk, by covering for the losses on defaults. Financiers may at any time expand their credit, thereby increasing market share by increasing risk, without increasing profitability. Competition over market share is therefore associated with excessive risk-taking and over-expansion of credit, but competition is also a key driver of efficiency.

Without lending and borrowing, a firm must first save in order to invest, but in order to save, the firm must first have a stream of revenue from which to save. To generate revenue, the firm must first invest, meaning that the firm must invest in order to save, while saving to invest. This chicken-and-egg conundrum of savings and investments is solved by inter-temporal trade (Obstfeld and Rogoff, 1996), where individuals with savings but no profitable investment opportunities lend to individuals with profitable investment opportunities but no savings to carry them out. The bank is mobiliser of capital, collecting savings, and investing it into businesses that need capital, while keeping the savings available for the savers (Diamond, 1984).

Investments, once made, become iliquid capital, meaning that it must be sold under market price to be quickly turned into money. It is by the converting of liquid savings into capital that the bank earns its profit (Diamond and Rajan, 1999). Banks create liquidity through fractional reserves, meaning that banks lend out more money than is deposited, which is why panics cause bank runs that deposit insurance attempts to prevent (Diamond and Dybvig, 1983). Therefore, and especially after the financial crisis, there is strict banking regulation, which does not apply to fintech lenders, as they do not take deposits. The regulations put in place to decrease risk in the financial industry (Basel III), have made it harder for small banks to operate (Cortes et al, 2020).

Over-expansion of credit occurs when credit is expanded without underlying economic growth, thereby incentivising the funding of worse projects. According to several scholars, this increases macroeconomic volatility and the frequency of financial crisis through the higher levels of risk taken on by the financial industry (Berger, Molyneux, and Wilson, 2020).

Regulators must continuously balance credit expansion and credit risk, counteracting and enforcing innovations as they come along. Historically, regulators have not been able to foresee the future outcomes of innovations nor regulations, passing through a regulatory cycle of innovation and/or deregulation, boom-and-bust, re-regulation, and regulatory failure (Vives, 2019). Many innovations are aimed to walk around existing regulations, making a cat of the regulatory bodies in a cat-and-mouse race with financial innovators. The non-bank financial institutions in particular, not being bound by banking regulation, have been an issue for regulators, which cannot regulate as fast as these regulatory arbiters navigate around the regulations (Cornelli et.al., 2022).

Competition in lending increases the supply of capital, but does not affect the business demand for capital, meaning that the cost of capital must decrease (Rice and Strahan, 2010). Up to the financial crisis, the industry was characterised by fierce competition across the world, but due to the regulations put in place after the crises in Asia, Europe, and the US, the industry has converged toward fewer and larger financial insitutions (Gopal and Schnabl, 2022). Competition has therefore stagnated, and the supply of credit has decreased while the cost of capital has increased, indicating that regulations have been too strict. Therefore, recent regulations have encouraged competition and innovation in information technology, which increases the supply of capital to the economy and reduces the cost of capital, but increases financial risk, though the total effect of increased lending due to gains in informational efficiency is positive (Bayluk, Berger, and Hackney, 2022).

In a perfect capital market, every profitable investment opportunity is realised, and there is no external finance premium (Miller and Mogdiliani, 1958). In this market, there is full information, but due to information asymmetries between the bank and the firm, the bank is unable to accurately identify firms with high credit-default risk (Eklund, 2010). All firms pay for the bank's uncertain identification through the external capital premium, but the "financially constrained" firm is most disadvantaged as it has profitable investment opportunities which are not carried out due to the bank's uncertainty (Gugler, Mueller, and Yurtoglu, 2004).

The cheapest form of capital for the firm is capital generated internally through business, but retained earnings are therefore in low supply (Myers, 1984; 2001). The cost of internally financed investments is equal to the opportunity cost, the earnings the capital would have earned elsewhere. Small and young firms face external capital premiums that are relatively higher, since they are more informationally opaque and cannot pose as much collateral, meaning that their return on externally financed investments will be relatively smaller (Bernanke and Gertler, 1995).

2.1.3 Economic Development

Innovations are the novel combinations of existing input factors (Schumpeter, 1911). The business cycle begins with innovation and competition, the businesses with higher return on investment growing faster. As the business cycle continues, implementation of the innovation takes place, as other businesses see the value in the original innovation. It may be argued that the banking sector is currently in the beginning phases of a new business cycle as a result of the innovation of fintech lenders, which are increasing the output of a stagnating financial industry. Previous research and historical evidence teach us that innovation drives competition, but not only between entrants and incumbents, as incumbents that must implement innovations compete with each other in order to stay on the technological frontier (Solow, 1957). The incumbents that do not reach the new technological frontier, will be replaced by innovating entrants and implementing incumbents in the latter stages of the business cycle (Mansfield, 1962).

Yet, the business cycle is only one of the multiple simultaneous cycles that drive the economy at all times (Schumpeter, 1939). Another driver is the financial cycle, which describes the credit creation by the financial industry, often driven by regulatory changes (Rajan, 1994). Deregulations are put in place to boost the economy (Levine, 1997; 2004), but if the deregulation leads to excess money in the economy, a bust usually follows, as seen in Norway in 1988, the United States in 2001, and across the world in 2008 (Bernanke, 2018; 2020). The prolonged recession we face today seems to be, at least, a hangover from the massive creation of money by the central banks and governments during the Corona pandemic. Increased financial activity without development only serves to increase the swings of the business cycle, leading to larger booms and harder busts (Mises, 1912), due to the phenomena of excess money (Werner, 2016).

Bernanke, Gertler, and Gilchrist (1999) give an explanation for the accelerating effect the financial market has on the economy. Economic upturns are followed by an equivalent upturn in the supply of capital, but no excessive fluctuation in the cycle is necessarily generated. Increases in the capital supply do not lead to excess money if it is met by an equivalent increase in economic productivity. Still it is known that finance accelerates the business cycle. When applying for a loan, the firm must pose collateral as a security for the bank. The net worth of the firm determines how much collateral the firm can pose as security, and this net worth will increase when business is good, as its capital is more productive. With its higher net worth, firms are able to take on larger loans than before. That is, if the firm does not anticipate the devaluation of its net worth when business is not good. Therefore, businesses over-invest in good times and under-invest in bad times, causing larger fluctuations in the business cycle (Bernanke, Gertler, and Gilchrist, 1999).

On this point, it is of relevance to note that business cycle fluctuations in the financial industry causes fluctuations in the financial cycle, which affects all firms in the economy since the output of the financial industry is capital. This expansion of credit is good, if it is the innovation driving the business cycle that drives the expansion of the capital supply, instead of changes to the credit risk, which drives the boom-and-bust dynamic (Werner, 1997). The business-and financial cycles for the financial industry are very hard to distinguish, since the financial industry may always increase output by increasing risk. In terms of causing economic development through distributing capital to firms in ever-continuing business cycles, banks and the wider financial industry may only be said to be contributing if it is increasing the efficiency of capital. That is, by issuing credit to projects with positive marginal rates of return on investment (Schumpeter, 1939).

2.1.4 Questions of Causality and Dynamics

Schumpeter, in his first and second volumes of Business Cycles (1939), details how banks steer the economy through economic development. Essentially, besides mobilising savings, banks function as investors, which maximise their return on investment by identifying the entrepreneurs with the highest chance of repaying their debt. The business cycle is continuous, but an upwards cycle begins with a "wave" of innovation. The upswing ends when the innovation is fully implemented, at which point those who have fallen behind the technological frontier exit the market in the following downswing. The downswing is at is lowest in the trough, swinging up again with a new wave of innovation. The task of the financial sector is to identify and lend to the firms that are going to be close enough to the technological frontier that they survive in the market. By identifying the more profitable firms, the financial sector contributes to economic growth.

In the literature, the big question has been: what comes first, finance or growth? In order to answer this, later authors employed the VEC model's estimation of both directions of causality in both short-and medium runs, finding that the dominant direction of causality is from financial development to economic growth (Rousseau and Wachtel, 1998), that financial development drives growth through investment in the long run (Xu, 2000), and that the long-run effect of finance on growth is stronger than the short run effect (Christopoulos and Tsionas, 2003). These authors' research give strong evidence of King and Levine's (1993) article "Schumpeter Might Be Right", but Rajan and Zingales' *post hoc, ergo propter hoc* contension still stands (Rajan and Zingales, 1996). They argue that since credit expands when *expected* economic growth increases, credit expansion will happen *before* economic growth, but not *cause* economic growth. The "causality" estimated by the VAR and VEC models, is a *before therefore after* interpretation of causality, meaning that this "Granger-causality" is only equal to true causality if causality only leads from before to after. The question of causality that the literature discusses can be boiled down to a question of whether the financial sector generates economic growth, or, whether it is a facilitator of economic growth. In my model, I assume such a difference between investments and expenses, where I describe investments as *generating* growth through expanding supply and demand, while expenses *facilitate* growth by fulfilling the demand for supply.

We may view firms as small open economies in a monetary union, which trade with other firms and sell to consumers. In doing this, they seek to maximise their return on investment, but they must choose between reinvesting or saving their return (Marris, 1964). Dynamic econometrics has been used to understand how firms grow through their choices, but through two generally separated areas of study, the first studying revenue and the second profits. In a late 2021 article, a VAR model was proposed to combine the two fields (Idsø, Opstad, and Valenta, 2022), since it simultaneously estimates the autoregressive and cross-terms between revenue and profits. The model in this master thesis is inspired by that model, although with several key differences. First, instead of including profits directly into the model, I estimate and define profits as the residual in an equilibrium between revenues and costs. The residual from this equation is then included in the revenue-and cost growth model. Second, in this study, I do not estimate the persistency of profits (Mueller, 1986; Hirsch, 2018), which can differentiate between the long run profits within sectors by segment (Opstad and Valenta, 2022a) and be used to explain market turnover (Opstad and Valenta, 2022b). The parameter, often denoted λ, is used to estimate the long run profit rate (Mueller, 1986), similar to in the current study. I do not include it here, however, since I use the cointegration coefficient β to estimate the long run return on investment instead. In this way my master thesis, building on previous research, combines two of Mueller's concepts and estimations: the long run rate of profit (Mueller, 1986) and the marginal return on investment (Mueller and Reardon, 1993).

Euler's constant is also of relevance to discuss as it describes growth through the percentage interpretation of logarithms, used to understand planetary movements, the growth of plants and animals, and normal distributions, as much as to understand economic growth (Stigler, 1986). In business economics, Euler's constant is applied in Gibrat's Law of Proportional Effect, which corresponds to all businesses in a sector growing relative to their size (Gibrat, 1931). The law holds when the autoregressive α parameter of revenue is equal to unity, the moving average component of revenue μ is equal to zero, and the variance of a firm's revenue σ is relative to its revenue (Mansfield, 1962). By some interpretations in the literature, the current VEC model also estimates this α, but I partly disagree, as the α of revenue growth is much different than the α of total revenue (Valenta, Idsø, and Opstad, 2021). In the current study, I do not use panel data methods, since the "fixed effects" at its core (Wooldridge, 2013), are at odds with the EC model's deliberate exclusion of constants in the equilibrium equation. I do show the steps of the VEC model with panel data for convenience (in table 3), but I do not use panel data to reach any of main results.

2.2 The Return on Investment

In the second and final section of this chapter, I review the dominant theories of investment, and how they relate to my model. The general theory of investment follows in the footsteps of Keynes (1936), but his theory is a simplification of Fisher's (1930) investment model, where multiple investment choices are ranked according to their return on investment. Keynes' simplification is great for estimating the return on investment between firms, and it is a strong building-stone for further analysis, in particular through Tobin's (1969) framework. My investigation into the return on investment is more in spirit to the direct estimation of return on investment proposed my Mueller and Reardon (1993) and the long run rate of profit (Mueller, 1986). It is Fisher's original (1930) model, however, that best encapsulates my model, since it details the ranking of investment opportunities within firms. The model I propose can, in the Keynesian Tobin's q fashion, rank firms by their return on investment, but I extend the model to enable the ranking of investment choices within firms per Fisher's model. This enables the identification of factors that make some firms successful and others not, or equivalently, the identification of areas for improvement in the business.

2.2.1 General Theory of Investment

The return on investment is central in economics, determining which businesses grow, which shrink, and which leave the market (Jovanovic, 1982). Positive returns on investment generate earnings, which may be saved or reinvested into the business. Savings have a return equal to the interest rate offered by the bank where the savings are held. The rational firm should keep investing until the return on investment is equal to the interest on savings, at which point it is indifferent between investing and saving (Eklund, 2013). Profitable businesses earn profits on their investments, equal to difference between the revenue generated by investment and the cost of investment. The difference between the expected revenue generated and cost induced by investment, is the present net value of an investment option (Mueller and Reardon, 1993).

The present net value of saving the money is equal to the interest rate on savings, meaning that any investment must be compared to the rate of interest at the bank. A higher interest rate will reduce the present net value of all other investment equivalently to the increase in the interest rate. The marginal efficiency of capital is equal to the interest rate that equates the revenue generated by investment and the opportunity cost of the investment (Keynes, 1936). A higher interest rate reduces consumption demand too, meaning that both the opportunity cost and the revenue generated are affected by the interest rate, being increased and decreased, respectively.

The Keynesian interpretation of the productivity of capital is convenient in understanding his description of the central bank's counter-cyclical role in the macro economy. When firms invest, however, they do not only compare investment and interest rates, rather, they compare investments with each other. Therefore, an investment model must incorporate the different investment opportunities. Fisher's rate of return over cost is therefore defined as the difference in net present value between investments, both discounted by the interest rate on savings (Fisher, 1930). The marginal rate of return over cost is then the interest rate that sets the difference between the two investment opportunities equal to zero. Though Keynes claimed that his marginal efficiency of capital and Fisher's rate of return over cost where precisely the same, it seems, as Alchian (1955) observes, that he must have overlooked Fisher's clear description of investment opportunities as a choice between two alternative investments.

In the spirit of Fisher's return on investment, I compare the rate of return of investment choices, comparing the return on investment on inventory purchases and marketing efforts. At any point in time, the advantage of one investment over another may be described as $(R_1 - C_1) - (R_2 - C_2)$ where R denotes the revenue generated by investment, C is the cost of investment, and the subscript $i = 1, 2$ denotes the investment alternative. My comparison of return on inventory and marketing may be understood similarly, but the stream of revenue that will result from investment cannot be known a priori. Here lies the main goal of this master thesis: to estimate the revenue that is generated by investment, in order to evaluate investment opportunities within and between firms.

2.2.2 Expected Return on Investment

Most industries have marginally decreasing returns to scale, meaning that each new investment will be less profitable than the last. All capital has a return on investment, and each sector has a demand for capital (Tobin, 1969). The higher the return on capital, the higher the demand for capital, the higher the price of capital, and the lower the return on investment. Through market forces, a higher return on capital reduces the return on investment (Schumpeter, 1939).

Measuring the return on investment on all investments is a difficult task, but if all returns on investment converge toward a common economic return on investment, the Keynesian approximation of Fisher's opportunity cost may be a useful one. The key feature in measuring the return on investment is the firm's maximisation function. This is true whether the firm saves, invests, or withdraws profits, the firm will always maximise its return on investment. That is, to the best of its ability. As Jorgenson (1963) points out, the firm will, at all times, maximise its long run return on investment, thereby maximising its long run profit and net worth.

The present value of the investment must be higher than the cost of the capital needed for the investment to be carried out (Schumpeter, 1911). The present value decreases with increasing discount (interest) rates (Keynes, 1936). For the present value of the investment to be greater than the opportunity cost, the return on investment must be higher than the interest rate. The market value of a firm is equal to the market value of the previous period, together with the present value of the investment made in this period, minus depreciation, where the market value is defined as the sum of physical capital, intangible capital due to marketing, research and development, and any other capital (Mueller and Reardon, 1993). This definition of the market value allows for the evaluation of different investment opportunities within each business, following Fisher's (1930) description of investment opportunities as choices, which is in the same spirit as my estimation of the return on investment *choices*. Mueller and Reardon's (1993) "c" is equal to the ratio of the perpetual (long run) return on investment "r" over the discount (interest) rate "i". If c is greater than unity (1), the investment has a positive net present value, meaning that it should be carried out. With a c less than 1, the firm will be losing money by carrying out the investment. This similarity with the model proposed in my thesis and the Mueller and Reardon (1993) paper begins and ends with the foundation in Fisher's (1930) investment model, except that both use regressions, though the cointegration coefficient I estimate has a similar interpretation to Mueller's (1986) long run return on (or rate of) profits.

To estimate the expected return on investment of a firm, the investor must have extensive information on the firm's business, but the firm may itself find it challenging to determine its return on investment. Jovanovic (1982) puts forth the assumption that a firm entering a new market has a limited idea of how profitable and efficient its business will be relative to the incumbents in the sector. As the firm conducts its business, it begins to understand how efficient it is, in other words what its individual efficient scale (IES) is. Until reaching the IES, the rational firm keeps reinvesting all earnings, but when the IES is reached, the rational firm retains all earnings, keeping the businesses running at status quo. Each sector has a minimum efficient scale (MES). Firms with an IES below MES exits the market, and through this process of self-selection the market evolves (Jovanovic, 1982). With innovation and the consequent start of a new business cycle, there are many entrants into the market. To survive in the sector over the long run, the firm must adapt, implement innovation, and learn by doing (Ericson and Pakes, 1992).

Perspectives on investment in the context of unknown information are also of relevance for my thesis because I aim to gain insight into their "true" return on investment. That is the task of every lender and every firm, but the firm has traditionally been understood to have the informational advantage over the bank. The asymmetric information between bank and firm causes the external finance premium through its agency costs (Bernanke and Gertler, 1995). With open banking, lenders gain great insight into firms, and I will explore through the application of the model, whether lenders can gain *greater* insight into the return on investment of a business than the firm itself can.

2.2.3 The Revenue-Profit-Investment Channel

New small firms are often viewed as being financially constrained, in that they have profitable investment opportunities, but not the funds to carry them out (Fazzari, Hubbard, Petersen, 1987; 2000). In this case, the firm may apply for external funds, if the return on investment exceeds the discount rate plus the interest paid on the debt. This external finance premium is driven by informational asymmetries between the firm and bank, the firm not being able to convincingly convey the return on their investment opportunities to the bank (Gugler, Mueller, and Yurtoglu, 2004). Smaller firms are more opaque, the smallest lacking even financial statements, in addition to having little existing wealth to pose as collateral, leading to the common finding that smaller firms face larger financial constraints than larger firms (Berger and Udell, 1998). Financial development, described as gains in efficiency of the financial sector, has been observed to disproportionately benefit small firms (Beck, Demirguc-Kunt, Laeven, and Levine, 2008). The excess sensitivity small firms have to improvements in financial efficiency is taken as an indication that small firms face larger financial constraints, although the finding may also indicate that financial developments have targeted the financial inclusion of small firms.

Several perspectives on cash-flow sensitivity have also guided my study, combined with literature on the dual causality concerns (Fazzari, Hubbard, and Petersen 1987; Kaplan and Zingales 1997; Gilchrist and Himmelberg 1998). It is maintained that internal finance is cheaper than external finance, but that external finance is cheaper for firms that can more convincingly confirm their positive return on investment (Akerlof, 1962). The saving accumulated through the retention of earnings generated by profitable investments is the only source of funds for firms that cannot fund their investment externally. With profitable investment opportunities, but not the funds to carry them out, a firm will invest as soon as funds are made available to it. Its investments are therefore "cash-flow" sensitive. Firms that have sufficient funds to carry out all their profitable investment opportunities will not reinvest their earnings, as there are no further profitable investments to carry out. Therefore, a financially constrained business may be identified by the sensitivity of its investments to its cash-flow (Fazzari, Hubbard, and Petersen, 1987).

Increases in the revenue generated by investments, however, indicate that the return on investment has increased, meaning that the cash-flow sensitivity of investments may capture the firm's adjustment to a new optimal capital accumulation, invalidating the interpretation of investment cash-flow sensitivities as a measure of financial constraints (Kaplan and Zingales, 1997). Therefore, Tobin's Q must be added to the model, to control for the increased present value of investment suggested by increased cash-flows (Fazzari, Hubbard, and Petersen, 2000). The inclusion of Tobin's Q as a control variable could validate the cash-flow sensitivity strategy of identifying financially constrained firms, but there is no way of cementing Tobin's Q complete identification of changes in the profitable investments. As I see it, the core of this discussion is the fact that revenues are generated by investments, but that revenues also determine the available funds the firm has to carry out investments. Investments generate revenue, so it may be argued that investments and revenue both lead to each other. Therefore, the central question is one of causality, where the best course of action to estimate both directions of causality simultaneously, as through a VAR or VEC model (Lutkepohl, 2005).

By including the signal cash-flows send about the return on investment explicitly into the investment model, the cash-flow sensitivity of investments can be more accurately interpreted as financial constraints (Gilchrist and Himmelberg, 1995; 1998). The vector framework effectively distinguishes between the response of investments to fundamental factors signalled by the revenue, and the effect of available internal funds, finding that the causality goes both ways (Love and Zicchino, 2006).

2.2.4 Marginal Return on Investment

The applications following in Tobin's footsteps, such as in the cash-flow sensitivity literature (Fazzari, Hubbard, and Petersen 1987; 2000; Kaplan and Zingales, 1997; Gilchrist and Himmelberg 1995; 1998), measure the average return on investment. The average return on investment is only equal to the marginal return on investment if, and only if, there are constant returns to scale (Hayashi, 1982). If the return to scale is decreasing, as it is in most industries, the marginal return on investment will be less than the average return on investment. The average return on investment is the mean of the return on investment on all previous investments, while the marginal return on investment is the return on investment on the *next* investment.

I do not estimate the marginal, nor the average, return on investment. Rather, I estimate the long run return on investment, as defined by the cointegration coefficient describing the cumulative effect of investments on revenue β_I (Johansen, 1988). The cost of investment is normalised to one, meaning that the long run profit of investment is $\beta_I^* = \beta_I * C_I - C_I = \beta - 1$, where C_I is the cost of investment. This parameter is similar in spirit to Mueller's (1986) long run profit, defined by the autoregression of the persistency of profits (Hirsch, 2018), and his definition of the return on investment (Mueller and Reardon, 1993). My methodology is much different from Mueller's, but also to all other studies of my knowledge, due to the use of open banking data for econometric analysis and the use of the VEC model to estimate the return on investment. The simplified multivariate VEC model I propose can be used for firms with very few observations, and the daily data afforded by open banking is not necessary.

I simplify Fisher's (1930) model, by not including a time-horizon in his equation for the marginal rate of return over cost, focusing on his ranking of investments by their return. By finding that the firm as having multiple equilibria, one between revenue and each type of investment, the VEC model can estimate the realisation time of investments, meaning that this extension would complete Fisher's investment model, after a expected discount-or interest rate on savings is included, but I choose to simplify rather than complicate the model. The power of simplification and generalisation is that many variables can be included in a model, each variable having a clear and understandable role in the system (Keynes, 1936). By constructing clear and observable models, the variables can be measured and the parameters in the model can be estimated. As will be shown in the next chapter, the model I propose has the same benefits. The main part of my thesis is in confirming that the model is estimating the return on investment, for which I construct hypotheses with logical expected outcomes, that are rejected or not on the basis of simple statistical tests.

The model I construct estimates the long run return on investments, and it can be used to rank the return on investment between firms, but also between investments *choices* within firms. In the long run, the investment has fully turned over into revenue. This is different from the average return on investment, which can in the most simple of examples be defined as excess revenue after costs, as I do when separating high and low profit-and growth firms. This is *not* the "true" return on investment, which is better described by the long run or marginal return on investment, but firms that have had high excess revenue generally have a higher return on investment. The average and long run return on investment will only be equal if inventory firm is constant, but firms actively use their inventory as a store of value. The long run return on investment β is also different from the marginal return on investment, though less obviously. The cointegration coefficient estimates the cumulative effect of a unit increase in a type of investment on revenue, while controlling for other costs of the business. The marginal return on investment is the return on the *next* investment, but the long run return on investment is how much each investment *choice* earns in the long run. That is, where should the firm spend its next dollar, and which firm should be funded?

3 Model

In this chapter, I construct the model that I use to estimate the long run return on investments, the cash-flow sensitivity of costs, and the realisation-time on investment. The chapter begins by describing the data. I then present the descriptive statistics of all variables in Table 1, where I split the sample into various categories to gain insight into the data. I describe the profit-and growth rate per sector per quarter in Table 2.

In the second section of this chapter, I present the most important concept in my master thesis: The equilibrium between revenue and cost that is profit. It is through profits that revenue and cost affect each other. Profits are the excess revenue after cost, and they are the internal funds of the firm, which are used to pay for investments and expenses. Revenue affects costs in the short-run through the financial constraint mechanism, but investments generate revenue in the long run, and I argue that investments *cause* revenue. The model incorporates short-run signalling effects and medium-run cash-flow sensitivity, meaning that these effects are controlled for in the estimation of the long run return on investment. In the last section of this chapter, I extend the model to enable the ranking of investment opportunities, where I also discuss the possibility of multiple equilibria. Instead of complicating the model with multiple equilibria, I simplify the model, which allows for a comparison between moving average and aggregate data.

3.1 Data Description

My data consists of MXNEY's live open banking connections. The data have been made available to me from MXNEY through their graduate program. It is categorised by the description of each transaction, such as "google ads" being marketing, which is partially done by manual labour and partially done by MXNEY's automatic categorisation. In the second step of the construction, I aggregate the data to daily intervals in order to apply econometric analysis. In this step, the description of the transaction is replaced by its categorisation, and the firms' names and identifiers are replaced by numbers, meaning that the data is completely anonymous.

MXNEY is a revenue-based finance lender, based in Copenhagen. It specialises in lending to small firms, but specifically to ecommerce. These are the firms in my sample, but I do not inquire into ecommerce specifically, except if the data reveals relevant aspects of the sector. The big picture of fintech lenders in the economy, as an innovative competitor in the financial industry, is of more relevance, as the model I propose can be used by anyone with an open banking connection and statistics software.

Ecommerce businesses must have a webshop, but they can also have a physical store. Pure ecommerce businesses can operate without fixed costs in theory, meaning that they have a low probability of default. All the sales of an ecommerce are recorded in their web-shop, which makes

MXNEY's revenue-based finance model possible from a technical standpoint. One of the main advantages of revenue-based finance over traditional debt finance is that the repayment is relative to the firm's revenue (Felix, 2018).

MXNEY started in the first quarter of 2020, and it now lends to between one and two hundred active firms. My inquiry is into these firms, which as of 31.03.23, consist of 232 open bank account connections for 146 individual firms. Of the 232 connections, 113 are inactive, meaning that their data was not live on the cut-off date (31.03.23). The 119 active connections describe 78 individual firms. Many of the businesses are physical first but with webshops, although the distinction is hard to make in the data. 23 of the 72 active firms are sole proprietorships, that use their personal accounts to do business, unlike the limited liability companies that have business accounts. Personal accounts describe people better than businesses, but since this is not a psychology paper, I choose to remove these from the sample. The 52 business banks accounts of limited liability companies (traditional firms) that are connected to the MXNEY server as of 2023Q1, describe each and every transaction the business has conducted using the account, with an amount, a date, and a description of the counter-party of the transaction. That is, as far back as the data goes. To make any quantitative analysis, sufficient observations are needed. How much is sufficient is harder to determine, but I set the arbitrary cut-off at one year, meaning that the 9 firms with data stretching less than a year back in time, are excluded from the sample.

Bank account data is infamously tedious to work with, as all (most) transactions must be categorised before analysis can take place. Automatisation helps, but only partially, since descriptions may be cryptic or misguiding. Therefore, considerable human hours are necessary to make distinctions that the computer cannot. Still, there is a limit to what can be adequately categorised. To avoid measurement error, I removed an additional 8 firms from the sample, those with data that I could not categorise more than 50 percent of. The remaining 35 firms in the sample contain a total of 189 thousand observations, all described by a category and a sub-category, in addition to a date and an amount.

It should be noted that my thesis is less about ecommerce, than it is about the methodology I propose: to use open banking data to estimate the cointegrating coefficient β, which I interpret as the long run return on investment. To apply the VEC model appropriately, I investigate the descriptive statistics of the sample, to gain insight into the general market forces, trends, dynamics, and distributions of the data that form the observational foundation of my model, and the economic interpretation I give to its estimated parameters. The model is not restricted to application in ecommerce in any way, and may be employed by anyone with business bank account data and a statistics software package. The thesis applies the open banking data from MXNEY in order to introduce a combination of information sharing technology and statistical methodology to the field of business lending.

3.1.1 Cash Flow, Input Factors, and External Finance

I begin by collapsing the data, so that I have daily observations, rather than per transaction data. This reduces the sample from 189 thousand observations, to a bit more than 18 and a half thousand. I categorise into 5 variables: Revenue, Costs, Taxes, Transfers, and Finance. Revenue is the main inflow and costs are the main outflow. Without external finance, profits are equivalent to the difference between revenues and costs:

$$P_{it} \equiv R_{it} - C_{it} \tag{1}$$

where P is profits, R is revenue, C is costs, i stands for firm, and t stands for time. The equivalence is universal, holding for every firm at any point in time. Equation 1 is fundamental for understanding the remainder of the findings in this model, method, and results of this thesis. In a day where the firm has more revenue than costs, the firm is profitable, earning the excess revenue as profit. If a firm spends more than it sells one day, it was not profitable that day, losing the excess costs. Revenue is more stable day-to-day than costs, meaning that firms usually have many days with small positive profits and few days with large negative profits. In a traditional regression of revenue on cost, the residual is equal to the profit if the coefficient is equal to 1 (equation 5).

The input factors of a businesses are varied, but can be broken into two basic categories: Investments and expenses. Investments generate revenue, while expenses facilitate revenue. The most traditional investment is into means of production, but ecommerce businesses rarely produce. In ecommerce, investments are mainly into inventory and marketing, which expand the firm's supply and demand. An investment the ecommerce business might do, is into software, but software costs are usually better categorised as an expense. It is the second largest expense, after shipping and logistics. Expenses have in common that they do not generate any revenue themselves, but rather function to facilitate for the supply and demand that the investments generate. I categorise all costs that are not investments into inventory or marketing as expenses, to ensure that the reality of total costs is not lost. In general, there are two types of expenses: fixed and variable. Fixed costs remain the same as the business grows, but they can pose a treat for the business financially if it struggles. I categorise and sub-categorise costs as:

$$C_{it} \equiv I_{it} + E_{it} \equiv II_{it} + MI_{it} + FE_{it} + VE_{it} \tag{2}$$

where I is investment, E is expense, II is inventory investment, MI is marketing investment, FE is fixed expense, VE is variable expense, $I = II + MI$, and $E = FE + VE$. All costs that cannot be classified as II, MI, or FE, are classified as VE, which functions as an "other" costs variable. Therefore, the sum of the four cost categories will be equal to costs (see table 1 below), with an error equal to the *unknown* transactions that could not be categorised.

Table 1: Descriptive Statistics and Split

The values below are in daily average Danish crowns (DKK), meaning that the average firm in the sample had a daily revenue of 23274 DKK. The average excess revenue, the revenue after costs, which I call profit, is equal to 23274-20459=2815 DKK. All firms have been a non-MXNEY customer, before the first round of finance. Strategies are mutually exclusive and based on the quarterly efficiency rates in the third section.

Descriptive Statistics and Comparisons	**All Sample Firms**	Non-MXNEY Customers	Active MXNEY Customers	Before Inflation 2020-21	After Inflation 2022-23	High Profit Firms	Growth Maximising Strategy	Unclear Maximising Strategy
Cash Inflow	**27 605**	22 342	41 748	21 218	32 065	34 072	27 002	22 763
	(62 624)	(43 516)	(95 385)	(43 521)	(72 687)	(52 895)	(87 767)	(49 173)
Cash Outflow	**- 27 982**	- 22 922	- 41 577	20 738	33 004	- 34 559	- 26 764	- 23 511
	(69 644)	(51 503)	(102 488)	(47 172)	(81 289)	(63 647)	(96 263)	(50 700)
Revenue	**23 274**	19 332	33 866	**18 507**	**26 648**	**30 017**	**21 073**	**19 080**
	(52 924)	(38 978)	(78 053)	(39 194)	(60 513)	(47 876)	(69 120)	(43 999)
Total Costs	**- 20 459**	- 17 355	- 28 802	**- 15 761**	**- 23 708**	**- 23 962**	**- 18 809**	**- 19 197**
	(56 298)	(42 406)	(82 217)	(39 469)	(65 300)	(48 121)	(77 913)	(45 431)
Taxes	**- 1 957**	- 1 376	- 3 520	- 1 086	- 2 580	- 3 045	- 2 074	- 811
	(24 328)	(16 432)	(38 127)	(16 754)	(28 423)	(26 907)	(31 623)	(13 172)
Transfers	**- 228**	- 264	- 133	- 257	- 213	- 641	245	- 154
	(9 322)	(9 915)	(7 499)	(11 088)	(8 134)	(6 582)	(10 788)	(10 844)
Cost Categories	**All**	Non-MX	MXNEY	2020-21	2022-23	Profit-Max	Growth-Max	Unclear-Max
Inventory Inv.	**9 605**	8 551	12 442	8 094	10 666	**12 220**	**7 585**	**8 889**
	(39 657)	(31 881)	(55 292)	(29 689)	(45 344)	(33 586)	(51 556)	(36 357)
Marketing Inv.	**4 779**	3 335	8 658	2 662	6 165	**4 119**	**6 896**	**4 197**
	(26 469)	(10 707)	(47 486)	(10 421)	(33 103)	(14 722)	(46 649)	(12 123)
Fixed Exp.	**2 437**	2 393	2 556	2 254	2 576	3 754	885	2 337
	(11 042)	(11 360)	(10 141)	(11 986)	(10 534)	(16 155)	(4 500)	(7 176)
Variable Exp.	**3 480**	2 951	4 901	2 564	4 158	3 445	3 447	3 774
	(15 768)	(13 594)	(20 431)	(10 912)	(18 340)	(15 894)	(18 636)	(13 600)
Unknown	**1 235**	1 014	1 831	1 106	1 354	2 734	441	287
	(16 825)	(17 661)	(14 324)	(18 807)	(15 676)	(25 487)	(6 285)	(9115)
Efficiency Rates	**All**	Non-MX	MXNEY	2020-21	2022-23	Profit-Max	Growth-Max	Unclear-Max
Growth Rate	**7.59%**	5.05%	12.31%	13.44%	4.67%	**- 0.21%**	**17.94%**	**9.76%**
	(40.46%)	(36.14%)	(47.32%)	(36.73%)	(42.01%)	(33.25%)	(49.30%)	(40.28%)
Profit Rate	**9.61%**	8.18%	13.02%	10.24%	9.22%	**23.60%**	**- 0.09%**	**2.25%**
	(27.09%)	(27.72%)	(25.40%)	(30.48%)	(24.62%)	(24.64%)	(30.31%)	(19.56%)
External Finance	**All**	Non-MX	MXNEY	2020-21	2022-23	Profit-Max	Growth-Max	Unclear-Max
Debt Finance	**87 605 [42]**	75 463 [29]	114 690 [13]	34 155 [10]	104 307 [32]	128 331 [3]	35 300 [16]	118 678 [23]
	(106 738)	(104 770)	(110 284)	(28 869)	(116 690)	(68 936)	(28 281)	(130 498)
Internal Equity	**56 940 [58]**	62 368 [50]	23 017 [8]	63 174 [29]	50 706 [29]	82 639 [18]	26 461 [21]	66 282 [19]
	(68 422)	(72 159)	(12 097)	(53 597)	(81 099)	(102 808)	(17 712)	(51 551)
External Equity	**241 836 [11]**	175 000 [4]	280 028 [7]	187 500 [2]	253 911 [9]	0 [0]	241 836 [11]	0 [0]
	(268 745)	(61 237)	(337 369)	(88 388)	(297 323)	(0)	(268 745)	(0)
Revenue-Based Finance	**90 477 [158]**	0 [0]	90 477 [158]	66 293 [5]	91 267 [153]	67 757 [56]	110 496 [46]	96 752 [56]
	(89 326)	(0)	(89 326)	(55 462)	(90 226)	(68 944)	(121 514)	(71 004)
Firms	**35**	35	29	18	35	14	11	12
Days	**18 627**	13 575	5 052	7 166	11 142	6 969	4 857	6 801

Internal equity is mainly wealth, but generally non-revenue internal funds. The number of observations for each type of external capital inflow is denoted in the [brackets], no non-MXNEY customers have received RBF.

Logarithmic transformations are the most appropriate for monetary models, as I will discuss in Section 4.2. It is based on Euler's constant, which enables the forming and testing of many economic hypothesis. Specifically, the logarithmic transformation of X is the number Y that solves $e^Y = X$, where e is Euler's number, equal to 2.7183. The change in a logarithmicly transformed variable is *approximately* equal to the rate of growth, and the difference between logarithmic revenue and cost is *approximately* equal to the rate of profit:

$$Log(P_{it}) \equiv Log(R_{it}) - Log(C_{it}) = (R_{it} - C_{it})/R_{it} \equiv PR_{it} \tag{3}$$

where Log stands for the logarithmic transformation of X, and PR stands for the rate of profit. The equality between Log(P) and PR is approximate, but it is what I use for splitting the firms into three categories, according their growth and profit. I classify a highly profitable firm as one that has more than a 15 percent average logarithmic profit (rate). High growth firms have more than a 15 percent average revenue growth per quarter, defining growth as:

$$Log(G_{it}) \equiv Log(R_{it}) - Log(R_{it-1}) = (R_{it} - R_{it-1})/R_{it} = GR_{it} \tag{4}$$

where G stands for growth, GR is the growth rate, and the time interval is quarterly ($t = t-1 = 91$ days). No firm averaged both a profit rate and a growth rate above 15 percent per quarter in the sample. Those who averaged less than both, form the third category, which I use a bench-line. The two rates in equation 3 and 4, profit and growth, are not used directly in any analysis, but only serve as a split of firms in segments.

The firms with high profit have an average profit of 23.6 percent, combined with a growth rate of negative 0.2 percent. The standard deviation of the growth rate, describing the volatility of the growth, for high profit firms is 33 percent. High growth firms have a profit of negative 0.1 percent, but an average quarterly growth of almost 18 percent. Their growth volatility is 17 percent higher than for the high profit firms. The bench-line firms had an average quarterly growth rate of a bit less than 10 percent, and an average profit rate of 2 percent. The high profit firms are the largest, followed by the high growth firms, and the then the bench-line firms, with respective daily average revenues of 30K, 21K, and 19K.

The build up of input factors is strikingly different among the three categories. High profit firms invest three times as much in inventory (12.2K) as in marketing (4.1K), while the high growth firms invest almost as much in marketing (6.9K) as in inventory (7.6K). High profit almost spend as much on fixed expenses (3.8K) and variable expenses (3.4K) as on marketing, while the high growth firms have close to none fixed expenses (0.9K), although their variable expenses (3.4K) is equal to that of the high profit firms. Note that variable "expenses" function as "other" category of costs, and will therefore be inflated.

Only high growth firms have received external equity in the sample. These firms have received an average of 241K in external equity over 11 equity investments. The high growth firms also receive the largest tranches of revenue-based finance from MXNEY, averaging 110K over 46 tranches. The bench-line firms have received as many tranches as the high-profit firms, but the high profit firms received 68K per tranche, while the baseline received 97K per tranche. Though the high profit firms have the highest revenue, they receive the smallest loan tranches, contrary to the logic of (revenue-based) finance. If we view the high-profit firms as maximising profit, as indicated by their average 23.6 percent profit rate and average 0.2 percent growth rate, one may pose a question as to why they want to receive external finance at all. Only businesses that are financially constrained demand external funds, but these high profit firms must have a surplus of retained earnings, as they have not reinvested them.

MXNEY requires at least six months of data from its potential borrowers, meaning that every firm in the sample has at some point in time *not* been a borrower from MXNEY. I split the firms in the sample into a pre-and post MXNEY period. Since every firm has been unaffiliated with MXNEY in the past, the comparison is roughly indicative of the effect that MXNEY has on the firms. The average MXNEY-firm has a 14K higher revenue than the non-MXNEY firm, but firms generally had 8K higher revenue in 2022-23 than in 2020-21. Ecommerce was booming during the pandemic and the ensuing lockdown, as business went online. The growth of the firms in the sample was 13.4 percent in 2020Q2-21Q4, compared to 4.7 percent in 2022Q3-23Q4. The profit rate stayed the same, at about 9-10 percent. Firms, in general, grow by 5.1 percent a quarter before they receive funds from MXNEY, but 12.3 percent after they receive money. Their growth increases from 8.2 percent to 13.0 percent after receiving money from MXNEY.

3.1.2 Booms, Cycles, and Strategies

To investigate the ecommerce business cycle closer, I split the sample into quarters, from 2020Q2 to 2023Q1, and into sectors, identifying the average profit-and growth rates per sector per quarter (see Table 2 below). The highest growth in the sample was in 2021, during the lockdown-phase of the pandemic, but also in the fourth quarter of each year, with Christmas shopping. Profitability was highest in the first quarter of 2021, falling from 39 percent and stabilising at 10-15 percent. The most common sector in the sample is the "clothing" sector, which includes everything from secondhand stores, to shoe sellers, to sporting equipment (in textile), and hats. Most businesses in the clothing sector are B2B (business to business) although one or two may be classified as D2C (direct to customer). The clothing industry sticks out as the most stable of the sectors, with the lowest standard deviation in terms of growth, although this may be due to the sector's large sample size. In the recorded period, no quarter was unprofitable for the clothing sector as a whole, but their quarterly growth (10.9 percent) has only been half that of the electronics sector (19.5

Table 2: Evolution of Sectors per Quarter

Below, I have detailed the mean growth rate and profit rate of the firms in my sample, by quarter and sector. The common high-season in ecommerce is during the fourth quarter.

Evolution of Sectors by Quarter	**All Firms [35 Firms]**	Electronics [8 Firms]	Clothing [11 Firms]	Pets [3 Firms]	Design [4 Firms]	Other [9 Firms]
Growth Rate All Time	**9.92% (54.57%)**	**19.50% (80.46%)**	**10.90% (35.75%)**	**6.93% (41.90%)**	**9.27% (54.35%)**	**5.32% (45.13%)**
2020-Q2	**- 13.83% [4]**	- 71.20% [2]	N/A	N/A	30.89% [1]	56.18% [1]
2020-Q3	**- 45.57% [4]**	- 27.23% [2]	N/A	N/A	- 66.71% [1]	- 61.03% [1]
2020-Q4	**24.00% [7]**	- 37.50% [2]	7.94% [2]	N/A	73.66% [1]	76.72% [2]
2021-Q1	**21.57% [13]**	67.47% [4]	9.72% [3]	N/A	- 51.73% [2]	21.20% [4]
2021-Q2	**30.43% [18]**	56.73% [4]	11.48% [5]	N/A	38.59% [3]	24.62% [6]
2021-Q3	**- 7.46% [24]**	- 26.73% [5]	- 0.72% [8]	73.76% [2]	- 1.01% [3]	- 30.69% [6]
2021-Q4	**29.22% [28]**	40.02% [6]	30.82% [8]	- 12.33% [3]	49.16% [4]	24.52% [7]
2022-Q1	**- 1.54% [32]**	-7.36% [8]	16.95% [9]	- 3.72% [3]	- 34.12% [4]	6.00% [8]
2022-Q2	**28.69% [34]**	44.93% [8]	26.08% [10]	28.38% [3]	.38.71% [4]	12.79% [9]
2022-Q3	**- 10.85% [35]**	17.94% [8]	- 11.88% [11]	- 4.32% [3]	- 4.82% [4]	- 40.05% [9]
2022-Q4	**31.32% [35]**	50.87% [8]	25.56% [11]	10.37% [3]	16.24% [4]	34.65% [9]
2023-Q1	**- 2.35% [35]**	10.62% [8]	- 5.21% [11]	- 21.33% [3]	4.16% [4]	- 6.94% [9]
Profit Rate All Time	**14.81% (40.12%)**	**- 2.78% (40.24%)**	**14.72% (30.17%)**	**15.46% (24.12%)**	**4.87% (25.47%)**	**35.15% (53.72%)**
2020-Q2	**- 2.72% [4]**	- 29.10% [2]	N/A	N/A	25.40% [1]	21.91% [1]
2020-Q3	**- 3.80% [7]**	1.65% [2]	7.51% [2]	N/A	11.35% [1]	- 28.15% [2]
2020-Q4	**14.73% [13]**	- 42.26% [4]	21.04% [3]	N/A	- 38.22% [2]	93.45% [4]
2021-Q1	**39.06% [18]**	- 13.18% [4]	46.07% [5]	N/A	- 5.17% [3]	90.16% [6]
2021-Q2	**25.26% [24]**	- 12.31% [5]	27.48% [8]	- 0.00% [2]	2.88% [3]	73.26% [6]
2021-Q3	**17.90% [28]**	7.58% [6]	20.22% [8]	18.94% [3]	13.55% [4]	26.11% [7]
2021-Q4	**10.90% [32]**	4.26% [8]	7.63% [9]	27.03% [3]	0.19% [4]	20.52% [8]
2022-Q1	**10.38% [34]**	- 0.57% [8]	6.66% [10]	13.85% [3]	4.34% [4]	25.76% [9]
2022-Q2	**15.60% [35]**	- 8.22% [8]	17.35% [11]	22.93% [3]	6.22% [4]	36.38% [9]
2022-Q3	**11.38% [35]**	2.54% [8]	10.82% [11]	27.56% [3]	1.70% [4]	18.84% [9]
2022-Q4	**13.04% [35]**	4.60% [8]	11.51% [11]	6.68% [3]	14.01% [4]	24.10% [9]
2023-Q1	**14.52% [35]**	4.65% [8]	13.46% [11]	17.59% [3]	17.90% [4]	22.05% [9]

The sample expands with time. The number of firms in each sector in each quarter is denoted in the [brackets] behind each mean value. N/A denotes no observations. "Other" firms contain a variety of businesses: yarn, bicycles, jewellery, hearing aids, cameras, kitchen supplies, and general marketplaces.

percent). From 2020Q2 to 2023Q1, this trend corresponds to the difference between a three-fold and nine fold increase in quarterly revenue.

Electronics ecommerce are the only with a negative average profitability (-2.8 percent), but their 19.5 percent quarterly growth is twice as high as that for the firms in general (9.9 percent). Two of the four firms active since the first year of the sample are in the electronics sector. These two firms struggled tremendously in the corona-quarters of 2020, recording -71.2, -27.2, and -37.5 percent growth rates in the second, third, and fourth quarters, along with profit rates of -29.1, 1.7, and 42.3 percent, respectively. This corresponds well with the supply shortage from China in the isolation phase of the early pandemic. The pet sector has been stable, but I do not have data from before the third quarter of 2021. From the third quarter of 2021 to the first quarter of 2023, the three businesses representing the sector has consistent rates of profitability between 10 and 30 percent, with a mean growth rate of 6.9 percent. These businesses value profits over growth, sustaining rather than expanding the business, and retaining earnings. This form of business strategy is usually characterised by many small and scheduled investments, which are better seen as re-stockings and maintaining market presence than actual investments. The same is true for the average "other" sector businesses, but they are much more successful, earning an average profit of 35 percent. This is twice as large as the average (14.8 percent). These firms specialise, selling yarn, bicycles, jewellery, hearing aids, cameras, and kitchen supplies. The four businesses in the "design" sector sell paintings, art, and frames, are the next-to-least profitable with average growth.

The differences between trends, cycle, and strategies between firms in different sector may bias the estimates of a model, especially if the goal is to understand the differences between sectors, years, or strategies. Economic factors, such as pandemic, war, inflation, and the exchange-and interests rate, may be included in the VEC model, but since every firm in the sample is in the same economic sector, and I aim to generalise the model, I do not include any exogenous variables. There exists a possibility that the strategies of firms with regard to growth-versus profit maximisation can bias the comparison between groups, but I do not investigate this matter in much depth. Firms may have different strategies, maximising profit, growth, or both, but nonetheless have the same level of *general* profitability or return on investment. In the latter parts of this thesis, to avoid the potential bias this can ensue through unknown statistical correlations, I differentiate the firms by their sum of quarterly profit and growth, though I only use this a rough measure to construct a test for the VEC model's cointegration coefficient β.

3.2 The Equilibrium

3.2.1 Components

Each day, the difference between the stream of revenue and the stream of costs will be the profits of the business, as defined by equation 1. In a regression with two variables, a bivariate regression, there is only one parameter, the cointegration coefficient θ

$$R_{it} = \hat{\theta} C_{it} + \hat{\epsilon}_{it} \tag{5}$$

where R is revenue, C is costs, $\hat{\theta}$ is the *estimated* correlation coefficient between revenue and cost, and $\hat{\epsilon}_{it}$ is the residual. A regression will always minimise the error term ϵ, resulting in the estimated error term, the residual $\hat{\epsilon}_{it}$. If the correlation coefficient $\hat{\theta}$ is equal to one, equation 5 reduces to $R_{it} = C_{it} + \hat{\epsilon}_{it}$, meaning that the expression can be restructured to

$$\hat{\epsilon}_{it} = R_{it} - C_{it} = P_{it} \tag{6}$$

where P is the profit from equation 1, which is equal to the residual in equation 5 if $\theta = 1$. If the estimate $\hat{\theta}$ deviates from 1, equalling say 1.2, this "regression premium" must be taken into account when interpreting the results. I assume, in this master thesis, that equality in equation 6 holds at all times, and I show that this is a fair assumption when the firm has an equilibrium between revenue and cost.

The mathematical possibilities opened by functions, such as parameters, being equal to unity are vast, and fundamental in time series statistics (Brooks, 2019). Variations of Euler's equation is used to describe the growth of business (Gibrat, 1931), the cost of debt (Obstfeld and Rogoff, 1996), and the efficient capital market hypothesis (Fama, 1970). The interpretation of a unity function of logarithmic variables, is always related to the equality of relative variables. The logarithmic formulation of equation 5 is:

$$Log(R_{it}) = \hat{\theta} Log(C_{it}) + \hat{\epsilon}_{it} \tag{7}$$

The unity of $\hat{\theta}$ in the logarithmic regression produces a residual $\hat{\epsilon}$ which equal to the definition of profits in equation 4, which is approximately equal to the rate of profit. The actual Euler equation is the base of natural logarithms to the power of π times the imaginary matrix convergence constant i, is exactly equal to negative unity. That is: $e^{i\pi} + 1 = 0$.

The regression of a variable on its lagged value is an auto-regression, structured as

$$Log(R_{it}) = \delta + \alpha Log(R_{it-1}) + \epsilon_{it} \tag{8}$$

where R_{it} is the revenue today and R_{it-1} is the revenue yesterday. If $\hat{\alpha} = 1$ then the variable grows relatively to its size, the expected value of its revenue is whatever its revenue was yesterday, plus the constant growth $\hat{\delta}$ which equal zero if the average growth of the firm is zero. If α is above unity, revenue grows faster the larger it is, and vice versa if α is less than unity. If $\hat{\alpha} = 1$, the residual ϵ_{it} will be equal to the growth rate defined in equation 5. In economics, a variable with a unity parameter in an autoregression is termed an integrated variable, as having a unit root, or following a random walk. The is latter is an apt description of the process, since the variable will grow and shrink continuously, reverting to its mean only occasionally, if it has a traditional "mean" at all (Brooks, 2019).

The vector auto regression model (Sims, 1980) was the start of a leap forward in data science and econometrics. It encapsulates what the economic science begun by Frisch (1933), is all about: describing the relationship between variables. Frisch is the father of regressions, which "partial out" the covariation of multiple variables in estimating the effect of a variable of interest on a "dependent" variable. That is, by including socioeconomic background in an equation describing the effect of scholastic achievements on realised earnings, one can exclude the fact that socioeconomic background increases both achievement and wages (Angrist and Pischke, 2009). Frisch's insight solidified economics as a science, and the regression is constantly used in economics, but also in medicine, the social sciences, engineering, sports, and gambling.

The traditional simultaneous equation model (SEM) uses instrumental variables (IV) to untangle causality issues due to simultaneity to reach a final model which can (hopefully) unravel the actual relationships. The limitation of these models are the same as across all statistics, in that one can never test the truthfulness of estimates directly, only through arguments based on statistical tests. The vector auto regression (VAR) model combines concepts from SEM and IV to enable the treatment of all variables as endogenous:

$$R_{it} = \delta_R + \alpha R_{it-1} + \tau C_{it-1} + \epsilon_{Rit} \tag{9}$$

$$C_{it} = \delta_C + \alpha C_{it-1} + \tau R_{it-1} + \epsilon_{Cit} \tag{10}$$

where δ is the growth constant, α is the trend variable, τ is the cross-effect between revenue and cost, and t is days. The subscripts R and C identify parameters affecting revenue and cost, respectively. All parameters come in pairs, one for revenue and one for costs. The number of parameters increases with the square of the number of parameters, and geometrically with the number of lags, though I never employ more than one lag. In the latter parts of the study, I formulate a simplified VEC model, with no lags, meaning that the parameters in equation 9 and 10 are excluded from the model. The vector framework is a cornerstone to the estimation of the return on investment, and though the VAR parameters are later excluded, the vector framework is always a part of the VEC model.

Regressions estimate parameters by finding the parameter that minimises the error term of the model. Integrated variables have an expected value that is equal to their present value, corresponding to $\alpha = 1$ in equation 8 and 9. My hypothesis is that both revenues and costs are integrated, but not only integrated, as I must assume that revenue and cost cointegrated. This is the case if there exists a long run relationship between revenue and cost, defined as the existence of a parameter that generates a stationary (non-integrated) residual from a regression of revenue and costs, where both where integrated to begin with (Engle and Granger, 1987). If this parameter θ is at unity, the residual is equal to the rate of profit, but this is neither necessary nor sufficient for cointegration. A regression of cointegrated variables is super-efficient due to the drastic reduction in error achieved by estimating the cointegrating parameter, defined as generating a stationary residual out of two or more non-stationary (integrated) variables. This residual moves around its mean closely and frequently, per its stationarity.

The error, or equilibrium, correction model and includes the residual from equation 5 into the same revenue-cost regression, but in growth terms:

$$\Delta R_{it} = \delta + \tau \Delta C_{it} + \gamma \hat{P}_{it} + \mu_{it} \tag{11}$$

where Δ stands for "change in" revenue and costs, and μ is the residual, but not the residual from before. The residual that equals the profit is $\hat{P}$, defined as $\hat{P} = R_t - \theta C_t$ from the regression in equation 5, where $\hat{P} = P$ if $\hat{\theta} = 1$. Here, the parameter γ describes the effect of profits on the change in revenue. I interpret γ as the time it takes for profits to generate revenue through reinvestment. The parameter estimate $\hat{\tau}$ may capture the same effect. It is interpreted as the effect of changes in costs on changes in revenue, but τ and γ act on different time horizons, τ being the short-run effect and γ being the medium term adjustment or convergence of revenue to the profit equilibrium, which is equal to zero if the firm has reinvested all profits.

The simplicity of the equilibrium correction model makes it particularly attractive. Its first step is the correlation coefficient, estimated as:

$$\theta = \frac{\sum[(R_t - \bar{R})(C_t - \bar{C})]}{\sum(C_t - \bar{C})^2} \tag{12}$$

where R is revenue, C is costs, and the bar in $\bar{x}$ is the average value of the variables. The regression equation is interpreted as the covariance between revenue and costs, divided by the variance of costs. The parameter θ replicates a residual exactly equal to the equation 1 definition of profit if θ is equal to one, but this is not necessary nor sufficient for the error correction (EC) regression:

$$\Delta R_t = \hat{\delta} + \frac{\sum[(\Delta R_t - \bar{\Delta R})(\Delta C_t - \bar{\Delta C})]}{\sum(\Delta C_t - \bar{\Delta C})^2} \times C_t + \frac{\sum[(\Delta R_t - \bar{\Delta R})((\hat{P}_t - \bar{\hat{P}})]}{\sum(\hat{P}_t - \bar{\hat{P}})^2} \times \hat{P}_t + \mu_t \tag{13}$$

where $\hat{P}_t = R_t - \theta C_t$ from the regression in equation 5, and the first term corresponds to τ and the second term corresponds to γ. This is a simplified version of the error correction model, that does *not* incorporate Frisch's formula for regression coefficients. The correlation coefficients in equation 12 are not used for any of the analysis, except for in the first step of the VEC regression described in equation 5, and equation 13 serves only for presentation.

In a multivariate regression, the covariance (numerator) between explanatory variables is controlled for. The intuition is best understood through sector diagrams, imagining a sector for the endogenous variable (R) and a sector each for the explanatory variables (C and P). The intersection between R and C corresponds to the parameter τ, while the intersection between R and P corresponds to the parameter γ. The intersection between R, C, and P, is "partialled out", meaning that effects on R that come from C and P together are removed from the parameter estimate. This is to ensure that the estimated effect of C on R and P on R are as accurate as possible. The common effect that is partialled out, may be included in the model through an interaction term, where C and P are multiplied, but this is not relevant for my model.

The cointegration coefficient β described in the VEC model as the cumulative effect of one variable on the other, retains the concept of the regression coefficient as a partial effect. That is, when the return on multiple types of investments are estimated simultaneously, each return on investment is interpreted as a partial effect, meaning: the common effect of investments, expenses, marketing and inventory on revenue is removed when estimating each effect on revenue.

3.2.2 Vector Equilibrium Correction

To combine the equilibrium correction and vector auto-regression frameworks, one takes the residual from equation 5, and includes it as an exogenous variable in a differenced VAR model. This corresponds to including the profit in a simultaneous equation of the growth in revenues and costs. This combination allows us to disentangle the effects of the profit on revenue and on cost, while estimating the trends and short-run effects in revenue and cost:

$$\Delta R_t = \delta_R + \alpha_R \Delta R_{t-1} + \tau_R \Delta C_{t-1} + \gamma_R \hat{P}_{t-1} + \epsilon_t \tag{14}$$

$$\Delta C_t = \delta_C + \alpha_C \Delta C_{t-1} + \tau_C \Delta R_{t-1} + \gamma_C \hat{P}_{t-1} + \epsilon_t \tag{15}$$

where Δ is growth, R is revenue, C is cost, and $\hat{P}$ is the estimated residual from regression 5. The parameters all come in pairs, starting with δ_R and δ_C that are the average growth in R and C, α_R and α_C are the trends in the growth of R and C, while τ_R and τ_C describe the short-run "shock" effects of C on R and R on C, γ_R and γ_C describes effect of P on the growth in R and costs, and both R and C have an error term ϵ_R and ϵ_C.

Profits affect the revenue through investment. When a business invests, its profit decreases with the the cost of the investment, but when the investment begins to generate revenue the profit increases. If the return on investment is positive, the initial drop in profits results in an increased profit. Whether the investment was profitable or not, however, the initial investment will cause an immediate drop in profits, which is followed by a gradual generation of revenue. In the equilibrium between revenue and cost, with an $\theta = 1$, there is no residual and therefore no profit. The parameter γ_R in equation 14 estimates how long it takes for revenues to converge toward this equilibrium, which I interpret as the time it takes for the investment to realise fully.

The cash-flow sensitivity of investments is interpreted through γ_C, which estimates the time it takes for profits to be spent. This is the essence of the literature following Fazzari, Hubbard, and Petersen (1988), who argue that if a firm spends internal funds as it generates revenue, then that firm must be financially constrained. If the firm was not financially constrained, it would not matter when it generated revenue, since the financially *un*constrained firm has fulfilled all its profitable investment opportunities. The literature contending this view, following Kaplan and Zingales (1997), argues that revenue also signal increased return on investment, meaning that the firm will respond by investing the generated funds, thereby making the cash-flow view invalid due to omitted variable bias. The argument then boils down to whether an adequate controlling variable can be constructed for investment opportunities, so that financially constrained firms can be identified through the cash-flow sensitivity of investments without bias.

The vector equilibrium correction model is powerful in its ease of interpretation. In my application of the VEC model, there is only one relationship that is estimated, and that is the effect of investments on revenue. But, investments effect revenues through multiple channels, of which I identify three. This is made possible by the equilibrium between revenues and costs, which allows the estimation of the short-run VAR parameters, the medium-run realisation time of revenue to investments, and the long run cumulative effect of investments on revenue.

3.2.3 The Long Run Return

The cointegrating coefficient β of the VEC model is interpreted as the cumulative effect of one variable on the other. It is defined through the converge parameters γ, but to identify the parameter, we must restrict the matrix space into a regression framework. The properties of the cointegrating coefficients, and the VEC model in general, contains complex matrix calculation described by Søren Johansen in a number of articles. I draw mostly on his seminal 1988 article (Johansen, 1988), his 1991 article in Econometrica (Johansen, 1991), his 1997 working paper (Engsted and Johansen, 1997), his 2000 article in Economic Modelling (Johansen, 2000), and his cointegration overview at the University of Copenhagen in 2004 (Johansen, 2004). The conclusion from his extensive research and matrix calculation is that his β is interpreted as a regression parameter like Frisch's regression parameter, but as cumulative long run partial effects. This means that we have to introduce restrictions to the matrix space, per Johansen (1998;1991;1997;2000;2004), so that we have one endogenous variable, while the other variables are exogenous. In my model, it is always revenue (R) that will be the endogenous variable, while the cost variables (C=I+E=II+MI+E) will always be exogenous. The cointegration coefficients are represented as a sub-set of the VEC model, similar in structure as a traditional regression:

$$R_{it} = \beta_C C_{it} + \varepsilon_{it} \tag{16}$$

where β_C is the long run return of costs on revenue, and ε is an error term. The parameter β_C is the cumulative effect of one unit increase in costs on revenue. If $\beta = 1$, the investment was break-even, generating as much revenue as the cost incurred. For an investment to be worth investing in, the return on investment must be positive, meaning that the cumulative return of revenue must be higher than the cost of the investment, corresponding to $\beta > 1$. If the parameter is estimated to be, say $\hat{\beta} = 1.2718$, then it means that an investment of 1 unit will return 0.2718 excess revenue, equivalent to a 27.18 percent return on investment. In the bivariate model, costs are not distinguished into investments and expenses. The general return on costs in the bivariate model gives the most easily interpretable estimate for the general profitability of the firm.

The maximum likelihood estimator restricts the cointegrating space, $\prod = \gamma\beta'$, to identify the cointegration coefficient β (Johansen, 1991). With no restrictions, the model will be over-parameterisized, meaning that neither the adjustment parameter γ nor the cointegration parameter β will be identifiable. Interpreting the adjustment of revenue to the profit-equilibrium as the incremental return on investment (Johansen, 2004), the long run return on investment may be defined as in **cointegration coefficient**

$$\beta_i = \sum_{t=1}^{\infty} (\gamma_{it}) \tag{17}$$

Meaning: β is the sum of all incremental adjustments to equilibrium. The return on and investment i made a time t is equal to the response of revenue R to being below the equilibrium as a result of an increase in costs C due to investment. This is a simplified expression for the cointegration coefficient β in the comprehensive maximum likelihood vector matrix mathematics (Johansen, 1988; 1991; 1997; 2001; 2004), but I believe that I have encapsulated the spirit of β's *meaning* with the expression in equation 17. That is, β is the long run cumulative partial effect of one exogenous cointegrated variable (C) on the endogenous cointegrated variable (R).

3.3 Extensions

3.3.1 Multivariate Model

I estimate the VEC model in a number of different specifications, but in general, I apply the model in three ways. First, I compare the return on investment to the return on expenses, where Investments generate revenue while expenses facilitate for revenue, to verify that $\hat{\beta}$ can differentiate between the high return on investment and the low return on expenses. Second, I pre-classify firms by their *rough* return on investment, as defined by the sum of profit and growth rates, to confirm that β can differentiate between firms with high and low sums of profit and growth. The trivariate model is needed to make the first validation, and it begins with the same equilibrium regression as the bivariate model, but with two exogenous variables:

$$R_t = \theta_I I_t + \theta_E E_t + \epsilon_t \tag{18}$$

where I stands for investments, E stands for expenses. In this first step regression, $\hat{\theta}_I$ and $\hat{\theta}_E$ are contemporaneous regression coefficients, describing partial effects (Frisch, 1933). The return on investment β is estimated through the sub-matrix of the vector equation system in the first of the three equations (19-21 below) in the vector model. The return on both investments and expenses is defined with the parameter γ_R. Since the two β parameters are defined in the same matrix space with only one γ parameter in $\prod = \gamma_R \beta_k'$, the β parameters must be restricted to gain the partial effect interpretation and avoid overparametrisation (Johansen, 1991).

In this application of the VEC model, the residual ϵ is still interpreted as profit as I assume $\theta = 1$, and the profit enters in level form in the growth model of revenue, investments, and expenses:

$$\Delta R_t = \delta_R + \alpha \Delta R_{t-1} + \tau_{IR} \Delta I_{t-1} + \tau_{ER} \Delta E_{t-1} + \gamma_R \hat{P}_{t-1} + \mu_t \tag{19}$$

$$\Delta I_t = \delta_I + \alpha \Delta I_{t-1} + \tau_{RI} \Delta R_{t-1} + \tau_{EI} \Delta E_{t-1} + \gamma_I \hat{P}_{t-1} + \mu_t \tag{20}$$

$$\Delta E_t = \delta_E + \alpha \Delta E_{t-1} + \tau_{RE} \Delta R_{t-1} + \tau_{IE} \Delta I_{t-1} + \gamma_E \hat{P}_{t-1} + \mu_t \tag{21}$$

where $\hat{P}_t$ is the residual ($\hat{\epsilon_t}$), and the parameter γ_R is of primary interest. By restricting the matrix space, by defining revenue (R) as the *only* endogenous variable, contrary to the vector framework which treats *all* variables as endogenous. The cumulative effect of revenue on itself is always equal to itself, equal to 1. Only the endogenous variables can be restricted, meaning that there is no cointegration coefficient describing the effect of profits, as *profits are created* by investments through revenue which generates profits. Therefore, γ does not appear in equation (22) below, only the two cointegration coefficients β_I and β_E are described, in the **long run regression equation**:

$$R_t = \beta_I I_t + \beta_E E_t + \varepsilon_t \tag{22}$$

where β_I is the long run return on investment, β_E is the long run return on expenses. The interpretation of θ and β must not be confused, as θ is the regression coefficient (Frisch, 1933) estimated super-efficiently (Engle and Granger, 1987), while β is the *cointegration* coefficient (Johansen, 1988). The cumulative effect estimated by $\hat{\beta}$ has a regression interpretation (Frisch, 1933), and is based on a combination of the vector (Sims, 1980) and cointegration (Engle and

Granger frameworks, 1987), estimated in a restriction of the VEC model (Johansen, 1988).

The return on investment β_I in equation (22) is perhaps the most discussed concept in this thesis, as I hypothesise that it *must* be larger than β_E, and that it (β_I) *must* be larger for firms that have more excess revenue after costs. These are two of my three main hypotheses, the last main hypothesis extending the model by splitting investments into inventory and marketing, which constitute the primary investment types in ecommerce.

The parameters γ_I and γ_E are of secondary interest in this study, interpreted as the cash-flow sensitivity of investments and expenses. If investments are cash-flow sensitive, it may suggest that firm has profitable investment opportunities but not the funds to carry them out (Fazzari, Hubbard, and Petersen, 1987). However, if expenses are cash-flow sensitive, I argue that it may suggest that the firm is financially constrained in the sense that it is struggling to make ends meet, spending revenue on covering expenses. There could be other reasons for the finding significant estimates for γ_I, γ_E, and γ_R, either due to other dynamics within business or due to statistical biases in the model and estimation.

The method's underpinnings in the equilibrium between revenue and costs does not change when splitting costs into investments and expenses, nor when splitting the two further. The residual will continue to define profits as it did before, as long as all parts of the total costs are included, and the total parameter θ is in fact equal to 1, or at least approximately so. I continue by splitting investments into inventory and marketing in the quadvariate VEC model, beginning with the first step traditional regression:

$$\hat{P}_t = \hat{\epsilon}_T = R_t - (\theta_{II} II_t + \theta_{MI} MI_t + \theta_E E_t) \tag{23}$$

where $\hat{P}_t$ is the residual from the first step regression, which is included in as an exogenous variable in the four differenced (Δ)simultaneous VAR equations:

$$\Delta R_t = \delta_R + \alpha_R \Delta R_{t-1} + \tau_{IIR} \Delta II_{t-1} + \tau_{MIR} \Delta MI_{t-1} + \tau_{ER} \Delta E_{t-1} + \gamma_R \hat{P}_t + \mu_t \tag{24}$$

$$\Delta II_t = \delta_{II} + \alpha_{II} \Delta II_{t-1} + \tau_{RII} \Delta R_{t-1} + \tau_{MIII} \Delta MI_{t-1} + \tau_{EII} \Delta E_{t-1} + \gamma_{II} \hat{P}_t + \mu_t \tag{25}$$

$$\Delta MI_t = \delta_{MI} + \alpha_{MI} \Delta MI_{t-1} + \tau_{RMI} \Delta R_{t-1} + \tau_{IIMI} \Delta II_{t-1} + \tau_{EMI} \Delta E_{t-1} + \gamma_{MI} \hat{P}_t + \mu_t \tag{26}$$

$$\Delta E_t = \delta_E + \alpha_E \Delta E_{t-1} + \tau_{RE} \Delta R_{t-1} + \tau_{IIE} \Delta II_{t-1} + \tau_{MIE} \Delta MI_{t-1} + \gamma_E \hat{P}_t + \mu_t \tag{27}$$

which totals four growth constants δ_k, four autoregressive parameters α_k, twelve short-run cross-variable effects τ, four adjustment terms γ, and three cointegration coefficients. The latter, β_k, where $k = II, MI, E$, is estimated in the restricted matrix $\prod = \gamma_R \beta_k'$ and interpreted as the long run regression :

$$R_t = \beta_{II} II_t + \beta_{MI} MI_t + \beta_E E_t + \varepsilon_t \tag{28}$$

where β_{II}, β_{MI}, and β_E are the cointegration coefficient describing the long run cumulative partial effect of of inventory and marketing investments, and expenses, respectively. The estimates of the parameters β_k are interpreted as: the return on spending one dollar on input factor or variable k.

3.3.2 Multiple Equilibria and a simplification of the VEC model

When there are more than two variables in a VEC model, there may be multiple cointegrating relationships. In the trivariate model, this corresponds to separate equilibria between revenue and investments and between revenue and expenses, while the quadvariate model could contain three cointegrating equilibria. The argument for is stronger for multiple cointegration between revenue and inventory investments and revenue and marketing investments, as expenses are partially fixed.

The multiple equilibria VEC model is achieved through two first step regressions, one for revenue and inventory, and one for revenue and marketing:

$$R_t = \theta_{II} II_t + \epsilon_{IIt} \tag{29}$$

$$R_t = \theta_{MI} MI_t + \epsilon_{MIt} \tag{30}$$

where neither ϵ_{IIt} nor ϵ_{MIt} can no longer be interpreted as the general profit, but rather the "isolated" profit on the respective investments, if $\theta_{II} = \theta_{MI} = 1$. This special case giving ϵ_{IIt} and ϵ_{MIt} the interpretation if isolated profits is neither necessary nor sufficient for the VEC model, where both residuals are now included as equilibrium terms in the VAR model:

$$\Delta R_t = \delta + \alpha \Delta R_{t-1} + \tau \Delta II_{t-1} + \tau \Delta MI_{t-1} + \tau \Delta E_{t-1} + \gamma \hat{P}_{IIt} + \gamma \hat{P}_{MIt} + \mu_t \tag{31}$$

$$\Delta II_t = \delta + \alpha \Delta II_{t-1} + \tau \Delta R_{t-1} + \tau \Delta MI_{t-1} + \tau \Delta E_{t-1} + \gamma \hat{P}_{IIt} + \gamma \hat{P}_{MIt} + \mu_t \tag{32}$$

$$\Delta MI_t = \delta + \alpha \Delta MI_{t-1} + \tau \Delta R_{t-1} + \tau \Delta II_{t-1} + \tau \Delta E_{t-1} + \gamma \hat{P}_{IIt} + \gamma \hat{P}_{MIt} + \mu_t \tag{33}$$

$$\Delta E_t = \delta + \alpha \Delta E_{t-1} + \tau \Delta R_{t-1} + \tau \Delta II_{t-1} + \tau \Delta MI_{t-1} + \gamma \hat{P}_{IIt} + \gamma \hat{P}_{MIt} + \mu_t \tag{34}$$

where I have dropped the subscripts on the parameters for simplicity, while the equilibrium terms are from equation 29 ($\hat{P}_{IIt} = \hat{\epsilon}_{IIt}$) and equation 30 ($\hat{P}_{MIt} = \hat{\epsilon}_{MIt}$). Since both are included, we can now differentiate between the time to realisation of investments of both kinds to revenue through γ_{RI} and γ_{RM} parameters describing the convergence speed of revenue to its two equilibria with inventory-and marketing investments, separately. From the multiple cointegrating equilibria VEC model, the marginal rate of return over costs (Fisher, 1930) can be estimated, where the realisation time of investments must be discounted with the return the capital could have earned as savings instead of investments. Unfortunately, this model loses the interpretation of the residual as profits as the "partial" interpretation of the regression is lost. The variables are treated as exogenous to each other, meaning that the cointegration coefficient can be interpreted as being form separate long run regressions:

$$R_t = \beta_{II} II_t + \varepsilon_t \tag{35}$$

$$R_t = \beta_{MI} MI_t + \varepsilon_t \tag{36}$$

where the estimates $\hat{\beta}_k$ can be combined with the $\hat{\gamma}_k$ parameters from equation 31 to form a Fisher style marginal rate of profit with interest rate discounts, as both the full return on investment β_k and the realisation time of investment γ_k are estimated.

To estimate the model with another exclusive equilibrium, between revenue and expenses, one must estimate another first step regression, producing another residual to add to the model, the cumulative effect of which is represented as a third long run regression.

Instead of introducing complexity to the model, I seek simplicity, so that the model's parameters can be clearly interpreted and intuitively understood. Furthermore, overparametrisation issues arise in multivariate models. With four variables, there are four sets of δ, α, and γ parameters, along with twelve short-run τ parameters, and three long run parameters β, but is only the β and γ parameters that are needed to estimate the return on investments. Therefore, I propose a reduced version of the VEC model, where I restrict the VAR parameters to zero:

$$\Delta R_t = \delta_R + \gamma_R \hat{P}_t + \mu_t \tag{37}$$

$$\Delta II_t = \delta_{II} + \gamma_{II} \hat{P}_t + \mu_t \tag{38}$$

$$\Delta IM_t = \delta_{IM} + \gamma_{IM} \hat{P}_t + \mu_t \tag{39}$$

$$\Delta E_t = \delta_E + \gamma_E \hat{P}_t + \mu_t \tag{40}$$

where the equilibrium term $\hat{P}_t = (R_t - (\hat{\theta}_{II} II_t + \hat{\theta}_{IM} IM_t + \hat{\theta}_E E_t))_t$ as per equation 23. The long run regression representation of the cointegrating coefficients remains as is in the single equilibrium multivariate model in equation 28 ($R_t = \beta_{II} II_t + \beta_{MI} MI_t + \beta_E E_t + \varepsilon_t$).

The exclusion of short run effects from the model, focusing in on the convergence-and long run return parameters, reduces the number of parameters from 7 to 3 in the bivariate model, from 14 to 5 in the trivariate model, and from 23 to 7 in the model with four variables. This model allows for the inclusion of many variables, but it is applicable for thinner time series data too. In fact, since the cointegrating relationship is estimated super-efficiently, the model can be applied to the yearly data frequency of financial statements.

3.3.3 Investment Choice Ranks

Fisher (1930) compares investment opportunities, ranking them according to their difference in revenue and cost

$$(R_1 - C_1) - (R_2 - C_2) = (R_1 - R_2) - (C1 - C_2) \tag{41}$$

where R_i is the revenue and C_i is the cost, and $i = 1, 2$ denotes the investment alternative. The latter expression is used by Fisher to interpret the alternatives as the surplus of advantages over disadvantages between investment 1 and 2 (Alchian, 1955). The cost (C) of an investment (I) is observable, but the revenue generated by an investment must be estimated. I use the cointegrating coefficient β from the VEC model to estimate the revenue generated by investment, where costs are normalised to one:

$$(\hat{\beta}_I C_I - C_I) - (\hat{\beta}_E C_E - C_E) = (\hat{\beta}_I - 1) - (\hat{\beta}_E - 1) = \hat{\beta}^*_I - \hat{\beta}^*_E \tag{42}$$

where $\hat{\beta}^*_k$ is the expected revenue generated by investment minus the cost of investment, which I define as the return on investment. This simplification of Frisch' investment choice model can be applied to rank investments within firms, such as in equation 42, or between firms.

The return a firm can make on any investment must be evaluated against the interest rate. The longer an investment takes to realise into revenue, the larger a discount must be made on the return on that investment, since the capital earns the more interest the longer it is saved. Fisher's marginal rate of return over costs is the interest rate that equates the returns on different types investment. It identifies the discount rate that would equate the returns on the faster realisation investment and lower return on investment with the return of the slower-but higher return on investment. The VEC model estimates the time it takes for an investment to realise through the parameter γ_R, meaning that Fisher's marginal rate of return can be estimated within the VEC framework, but this requires multiple equilibria. I do not estimate the different time to realisation within firms, but I do between firms. Extending the framework I propose allows for the estimation of Fisher's marginal rate of return over cost, which would prove a much more detailed ranking of investment opportunities. Another possible extension, is the introduction of investment risk measures. Regressions always estimate a standard error $\hat{\sigma}$ for every parameter. The standard error is a measure of the estimate's certainty, confidence, or accuracy. Whereas the correlation-, regression-, and cointegration coefficients describes the mean covariance between variables, their standard errors describes the standard distance from the predicted value to its actual value. High standard errors can indicate that the return on an investment is uncertain, varying widely within and/or between firms. By incorporating the standard error on realisation time too, one may construct an uncertainty-corrected ranking of investments between and within firms, by their return.

4 Method

In this chapter, I apply Johansen's (1988) VEC model, that combines the EC (Engle and Granger, 1987) and VAR (Sims, 1980) frameworks, to the Fisher (1930) investment rank framework. I interpret Johansen's cointegration coefficient β_I in a model of with revenue, investments, and expenses, as the cumulative partial effect permanent changes of investment on revenue. It (β_I), minus the cost of investment that is normalised to one equals the long run return on investment $\beta^* = \beta - 1$, though I may use β and β^* interchangeably at times. I hypothesise that the return on investment must be higher than the return on expenses, which is estimated in the same fashion. I split investments into inventory and marketing to investigate their investmentment rank, and in a final step, I estimate the return on investment, split into inventory and marketing, for high-and low profit-and growth firms, in order to identify *why* some firms are successful while others are not.

4.1 Hypotheses

4.1.1 Verification Hypothesis

The objective of this study is not to investigate every possible application and extension of the model, but rather to introduce the application of the VEC model to business bank account data. Specifically, I aim to show that the return on investment can and is estimated by the cointegration coefficient β (Johansen, 1988), and that the VEC model can be used to rank investment opportunities both within and between firms. The theoretical underpinnings have been discussed thus far, but in order to verify the methodological approach and consequent interpretation, I construct several statistical tests based on the cointegration coefficients β.

As a first step in validating my approach, I use the reasoning that investments generate a higher return than expenses to construct a testable hypothesis of the model's validity. Specifically, I test whether the return on investment is statistically larger rather than the return on expenses, to confirm the same pre-conceived notion. Although expenses are necessary to facilitate for revenue, it is investments that actually generate revenue. Therefore, the revenue generated by expenses must be close or equal to zero ($\beta_E = 0$). If the revenue generated by investment is equal to unity ($\beta_I = 1$), then the return on investment is zero ($\beta^*_I = 0$). The earnings generated by investments, however, must cover for expenses too, meaning that a business with no return on investment will be losing money. Having a return on investment higher than zero ($\beta^*_I > 0$) is thus necessary, but not sufficient to be a profitable business. Therefore, I pose the dual null hypothesis that the revenue generated by investment is equal to the cost of the investment, while expenses do not generate any revenue at all:

$$H_{01I} : \beta_I = 1 \tag{43}$$

$$H_{01E} : \beta_E = 0 \tag{44}$$

where $\beta_I = 1$ corresponds to a zero return on investment, $\beta_I^* = 0$, and $\beta_E = 0$ equates to $\beta_E^* = -1$. Implicit in the null hypothesis is that the return on investment is higher than that for expenses. The non-rejection of both hypotheses is a step in validating the approach to estimating long run returns in investment I propose, as it confirms the theoretical argument that it is investments, not expenses, that generates revenue. For a business to be profitable, its return on investment must cover for the expenses that keep the business running. I do not pose any specific alternative hypothesis for expenses, but as long as $\beta_E < \beta_I$ the return on investment is higher than that for expenses. I extend the hypothesis to inventory-and marketing investments, with the same hypotheses and expectations as for investments generally, presented as

$$H_{01II} : \beta_I = \beta_{II} = \beta_{MI} = 1 \tag{45}$$

$$H_{A1II} : \beta_I = \beta_{II} = \beta_{MI} > 1 \tag{46}$$

where $\beta_I > 1$ indicates that investments generate earnings by returning more revenue than the initial cost of the investment, and II stands for inventory investments and MI stands for marketing investments. I am ambivalent to which of the investments have the highest return. Here, I will rely on the empirical evidence to make an assessment, letting the data decide whether $\beta_{II} > \beta_{MI}$ or $\beta_{II} < \beta_{MI}$ in the framework of ranking investment choices (Fisher, 1930).

The equilibrium term in the VEC model yields parameters interpreted as the convergence to the equilibrium. In the financial development literature, this parameter γ has been interpreted as the long run effect of finance on growth and vice versa, using longer horizon yearly financial statement data (Xu, 2000; Christopoulos and Tsionas, 2014). In the much higher frequency data I use, I interpret the converge parameter to estimate the time to realisation of investment into revenue γ_R and the cash-flow sensitivity of investments to revenue γ_C. The equilibrium (profit) is positive by construction when revenue is higher than cost, while it is negative when costs are higher than revenue. This is not an assumption or a hypothesis, the residual from equation 5 is, per construction positive when revenue is higher than cost if θ is positive. Confirmatory parameter estimates correspond to γ_R being negative and all γ_Cs to being positive. If there is no convergence, both parameters are statistically equal to zero:

$$H_{02} : \gamma_R = \gamma_C = 0 \tag{47}$$

where C stands for costs, but the same interpretation is true for investments (into marketing and investment) and expenses. Convergence takes place following a shock that moves profits from its mean value. Investments are typical negative shocks, while external funds, internal equity, and large sales are typical positive shocks. A shock in costs due to investments causes a negative profit, but as the investment realises the revenue increases. That is, if the revenue is bound by the equilibrium, in which case a negative (positive) shock in the equilibrium causes a positive (negative) convergence in revenue toward equilibrium. Therefore, the convergence of revenue may be interpreted as an estimate for the duration of return on investment.

The convergence of costs (investments) to the equilibrium tells us how fast costs converge toward the mean profit. Firms with profitable investment opportunities invest, but not all firms have sufficient funds to carry out their profitable investment opportunities. These firms will invest as soon as they have available funds, and are known cash-flow sensitive (Fazzari, Hubbard, Petersen, 1987; 2000; Kaplan and Zingales, 1997; Gilchrist and Himmelberg, 1995; 1998). I argue that this is estimated by γ_C, and my interpretation of γ_R is that it measures the realisation time of investments, which requires that

$$H_{A2a} : \gamma_R < 0 \tag{48}$$

$$H_{A2b} : \gamma_C > 0 \tag{49}$$

corresponding to convergence towards the equilibrium, where C stands for costs, but also its sub-categories. Firms that are financially constrained have a higher return on investment than they are able to communicate to financial companies, but it is unclear whether financially constrained are more or less profitable than other firms. If $\gamma_R < 0$, revenues are expected to increase following a increase in costs, identified as the effect of (decreased) profits on revenue.

It may be argued, to put it simply, that business may be financially constrained for one of two reasons. In the traditional sense, businesses are financially constrained if they have profitable investment opportunities, but not enough funds to carry them out. These firms are identified by the pace at which they invest once they get funds, but I argue that there is another interpretation of financial constraints. That is, firms struggling to make ends meet, covering expenses when they have the money, marketing to sell of lagging inventory, and restocking to meet the existing demand. In order to distinguish firms that are creditworthy from those who are not, I estimate the return on investment directly, but I also identify which segment of the sample is more cash-flow sensitive.

4.1.2 Identification Hypothesis

To prove useful, the estimated parameter $\hat{\beta}$ must not only be sensible, but it must by able to identify businesses with high return on investment, by distinguishing them from the firms that are relatively less profitable. Firms may maximise growth or profits, but the return on investment among firms need not differ between firms choosing different strategies. Surplus revenue may either be reinvested or saved, but my main concern is with the return on investment, not necessarily with the strategy of firms. The expansion and depletion of inventory stocks distorts any direct observation of profitability in the data, as firms expanding and depleting their stocks will appear the same, though they are completely different. To verify the VEC model's precision in distinguishing between firms with high-and low return on investment, however, I define a measure to compare segments of the sample. The measure I construct is simple but indicative, defined as:

$$q_{it} = \Delta Log(R_{it}) + Log(P_{it}) \tag{50}$$

where q is the sum of quarterly growth and profit, $\Delta Log(R)$ is the percentage growth in revenue per quarter, and $Log(P)$ is profits defined as $Log(R) - Log(C)$. This is not Tobin's q (1969), but it does measure the same concept: the return on investment.

I use the separation of high profit-and growth firms and low profit-and growth firms to construct a test that can verify whether the cointegration coefficient β differentiates between firms with high and low return on investment. If the parameter β does to distinguish between, the two returns on investments would be statistically identical across the two groups, and my interpretation of β as the long run return on investment would be erroneous. The estimated return on investment must be higher for firms with high return on investment than for firms with low return on investment:

$$H_{03a} : \beta_{IH} = \beta_{IL} \tag{51}$$

$$H_{A3a} : \beta_{IH} > \beta_{IL} \tag{52}$$

where the subscript H denotes the firms with above average q and L denotes those with low q. The only logical conclusion of this hypothesis is that the null hypothesis is rejected in favour of the alternative.

Financial constraints occur when the business does not have enough funds to carry out the investments it wishes to carry out. This may be due to high or low return on investment alike, so I pose null and alternative hypothesis 3b without any preconceived notions:

$$H_{03b} : \gamma_{CH} = \gamma_{CL} = 0 \tag{53}$$

$$H_{A3b} : \gamma_{CH} \neq \gamma_{CL} > 0 \tag{54}$$

where $\gamma_{CH} > \gamma_{CL}$ suggests that high-return firms are financially constrained, supporting the idea that the parameter γ_C identifies firms that have profitable investment opportunities, but not the funds to carry them out. If $\gamma_{CH} < \gamma_{CL}$, it is the low-return firms that have been identified as (more) financially constrained. In this case, I would argue that the parameter identifies firms that are financially constrained since their investments are not profitable, the firm struggling to make ends meet, and having to spend its money as soon as it gets it.

Firms with higher return on investment may have faster return on investment too, although these firms may also invest larger amounts at a time to decrease the marginal cost, meaning that their realisation of investment can be longer. A hypothesis for the realisation time of investments can be made, and its result may be important for revenue-based financiers, but at this point, I have no theoretical or logical alternative hypothesis, and I do not extend the model to the multiple equilibria required to test such hypotheses.

4.1.3 Ranking Hypothesis

For the non-producing firm operating online, there are two main types of investment: inventory and marketing. Ecommerce businesses only market themselves online, having no physical store or location that customers can visit. Marketing drives revenue through the demand-channel, contrary to investments into inventory that drive revenue through supply. From experience, businesses within the same sector, especially if they follow the business-to-business (B2B) model, often source their inventory from the same suppliers. Therefore, their return on inventory investment will likely be relatively similar. By investing in inventory in bulk, the unit price decreases, especially if the supplier is a long geographical distance away, since the shipping prices per unit also decrease. Firms that follow the direct-to-customer (D2C) business model, often specialise in one or a few products, often electronic products, often from China. The inventory must be shipped from long distances, but it also costs the producer less to manufacture more at a time, meaning that cost per unit is expected to decrease with an increase in units ordered. Therefore, the return on inventory investment for firms that purchase in bulk may be expected to have higher returns on investment.

Marketing is increasingly important for a firm's value (Joshi and Hanssens, 2010), and as Hoffman and Novak already argued in 1997, the world wide web is a "new marketing paradigm for electronic commerce" (Hoffman and Novak, 1997). The largest marketing platforms in my sample of ecommerce firms is Google and Facebook. Marketing platforms, and shop-system platforms, continuously generate data on marketing and its relation to sales, constructing measures such as the "customer acquisition cost" and the "return on ads spent", which relate clicks on ads to the sale of the good to calculate a return, used in the credit assessment of my sample fintech lender, MXNEY. The use of data to gain insight to marketing performance, can allow for the improvement of marketing efforts (Troisi et al., 2020; Conway and Hemphill, 2019). Therefore, fintech companies such as MXNEY provide marketing expertise to its customers by looking at their marketing performance statistics.

An explanation for the perceived importance of marketing may be the role of good products, that are attractive to the customer base. To put it simply, selling a product that is more attractive to the customer is much easier than selling a product that is not. Therefore, less marketing efforts are needed to sell the product, meaning that the firms with appealing products will have higher returns on marketing, making them more profitable. That is, the return on marketing investment may indicate the appeal of the product more than the quality of the marketing effort itself. The inventory is bought at similar cost across the sector, but it is the product's appeal to the customer that matters.

Assuming that the alternative hypothesis in equation 52 is confirmed, investments in both inventory and marketing can be assumed to be higher for high Q firms than low Q firms:

$$H_{04a} : \beta_{IIH} > \beta_{IIL} \tag{55}$$

$$H_{04b} : \beta_{MIH} > \beta_{MIL} \tag{56}$$

Following the intuition from ecommerce that successful firms differentiate themselves through high return on marketing, I pose the alternative hypothesis:

$$H_{A4a} : \beta_{IIH} = \beta_{IIL} \tag{57}$$

$$H_{A4b} : \beta_{MIH} >> \beta_{MIL} \tag{58}$$

Meaning: the return on inventory investment is equal between high-and low return firms, while the return on marketing is much higher for high return firms. The null and alternative hypothesis in equations 43-45 and the alternative hypothesis equation 52, are based on logic reasoning, and a failure to confirm $\beta_I > \beta_E$ and $\beta_{IH} > \beta_{IL}$ is a detriment to the functionality of the cointegration coefficient as a measure for the return on investment. The hypotheses in 55-58 are based on industry knowledge, and it would make sense if the alternative hypothesis was confirmed, though its logical foundation is not as strong.

4.2 Specification, Selection, and Presentation

To reach the goal of estimating these sets of hypotheses, there are several considerations that must be evaluated first. In this section, I begin with a discussion of the moving average transformation that I use for the model, and of the numeric (monetary) formulation as opposed to the logarithmic formulation. Then, I conduct the unit root and cointegration tests, which serve as selection criteria for the VEC model. If the reader should at any time wonder why the sample is smaller in the results chapter than in this descriptive statistics, the answer lies within the selection criteria described in the concluding parts of this chapter. I close the methodological chapter with a description of the calculation of the parameter estimates presented in the tables, and a description of the significance test used for the hypotheses.

4.2.1 Moving Average Transformation

The VEC model is based on the concept of the cointegrated relationship. This is defined as the existence of parameter that produces a stationary residual from a regression of two integrated variables (Engle and Granger, 1987). Revenue is integrated in Gibrat's (1931) proposition, stating that if revenue is integrated, firms grow relative to their size. If a business grows relative to its size, it is likely that we have cointegration too, as the revenue and cost of a firm are in a long-run relationship. The entries in business bank accounts often described revenue every day, but investments are much less frequent. Variables that are zero at any point in time, cannot be said to be continuous, and will be treated as a missing observation in the logarithmic transformation. The cointegration tests and the VEC model does not allow missing observations, since a basic requirement of the concept is that the variables must be at least be continuous. Therefore, I employ a moving average transformation, on the form of:

$$X_t^S = (X_{t-1} + X_{t-2} + ... + X_{t-s})/S \tag{59}$$

where S is the length of the moving average, s is denotes the number of days previous to t, and t is days. Every variable is thus an average of its S past daily values. The most important short-run cycle in any business, after the daily open-closed cycle, is the weekly weekday-weekend cycle. By having a weekly moving average, the weekly cycle is completely controlled for. A week is not enough to ensure that there are no missing values, however, but when extending the moving average, I do it in weekly intervals, to ensure that the weekly cycle is completely controlled for. I use a two-, three-, and four-week moving averages:

$$X_t^{14} = (X_{t-1} + ... + X_{t-14})/14 \tag{60}$$

$$X_t^{21} = (X_{t-1} + ... + X_{t-21})/21 \tag{61}$$

$$X_t^{28} = (X_{t-1} + ... + X_{t-28})/28 \tag{62}$$

In a verification-sequence, I compare the estimates produced by the four-week moving average with that produced by monthly aggregates. The literature that does not use financial statements, often use quarterly data of entire sectors, such as that published by Statistics Norway (SSB). Therefore, I extend this a comparison with quarterly data, with a quarterly moving average

$$X_t^{91} = (X_{t-1} + ... + X_{t-91})/91 \tag{63}$$

In the comparison of the moving average and aggregate variables, the aggregate variables are defined as the total sum for the respective month or quarter. I have not seen the application of moving average variables in econometric analysis in the literature, but as I have noted, I have not seen any analysis using bank account data either. Since the methodology is untried in the literature, however, I undertake extensive robustness-testing throughout my analysis. I view the moving average transformation as a necessary step to fully exploit the high-frequency data, but the exact statistical properties of using a moving average transformation are unknown.

4.2.2 Interpretation of Values and VEC Steps

I experience that the monetary (numerical) variables give a clearer interpretation of the parameters, but that the logarithmic model is a better fit for my model, eliminating heteroskedasticity by viewing the process as evolving by the laws of growth. In order to assess which formulation is best, and to evaluate their advantages and disadvantages, I estimate both models (see table 3 below). I present the steps that must be taken to estimate a bivariate panel VEC model using daily open banking data in the statistical software Stata, based on the panel VAR model (Abrigo and Love, 2016). To enable precise estimates, a moving average transformation of the the variables (revenue and cost) is undertaken, and since the equilibrium does not contain a constant, the regression cannot contain a constant either. Then, I save the residuals and include the variable exogenously and lagged in a differenced panel VAR model.

The best practice in determining relationships within no-stationary variable, that grow continuously over time, is to perform a logarithmic transformation of the variables. In logarithms, the variables are relative in size. If a business is continuously growing, its variation in nominal values will also continuously increase. When the variance of a variable is correlated with the mean of that variable, the variable is heteroskedastic (has different variance). Heteroskedasticity does not change the beta-parameter in an estimation, as it is determined by the covariation from the variables' means. It is only the standard error, which measures the accuracy of the estimation that is larger with higher heteroskedasticity. The logarithmic transformation solves this heteroskedasticity, as the variance of the variable will grow relative the value of the variable. That is, if Gibrat's third proposition of relative heteroskedasticity holds. A common finding is that large firms have relatively less volatile revenue (Mansfield, 1962; Idsø, Opstad, and Valenta, 2021a), supporting the self-selection (Jovanovic, 1982) and learning-by-doing (Ericson and Pakes, 1993) hypotheses.

In model E and F, section 1, table 3 (below), the estimate for θ is found, equalling 0.943 in the monetary model (E) and 1.012 in the logarithmic model (F). Since the parameter θ is found super-efficiently by the regression of cointegrated variables, the corresponding standard errors are very small, less than 0.01 in both models. In model C and D, section 2, table 3 (below), I estimate the VEC model, finding that both convergence parameters are of the correct sign: $\hat{\gamma}_R = -0.0291$ and $\hat{\gamma}_C = 0.0777$, respectively, across all firms in the logarithmic formulation. This does not change much when splitting the firms according to the maximisation criteria in table 1. The general impression of the short-run and auto-regression VAR parameters, is that they are insignificant and stable, indicating that the other parameters are valid.

The long run return on total costs estimated in table 3 shows no difference from unity. This can be taken as a suggestion that the perfect capital market hypothesis holds, as there is no evidence that there exists profitable investments that have not been carried out. These *introductory* results must be taken with a pinch of salt, however, as these are only introductory results on which I have not conducted any cointegration tests.

It may be argued that when estimating the internal process of the business, the logarithmic transformation does not lend to the best interpretability. Each part of total costs is less than the total costs. The parameter interpretation of "a percentage increase in inventory investment is expected to increase revenue by β percent" does not give insight into the profitability of a business. The relative long run return on investment β will never be above unity, if the β of total costs is equal to unity. For the ease of interpretation, most of my analysis uses the monetary formulation, which lends the "cents earned per dollar spent" interpretation, well knowing that the logarithmic model is a better "fit" for the model.

4.2.3 Cointegration Tests

All tests based on the parameter beta and its relation to the standard error, will be biased toward non-rejection of the null hypothesis when using numerical (monetary) values, especially for businesses which exhibit monetary heteroskedasticity (Wooldridge, 2013). It must therefore be expected that the monetary model will reject more null-hypotheses than the logarithmic model. Since the logarithmic parameter estimates relative return on investment, I cannot test against the nullhypothesis that a parameter is equal to unity. The student's t-statistic is used to evaluate the hypotheses I have laid out, which is calculated by **t-value 1**:

$$t - value = \frac{(\hat{\beta} - \beta_0)}{s.e.(\hat{\beta})} \tag{64}$$

where $\hat{\beta}$ is the parameter estimate, β_0 is the parameter null hypothesis, and $s.e.(\hat{\beta})$ is the standard error of the parameter estimate. The normal distribution is used to evaluate the significance of the result, where a t-stat over 2 is the rule-of-thumb for a statistical significance of 95 percent, meaning that the rejection of the null hypothesis will be incorrect in less than 5 percent of cases (Type I error). There are three types of student's t-statistic tests, all based on equation 64. The first tests the significance of a parameter against zero ($\beta_0 = 0$)), the second test the difference of the parameter from a hypothesised value($\beta_0 = X$), and the third tests the difference between parameters($\beta_1 = \beta_2$). Heteroskedasticity increases the $s.e.(\hat{\beta})$, meaning that the test will be biased toward not rejecting the null hypothesis in the monetary formulation of the model.

Table 3: Introductory Results Step-by-Step

Below, I walk through the steps of the VEC model. The first step is a regression with no constant, where a parameter estimate equal to unity (1) generates residuals exactly equal to profits. The VEC model includes the residual from the regression into a differenced VAR model. The effect of profit is interpreted as equilibrium convergence. The long run return is estimated per firm, the presented values are the averages of the individual estimates.

Section 1	(A)	(B)	(C)	(D)	(E)	(F)
First-Step	Level	Logarithm	Moving Avg.	Log.MA	MA. No Constant	Log.MA No Cons.
Regression	Revenue	Revenue	Revenue	Revenue	Revenue	Revenue
Costs	0.243***	0.355***	0.849***	0.890***	**0.943*** **	**1.012*** **
	(0.00665)	(0.00667)	(0.00416)	(0.00319)	(0.003534)	(0.000524)
Constant	18,298***	6.174***	5,847***	1.148***		
	(398.6)	(0.0604)	(149.5)	(0.0296)		
Observations	18,627	12,840	26,977	26,386	26,977	26,386
R-squared	0.067	0.180	0.607	0.747	0.724	0.993

Section 2	(A)	(A)	(B)	(B)	(C)	(C)	(D)	(D)
VAR and VEC	D.MA	D.MA	D.Log.MA	D.Log.MA	D.MA	D.MA	D.Log.MA	D.Log.MA
Comparison	Revenue	Costs	Revenue	Costs	Revenue	Costs	Revenue	Costs
Lag1.Revenue	0.0320	-0.0133	0.00884	0.0446***	0.0349	-0.0190	0.0236***	0.00511
	(0.0227)	(0.0250)	(0.00837)	(0.0128)	(0.0232)	(0.0258)	(0.00878)	(0.0129)
Lag1. Costs	0.0170	-0.0352	0.0228***	-0.0535**	0.0224	-0.0321	0.0104	-0.0203
	(0.0275)	(0.0308)	(0.00738)	(0.0236)	(0.0269)	(0.0309)	(0.00771)	(0.0211)
Lag1. Profit					**-199.4*** **	**722.6*** **	**-0.0291*** **	**0.0777*** **
					(67.31)	(74.18)	(0.00523)	(0.00774)
Observations	26,858	26,858	26,207	26,207	26,300	26,300	26,207	26,207

Section 3	(A)	(A)	(B)	(B)	(C)	(C)	(D)	(D)
VEC Model	All Firms	All Firms	Profit Max	Profit Max	No Max	No Max	Grow Max	Grow Max
By Strategy	Revenue	Costs	Revenue	Costs	Revenue	Costs	Revenue	Costs
Lag1.Revenue	0.0236***	0.00511	0.0288	-0.0146	0.0333	0.0225	0.125***	0.0925***
	(0.00878)	(0.0129)	(0.0262)	(0.0366)	(0.0206)	(0.0260)	(0.0261)	(0.0334)
Lag1.Costs	0.0104	-0.0203	0.0140	-0.0498	0.00448	-0.0579*	0.00726	0.00447
	(0.00771)	(0.0211)	(0.0115)	(0.0566)	(0.0227)	(0.0333)	(0.0134)	(0.0200)
Lag1.Profit	**-0.029*** **	**0.0777*** **	**-0.0105**	**0.120*** **	**-0.060*** **	**0.0638*** **	**-0.049*** **	**0.0954*** **
	(0.0052)	(0.00774)	(0.00641)	(0.0213)	(0.0194)	(0.00978)	(0.00625)	(0.0149)
Observations	26,207	26,207	6,848	6,848	5,071	5,071	6,315	6,315
Long Run								
Revenue	1		1		1		1	
Costs	**1.048**		**0.886**		**1.133**		**1.107**	
	(0.136)		**(0.147)**		**(0.145)**		**(0.114)**	
Observations	26,207		6,848		5,071		6,315	

The return of revenue on revenue is always equal to unity (1). None of the return estimates are significantly different from unity in this introductory estimation.

The inclusion and exclusion of firms into and from the VEC model is determined by three tests. The first is the Dickey-Fuller unit root test, the second is the Johansen ”trace” cointegration test, and the third is a one percent parameter outlier test. The framework of the significance test from unity, t_2, is implemented in the most common unit root test, the Dickey-Fuller test, which has the null hypothesis that the variable is integrated. I incorporate a trend in the test structure, to control for the growth in business. I conduct the unit root test for each of the 35 firms in the sample, for both the monetary-and logarithmic model, and for the two-, three-, and four, week moving averages transformations. I only perform the unit root test on the cost-categories of firms whose unit root test for revenue is not rejected, since the integration of both revenue and the respective cost-categories are necessary for cointegration and VEC-estimation.

Only 17 and 12 firms passed the unit root for revenue using the two-week moving average for the monetary and logarithmic model, respectively. For the three-and four-week models, the more precise logarithmic model is stricter, rejecting the null hypothesis of unit root more often than the monetary model, which is in line with the heteroskedasticity discussion. The formulation that generates the most non-rejections of integrated revenue is the four-week monetary model. Though null-hypotheses are never accepted, the test indicates that 26 of the 35 firms have integrated monetary monthly moving average revenue. Within these 26 firms, 20 had integrated costs, and 20 had integrated investments. Only 11 had a unit root in their expenses, suggesting that expenses do not grow with the business in 9 of the 20 cases that had costs growing with the business (as defined by being integrated). The further splitting of investments and expenses does not change the number of firms with integrated variables much.

The firms that have variables that pass the unit root test for the respective formulation of the model, is tested for cointegration using Johansen's trace cointegration test (see table 4 below). For the bivariate model, this test is identical in spirit to the t-test, testing whether a statement is true or not: are the variables cointegrated or not. However, it is possible that there exists multiple cointegrating relationships θ when there are more than two variables. The Johansen trace test begins with the assumption that the variables are integrated, but that there is no cointegrating relationship between them that generates a stationary residual. If this is rejected, the test assumes that there exists one cointegrating relationship between the three or more variables, which generates a single stationary residual. If this is rejected, the model assumes that there exists two cointegrating relationships, that generate two stationary residuals. If this test is rejected in the trivariate case, every variable has a stationary relationship with itself, meaning that the variables themselves are stationary. The latter conclusion is a rejection of Johansen's cointegration test, but so is the very first hypothesis, that there exist no stationary parameter that generates a stationary relationship between the integrated variables. That no stationary equilibria is found and that a full stationary model is found, both result in a rejection of the test. It is the non-rejection of the intermediate hypothesis that results in passing of the cointegration test.

Table 4: Unit Root and Cointegration Tests

To be included in the VEC estimation, the variables in the system must be cointegrated, but to be cointegrated, the variables must first be integrated (have a unit root). I test for a unit root in the revenue of each business, using the Augmented Dickey-Fuller test with a trend and one lag. The number of businesses that pass the test is noted before the "/" below, the total number of tested business being behind the "/". I only test for unit roots in the other variables if the business has a unit root in its revenue. Only those who have a unit root in revenue and the relevant cost-variables are tested for cointegration.

Unit Root and Cointegration Tests Selection Criteria	Monetary 2 Week M.A.	Monetary 3 Week M.A.	**Monetary 4 Week M.A**	Logarithmic 2 Week Log M.A.	Logarithmic 3 Week Log M.A.	Logarithmic 4 Week Log M.A.
Unit Root Test						
Revenue (R) (Main Variable)	17 / 35	23 / 35	**26 / 35**	12 / 35	18 / 35	22 / 35
Total Cost (C)	5 / 17	15 / 23	**20 / 26**	4 / 12	10 / 18	12 / 22
Investment (I)	6 / 17	16 / 23	**20 / 26**	4 / 12	10 / 18	13 / 22
Expenses (E)	2 / 17	5 / 23	**11 / 26**	1 / 12	5 / 18	6 / 22
Inventory Inv. (II)	2 / 17	13 / 23	**18 / 26**	7 / 12	9 / 18	12 / 22
Marketing Inv. (MI)	6 / 17	12 / 23	**17 / 26**	4 / 12	7 / 18	13 / 22
Fixed Exp. (FE)	2 / 17	7 / 23	**10 / 26**	2 / 12	2 / 18	4 / 22
Variable Exp. (VE)	2 / 17	5 / 23	**12 / 26**	3 / 12	5 / 18	9 / 22
Cointegration Test						
Bivariate Model	2 / 5	10 / 15	**12 / 20**	2 / 4	6 / 10	7 / 12
Trivariate Models	1 / 2	5 / 6	**9 / 12**	1 / 1	4 / 5	6 / 6
(I-E and II-MI)	1 / 1	7 / 8	**10 / 14**	2 / 3	3 / 4	5 / 6
Multivariate Models	1 / 1	3 / 3	**5 / 6**	1 / 1	2 / 2	4 / 4
(II-MI-E and Full)	0 / 0	1 / 1	**3 / 3**	1 / 1	1 / 1	2 / 3

I run the tests for six model specifications, where the first three are with numeric (monetary) values for the 2-, 3-, and 4-week moving averages, and the following three specification are with logarithmic values. In general, the logarithmic specification is stricter (due to heteroskedasticity), and more businesses pass the selection criteria when the moving average is longer. Only the business that pass the selection criteria are used in subsequent estimations.

Only the firms with variables that pass all respective unit root tests for the model in question, are tested for cointegration. The same 20 firms have integrated revenue and cost in the monetary monthly moving average formulation, of which 12 pass the cointegration test. In the logarithmic monthly moving average formulation, 12 firms are eligible for the cointegration test, of which 7 pass. In general, more firms stay in the sample the longer the moving average is, if the formulation is logarithmic, and the fewer variables are included in the model. In the logarithmic formulation of the two trivariate specifications containing investments and expenses, and inventory and marketing investments, respectively, six and five firms pass both tests. In the monetary model equivalent, 9 firms pass in the first trivariate model, and 10 pass the second trivariate model. I prefer the interpretation of the models that contain all costs, since the interpretation of the first step residual with investments split into two, but no "other costs" variable, is not equal to profits.

To enable the ranking of investments, while controlling for the entire spectra of firms' costs, I formulated the quadvariate model, in which 5 firms pass both the unit root test for all four variable and the cointegration test. In the model where expenses are split into variable and fixed expenses, only 3 firms pass both tests. The VEC model requires cointegrated variables, meaning that only the firms that pass the cointegration test may be kept in the sample. In the most charitable of models, the bivariate monetary monthly moving average model, 12 of 35 firms pass both tests, but I aim to rank investment opportunities both within and between firms. Therefore, I extend the moving average to estimate a model with quarterly moving averages in the latter parts of the paper. This includes 14 firms in the quadvariate model of revenue, inventory-and marketing investments, and expenses, compared to the 5 in the monthly moving average case. With this increase in the sample, I split the firms into high and low profit-and growth firms.

I estimate the VEC model for each firm that passes the cointegration test separately. In the presentation of the estimated parameters, I calculate the mean parameter estimates of the individual regressions. In each model, the return of every variable in the model on revenue is estimated. The parameter estimates and standard errors are the average estimates for each variable and each model specification. To ensure that the results are not driven by outliers, I conduct a final one percent outlier test. I define an outlier as an estimate that is 3 times as large in magnitude as the mean parameter estimate. That is, for the bivariate case in which we observed a mean estimate of unity, I exclude the firms with returns estimated to be more than 3 in absolute values. The same is true for firms with standard errors.

The number of firms is at times strongly reduced from the starting sample of 35 eligible firms, but after finishing the estimation of my proposed models, and the various formulation and specifications I have tried, I must say that I am impressed by how consistent the estimates are. Engle and Granger (1987) showed that the regression is super-efficient with cointegrated variables, and my impression is that Johansen's (1988) cointegration coefficient is estimated very efficiently too.

5 Results

The strategy I have chosen to validate the interpretation of the VEC cointegration coefficient β as the long run return on investment, is formulated through three main hypotheses. First, I investigate whether the estimated return on investments is higher than the estimated return on expenses. Second, I test whether the model estimates higher (lower) return on investment for firms that have higher (lower) sums of profit and growth. Third, I rank investment choices to compare the types of investments between firms, but also within firms. A total of four tables are presented in this section, which present: a comparison of return on investments, expenses, and their sub-categories across the six initial model specification (Table 5), the full VEC model (Table 6), the simplified VEC model for monthly and aggregate data (Table 7), and finally the high/low return split (Table 8). The first three tables (5-7) are partially used as steps to find the best specification of the model, but they also show the robustness and consistency of the parameter estimates across different model specifications. These three tables confirm the first hypothesis, but it is in the final table that the two latter hypotheses are confirmed. I begin the chapter with a summary of the three hypotheses.

5.1 Summary of Hypotheses

5.1.1 Main Hypotheses

Investments generate revenue through marketing (demand) and inventory (supply), while expenses must be paid to facilitate for the revenue. Since expenses do not generate revenue themselves, investments must generate enough revenue to cover for the cost of the investment *and* the cost of expenses needed to facilitate for the revenue. For the logarithmic model the simplified hypothesis $H_I : \beta_I > \beta_E$ is tested, but in the monetary model, a verification of the cointegrated coefficient β's interpretation as the return on investment must fulfil the extended verification hypothesis:

$$H_I : \beta_I > 1 > \beta_E = 0 \tag{65}$$

Meaning: The return on investment is positive and higher than the return on expenses, which do not generate revenue. This is based on the intuition that it is investments, into inventory and marketing in ecommerce, that drive, cause, generate, and determine revenue. Expenses, on the other hand, are necessary for the operation of a business, but in the facilitation for revenue.

I assume that firms with high return on investment grow faster and have more profits. Therefore, the parameter β must differentiate between firms with high and low return on investment, estimating higher β for firms with higher (H) growth-and profit than for firms with low (L) growth-and profit. This is the identification hypothesis:

$$H_{II} : \beta_{IH} > \beta_{IL} \tag{66}$$

Meaning: That the estimated return on investment is higher for high growth-and profit firms.

Based on industry intuition, I speculate further that high return firms differentiate themselves from low return firms with more efficient marketing. This may indicate efficient marketing, or a "appealing products" effect, but successful firms are hypothesised to have much higher return on marketing. The return on inventory investments, on the other hand, is not much different as the firms source their inventory from similar suppliers. The returns on the different types of investment are "partial", meaning that the increased return on inventory due to higher return on marketing is controlled for in the ranking hypothesis:

$$H_{III} : \beta_{MIH} >> \beta_{MIL} \tag{67}$$

$$\beta_{IIH} = \beta_{MIL} \tag{68}$$

Meaning: That high return firms have a much higher return on marketing investment, but that both types of firm have similar return on inventory investment.

5.1.2 Secondary Hypotheses

The long run return on investment is the main concern of my thesis. It is estimated through a restricted sub-matrix of the vector space, giving a "partial" interpretation of the cointegrating coefficient β, as the cumulative effect of a unit investment in the respective input factor. Its counterpart is the adjustment or convergence parameter γ, the traditional error correction term, explaining the effect of the profit (equilibrium term) on every other variable. If the model is correctly specified, γ_R must be negative while γ_C must be positive:

$$\gamma_C > 0 > \gamma_R \tag{69}$$

I interpret γ_R as the realisation time of investments, a larger parameter indicating that investments are turned into revenue faster. An extension of the model to multiple equilibria could compare the realisation time of different *types* of investments, but I do not make that step, and I do not have any specific hypothesis concerning γ_R, except that it must be negative. The variable "costs" is split into investments and expenses, but in general γ_C estimates the sensitivity of costs to profits, which I interpret as the cash-flow sensitivity discussed in the literature. The conclusion in the literature that the financial constraints indicated by cash-flow sensitivity are caused by lacking funds to fulfil investment opportunities, however, is not evident. I estimate the cash-flow sensitivity of inventory and marketing investments, and expenses, but I argue that their sensitivity to available funds may also indicate firms that struggle to make ends meet, in particular for the case of expenses. I am ambivalent to the which interpretation of financial constraints per cash-flow sensitivity is more correct, but the γ_C parameters must be positive for convergence:

$$\gamma_C = \gamma_I = \gamma_E = \gamma_{II} = \gamma_{MI} > 0 \tag{70}$$

Only in one of the four tables do I estimate short-run VAR parameters, and I do not have any particular hypothesis concerning them. In general, the short run parameter estimates are close to zero, and at least below one, confirming the stability of the model. The "short-run" in the model is day-to-day, which I assume is too short for any "real" effects.

5.2 Verification Hypothesis

5.2.1 Return on Investment

The mean parameter estimate of the cointegrating coefficient β_I, for the firms that pass the unit root and cointegration tests, is higher than the mean parameter β_E (see Table 5 below). The finding is consistent, and in both an economic and a statistic sense, highly significant. This is indicative of a validation of my verification hypothesis that the cointegrating coefficient β is interpreted as the (long-run) return on investment. In the monetary three-and four-week formulations, $\hat{\beta}_I$ is significantly larger than unity ($\hat{\beta}_I^{21} = 1.513$ and $\hat{\beta}_I^{28} = 1.444$) suggesting returns on investment of ($\hat{\beta}_I^* = \hat{\beta}_I - C_I) = .513$ and ($\hat{\beta}_I^* = \hat{\beta}_I - C_I) = .444$, corresponding to 51.3 and 44.4 cents earned per dollar, respectively, since the cost of investment is normalised to one. The return on expenses is not only negative, but the revenue generated by expenses is statistically equal to zero in the monetary formulations, which for the corresponding formulations is $\hat{\beta}_E^{21} = -0.733$ and $\hat{\beta}_E^{28} = 0.425$, both with mean standard errors larger than the cointegration coefficients ($s.e(\hat{\beta}_E = 1.629$ and 0.898). The monetary monthly moving average estimates from Table 5 can be presented in the verification hypothesis form as:

$$\hat{\beta}_I > 1 > \hat{\beta}_E = 0 \tag{71}$$

$$1.444 > 1 > 0.425 = 0 \tag{72}$$

confirming H_I in equation 65, by rejecting equation 43 in favour of equation 45 and not rejecting equation 44, meaning: the return on investment is positive and higher than the return on expenses, which do not generate revenue. The estimates suggest that spending a dollar on investments rather than on expenses, will result in a $1.444 - 0.425 = 1.019$ dollar increase in earnings over the long run. Whether the return on investment $\beta_I^* = 1.444 - 1 = 0.444$ is enough to cover for the the return on expenses $\beta_E^* = 0.425 - 1 = -0.575$ must be calculated using the amount spent on each, but the minimum requirement for a profitable firm can be stated as $\beta_I^* X_I + \beta_E^* X_E > 0$ where X_k

is the amount spent on each input factor, bound by $X_I + X_E = 1$. If the representative firm in the sample spends twice as much on investments than on expenses, then the representative firm has a general profitability of $0.444 * \frac{2}{3} - 0.575 * \frac{1}{3} = 0.0877 = 8.77$ cents per dollar, whereas if the firm's expenses rose to be as large as investments, the firm would lose 9.05 cents per dollar spent.

For the logarithmic formulation, the same ranking holds, but the long run effect of expenses on revenue is negative, corresponding to a $\hat{\beta}_E^{14} = -0.24$, $\hat{\beta}_E^{21} = -0.32$, and $\hat{\beta}_E^{28} = -0.26$ percentage decrease in long run revenue as a result of a percentage increase in spending on expenses. An increase in expenses, when controlling for the effect of investments, causes a long-run relative decline in revenue. The percentage effect of investments on revenue is consistently positive and close to unity, a percentage increase in investment spending being estimated to cause a $\hat{\beta}_I^{14} = 0.90$, $\hat{\beta}_I^{21} = 0.84$, and $\hat{\beta}_I^{28} = 0.77$ cumulative percentage increase in revenue over the long run.

In the bivariate model, the only formulation which does not indicate that firms are generally profitable, as indicated by the long run return on total costs, is the two-week monetary formulation. Neither this formulation nor any the other formulations, estimate a long run return on costs statistically different from zero, which means that the model cannot reject the benchmark perfect capital market hypothesis

$$\hat{\beta}_C^* = \hat{\beta}_C C_C - C_C = 1.172 - 1 = 0.172 = 0 \tag{73}$$

since the mean standard error is 0.125, meaning that $\frac{1.172-1}{0.125}$ is less than 2, and therefore not statistically significant.

The revenue generated by marketing investments is significant and higher than unity in the monthly monetary formulation, with an estimated long run return on marketing of

$$\hat{\beta}_{MI}^* = \hat{\beta}_{MI} - C_{MI} = .603 > 0 \tag{74}$$

corresponding to 60.3 units earned per unit spent. The return on inventory investments is estimated to be statistically equal to zero

$$\hat{\beta}_{II}^* = \hat{\beta}_{II} - C_{II} = 0.034 = 0 \tag{75}$$

in the monetary formulation. The estimated revenue generated by inventory investments is at or close to zero in the logarithmic formulation too. Between the three logarithmic estimates of the long run revenue generated by marketing investments, the average estimate is 672, corresponding to a percentage increase in marketing investments generating a .672 percentage increase in long run revenue. To calculate the monetary value of this estimate, one must assess the monetary value of a percentage increase in marketing and revenue respectively, and then multiply the monetary value of a percentage increase in revenue by .672. This complication is a downside of the logarithmic

Table 5: Long Run Return on Investment

Johansen's long run VEC parameter is estimated per business, the average estimate and standard error is presented below, with the number of observations for each parameter. The return on costs is consistently above unity, though never significantly so. The model 2a estimates are consistently larger for investments than returns. In the monetary formulation, return on investment is significantly above unity for the 3-and 4-week moving averages, while the return on expenses is statistically equal to zero.

Long Run Return on Investment	Monetary 2 Week M.A.	Monetary 3 Week M.A.	Monetary 4 Week M.A	Logarithmic 2 Week M.A.	Logarithmic 3 Week M.A.	Logarithmic 4 Week M.A.
Revenue (R)	1	1	1	1	1	1
Model 1						
Total Costs (C)	0.975	1.254	1.155	1.254	1.217	1.171
(Std. Error)	(0.073)	(0.145)	(0.099)	(0.171)	(0.130)	(0.125)
[Observations]	[2]	[10]	[12]	[2]	[6]	[7]
Model 2a						
Investments (I)	**0.121**	**1.513*****	**1.444*****	**0.904*****	**0.844*****	**0.773*****
	(9.124)	(0.208)	(0.158)	(0.103)	(0.102)	(0.111)
	[1]	[4]	[9]	[1]	[5]	[6]
Model 2a						
Expenses (E)	**0.52**	**- 0.733**	**0.425**	**- 0.240***	**- 0.320***	**- 0.258***
	(0.358)	(1.629)	(0.898)	(0.136)	(0.134)	(0.133)
	[1]	[3]	[9]	[1]	[5]	[6]
Model 2b						
Inventory Inv. (II)	0.696	1.144	1.034	0.178	0.041	0.277**
	(0.167)	(0.121)	(0.231)	(0.121)	(0.152)	(0.074)
	[1]	[5]	[5]	[2]	[1]	[4]
Model 2b						
Marketing Inv. (MI)	1.201	1.385	1.603***	0.733**	0.753**	0.529**
	(0.169)	(0.564)	(0.247)	(0.072)	(0.087)	(0.055)
	[1]	[5]	[5]	[2]	[1]	[4]
Model 3						
Fixed Exp. (FE)	N / A	N / A	-4.497*	N / A	N / A	0.179
			(0.838)			(0.064)
			[3]			[2]
Model 3						
Variable Exp. (VE)	N / A	N / A	0.475	N / A	N / A	0.104
			(0.321)			(0.069)
			[3]			[2]

In the logarithmic formulation, we must interpret the parameters as percentage changes. Since each cost category is less than total costs, no parameter will be higher than that for total costs. Nonetheless, the logarithmic model confirms the higher return on investment than expenses, the former being close to unity, while the latter is significantly negative.

formulation in my model, but the upside of the logarithmic formulation is its precise estimates, and its interpretation of relativity. The logarithmic formulation is confident that investments generate more revenue than expenses, but also that expenses have a long-run negative effect on revenues, and that the return on marketing investments is higher than the return on inventory.

In the remainder of the results, following the findings in Table 4 and 5, I continue with the trivariate model with revenue, investments, and expenses, and the quadvariate model where investments have been split, leaving the split into fixed and variable expenses behind.

5.2.2 Investment Realisation and Cash-Flow Sensitivity

The convergence parameter γ_C estimates the speed at which firms spend money once they get money, which is the essence of the cash-flow sensitivity view of financial constraints. In the trivariate model with investments and expenses (see table 6 below), I find that investments are sensitive to profits, but that expenses are not.

$$\hat{\gamma}_I > \hat{\gamma} = 0 \tag{76}$$

$$.0223 > .003 = 0 \tag{77}$$

In the quadvariate model, the estimates are significant, but of the incorrect sign, indicating that more is spent on marketing and expenses the less internal funds the firm generates:

$$\hat{\gamma}_{II} = 0 > \hat{\gamma}_{MI} > \hat{\gamma}_E \tag{78}$$

$$.0027 = 0 > -.0036 > -.0047 \tag{79}$$

which is contrary to my hypothesis that costs converge toward their equilibrium with revenue.

In the short-run VAR parameters in the quadvariate model, a economically and statistically significant short run effect of marketing on revenue equal to $\hat{\tau} = 0.32$ is estimated, corresponding to the expected increase in growth tomorrow as a result of a dollar increase in marketing expenditure today being 0.32 cents. A weak, $\hat{\tau}_{ER} = -0.11$, effect of expenses on revenue is estimated, along with weak positive effects of marketing on expenses, of inventory on marketing, and of revenue on inventory. These suggest that spending often comes in bulk, the increase in one category being related to increases in the other, and that inventory stocks are often replenished when goods are sold. The convergence of inventory investments also captures the latter, and γ_{II} is of the correct positive sign, but statistically insignificant. In general, the VAR parameters in equation 9 and 10 are approximately equal to zero:

$$\hat{\delta}_k \approx \hat{\alpha}_k \approx \hat{\tau}_k \approx 0 \tag{80}$$

where k stands for revenue, investments, expenses, inventory, and marketing.

Table 6: Full VEC Model Presentation

The VEC model contains short-run parameters, which may cause over-parametrisation in the model, as the number of short-run parameters increases with the square of variables in the model. No significant short-run effects are estimated for the trivariate model (A). The strongest short-run effect in the multivariate model (B) is the positive effect from marketing to revenue. In order from strongest to weakest significant effects: a positive effect of marketing on expenses, a negative effect of expenses on inventory, a positive effect of marketing on marketing, and a positive effect of inventory on marketing. Revenue converges to the equilibrium, and so do investments in (A), suggesting cash-flow sensitivity.

Full VEC Model Result Presentation	(A) Revenue Month MA	(A) Investment Month MA	(A) Expenses Month MA	(B) Revenue Month MA	(B) Inventory Month MA	(B) Marketing Month MA	(B) Expenses Month MA
L1.Revenue	0.0446	0.0426	0.0052	-0.0079	0.0459	0.0008	-0.0125
	(0.0408)	(0.0349)	(0.0175)	(0.0230)	(0.0173)	(0.0033)	(0.0134)
L1. Investment	0.0403	0.0081	-0.0035				
	(0.0522)	(0.0407)	(0.0204)				
L1.Inventory				0.0261	-0.0364	**0.0252*** **	-0.0304
				(0.0332)	(0.0225)	(0.0057)	(0.0174)
L1.Marketing				**0.3198****	0.0350	**0.0527****	**0.1338****
				(0.1306)	(0.1020)	(0.0241)	(0.0632)
L1.Expenses	0.0224	-0.1186	-0.0208	-0.0343	**-0.1077****	-0.0110	-0.0284
	(0.1600)	(0.1212)	(0.0402)	(0.0586)	(0.0418)	(0.0082)	(0.0223)
EC1.Profit	**-0.0282*** **	**0.0223*** **	**0.0030**	**-0.0436*** **	**0.0027**	**-0.0036*** **	**-0.0047*** **
	(0.0094)	(0.0076)	(0.0039)	(0.0054)	(0.0032)	(0.0007)	(0.0019)
Constant	36.8888	14.1600	12.8109	16.5420	-0.9120	10.3968	10.7084
	(98.4632)	(81.0146)	(29.6303)	(57.6186)	(37.6310)	(13.8324)	(31.1566)
Firms	9	9	9	5	5	5	5
Observations	6142	6142	6142	3325	3325	3325	3325
Long Run							
Revenue	**1**			**1**			
Investment	**1.3036***						
	(0.1874)						
Inventory Inv.				**1.2260**			
				(0.2472)			
Marketing Inv.				**1.7218**			
				(0.5902)			
Expenses	**-0.1767**			**-0.1592**			
	(0.9857)			(0.4752)			
Firms	9	9	9	5	5	5	5
Observations	6142	6142	6142	3325	3325	3325	3325

The long run return on investment is significantly above unity in (A), indicating positive earnings on investment, while the return on expenses is statistically equal to zero, expenses generating no revenue in both (A) and (B). The return on inventory and marketing is above unity in (B), but not significantly so.

The realisation time on investment, defined by γ_R, is estimated to be significant and of the correct sign, profits have a negative effect on revenue growth. Though significant and of the correct sign in both models, the parameter is also significantly different between the trivariate () and the quadvariate ($= -0.044$) models:

$$|\hat{\gamma}_R^4| > |\hat{\gamma}_R^3| = |-0.044| > |-0.028| = 0 \tag{81}$$

The statistical reason for this difference may that the quadvariate model controls for the short run effect of marketing on revenue, which biases the trivariate $\hat{\gamma}$ toward zero. To interpret the parameter estimates, one must understand that as the distance to the profit-equilibrium diminishes, due to the revenue-correction described by γ_R, the effect of the dis-equilibrium also diminishes. Since the effect of the investment that caused the negative shock to profits reduces, the time to realisation cannot be calculated as $\gamma_R * t = 1$ where the t that solves the equation is the realisation time. Rather, the realisation time calculation must account for the decreasing pull profits have on revenue as the revenue increases.

The negative estimates of γ_R and the positive estimate of γ_I and γ_{II} are taken as indications for the validity of the model, while the negative estimates of γ_{MI} and γ_E are taken as indications against the validity of the model. The null hypotheses that $\gamma_R = \gamma_C = 0$ is thus fully rejected in favour of alternative hypothesis 2a as $\gamma_R < 0$, but only rejected in favour of alternative hypothesis 2b for investments in the trivariate model $\gamma_{I_{M3}} > 0$. Null hypothesis 2b is not rejected for the parameters $\gamma_{E_{M3}}\gamma_{II_{M4}} = 0$, but it is rejected for $\gamma_{MI_{M4}}\gamma_{E_{M4}} < 0$ in favour of the wrong alternative hypothesis. This indicates that the model is not stable, though the estimates are not economically significant, only amounting to a daily dis-convergence of 0.36 and 0.47 percent, meaning that if the firm gets 100 dollars in profits it will spend 36 and 47 cents less on marketing and expenses. Convergence parameters of the correct sign would indicate that the firm spends more when it gets more, such as for sensitivity of investments to profit in the trivariate model, where increased profits are expected to cause a daily investing of 2.2 percent of that profit, until there are no profits left.

The insignificant and significantly wrong estimates suggest that the equilibrium is perhaps a bit weak, but the full model could also be suffering from overparametrisation. The inflation of the standard errors in the estimation of the long run cointegration coefficients β are a sign of over-parametrisation, which is the reason why neither inventory-nor marketing investments have statistically significant long run return in the full model. Both have a significant cumulative effect on revenue, but neither effect is statistically higher than the cost of investment, though the return on marketing is economically large:

$$\hat{\beta}_{MI} > \hat{\beta}_I > \hat{\beta}_{II} > 1 > 0 > \hat{\beta}_E \tag{82}$$

$$1.7218 > 1.3036 > 1.2260 > 1 > 0 > -0.1680 \tag{83}$$

where only $\hat{\beta}_I$ is significant due to the high standard errors of the estimates. The cumulative effect of expenses on revenue is found to be zero in both models, the average is presented equation 83.

The insignificance of all autoregressive parameters α strongly suggest that the model is stable and that the integrated variables are stationary in differences. The insignificance of the constants is expected, and they too indicate that the model is correctly specified. Though the full VEC model seems to be suffering from overparametrisation, the findings are in support of the verification hypothesis: the cumulative revenue generated by investments are at or above unity, while the cumulative revenue by expenses is zero or even below zero.

5.2.3 Simplification and Aggregation

By removing the short-run and auto-regressive VAR parameters, the issue of overparametrisation is solved, as nine and sixteen parameters are removed in the trivariate and quadvariate models, respectively. The simplified VEC only keeps the intuition of the VAR framework in estimating the EC parameter γ, and the cointegration coefficient β from the restricted sub-matrix. I employ the simplified VEC model in a comparison of the moving average formulation with an aggregate data formulation. My motivation for using the moving average transformation was to exploit the frequency of the data, thereby allowing for a more accurate estimation of the parameters described in the model. As seen in Table 7, however, the monthly aggregate formulation for both the trivariate and quadvariate models leads to more firms (13 and 10) passing the cointegration test, than for the monthly moving average formulation (9 and 5). With quarterly aggregates, three firms were confirmed to have cointegrated revenues, investments, and expenses. The cointegration test can thus be passed using an average of only 10 quarters, though more pass with shorter frequency data.

There are few significant differences between the estimates of the monthly aggregate and the monthly moving average model, if any. The statistical differences are between the return on investment and the return on expenses, which is clear, consistent, and robust to either specification. In the aggregate specifications, the convergence parameters γ_k are easier to interpret. None of the cash-flow sensitivity parameters are significant, but the parameter γ_R that I interpret as the realisation time of investments is, and it is estimated to by $\hat{\gamma_R} = -0.632$ and $\hat{\gamma_R} = -0.561$ in the two monthly aggregate models, corresponding to a realisation time of between 2 and 4 months. The return on investment is the full, total, and long run realisation of investments to revenue, which is the time between buying and selling the good, and the estimated realisation time of 2 to 4 months sounds about right in this perspective.

The consistency of the monthly moving average and aggregate estimates of the reduced VEC model, inspired me to go further. Using the sample of 3 firms that passed the cointegration test

Table 7: Moving Average and Aggregates

Moving average transformations of the data are rarely seen in economic studies, since the data is usually too infrequent for the transformation to be relevant. Bank account data is per transaction, which I collapse to daily intervals and then "smooth" with moving averages. To ensure that the MA transformations does not incur any crucial statistical biases, I compare the reduced VEC model with MA data with the VEC model using aggregate data, monthly and quarterly. Monthly, the MA and AG estimates are, in broad strokes, similar, showing the same general result: Returns on investment are higher than returns on expenses.

Reduced VEC Model	(A) Revenue	(A) Investment	(A) Expenses	(B) Revenue	(B) Inventory	(B) Marketing	(B) Expenses
Monthly MA							
EC1.Profit	-0.0256***	0.0249***	0.0033	-0.0495***	0.0057	-0.0035	-0.0056
	(0.0098)	(0.0081)	(0.0039)	(0.0093)	(0.0069)	(0.0028)	(0.0041)
Long Run Return	1	**1.3552****	**0.2578**	1	**1.2592**	**2.2432*****	**-0.8643**
		(0.1774)	(0.7996)		(0.2332)	(0.5190)	(0.4139)
Firms	9	9	9	5	5	5	5
Observations	6142	6142	6142	3325	3325	3325	3325
Monthly AG							
EC1.Profit	-0.6316***	0.3310	-0.0177	-0.5611***	0.3486	0.0405	0.0657
	(0.2777)	(0.2136)	(0.0915)	(0.2912)	(0.1893)	(0.0723)	(0.0893)
Long Run Return	1	**1.6105*****	**-0.1121**	1	**1.1846**	**1.5668**	**0.2144**
		(0.1871)	(0.3035)		(0.1587)	(0.5661)	(0.3020)
Firms	13	13	13	10	10	10	10
Observations	286	286	286	260	260	260	260
Quarter MA							
EC1.Profit	-0.0768***	0.0171***	0.0193***	-0.0434***	0.0166***	0.0000	0.0064
	(0.0109)	(0.0075)	(0.0059)	(0.0086)	(0.0063)	(0.0027)	(0.0035)
Long Run Return	1	**1.190**	**0.4229**	1	**1.7600*****	**2.1794*****	**0.7043**
		(0.0993)	(0.2462)		(0.1622)	(0.4255)	(0.3088)
Firms	11	11	11	14	14	14	14
Observations	8585	8585	8585	11434	11434	11434	11434
Quarter AG							
EC1.Profit	-3.2669***	-0.9545	-0.3773	-	-	-	-
	(1.4398)	(0.9125)	(0.2377)	-	-	-	-
Long Run Return	1	**1.5111*****	**0.4291**	-	-	-	-
		(0.1166)	(0.2023)	-	-	-	-
Firms	3	3	3	0	0	0	0
Observations	31	31	31	0	0	0	0

The quarterly aggregates stretch the data very thin, the longest time series being 12 quarters. No cointegration is found in the multivariate model, but the three firms that are cointegrated in the trivariate model (A) with quarterly AG data, the results follow the same general story. With quarterly MA data, 14 of the 35 firms are cointegrated, the most of any specification. In the multivariate (B) model, the return on both inventory and marketing are statistically and economically larger than unity. The return on expenses is insignificantly different from 0.

for the quarterly aggregate data, the model is still consistent, estimating a strongly positive return on investment and a close-to-zero estimate for the revenue generated by expenses

$$\hat{\beta}_I^* = \hat{\beta}_I - C_I = 1.511 - 1 = .511 \tag{84}$$

$$\hat{\beta}_E = 0.422 \tag{85}$$

The difference in the estimates corresponds to a difference in return on investment expenses of more full dollar per dollar spent:

$$\hat{\beta}_I - \hat{\beta}_E = 1.511 - 0.422 = 1.11 \tag{86}$$

That is, the average firm is expected to earn more than a dollar more by investing a dollar rather than covering expenses with the dollar, similar to the finding in Table 5 and Table 6. This evidence continues to be strong in favour of my verification hypothesis, and I therefore conclude that the correlation coefficient β in Johansen's VEC model estimates that:

H_I: The return on investment is higher than the return on expenses

Confirming the first of my three main hypotheses, which suggests that the cumulative partial effect of investments on revenue β_I does estimate the long run return on investment. This hypothesis is necessary for the interpretation of β_I as a measure for the creditworthiness of a firms, but it is not sufficient. The parameter β_I must estimate higher returns for higher return firms and for high return investments. The general impression thus far is that of the estimates is that the cost types can be ranked by their cumulative partial effect on revenue:

$$\bar{\bar{\beta}}_{MI}^* = .963 > \bar{\bar{\beta}}_I^* = .399 = \bar{\bar{\beta}}_{II}^* = .401 > 0 > \bar{\bar{\beta}}_E^* = -.608 \tag{87}$$

where costs are normalised to one and subtracted from β_k, and the $\bar{\bar{\beta}}$ is the average of the estimated long run returns between the seven formulations. In the counterfactual of spending a dollar on investments rather than on expenses, the business would earn 40 cents instead of losing 60 cents. Furthermore, if the firm were to invest a dollar in marketing rather than in inventory, it would earn an additional 56 cents.

In the seventeen estimates of costs' sensitivity to profits, four significant results are found, all of which of the correct sign. It is investments that are sensitive to profits in these estimations (2 of 3), and in particular inventories investments (1 of 3), while expenses are found to be sensitive to profits only once of its seven iterations (1 of 7). The convergence of revenue is much stronger, found to be both statistically and economically significant in all seven estimations (7 of 7). An interpretation of this finding is that the business drives revenue through investments on both supply and demand sides, rather than exogenous demand driving revenue and supply having to catch up to demand. Revenues strongly follow investments, but investments follow revenue only weakly.

The quarterly quadvariate model estimates that inventory generate a long-run increase of revenue, but that the return on marketing investments is significantly higher

$$\hat{\beta}^*_{MI} > \hat{\beta}^*_{II} = (2.18 - 1) > (1.76 - 1) = 1.18 > .76 \tag{88}$$

That is, a dollar spent on marketing returns two dollars in revenue, where one dollar is profit. The standard error on the estimated return on marketing investments is consistently the most inaccurately estimated parameter This indicates that marketing investment have a relative high return, but that it is also carries higher risk than inventory investments, which are relatively safer but lower return investments. This is the case in all specification including marketing in tables 5 through 7.

To summarise, the general result from the comparison of monthly-and quarterly aggregates-and moving averages is that: (i) the cointegration coefficient β is robust to different formulations, (ii) the return on investments in both inventory and marketing is more than the initial cost of the investment, (iii) expenses generate no revenue at all, and lastly (iv) the highest, but most variant return on investment is in marketing.

5.3 Identification and Rank Hypotheses

5.3.1 Identification of Profitable Firms

In this section, I test the second and third hypotheses: does the cointegration coefficient β sort the high return firms and investments from the low return firms and investments? In this investigation, I split costs into investments and expenses, and investments into inventory and marketing, as in the quadvariate model before. I add by splitting firms according to their sum of profit and growth, which I argue gives an adequate approximation of return on investment, to investigate whether the parameter β is higher for firms and investments with higher return. I use the *quarterly* moving average model from above, which includes 14 firms. They are split 50/50 according to their average quarterly sum of profit and growth.

In table 8 (below), I present descriptive statistics for the two groups, and their respective first step regressions, with the reduced VEC estimates. High return firms are drastically different from low return firms. Their average quarterly revenue is 2.6 million DKK, while the average low return firm turns over 0.9 million DKK a quarter. The high return firms grow faster than the low return firms, per construction, and the high return firms are also left with more excess revenue after cost.

Table 8: Identifying Investment Opportunities

High return (Q) firms are defined as having a combined average quarterly growth-and profit rate of above 25%, low return firms having less than that. I present quarterly descriptive statistics for the two types of firms, where the share of revenue is defined as the share of X over revenue, and the growth rate being the relative growth of X. High return firms are larger, spend relatively less on inventory, relatively more on marketing, and relatively less on expenses. The first step regression with three explanatory variables is shown, all variables are highly significant and together explain 97-99% of the variation in revenue.

Identifying High-Return Businesses	**High Q** Revenue Quarter MA	**High Q** Inventory Quarter MA	**High Q** Marketing Quarter MA	**High Q** Expenses Quarter MA	**Low Q** Revenue Quarter MA	**Low Q** Inventory Quarter MA	**Low Q** Marketing Quarter MA	**Low Q** Expenses Quarter MA
Des. Statistics								
Level	**2.614.085**	1.026.671	608.360	520.787	**920.754**	390.906	193.581	297.078
	(3.307.943)	(1.362.012)	(1.085.408)	(613.556)	(1.383.907)	(612.182)	(365.597)	(447.392)
Growth	**306.658**	76.552	146.266	49.130	**-16.079**	11.130	-9.094	-4.768
	(1.762.936)	(889.195)	(550.258)	(371.865)	(741.529)	(148.312)	(148.312)	(270.269)
Share of Revenue	1	**36.97%**	**23.60%**	**22.64%**	1	**47.97%**	**21.17%**	**37.46%**
		(16.75%)	(17.50%)	(10.65%)		(35.15%)	(12.42%)	(39.47%)
Growth Rate	20.35%	-2.03%	0.50%	0.14%	15.97%	-5.39%	0.75%	0.43%
	(31.77%)	(12.33%)	(7.91%)	(8.63%)	(35.77%)	(34.98%)	(9.46%)	(4.50%)
Firms	7	7	7	7	7	7	7	7
Observations	69	69	69	69	75	75	75	75
1st Regression								
Revenue		1.381***	0.745***	1.211***		1.129***	0.751***	1.016***
		(0.101)	(0.110)	(0.247)		(0.075)	(0.174)	(0.115)
R-Squared	**97.61%**				**98.76%**			
Firms	7	7	7	7	7	7	7	7
Observations	5239	5239	5239	5239	6195	6195	6195	6195
Vector Regression								
EC1. Profit	**-0.0644*****	**0.0094**	**-0.0075**	**-0.0025**	**-0.0224*****	**0.0238*****	**0.0076*****	**0.0153*****
	(0.0102)	(0.0066)	(0.0027)	(0.0031)	(0.0070)	(0.0060)	(0.0026)	(0.0039)
Long Run Return	**1**	**1.9915*****	**2.9748*****	**-0.5162**	**1**	**1.529*****	**1.3840**	**-0.8925*****
		(0.1682)	(0.4272)	(0.37714)		(0.1563)	(0.4238)	(0.2404)
Firms	7	7	7	7	7	7	7	7
Observations	5239	5239	5239	5239	6195	6195	6195	6195

The VEC estimates are clear. The return on investment is significantly higher for high return firms, the earnings on inventory investment being twice as high as for low return firms. The earnings on marketing investments for high return firms is six times that of low return firms, whose earnings on marketing investments is insignificantly different from 0 (return of 1). The investments and expenses of low return firms is cash-flow sensitive, suggesting that they are financially constrained. Investments generate revenue faster for high return firms.

The smaller less profitable firms, use much more of their revenue on expenses (37.5 percent) than the larger more profitable firms do (22.6 percent), and they spend relatively more of their revenue on inventory (48 percent) than their high return counterparts (37 percent). The high return firms spend 2.5 percent more of revenue on marketing than the low return firms. The difference between revenue, normalised to 1, and the cost categories defines a descriptive profit rate:

$$PR_L = 1 - (.480 + .212 + .375) = -.067 \tag{89}$$

$$PR_H = 1 - (.370 + .236 + .226) = .168 \tag{90}$$

for the Low q and High q firms, respectively, as defined by equation 50. The Low q firms are left with a (negative) profit rate of $PR_L = -6.7$ per cent per quarter, while the High q firms have a quarterly profit rate of $PR_H = 16.8$ per cent. The High q firms are more profitable than the Low Q firms, and their average quarterly growth rate is 307K DKK, compared to the Low q firms' -16K. Note that this is *not* Tobin's q, though it describes the same concept: the *average* return on investment.

The high return firms' investments realise almost three times as fast as the investments of the low return firms realise, and this difference is significant at the 99 percent significance level, meaning that null hypothesis 3b is rejected in favour of its alternative hypothesis of significance and inequality:

$$|\hat{\gamma}_{RH} = -.0644| > |\hat{\gamma}_{RL} = -.0224| > 0 \tag{91}$$

Meaning: there exists a convergence of revenue to the profit equilibrium, which is three times faster or stronger for high growth-and profit firms than for low growth-and profit firms.

The mean of the two types of investments, marketing and inventory, is substantially higher for High q firms than Low q firms:

$$\bar{\hat{\beta}}^*_{IH} = (\hat{\beta}^*_{IIH} + \hat{\beta}^*_{MIH})/2 = (.992 + 1.975)/2 = 1.483 \tag{92}$$

$$\bar{\hat{\beta}}^*_{IL} = (\hat{\beta}^*_{IIL} + \hat{\beta}^*_{MIL})/2 = (.529 + .384)/2 = .457 \tag{93}$$

corresponding to an average estimated return on investment $\bar{\hat{\beta}}^*_I$ that is three times as High for the High return firms as for the Low return firms. Equivalently, the High q firms earn more than a full dollar more on investments than Low q firms do. The difference is economically large in magnitude, strongly suggesting that the estimated cumulative partial effect of investments on revenue β_I is *much* larger for the firms pre-classified as High q compared to the Low q firms. I thereby reject the null hypothesis in equation 51 in favour of its alternative hypothesis in equation 52, through the route of the extended model hypotheses in equation 55 and 56, thereby confirming the second main hypothesis of the thesis in equation 66:

H_{II}: **The estimated return on investment is higher for firms with higher growth in revenue and higher excess revenue after cost**

The confirmation of this hypothesis is perhaps the most crucial for the application of the proposed model, since it indicates that the cointegration coefficient β can differentiate between firms based on their return on investment. The core principle in lending is to identify firms that will use the borrowed funds for profitable investments, as this is the best way to ensure that the debt is repaid. The model's efficiency in identifying firms with profitable investment opportunities has not been tested against other strategies, but the fact that it *can* sort the creditworthy firms from those who are not, is the second hypothesis of this master's thesis.

5.3.2 Financial constraints

Before heading on to the third and final main hypothesis of my thesis, I discuss the secondary hypothesis concerning financial constraints. It is identified by firms who react to revenue streams by spending it, which is argued to indicate that these firms have profitable investment opportunities but not sufficient funds to carry them out. I find that it is the firms with Low q h that are the financially constrained ones. The firms with Low q have a cash-flow sensitivity in all three types of costs, ranked by magnitude as:

$$\hat{\gamma}_{IIL} = .0238 > \hat{\gamma}_{EL} = .0153 > \hat{\gamma}_{MIL} = .0076 \tag{94}$$

$$\hat{\gamma}_{IIH} = \hat{\gamma}_{EH} = \hat{\gamma}_{MIH} = 0 \tag{95}$$

I interpret this as small firms, potentially below the minimum efficient scale of their sector, struggling to make ends meet. The estimates suggest that the first place newly available funds go to for financially constrained (Low q) firms is inventory, followed by expenses, and then marketing. Considering the strong negative effect of expenses, the relatively high revenue-share of expenses, and the low level of marketing investments for the low return firms, it may be argued that the low return on investment of these firms is due high "expenses", which includes shipping, logistics, storage, software, delivery, and everything else excluding investments into inventory and marketing.

According to the ranking of cash-flow sensitivities, the low return firms invest in inventory first and marketing last, when they receive a stream of revenue, paying for expenses in-between the two. Therefore, if the (Low q) firm must spend 48 and 37 percent of their revenue on inventory and expenses, there is not much left for marketing. The average low return firm spends 21 percent on marketing, meaning that they lose 6 cent per dollar of revenue. The return on marketing for the Low q firms is positive, but not statistically different from break-even.

5.3.3 Ranking of Investment Choices

Both the return on inventory and the return marketing is higher for the firms with sum of quarterly profit and growth higher than average, but their return on marketing is drastically higher than for the low return firms. The return on both types of investment for both types of firms can be represented in Fisher's (1930) investment choice model, as:

$$\hat{\beta}^*_{MIH} - \hat{\beta}^*_{IIH} = (\hat{\beta}_{MIH} - 1) - (\hat{\beta}_{IIH} - 1) = 1.975 - 992 = 0.987 \tag{96}$$

$$\hat{\beta}^*_{IIL} - \hat{\beta}^*_{MIL} = (\hat{\beta}_{MIL} - 1) - (\hat{\beta}_{IIL} - 1) = .384 - .529 = -.145 \tag{97}$$

Meaning: the return on marketing is twice as high as the return on inventory for high return firms, while low profit-and growth have a relatively higher return on inventory than on marketing. The former is equivalent to almost a full dollar in profit, while the difference in the latter is less than 15 cents. Rearranging the expressions, we can compare the difference in returns between firms:

$$\hat{\beta}^*_{IIH} - \hat{\beta}^*_{IIL} = (\hat{\beta}_{IIH} - 1) - (\hat{\beta}_{IIL} - 1) = .992 - .529 = .463 \tag{98}$$

$$\hat{\beta}^*_{MIH} - \hat{\beta}^*_{MIL} = (\hat{\beta}_{MIH} - 1) - (\hat{\beta}_{MIL} - 1) = 1.975 - .384 = 1.589 \tag{99}$$

Meaning: the high return firms are estimated to earn more than a dollar and a half more per dollar spent on marketing than low return firms, and almost 50 cents more on inventory. This fits well with the understanding of the ecommerce sector, that it is through marketing efficiency that ecommerce firms distinguish themselves, but also with the reasoning that the effect of appealing products is captured by ease of marketing. The return on inventory is also higher for high return firms, but the difference is not as substantial. This suggests that where the firm sources the goods and what price they pay, is *not* as important as the efficiency of the marketing or the appeal of the product. Implementing a simplified version of Fisher's investment choice model, I hypothesised that the return on inventory was equal across High and Low q firms, which is not the case, but it is true that:

H_{III}**: The return on marketing is drastically higher for high return firms, and their return on inventory is higher than for low return firms too, but not as substantially**

The cointegration coefficients estimated in the VEC model sub-matrix (Table 8), can be ranked in the simplified investment choice model:

$$\hat{\beta}_{MIH} > \hat{\beta}_{IIH} > \hat{\beta}_{IIL} > \hat{\beta}_{MIL} > 1 > \hat{\beta}_{EH} > \hat{\beta}_{EL} = 0 \tag{100}$$

$$2.98 > 1.99 > 1.53 > 1.38 = 1 > -.516 = 0 > -.893 \tag{101}$$

where equation 100 represents the hypothesised rank and equation 101 ranks the estimated parameters from table 8. The return on marketing is a partial factor determining the success of ecommerce firms, but investments into marketing are also the most risky. Unlike investments into inventory, investments in marketing may generate no revenue at all. With inventory, losses can always be cut by lowering the selling price. If a marketing investment fails, there is no insurance on losses. In fact, a dollar spent on a completely failed marketing effort will lose more than a dollar due to the opportunity cost incurred. The return on a marketing investment a standard error σ below the mean for the low return firm is negative, meaning that there is a rough 1 in 6 chance that this type of firm will lose money by investing in marketing.

5.3.4 Summary of Results

The model has confirmed the main hypotheses I constructed to test the validity of the interpretation of the cointegration coefficient β from the VEC model as the (long run) return on investment. The findings of this study are based on data made available to non-bank financial companies by open banking, and they contribute to the innovation in financial information processing that the initiative encourages. The insight the model gives can be found by anyone with business bank account data and a statistics software package. The estimates yielded by the model can be used to rank investments between firms and within firms between investment types, by their return.

The model I have developed is a general model, but it has several limitations, as I do not: (i) build a selection framework that can specify the most efficient formulation of the model, (ii) extend the model to include more than three categories of costs, (iii) estimate the different realisation times on investments through multiple equilibria, (iv) incorporate other data sources, (v) estimate the *marginal* return on investment, (vi) incorporate risk on return on investments into the investment choice model, nor any interest rate discount, nor do I (vii) compare with other methods of identifying the return on investment.

To summarise my results: (i) there exists a cointegrating relationship between revenue and costs, where the stationary residual is the profit in a special case, (ii) the cointegration coefficient $\hat{\beta}$ shows a higher cumulative long run effect on revenue by investment than by expenses, (iii) the long run return of investments $\hat{\beta}_I$ is higher for firms with higher growth-and profit rates than for low profit-and growth firms, (iv) firms with low return on investment are also cash-flow sensitive, indicating that they are financially constrained, in the sense that they are struggling to make ends meet, (v) high return firms have disproportionately higher return on marketing $\hat{\beta}_{IM}$ than inventory compared to low return firms and the return on inventory, and (vi) investments in marketing have a higher risk $\hat{\sigma}$ than investments in inventory.

6 Discussion and Conclusion

This master thesis has been methodological at heart, but the data I have used are part of the bigger picture of financial development. The combination of data and method I employed has not been used before, to the best of my knowledge, and the goal of my thesis was to show its validity. I have demonstrated that the econometric method that I have constructed can estimate the return on investment, both within and between firms. The model consistently yields higher estimates for investments with higher return, and higher estimates for firms with higher return on investment. To test the model's robustness, I have estimated various variations of the model, showing that it can be applied to lower frequency data too. I have estimated the long run return on investment with the cointegration coefficient β, from the restricted sub-matrix of the VEC model. The model consistently estimates higher return on investment than on expenses, and estimates higher return on investment for firms with higher return on investment, and higher estimates for investments alternatives with higher return on investment.

Evidence suggests that the reduced model is super-efficient, meaning that it can be applied to data aggregated to a lesser frequency than the daily (moving average) observations I implement. The monetary formulation of the model is the easiest to interpret, as it gives a "return on a dollar spent" interpretation, where the "return" is the excess revenue after cost of an investment, meaning: the profit. I found that firms that are pre-classified as high profit-and growth firms have a return on investment that is *more than a full dollar per dollar spent more* than the return on investment for the pre-classified low return firms. I found that return on investment was about a dollar higher than the return on expenses. Furthermore, I found that the successful ecommerce firms had drastically higher return on marketing than those who were not successful, though the groups had similar return on inventory.

The theoretical framework I have based this thesis on is three-fold. First and foremost, the thesis is fundamentally based on Johansen's combination of the concepts of cointegrations with the vector regression framework, which enables the estimation of the cumulative long run effect of investment on revenue. Second, the concept of the return on investment is based on the work of Mueller, from whom I borrow the concepts of the long run rate of profit and the marginal return on investment. The major difference from my model to most others in the investment literature, is that I incorporate Fisher's investment alternatives within firms, as opposed to the literature following Keynes' simplification where each firm has a general return on investment, such as the literature using Tobin's Q.

Tobin's Q is often used as a controlling variable, specifically in the literature attempting to identify financially constrained firms through their cash-flow sensitivity. The discussion in this field, between Fazzari, Hubbard, and Petersen on one side and Kaplan and Zingales on the other, has been on whether Tobin's Q can control for the fact that increased streams of revenue signal higher return on investment, in addition to providing the firm with more available funds. I do not engage in this argument, extending the VAR framework of Himmelberg and Gilchrist instead. I interpret the convergence of costs to the profit-equilibrium between revenue and cost, as the cash-flow sensitivity, arguing that it is profits, not revenue itself, that constitute available funds.

Open banking aims to increase capital market efficiency by decreasing the informational barriers to entry in finance. The initiative's strategy is to enable firms to share the data withheld in their bank accounts with other finance companies, thereby incentivising innovation in the processing of that data. Fintech lenders use open banking in the screening of potential borrowers, applying data-science strategies to identify credit risk and investment opportunities. Competition in lending is different, as lenders can always increase their supply of credit by increasing their credit risk, meaning that there is a fine balance between output and risk. Fintech lenders are bringing innovation, competition, and wider access to credit, through their edge in the processing of hard information over incumbent banks in the lending business. The general model I present can be implemented by any bank, financial institution, or fintech lender, in the evaluation of potential borrower's creditworthiness. The insight brought by the application of statistical models to identify investment opportunities and credit risk, is the type of innovation that the open banking initiative encourages.

The general model I proposed in this master thesis gives insight to several concepts for measuring business success, but it is only a general model. Therefore, there are several limitations of this study, leaving many interesting future research areas open. I found that the cointegration coefficient β_I estimates the long run return on investment both within and between firms, but I do not estimate the time it takes for investments to turn over into revenue for separate types of investments. I found that the return on marketing is drastically higher for successful firms, but also that the risk associated with investments into marketing is highest among types of investments. I do not incorporate any risk measures in the investment choice model, per Fischer (1933). A limitation of my study is the heavy simplification of Fischer's model, but the Johansen's VEC model (1988) can *expand* on Fischer's model by incorporating risk measures and partial effects (Frisch, 1934). My estimation of the return on investment is most similar to Mueller's long run rate of profit (1988), but I do not incorporate his *marginal* interpretation of the return on investment, though a Bayesian formulation of the model could find the long run return on the *next* investment.

7 References

Abrigo, M. R. M., & Love, I. (2016). Estimation of Panel Vector Autoregression in Stata. *The Stata Journal*, 16(3), 778–804.

Akerlof, GA. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3), 488-500

Alchian, A. (1955). The Rate of Interest, Fischer's Rate of Return over Costs and Keynes' Internal Rate of Return. *The American Economic Review*, 45(5), 938- 943.

Angrist, J. D. & Pischke, J.-S. (2009). Making Regression Make Sense. *In Mostly Harmless Econometrics: An Empiricist's Companion* (pp. 27–111). Princeton University Press.

Babina, T., Buchak, G. & Gornall, W. (2021). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. *Stanford University Graduate School of Business Research Paper, SSRN.*

Bayluk, T., Berger, A. & Hackney, J. (2022). What is Fuelling FinTech lending? The Role of Banking Market Structure. *Stanford University Graduate School of Business Research Paper, SSRN.*

Beck, T., Demirguc-Kunt., Laeven, L. & Levine, R., (2008). Finance, Firm Size, and Growth. *Journal of Money, Credit and Banking*, 40(7), 1379-1405.

Beaumont, P. Tang, H. & Vansteenberghe, E. (2022). The Role of FinTech in Small Business Lending. *Proceedings of the EUROFIDAI-ESSEC Paris December Finance Meeting in 2022.*

Berger, A. & Black, L. (2011). Bank size, lending technologies, and small business finance. *Journal of Banking and Finance*, 35(3), 724-735.

Berger, A., Molyneux, P. & Wilson, J. (2020). Banks and the real economy: An assessment of the research. *Journal of Corporate Finance*, 62(4), 101513

Berger, A. & Udell, G. (1995). Relationship Lending and Lines of Credit in Small Firm Finance. *The Journal of Business*, 68(3), 351-381.

Berger, A. & Udell, G. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking and Finance*, 22(6-8), 613-673.

Berger, A. & Udell, G. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal*, 112(477), 32-53.

Bernanke, BS. (2018). The Real Effects of Disrupted Credit: Evidence from the Global Financial Crisis. *Brookings Papers on Economic Activity*, 49(2), 251-342.

Bernanke, BS. (2020). The New Tools of Monetary Policy. *American Economic Review*, 110(4), 943-983.

Bernanke, BS. & Gertler, M. (1995). Inside the Black Box: The Credit Channel of Monetary Policy Transmission. *Journal of Economic Perspective*, 9(4), 27-48.

Bernanke, BS., Gertler, M. & Gilchrist, S. (1999). "The financial accelerator in a quantitative business cycle framework," *Handbook of Macroeconomics*, in: J. B. Taylor & M. Woodford (ed.), edition 1, volume 1, chapter 21, pp. 1341-1393, Elsevier.

Boyd, J. & Nicolo, G. (2005). The Theory of Bank Risk Taking and Competition Revisited. *The Journal of Finance*, 60(3), 1329-1343.

Brooks, C. (2019). Modelling Long- Run Relationships in Finance. *Introductory Econometrics for Finance*, 4th edition, Chapter 8. Cambridge University Press.

Cheng, M. & Qu, Y. (2020). Does bank Fintech reduce credit risk? Evidence from China. *Pacific-Basin Finance Journal*, 63, DOI:10.1016/j.pacfin.2020.101398

Christopoulos, D. & Tsionas, E. (2003). Financial development and economic growth: evidence from unit root and cointegration tests. *Journal of Development* Economics, 73 (1), 55-74.

Conway, T. & Hemphill, T. (2019). Growth hacking as an approach to producing growth amongst UK technology start-ups: an evaluation. *Journal of Research in Marketing and Entrepreneurship*, 21(2), 163-179.

Cornelli, G., Frost, J. Gambacorta, L. & Jagtiani, J. (2022). The Impact of Fintech Lending on Credit Access for U.S. *Small Businesses. Federal Reserve Bank Philadelphia, Working Paper*, 22-114.

Cortes, KR., Demyanyk, Y., Li, L., Loutskina, E. & Strahan, P. E (2020). Stress tests and small business lending. *Journal of Financial Economic*, 136(1), 260-279.

Demetriades, P. & Hussein, K. (1996). Does financial development cause economic growth? Time-series evidence from 16 countries. *Journal of Development Economics*, 51 (2), 387-411.

Diamond, D. (1984). Financial Intermediation and Delegated Monitoring. *The Review of Economic Studies*, 51(3), 393-414.

Diamond, D. & Dybvig, PH. 1983. Bank runs, Deposit Insurance, and Liquidity. *The Journal of Political Economy*, 91(3), 401-419.

Diamond, D. & Rajan, R. (1999). Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking. *National Bureau of Economic Research*. Working Paper 7430.

Engle, RF. & Granger, CWJ. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251-276.

Eklund, J. (2010). Q-theory of investment and earnings retentions – evidence from Scandinavia. *Empirical Economics*, 39(3), 793-813.

Eklund, J. (2013). Theories of investment: a theoretical review with empirical applications. *Swedish Entrepreneurship Forum*, 22,1-30.

Engsted, T. & Johansen, S. (1997). Granger's Representation Theorem and Multicointegration. *European University Institute*, Working Paper No. 97/15.

Ericson, R. & Pakes, A. (1998). Empirical Implications of Alternative Models of Firm Dynamics. *Journal of Economic Theory*, 79(1), 1-45.

Fama, EF. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383-417.

Fazzari, S., Hubbard, RG. & Petersen, B. (1987). Financing Constraints and Corporate Investment. *National Bureau of Economic Research.* Working Paper No.2387.

Fazzari, S., Hubbard, RG. & Petersen, B. (2000). Investment-Cash Flow Sensitivities are useful: A comment on Kaplan and Zingales. *The Quarterly Journal of Economics*, 15(2), 696-705.

Felix, J. (2016). Revenue-Based Financing. *Universidade Catolica Portuguesa*, Business and Economics.

Fisher, I. (1930). The Investment Opportunity Principles. *The Theory of Interest.* Part II, Chapter VII. New York: The Macmillan Company.

Frisch, R. & Waugh, F. (1933) Partial Time Regressions as Compared with Individual Trends. *Econometrica*, 1(4), 387-401.

Gibrat, R. (1931). *Les inegalite ´economiques.* Librairie du Recueil Sirey, Paris.

Gilchrist, S. & Himmelberg, CP. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics*, 35(3), 541-572.

Gilchrist, S. & Himmelberg, CP. (1998). Investment: Fundamental and Finance. *NBER Macroeconomics Annual*, 13, 223-262.

Gopal, M. & Schnabl, P. (2022). The Rise of Finance Companies and FinTech Lenders in Small Business Lending. *The Review of Financial Studies*, 35(11), 3859-4901.

Gugler, K., Mueller, DC. & Yurtoglu, B.B. 2004. Marginal q, Tobin's q, Cash Flow, and Investment. *Southern Economic Journal*, 70(3), 512-531.

Hayashi, F. (1982). Tobin's Marginal q and Average q: A Neoclassical Interpretation. *Econometrica*, 50(1), 213-224

Hirsch, S. (2018). Successful in the long run: A meta-regression analysis of persistent firm profits. *Journal of Economic Surveys*, 32(1), 23-49

Hoffman, D. & Novak, T. (1997). A New Marketing Paradigm for Electronic Commerce. *The Information Society*, 13(1), 43-54.

Idsø, J., Opstad, L. & Valenta, R. (2022). The dynamics of the profitability and growth of restaurants; the case of Norway. *Economies*, 10(2), 53. https://doi.org/10.3390/economies10020053

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231-254.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6), 1551-1580.

Johansen, S. (2000). Modelling of cointegration in the vector autoregressive model. *Economic Modelling*, 17(3), 359-373.

Johansen, S. (2004). Cointegration: an overview. Copenhagen. *University of Copenhagen*, Department of Applied Mathematics and Statistics.

Johansen, S. & Engsted, T. (1997). Granger's Representation Theorem and Multicointegration. *European University Institute Working Paper*, No.97/15.

Joshi, A. & Hanssens, D. (2010). The Direct and Indirect Effects of Advertising Spending on Firm Value. *Journal of Marketing*, 74(1), 20-33.

Jorgenson, D. (1963). Capital Theory and Investment Behavior. *The American Economic Review*, 53(2), 247-259.

Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, 50(3), 649-670.

Kaplan, S. & Zingales, L. (1997). Do Investment-cash flow sensitivities provide useful measures of Financing Constraints? *The Quarterly Journal of Economics*,112(1), 170-215.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money.* London: Macmillan.

King, R. & Levine, R. (1993). Finance and Growth: Schumpeter Might be Right. *The Quarterly Journal of Economics*, 108(3),717-737.

Levine, R. (1997). Financial Development and Economic Growth: Views and Agenda. *Journal of Economic Literature*, 35(2), 688-726.

Levine, R. (2004). Finance and Growth: Theory and Evidence. *National Bureau of Economic Research*. Working Paper No. 10766

Liberti, J. & Petersen, M. (2019). Information: Hard and Soft. *Review of Corporate Finance Studies*, 8(1), 1-41.

Love, I. & Peria, M. (2014). How Bank Competition Affects Firms' Access to Finance. *The World Bank Economic Review*, 29(3), 413-448.

Love, I. & Zicchino, L. (2006). Financial development and dynamic investment behavior: Evidence from panel VAR. *The Quarterly Review of Economics and Finance*, 46(2), 190-210.

Lu, Z., Wu, J., Li, H. & Nguyen, D. (2022). Local Bank, Digital Financial Inclusion and SME Financing Constraints: Empirical Evidence from China. *Emerging Markets Finance and Trade*, 58(6), 1712-1725.

Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Maggio, M., Ratnadiwakara, D. & Carmichael, D. (2022). Invisible Primes: Fintech Lending with Alternative Data. *National Bureau of Economic Research*. Working Paper No. 29840.

Maggio, M. & Yao, V. (2019). Fintech Borrowers: Lax Screening or Cream-Skimming? *The Review of Financial Studies*, 34(10), 4565-4618.

Mansfield, E. (1962). Entry, Gibrat's law, innovation, and the growth of firms. *The American Economic Review*, 52(5), 1023-1051.

Marris, R. (1964). T*he Economic Theory of Managerial Capitalism.* Free Press of Glencoe.

Mises, L. (1912). *The Theory of Money and Credit*. Ludwig von Mises Institute. Yale University Press, 2009 edition. Part 2 and 3.

Miller, M. & Mogdiliani, F. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, 48(3), 261-297

Mueller, D.C. (1986). *Profits in the Long Run.* Cambridge: Cambridge University Press.

Mueller, DC. & Reardon, EA. (1993). Rates of Return on Corporate Investment. *Southern Economic Journal*, 60(2), 430-453.

Myers, SC. (1984). Capital Structure Puzzle. *National Bureau of Economic Research*, Working paper No.1393.

Myers, SC. (2001). Capital Structure. *Journal of Economic Perspectives*, 15(2), 81-102.

Navaretti, GB. Calzolari, G. & Pozzolo, AF. (2017). FinTech and Banks: Friends and Foes. *European Economy: Banks, Regulation, and The Real Sector.* 3(2), 9-30.

North, D. (1982). *Structure and Change in Economic History.* W. W. Norton and Company.

Obstfeld, M. & Rogoff, K. (1996). *Foundations of International Macroeconomics.* Massachusetts Institute of Technology. Chapter 1. Intertemporal Trade and the Current Account Balance.

Opstad, L. & Valenta, R. (2022a). The Long-Run Equilibrium of Industry Population and Bottom Lines: Analysing the Near-Perfect Restaurant Industry. *Journal of Applied Business and Economics*, 24(3), 194-210

Opstad, L. & Valenta, R. (2022b) Non-Linear Economies of Scale in Food Production: The Case of Norwegian Bakeries. *International Journal of Applied Economics, Finance and Accounting*, 14(2), 182-193.

Petersen, M. & Rajan, R. (1994). The Benefits of Lending Relationships: Evidence from Small Business Data. *The Journal of Finance*, 49(1), 3-37

Petersen, M. & Rajan, R. (2000). Does distance still matter? The information revolution in small business lending. *National Bureau of Economic Research*, Working Paper 7685.

Rajan, R. (1994). Why Bank Credit Policies Fluctuate: A Theory and Some Evidence. *The Quarterly Journal of Economics,* 109(2), 339-441.

Rajan, R. & Zingales, L. (1996). Financial Dependence and Growth. *National Bureau of Economic Research*, Working Paper 5758.

Rice, T. & Strahan, P. (2010). Does Credit Competition Affect Small-Firm Finance? *The Journal of Finance*, 65(3), 861-888.

Rousseau, P. & Wachtel, P. (1998). Financial Intermediation and Economic Performance: Historical Evidence from Five Industrialized Countries. *Journal of Money, Credit, and Banking*, 30(4), 657-678.

Schumpeter, JA. (1911). *The Theory of Economic Development.* Harvard University Press. Cambridge.

Schumpeter, JA. (1939). *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process.* McGraw-Hill Book Company, Inc. Volume I, pp.72-219.

Sheng, T. (2021). The effect of fintech on banks' credit provision to SMEs: Evidence from China. *Finance Research Letters*, 39, https://doi.org/10.1016/j.frl.2020.101558

Sims, CA. (1980). Macro-Economics and Reality. *Econometrica*, 48(1), 1-48.

Solow, R. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, 39(3), 312-320.

Stigler, S. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* The Belknap press of Harvard University Press, Part One, The development of Mathematical Statistics in Astronomy and Geodesy before 1827.

Troisi, O., Maione, G., Grimaldi, M. & Loia, F. (2020). Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management*, 90, 538-557.

Tobin, J. (1969). A General Equilibrium Approach To Monetary Theory. *Journal of Money, Credit and Banking*, 1(1), 15-29.

Valenta, R., Idsø, J., Opstad, L. (2021). Evidence of a Threshold Size for Norwegian Campsites and Its Dynamic Growth Process Implications—Does Gibrat's Law Hold?. *Economies*, 9(4). https://doi.org/10.3390/economies9040175

Vives, X. (2019). Competition and stability in modern banking: A post-crisis perspective. International *Journal of Industrial Organization*, 64, 55-69.

Werner, R. (1997). Towards a new monetary paradigm: A quantity theorem of disaggregated credit, with evidence from Japan. *Credit and Capital Markets–Kredit und Kapital*, 30(2), 276- 309

Werner, R. (2016). A lost century in economics: Three theories of banking and the conclusive evidence. *International Review of Financial Analysis*, 46, 361-379

Wooldridge, J. (2013). Advanced Panel Data Methods. *Introductory Econometrics: A Modern Approach*, 6th edition, Chapter 14. Michigan State University.

Xu, Z. (2000). Financial Development, Investment, and Economic Growth. *Economic Inquiry*, 38(2), 331-344.

NTNU

Kunnskap for en bedre verden