The Use of Alternative Data by Fintechs to Provide Access to Credit for SMEs in South Africa

Mandlenkosi M. Khupe

2410076

Supervisor: Dr. J. Msimango-Galawe

A research report submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business

Johannesburg, 2024

ABSTRACT

In South Africa, access to finance remains one of the major contributors to SME failures. The International Finance Corporation estimates that the credit gap for MSMEs in South Africa is R510bn. Traditional lenders often impose stringent requirements and avoid certain segments due to high information asymmetry costs. This factor has led to the advent of fintechs that seek innovative ways to provide unsecured funding. This study examines the role of fintechs in leveraging alternative data to provide credit access to SMEs. The study adopts a standard qualitative approach, with the primary methodology used to collect data being semi-structured interviews. Thirteen participants with varied exposure to alternative data in the fintech sector were interviewed. The findings of the study reveal the transformative nature of alternative data, as it significantly reduces information asymmetry and enhances credit provision. Furthermore, the study concludes that alternative data should be used to supplement rather than replace traditional data. The research highlights the fact that while fintechs favour hard data over soft data for more objective decision-making, the use of alternative data in South Africa is still emerging. The study further reveals that soft data are less weighted by fintechs compared to hard data which enable fintechs to make objective decisions. The conclusion reached is the need for collaboration among fintechs, regulators, and other stakeholders in the financial ecosystem to foster and promote data sharing practices in the alternative data space. This collaborative approach is essential for addressing the prevailing issues of information asymmetry and for further enhancing SME access to finance.

KEYWORDS: Lending, SME, SMME, Alternative data, Credit, South Africa Credit, Fintech, Information asymmetry, Traditional data, Signalling, Screening, Adverse Selection

DECLARATION

I, Mandlenkosi M Khupe, declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Name: Mandlenkosi M Khupe Signature:

Signed at CAPE TOWN ……………………………………………………

On the 17 TH ……………………………. day of FEBRUARY ………………………… 20 24

DEDICATION

While writing this paper – I navigated through many highs and endured the deepest lows when my family was faced with a matter close to our hearts. As they say, the rest is history. We are here now!!

To my extended family, both Khupe and Makgoka your quiet support has not gone unnoticed. Your understanding and absence of guilt for the times I missed family gatherings have been comforting. It gave me space to focus. Thank you.

To my children, with one sitting on my lap as I type this, Papa has not always been available between balancing work and studying. While I have raised the bar with this milestone, I see it merely as a starting point for you. I will be your biggest cheerleader as you surpass the milestones and get our family to levels, we never imagined possible.

A special tribute to my mom and my sister, I dedicate this research paper to you. You were the ones that planted the initial seeds. Your countless nights of making me coffee mattered Mama, and Belinda for setting me up to do my undergrad. I want to tell you, the tree you planted more than a decade ago continues to bear fruit.

To my wife, Moyagabo, you have been a true soldier. You have not only shouldered a lot of family responsibilities, but allowed me the space to focus, all while caring for a newborn, come toddler. This accomplishment is as much yours, as it is mine. You have put into this research paper just as much work as I did. Thank you.

God Bless!!

ACKNOWLEDGEMENTS

"It takes a community to build a leader".

I have had a community support me through this – from my company, the participants, my supervisor and many more that make an academic ecosystem work.

I would like to extend my heartfelt gratitude to my company, Retail Capital, a division of TymeBank. Their financial support, flexibility and space provided for me to focus have been pivotal. I am especially grateful to my boss, whose encouragement and constant support has been invaluable.

This research paper would not have been possible without the insights provided by the participants. Their willingness to set time aside, share their deep insights and provide feedback is much appreciated.

To my supervisor, Dr. Jabulile Galawe (affectionately known as Dr J), you challenged me so much. Your comments were concise, but they could result in me deleting or reimagining pages and pages of work. I honestly could not have done it without you. I think I have been a demanding student and I appreciate your patience and your quick responses. To this day, I am still wondering how it is that you read my work faster than I could read my own work. Many thanks. Please continue shining your light with students that come after me. I too, will remember to pay it forward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

API:	Application Program Interface
BASA:	Banking Association South Africa
FMCG:	Fast Moving Consumer Goods
GDP:	Gross Domestic Product
IFC:	International Finance Corporation
MSME:	Micro, Small and Medium Enterprise
OECD:	Organisation for Economic Co-operation and Development
POPIA:	Protection of Personal Information Act
SARB:	South African Reserve Bank
SARS:	South African Revenue Service
SME:	Small Medium Enterprise

CHAPTER 1. INTRODUCTION

1.1 Statement of purpose

The purpose of this qualitative research report was to explore the use of alternative data by fintechs to provide access to credit for SMEs in South Africa.

1.2 Background of the study

Small and medium-sized enterprises (SMEs) play a vital role in the global economy, of significance to both gross domestic product (40%) and employment (50%) (World Bank, 2019). In South Africa, SMEs account for 34% of the country's gross domestic product (GDP) and employ a substantial labour force, ranging from 50% to 60% (International Finance Corporation, 2018). It is evident that ensuring the sustainability of SMEs is critical to promoting economic growth, job creation, stable tax revenues, inclusiveness and stable salaries (Charaia et al., 2021).

Defining SMEs, however, is complex as the classifications can differ across countries and industries. For instance, The Small Business Institute, uses size variables to categorise SMEs based on the number of individuals they employ (Stefan & Visser, 2021). According to Stefan and Visser (2021), micro enterprises employ 0-10 individuals, small businesses employ 11-50 individuals, medium-sized businesses employ 51-200 individuals, and large corporates employ more than 200 individuals. Remarkably, the spectrum of SMEs extends from medium-sized enterprises employing over 100 people to informal sole proprietors (Kalitanyi, 2019). The definitions of SARS, SARB, and BASA for SMEs differ, highlighting the inconsistency in SME classification (International Finance Corporation, 2021). Given this diversity in SME definitions, encompassing factors such as sector, size, age, asset value and profitability, this research paper adopts the definition provided by the South African National Act for Small Business of 1996, amended in 2003 and 2004, that defines small enterprises as having fewer than 50 employees and less than R10m annual sales revenue depending on industry. These businesses make up a large majority of small businesses found in South Africa. These businesses are represented by sectors such as health and beauty, general retail, backyard manufacturing and services, food and beverage, and construction (Kalitanyi, 2019).

Access to finance is a key determinant of SME survival, growth and productivity (Błach et al., 2020; Łasak, 2022; Msomi & Olarewaju, 2021). The likelihood of SMEs obtaining credit from traditional financiers, such as banks, is low (World Bank, 2019). These traditional institutions have typically perceived SMEs as risky or expensive to serve, largely due to information asymmetry, collateral deficiencies, and inadequate financial records (International Finance Corporation, 2018). There is recognition in the SME credit market that information access is imperfect and obtaining information is costly (Stiglitz, 2000). Lenders and borrowers only share information with each other to the extent that it creates incremental opportunities for their respective businesses (Yan et al., 2015). In response, fintech companies have introduced transformative innovations, expanding access to credit for marginalised businesses. These fintechs have shifted the lending landscape by utilising alternative data to evaluate creditworthiness and provide financing to SMEs.

Conventional banks often adhere to narrow interest rate ranges and binary credit approvals for approvals or declines, as opposed to finding the acceptable rate at which to lend (Mills, 2018). SMEs argue that when they submit credit applications to conventional lenders, they (lenders) often neglect to provide feedback on declined credit applications (Asah et al., 2020). In South Africa, the limitations of traditional banking systems in serving SMEs were notably exposed during the COVID-19 pandemic. Despite government efforts to implement a coronavirus loan guarantee scheme to mitigate losses, banks continued to rely on traditional lending methodologies and lacked cost effective distribution mechanisms, leading to low approval rates. According to International Finance Corporation (2018), at least 40% of SMEs have an unmet funding need in developing economies like South Africa. Due to the difficulties in accessing funding for SMEs, there is a massive gap that fintechs are bridging by utilising alternative data to provide access to credit for SMEs.

The global landscape has witnessed a surge in lending fintechs that utilise numerical data to automate credit decisions (Liberti & Petersen, 2019). South Africa has also experienced a similar trend, though to a lesser extent. Fintechs are defined as “advanced technology firms that have the potential to transform the provision of financial services spurring the development of new business models, applications, and whose products and services are directly applicable in the delivery of financial services” (Genesis Analytics, 2019, p. 2; International Monetary Fund, 2018, p. 14). In the credit sector, fintechs facilitate and leverage different forms of alternative data to provide funding to SMEs through internet-based, cloud-based, or app-based platforms (Genesis Analytics, 2019). The definition of alternative data is characterised as data collected from “non-traditional sources and not typically included in the traditional credit process. This alternative data may include unstructured and structured data” (Zou et al., 2020, p. 1). Other authors have defined alternative data as “proxy metrics or information originating from unofficial or noncompany sources” (Constable, 2019, p. 1). The International Finance Corporation (2021) defines alternative data more broadly, as any data that are not traditionally housed within a bureau environment.

Structured alternative SME data include e-commerce sales, platform sales, purchasing data, logistics and shipping data, online accounting data, SME billing and payment data, and inventory tracking data (Owens & Wilhelm, 2017). Unstructured data that hold potential utility encompass telco data, email data, social media data, and bank statement data.

1.3 Research problem

This research report addresses the challenges that SMEs in South Africa are faced with in accessing credit from traditional lenders due to information asymmetry. Access to credit for SMEs is crucial for the sustainability of the South African economy (credit and capital are used as synonyms in this document). According to Mills (2018), credit access is the lifeblood of SMEs, as it enables them to start, sustain, and expand their businesses. Demand for credit by SMEs is vast, however they struggle to raise capital from the traditional lenders as they tend to be more informationally opaque (Yan et al., 2015).

Traditional banks have been the primary data source for SME information regarding creditworthiness by reporting or providing credit data to bureaus (Liberti & Petersen, 2019). When assessing credit, banks primarily rely on conventional methods, which often require collateral, high turnover thresholds, audited financial accounts, monthly management accounts, an existing banking relationship, a good credit score, a proven management track record and equity contribution (Asah et al., 2020). Unfortunately, most SMEs do not have this information. This information is particularly relevant for a market like South Africa.

Globally, advances such as big data analysis are driving change in the lending industry. The study provides insight into the potential role alternative data present as leveraged by lending fintechs to reduce this problem. Fintechs are optimising collections, presentation, and evaluation of information for SMEs with thin credit files. SME lenders should be looking at an aggregate of different sources of data to get insight into the likelihood of repayment by SMEs (Asah et al., 2020). Asah et al. (2020) further highlight the potential role fintechs can play as pioneers of these financial breakthroughs when partnering with traditional banks.

1.4 Research objectives

The primary objective of the study is to explore the use of alternative data by fintechs to provide access to credit for SMEs in South Africa. To explore this, the study was broken into sub-objectives:

Sub-objective 1:

To explore how fintechs are transforming the SME credit landscape in South Africa through the use of alternative data.

This sub-objective is broken down into the following elements:

Element 1.1: To assess the types of alternative data being utilised by fintechs in credit scoring.

Element 1.2: To understand the extent to which alternative data complements traditional credit data in evaluating an SME’s creditworthiness.

Sub-objective 2:

To examine how fintechs address information asymmetry in the credit assessment process.

This sub-objective is broken down into the following elements:

Element 2.1: To evaluate the role of subjectivity in providing credit to SMEs

Element 2.2: To evaluate which data points are reliable to be an indicator of an SMEs' future sustainability

Element 2.3: To understand how specific data points are used to mitigate issues of information asymmetry

1.5 Research questions

There were two main research questions used in this study to explore uses of alternative data to provide credit access to SMEs in South Africa.

Research Question 1 (RQ1):

How are fintechs transforming the SME credit landscape in South Africa by using alternative data?

This research question has been sub-divided into two questions:

RQ1.1: What types of alternative data points do fintechs commonly rely on when credit scoring SMEs?
RQ1.2: How do alternative data complement or substitute traditional credit data in evaluating an SME's creditworthiness?

Research Question 2 (RQ2):

How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?

This research question has been sub-divided into two questions:

RQ2.1: How do fintechs incorporate subjective measures into their credit algorithms/models, transforming soft data into decision-making criteria?

RQ2.2: What data points do fintechs depend on as indicators of SMEs' future sustainability?

RQ2.3: What kinds of alternative data are fintechs utilising to limit issues of information asymmetry in the credit assessment process?

1.6 Rationale

The rationale for this research is grounded in the substantial challenges faced by small and medium-sized enterprises (SMEs) in South Africa, particularly those with an annual turnover of less than R10 million (Kalitanyi, 2019). These SMEs encounter significant hurdles in securing access to traditional banking lines of credit, a problem that is particularly acute for micro-enterprises. This paper provides insights regarding the importance of SMEs being more deliberate about their digital personas for funding access.

Traditional lenders work with hard and non-codifiable information such as borrower balance sheets and collateral (Fasano & Cappa, 2022). However, the South African SME market presents a unique scenario characterized by "thin files," where such critical information is frequently unavailable for the majority of SMEs. Consequently, this research contributes to the ongoing debate surrounding information asymmetry challenges faced by SMEs when seeking credit (Fasano & Cappa, 2022). Lenders, both fintech and conventional, can gain valuable insights into the utility of alternative data and understand how contributions from all players within the ecosystem can benefit everyone involved while addressing issues of information asymmetry.

Furthermore, this research delves into the transformative role played by lending fintechs in reshaping the credit landscape for SMEs through the utilisation of alternative data. Alternative data serve as a catalyst for the growth of SMEs and have the potential to reduce the overreliance on traditional banks as the primary lenders to SMEs. By leveraging such data, fintechs are shifting their decision-making processes from predominantly soft information to hard data-driven approaches (Fasano & Cappa, 2022). For example in the minibus taxi industry, SA Taxi provides access to credit using telemetry that tracks capacity and demand on key routes, TPN uses rental payment data, Credolab uses smartphone metadata to generate credit scores and both Yoco and iKhokha use mobile point of sale data (International Finance Corporation, 2021). Addendum uses SME invoices with corporates to offer supply chain finance. In this context, this paper contributes to the expanding body of knowledge on how alternative data can serve as both a substitute and a complement to mitigate the pervasive issue of information asymmetry in SME lending.

De la Torre et al. (2010) argue that banks in developed and developing countries view SMEs as strategic and have over time begun expanding operations to aggressively service this market. In the South African context, there is a noticeable absence of evidence indicating that domestic banks are aggressively servicing the SME market (Asah et al., 2020). This research highlights the transformative impact of fintechs in harnessing big data at a low cost and disrupting the credit landscape. Globally, traditional banks are recognising the pivotal role played by fintechs and have increasingly started forging partnerships to provide embedded funding solutions. Certain South African banks have begun to explore innovative approaches, such as utilising alternative data for credit decisions or forming strategic alliances with fintechs. As an example, GAP access by Nedbank and Pay As You Trade by Capitec are products that leverage mobile point of sale historic data to provide loans to SMEs. Other banks such as Standard Bank have collaborated with fintechs to provide a similar offering. Additionally, the emergence of disruptive neobanks like TymeBank, which have made strategic acquisitions of lending fintechs (acquired Retail Capital), demonstrates the evolving dynamics within the financial sector.

In summary, this research is motivated by the need to address the financial challenges faced by South African SMEs, particularly those with limited access to traditional credit lines. It seeks to contribute to the discourse on mitigating information asymmetry and explores the transformative potential of fintechs in leveraging alternative data to reshape the SME lending landscape. By highlighting the evolving credit landscape this research aims to provide valuable insights for the benefit of business owners, credit professionals, credit bureaus, CEOs (of companies that generate data servicing SMEs) and policymakers.

1.7 Delimitations of the study

The following areas were not explored in this study:

i) Type of credit: The study used the word credit to refer to all forms of borrowing by SMEs. The nuances between different types of credit (unsecured debt to secured debt) have not been explored.

ii) Credit assessment limitations: The specifications of the scoring models and inputs used by fintechs are subject to significant data limitations. These limitations are due to the sensitivity of the participants in sharing any data that provide insights into their intellectual property.

iii) Regulation: The role of regulation was considered to be beyond the scope of this study. A detailed discussion on the regulatory frameworks in the fintech and alternative space would be premature as the industry is still underdeveloped.

iv) Impact of alternative on SME success: The study does not measure the long-term impact of using alternative on SME growth and sustainability.

v) Data ownership: Ownership of data and the privacy issues associated with the alternative data were regarded as outside the scope of this research. Alternative data used by fintechs to evaluate SMEs for credit can be information generated by other fintechs, banks or other SMEs.

vi) Alternative data depth: This research study does not go into detailed analysis of specific types of alternative data and their relative predictive power.

vii) Limitations of the methodology: Lending fintechs tend to operate in a closed manner. The quantitative data regarding their performance and scoring models are often considered confidential and treated as intellectual property. These fintechs may also be hesitant to share the data since they generally do not own the data they utilize. Instead, they obtain it through partnerships.

1.8 Definitions of terms

Alternative data: Data collected from non-traditional sources and not typically included in the traditional credit process (Zou et al., 2020).

Fintech: This term is used to refer to companies that use technology to deliver and improve financial activities (Schueffel, 2016). It involves the many different ways in which finance and technology meet (Fasano & Cappa, 2022). Łasak (2022) states that the term fintech can be understood as follows: (1) it is a technology and a solution based on the technology which is used in financial services and (2) it is connected with entities (startups) based on financial technology. The industry segments identified in the fintech landscape include payments, lending, savings and deposits, insurtech, investments, financial planning and advisory, capital raising and B2B tech providers (Genesis Analytics, 2019). This research paper focused on the lending segment.

Fintech lenders: Fintechs that primarily ingest alternative data, conduct credit scoring and make credit decisions. Entities facilitating and leveraging different forms of alternative data to provide funding to SMEs through the internet, cloud, or app-based platforms (Genesis Analytics, 2019). Examples include online lenders, asset financing lenders, alternative scoring and lending marketplaces. In this paper, the terms "fintech lenders" and "fintechs" will be used interchangeably.

Hard information: Information that is reproducible in numbers and can be transmitted (Fasano & Cappa, 2022; Liberti & Petersen, 2019)

SME: In this paper, SMME and SME will be used interchangeably. International Finance Corporation (2018) defines SMMEs using turnover rather than the number of employees. The turnover bands used by International Finance Corporation have been included in Table 1 below.

Table 1: Annual SME Turnover Bands

MSME Size	Annual Turnover	Mothly Turnover
Micro	0 499 000	- 41 583
Very Small	500 000 999 999	41 667 83 333
Small	1 000 000 4 999 999	83 333 416 667
Medium	5 000 000 19 999 999 20 000 000	416 667 1 666 667 1 666 667

Source: International Finance Corporation (2018)

Soft information: Information that cannot be summarised in numbers and requires expertise in the field to become meaningful and is typically collected in person and usually recorded in text (Liberti & Petersen, 2019). This information is often difficult to standardise (Fasano & Cappa, 2022).

The definition used in this paper is in line with the definitions provided by South African legislation as reflected in Table below (page 11).

1.9 Assumptions

The primary assumption made in this research paper is that credit default risk from assessing businesses using alternative forms of data would be within an acceptable variance compared to that attained using traditional forms of data. The Gini Coefficient is used by lenders to assess the discriminatory power of default models (Frunza, 2013). However, given the nascency of alternative data models, it is acceptable for a model to have a low Gini Coefficient and would still be validated for use, for instance in portfolios with a low number of defaults. The basis of the credit default risk assumption is supported by the success of fintech businesses in the lending sector. Building a sustainable lending business is dependent on managing acceptable bad debt levels. Furthermore, there is reliance on signals provided from the continued capital raising by these fintechs from institutional funders. The funding raising implies the soundness of their business models and acceptable bad debt experiences. The sensitivity of the research outcomes to this assumption is low, as the primary focus is on usage of alternative data by fintechs.

Table 2: Definition of SMEs

Column 1	Column 2	Column 3	Column 4	Column 5
Sector or sub-sectors in accordance with the Standard Industrial Classification	Size or class	Total full-time equivalent of paid employees Less than	Total annual turnover Less than	Total gross asset value (fixed property excluded) Less than
Agriculture	Medium	100	R 4.00 m	R 4.00 m
	Small	50	R 2.00 m	R 2.00 m
	Very small	10	R 0.40 m	R 0.40 m
	Micro	5	R 0.15 m	R 0.10 m
Mining and Quarrying	Medium	200	R30.00 m	R18.00 m
	Small	50	R 7.50 m	R 4.50 m
	Very small	20	R 3.00 m	R 1.80 m
	Micro	5	R 0.15 m	R 0.10 m
Manufacturing	Medium	200	R40.00 m	R15.00 m
	Small	50	R10.00 m	R 3.75 m
	Very small	20	R 4.00 m	R 1.50 m
	Micro	5	R 0.15 m	R 0.10 m
Electricity, Gas and Water	Medium	200	R40.00 m	R15.00 m
	Small	50	R10.00 m	R 3.75 m
	Very small	20	R 4.00 m	R 1.50 m
	Micro	5	R 0.15 m	R 0.10 m
Construction	Medium	200	R20.00 m	R 4.00 m
	Small	50	R 5.00 m	R 1.00 m
	Very small	20	R 2.00 m	R 0.40 m
	Micro	5	R 0.15 m	R 0.10 m
Retail and Motor Trade and Repair Services	Medium	100	R30.00 m	R 5.00 m
	Small	50	R15.00 m	R 2.50 m
	Very small	10	R 3.00 m	R 0.50 m
	Micro	5	R 0.15 m	R 0.10 m
Wholesale Trade, Commercial Agents and Allied Services	Medium	100	R50.00 m	R 8.00 m
	Small	50	R25.00 m	R 4.00 m
	Very small	10	R 5.00 m	R 0.50 m
	Micro	5	R 0.15 m	R 0.10 m
Catering, Accommodation and other Trade	Medium	100	R10.00 m	R 2.00 m
	Small	50	R 5.00 m	R 1.00 m
	Very small	10	R 1.00 m	R 0.20 m
	Micro	5	R 0.15 m	R 0.10 m
Transport, Storage and Communications	Medium	100	R20.00 m	R 5.00 m
	Small	50	R10.00 m	R 2.50 m
	Very small	10	R 2.00 m	R 0.50 m
	Micro	5	R 0.15 m	R 0.10 m
Finance and Business Services	Medium	100	R20.00 m	R 4.00 m
	Small	50	R10.00 m	R 2.00 m
	Very small	10	R 2.00 m	R 0.40 m
	Micro	5	R 0.15 m	R 0.10 m
Community, Social and Personal Services	Medium	100	R10.00 m	R 5.00 m
	Small	50	R 5.00 m	R 2.50 m
	Very small	10	R 1.00 m	R 0.50 m
	Micro	5	R 0.15 m	R 0.10 m

Source: ("National Small Business Act 102 of 1996," 1996)

An assumption has been made on the SMEs' digital footprint in South Africa. It is assumed that most SMEs engage in some form of digital transactions or own mobile phone, that enable the generation of alternative data used in credit analytics by fintechs. Purely cash business SMEs have been excluded from the research findings. The sensitivity of the research outcomes to this assumption is high, as alternative data usage relies on the presence of digital footprints.

Invariably, fintech lenders do not have direct access to SME data for analysis and tend to access it through partnerships. It is often the case that SMEs are unaware that the data they generate are collected for credit decisions (Zou et al., 2020). For these reasons, an assumption has been made that all data gathered by the fintechs are in within the ambit of any applicable privacy laws, such as Protection of Personal Information Act (“POPIA’) in South Africa. The sensitivity of the research outcomes to this assumption is high, as non-compliance with privacy regulations could have significant implications for the fintech lending sector.

1.10 Chapter outline

The following chapter gives a review of the literature on traditional lending models, fintech business models and various alternative data uses.

Chapter 1 introduced the purpose of the study and the contextual background. The research problem and emerging questions were also covered in this section. The significance of the study for digital business, business practice and theory are fleshed out. Delimitations, assumptions and keywords were included to give perspective to the scope of the study.

Chapter 2 provides the literature review and theoretical framework. It explores the challenges related to credit availability and accessibility for South African SMEs and examines the alternative data landscape by fintech. In addition, the chapter delves into three key concepts – adverse selection, signalling and screening to construct the theoretical framework for the study.

Chapter 3 describes the qualitative research methodology followed in the study.

Chapter 4 presents the qualitative findings from the interviews illustrated with verbatim quotes from the participants. The chapter also discusses the emergent themes of the study.

Chapter 5 presents a discussion on the findings of the study and clearly shows the points of alignment and divergence with the existing literature.

Chapter 6 is the concluding chapter of the study. It highlights the recommendations and suggestions for further research.

CHAPTER 2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Introduction

In this chapter, an extensive review of the academic literature and theoretical framework pertaining to the status quo in SME credit is provided. The discussion is broken down into two primary sections. The first section delves into the existing literature on the traditional SME landscape and zooms into the challenges SMEs face in accessing funding. The second section explores the use of alternative data to determining the creditworthiness of SMEs and how there can provide a level of assurance similar to that of traditional or conventional funders.

2.2 Background discussion

South Africa's consumer credit market is well developed with sufficient information submitted by credit providers on both positive and negative data (International Finance Corporation, 2021). International Finance Corporation (2021) posits that there is no credit information sharing infrastructure for SMEs. This critical gap in information sharing for SMEs has a large impact on access to credit. According to the OECD (2020b) survey on South Africa, scaling up remains a challenge for South African SMEs. The G20 countries has dedicated resources to increase financial access for SMEs (Owens & Wilhelm, 2017). This situation is indicative of an acknowledgement that SMEs are critical to the well-being of economies as they account for the majority of firms and contribute a large share of employment and drive innovation (Abbasi et al., 2021; De la Torre et al., 2010).

Liquidity constraints hinder SMEs' growth rates (Chavis et al., 2011). Access to funding for expansion, working capital, inventory, hiring staff and equipment purchases is essential to the sustainability of SMEs and consequently economic development and growth. Msomi and Olarewaju (2021) found that external financing is an enabler for SMEs to compete and invest in business growth.

There is a stronger reliance by SMEs to finance their business needs using informal financing than bank financing (Chavis et al., 2011). This reliance is despite the high cost of raising financing through informal channels. The OECD (2020) report found that the low levels of SME financing by banks emanate from the demand side as SMEs indicate that they generally prefer not to borrow from financial institutions, particularly traditional banks. Bank finance is generally tailored to SMEs that demonstrate profitability, asset availability, a strong credit score and a proven management track record (Chavis et al., 2011). Most SMEs, especially in the micro space are unseen by banks (International Finance Corporation, 2021). Traditional bank loan exposure in South Africa at the end of 2017 to the SME sector was only 28% (OECD, 2020). As SMEs grow, they generally transition from informal financing to bank financing.

While information asymmetry exists between SMEs and lenders, there are technology-focused lenders with business models that put the use of SME digital data and advanced analytics at the centre of their processes (Owens & Wilhelm, 2017). It is estimated that the global stock of SME data will double every two years due to the accelerated growth in mobile, cloud, big data, electronic payments and social data (Owens & Wilhelm, 2017). Access to these data in real time and in verifiable form can be used by fintech lenders to score businesses for funding (Agarwal et al., 2020). The costs to serve and the costs to acquire clients are decreasing as analytic and processing capabilities of these data improve without negatively impacting the default rates (Agarwal et al., 2020; Owens & Wilhelm, 2017).

The regulatory framework is a catalyst to the growth of fintech business models. Most fintechs operate outside the ambit of legislation allowing them to innovate around data privacy, maximum pricing caps, consumer protection, credit information sharing and cyber security (Owens & Wilhelm, 2017). The regulatory framework will not be covered in any further detail in this document.

2.3 Availability and accessibility of credit for SMEs

Approaching a traditional lender for funding by an SME, entails submitting reams of financial documentation, providing collateral and long waiting periods for a decision. (Asah et al., 2020) found that collateral, annual business turnover and audited financial statements are factors primarily influencing whether SMEs can access funding from traditional lenders. Traditional data include financial ratios and depth of the SME's relationship with its bank. In this section, the literature is reviewed to determine the availability of credit for SMEs in South Africa.

The credit gap as highlighted by various studies is primarily caused by information asymmetry. A lack of information leads to increasing screening costs consequently leading to higher administration costs to service SMEs (Del Gaudio et al., 2022; Mpofu & Sibindi, 2022; Yan et al., 2015). The capital structure of small businesses in South Africa is not optimised as they rely more on internal funding (sourced from their own savings) given a lack of external funding (Mpofu & Sibindi, 2022). It becomes increasingly difficult for SMEs to sustain their operations and expand with without external funding.

2.3.1 Access to formal funding

This formal funding is credit typically provided by banks. The South African banking system is an oligopoly (Wanke et al., 2017) with four large banks dominating the SME business banking market, namely, Standard Bank, FirstRand, ABSA and Nedbank. Smaller banks that have an SME focus are Mercantile Bank (recently acquired by Capitec), Sasfin, Grindrod Bank and Ubank (placed under administration in 2020). However, studies have found that SMEs are less willing to be primarily banked by smaller banks for their transactional needs (Sheng, 2021). This author concluded that small banks are better at relationship lending and more suitable for handling the soft information on SMEs.

The approval rates for SMEs seeking funding from traditional banks is low. Banks tend to have rigorous regulatory frameworks that make them rigid and inflexible when serving SMEs (Mpofu & Sibindi, 2022). A lack of collateral (asset-backed lending), poor information transparency (relationship lending) and limited financial information (financial statement lending) are significant factors contributing to the low approvals of credit to SMEs from banks (Abbasi et al., 2021; Angilella & Mazzù, 2015; Asah et al., 2020; Mbedzi & Simatele, 2020; Sheng, 2021). Mpofu and Sibindi (2022) argue that SMEs' inability to have proper record keeping results in them failing to provide the required information to the banks.

Research by Asah et al. (2020) which covered interviews with over 106 bankers in South Africa, revealed that collateral is a critical input in lending to SMEs in the formal sector. Most common forms of collateral include land or property assets, investment accounts and sureties. Most SMEs have no collateral to provide, resulting in declined applications. The authors further claim that information asymmetry is high with SMEs which results in imperfect market conditions that ultimately leads to stringent criteria from traditional providers. Without access to formal financial records, SMEs lack a single source of verification for their finances.

The credit rationing theory states that banks can only provide credit to clients whose opportunity cost is less than the maximum that they can repay (Mpofu & Sibindi, 2022). Legacy systems hinder traditional providers' from aggressively lending to SMEs. They view smaller deal sizes as cost ineffective, which coupled with the heterogeneity of SMEs results in significant transaction costs (Asah et al., 2020). Sheng (2021) found that lending technology for traditional banks in China relies on transactional and relationship lending. The South African business banking market is built on similar lending technology. Large banks have a preference to lend to large firms rather than SMEs (Aleem, 1990; Mbedzi & Simatele, 2020).

Credit bureau coverage is patchy, especially when businesses have little or limited financial history with financial institutions (McEvoy, 2014). Although the author's paper was geared to a consumer rather than an SME market, there was a correlation in findings between individual and sole proprietary SMEs. In traditional evaluation models, credit bureau data play a significant role for formal lenders. Credit bureau information is usually dated, missing or limited (Owens &

Wilhelm, 2017). The coverage of this data in the SME space is limited. In South Africa, the leading providers are Experian, TransUnion and Compuscan. Despite South Africa's advanced consumer credit information sharing environment, there is a significant gap in its capacity to offer business-related information (International Finance Corporation, 2021). This gap has resulted in a heavy reliance on consumer data of business owners to assess credit risk which negatively impacts the business (International Finance Corporation, 2021). Fintechs consequently rely less on credit bureau scores in their SME scoring models (Jagtiani & Lemieux, 2019). Substitutes for traditional credit metrics states include mobile phone information, psychometric testing, social media activity and records of online transactions (McEvoy, 2014).

From the literature, formal funding is typically more accessible for more established SMEs with profitable operations, access to collateral, and ability to provide equity. This drastically limits the number of SMEs that can access credit.

2.3.2 Access to informal funding

Informal funding is used by SMEs that fail to access credit from the formal sector. Mpofu and Sibindi (2022) posit that it has been the backbone of small businesses, particularly in the informal sector. Informal funding is typically funding from private moneylenders or loan sharks. Among other issues, the reason SMEs fail to secure funding from the formal sector is information asymmetry, lack of collateral and perceived high default rates (Mpofu & Sibindi, 2022). The authors further argue that some entrepreneurs opt for informal finance even if they are eligible for formal finance as a result of its flexibility, speed, convenience and simple administrative procedures. There are less stringent qualification criteria and no collateral requirements. The downside to this type of funding is its exploitative nature. SMEs are charged exorbitant rates as highlighted by Mpofu and Sibindi (2022) in their paper.

2.3.3 Government funding

Governmental institutions that provide credit for SMEs include the Small Enterprise Development Agency, Khula Enterprise Finance, Small Enterprise Funding Agency, National Empowerment Fund, Industrial Development Corporation and Development Bank of South Africa. The lack of borrowing from government suggests a lack of confidence in these agencies as the requirements are as restrictive as formal funding (Msomi & Olarewaju, 2021). According to a survey conducted by Msomi and Olarewaju (2021), they found that less than 15% of SMEs are aware of the available sources of finance from the aforementioned institutions. Nanziri and Wamalwa (2021) found that self-exclusion and lack of awareness of credit opportunities impede the achievement of the benefits of financial inclusion policies for SMEs.

2.3.4 Fintech lending

Banks have been used as a focal point to determine how transformative fintechs have been as a substitute to traditional banks in supplying credit to SMEs (Irwin & Scott, 2010). Chavis et al. (2011) suggest that more and more SMEs are starting to place reliance on alternative financing with a move away from banks. Fintechs have a strong impact on the traditional providers of financial services, as they change the way interactions with customers occur and consequently what information is gathered (Fasano & Cappa, 2022). In the credit landscape, two important benefits of fintechs is that there are able to carry out credit operations from a distance and they enable lenders to standardise their lending frameworks (Fasano & Cappa, 2022). While it is not possible to completely eliminate bias, one could argue that the use of alternative data by fintechs has the potential to reduce discrimination (Ryan, 2020).

Fintechs rely on new technologies such as machine learning and big data to analyse and score SMEs for credit in a cost-effective manner (Beaumont et al., 2022) . According to Beaumont et al. (2022), fintech platforms hold a regulatory advantage over banks as most products on offer fall outside the ambit of regulation. The authors maintain that SMEs enjoy digital applications with quick turnaround, real time underwriting, unsecured products, and high chance of getting funded with fintechs.

According to a study by Beaumont et al. (2022), the authors found that fintech lending improves an SME's access to credit. The results were based on a sample of similar characteristic SMEs in the United States taking loans from a bank or a fintech. The conclusion reached was that SMEs that took a fintech loan experienced a 20% increase in bank credit in subsequent credit transactions. Ali Finance in China reports that their default rate for loans using alternative data are below 1% (McEvoy, 2014). There is limited research in South Africa on what alternative data points are available for SMEs and which institutions are using it to enable financial access to SMEs. Fintech businesses operating within the South African landscape include Retail Capital, Lulalend, Merchant Capital, Nomanini, iKhokha Cash Advance, Yoco Capital, SA Taxi, Cash Connect, Profit Share Partners etc.

2.3.5 Proposition 1

Alternative data are used by fintechs to effectively augment SME credit analysis where there are gaps in the traditional data points.

2.4 Usage of alternative data by fintechs

Fintechs gather alternative data in a way that enables real-time access to an SME's data. SMEs leave digital footprints of verifiable data whenever they utilise digital products, such as cloud-based services, digital payment transactions, or social media interactions (Owens & Wilhelm, 2017). The authors discovered that SMEs are willing to forego some degree of privacy in exchange for access to funding and other value-added tools. Alternative data provide a comprehensive view about an SME's strengths and weaknesses and offer valuable insights that are useful for credit assessment.

2.4.1 Alternative data

Alternative data in finance refers to proxy metrics or information originating from unofficial or noncompany sources that individuals can use to gain insight into a SME (Constable, 2019). These data may not necessarily be directly related to credit, but provide useful insights into the ability or willingness of a potential borrower to repay a loan (International Finance Corporation, 2021). For fintechs, the digital footprint from SMEs and their customers is increasing the accumulation of alternative data. SMEs are using cloud-based services, banking services, transacting through digital payments, using their smart phones, social media presence, buying or selling online, shipping and record keeping online (Agarwal et al., 2020; Owens & Wilhelm, 2017; Zou et al., 2020). Alternative data can also include rent, utility payments history, managers educational attainment, SME social media use and other behavioural information not traditionally factored into credit decisions (Ryan, 2020). Jagtiani and Lemieux (2019) state that fintechs can derive alternative data from local data such as local economic information, for example identifying whether loan applications are submitted from high-crime areas or in areas where factories are being shut down or relocated.

The figure below depicts how SMEs operate and create their digital footprint:

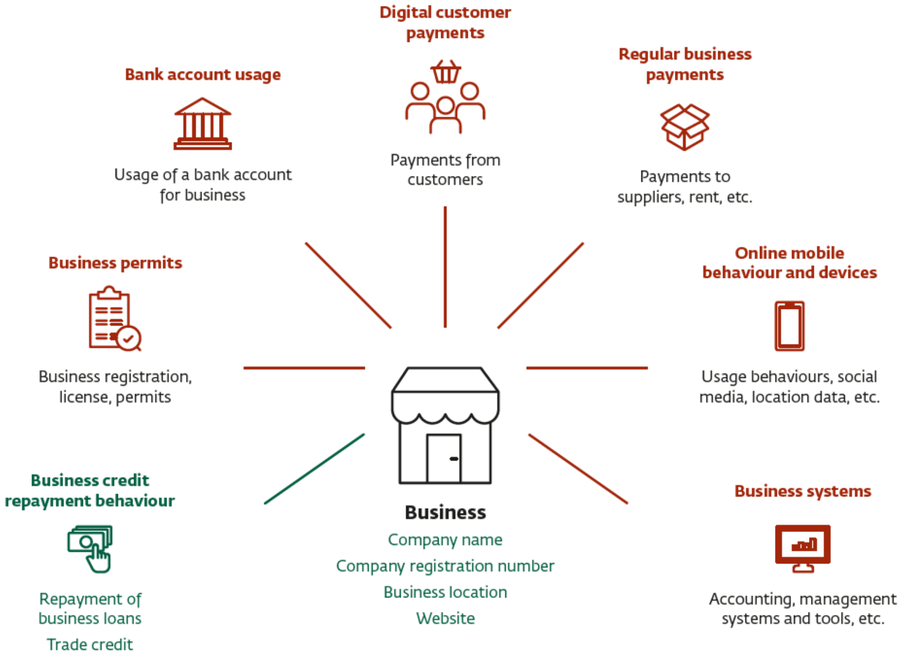

Figure 1: How SMEs create alternative data

Source: International Finance Corporation (2021)

The table below presents a variety of verifiable alternative data sources or verticals created by SMEs, although this list is not exhaustive.

Table 3: Alternative data verticals

E-commerce Data	Financial Data	Social Media Data	Mobile Data	Individual Data
E-Commerce Sales	Loans and Credit Card	Social Media Data	Mobile Call Pattern Data	Psychometric testing
Purchasing Data	Bank Statements	Search History	Mobile Business and Expense Data	
B2B Commerce	Investment Account	Website History	Mobile Recharge History	
Supply Chain Trade Flow Data	Insurance Data	Online Rankings and Reviews	Mobile E-Money Transactions	
Logistics and Shipping Data	Online Accounting		Smartphone Mobile App Analysis	
Performance Data of SME Business customers	Billing and Payment Data		Mobile Geo-Locational Data	
	Merchant POS			
	Marketing Data			
	Inventory Tracking Data			
	Economic and Industry Data			

Source: Owens and Wilhelm (2017)

2.4.2 Accessing alternative data (big data)

Alternative data allow fintech lenders to make decisions quickly (Jagtiani & Lemieux, 2019). Technological revolutions have enabled the generation of data that are accessible much faster and in a verifiable manner. Technology has allowed fintechs to serve small businesses without brick-and-mortar investments (Jagtiani & Lemieux, 2019). Data are generated in wide ranges from multiple sources (volume), in different kinds (variety) and can be collected and processed at certain speeds (velocity).

Fintechs ingest alternative data through proprietary complex machine learning algorithms that rely on big data to improve the availability and accuracy of the information used to analyse SMEs (Jagtiani & Lemieux, 2019; Sheng, 2021). Using alternative data, fintechs have been able to make credit available in a frictionless and convenient manner (Sheng, 2021). These credit products are embedded in various SME platforms.

Largely, fintechs lack the distribution required to gather the data. Fintechs need the buy-in of the data providers for access. There is an increased need to collaborate between fintechs and banks, fintechs and other fintechs. Fintechs partner with other fintech providers with direct access to transactional data via Automated Programming Interfaces (APIs); or the use of screen scraping (Owens & Wilhelm, 2017). As stated by Jagtiani and Lemieux (2019), alternative data are available through data aggregators and vendors that work directly with SMEs. Collaborations with banks are valuable given the distribution banks have. Data providers have no incentive to share the data as it creates a competitive advantage for them (International Finance Corporation, 2021).

2.4.3 Fintech ingesting alternative data

There has been an advancement of fintechs that access and leverage data through APIs and apply big data analytics to determine credit worthiness (Agarwal et al., 2020). APIs allow fintechs to access rich data sets from other platforms in a cost-effective manner and through efficient and proactive data collection methods (Sheng, 2021; Yan et al., 2015). This access to big data has reduced search costs for credit and improved underwriting processes (Sheng, 2021; Yan et al., 2015). Owens and Wilhelm (2017) have covered the following Fintech verticals as enablers of access to alternative data:

SME marketplace lenders – These are aggregators connecting SMEs to fintechs. The platforms are overlaid with alternative data points with wrapped leads of eligible businesses passed on to the lenders for the underwriting. Types of marketplace lenders include peer-to-peer lenders, online balance sheet lenders, payment innovators, analytic providers, identity information providers, cloud-based lending platforms supporting lenders and SME loan broker marketplaces. A fintech has to integrate with marketplaces to be able to ingest these leads.

E-Commerce and Mobile payments – They have access to transactional level data generated by SMEs using their platforms. A fintech ingests this data via API integrations and secure file sharing with e-commerce and mobile payment providers.

Supply chain platforms: These platforms digitise documents and transactions between SMEs and corporates. This digitisation is ideal for businesses that have monthly payment cycles and enables better tracking mechanisms through purchase orders, invoices, and receivables. A fintech ingests this data via API integrations with supply chain platforms.

Mobile data-based lending models: These models score businesses based on mobile usage patterns, mobile apps installed, SMS messages, emails, number of inbound vs outbound calls, geo-based locations and social network usage. A fintech ingests this data via API integrations and secure file sharing with telecommunications companies.

2.4.4 Alternative data and information asymmetry

Fintechs have proven that alternative data credit scoring business models can be run within acceptable risk levels and profitability (Owens & Wilhelm, 2017). Notably, there is limited reporting by alternative lenders on loan performance. Everett (2015) proposed three categories of determinants of default in fintech lending: loan characteristics, borrower characteristics and the borrower’s group characteristics. The higher the information asymmetry within each of those characteristics, the higher the default rate.

Alternative data have reduced information asymmetry across the three categories (Yan et al., 2015). In traditional credit evaluation lenders are reactive and rely on borrowers providing the information required for an assessment. The authors found that lenders in the big data era, can proactively search for a business’s online footprint and gain detailed insights into the business going back in time. Technologies such as alternative credit scoring have reduced information asymmetry and expanded credit availability for SMEs (Einav et al., 2013).

2.4.5 Proposition 2

Alternative data allow fintechs to depend on hard information derived from alternative data sources to make informed credit decisions.

2.5 Theoretical framework

2.5.1 Overview

This report unpacks the role alternative data play as enablers for SMEs to access credit. This process cannot be done conclusively without discussing scholarly theories that focus on the inhibitors of traditional banks advancing lending to SMEs. Credit risk evaluation is essential for lenders to make loan decisions and they rely on information provided in the application process (Yan et al., 2015). However, from the literature review discussed above, information economics has the potential of undermining the credit risk assessment process. The fact that different people know different things implies information asymmetry (Yan et al., 2015). Information asymmetry is one of the most prevalent factors affecting SME's access to credit (Del Gaudio et al., 2022).

Soft information asymmetry can be diminished through personal interactions between lenders and SMEs, while hard information asymmetry has been reduced by technological advances that facilitate the collection, processing and communication of standardised information (Fasano & Cappa, 2022). Both soft and hard information is crucial in providing credit (Fasano & Cappa, 2022). Yan et al. (2015) highlight the fact that to access funding, borrowers signal information about themselves and their SMEs characteristics (signalling), while the lenders search for credit information and screen the loan applicant (screening).

Fasano and Cappa (2022) argue that the threat of making decisions based solely on hard information reduces the amount available as a credit line. The authors argue that a lack of soft information incorporation in the assessment for credit hampers credit provision. In thin file environments, the screening challenges by the lender and limited signalling by the borrower causes information asymmetry. There is a reliance by fintechs on analysing big data to extract useful patterns or rules (Yap et al., 2011). Signalling refers to information asymmetry as a result of borrowers not disclosing complete information for concern over the potential negative signals they might be giving the lender or having limited means of showing their trustworthiness (Moro et al., 2015; Spence, 1973). Traditional banks have to expend significant time and resources on screening applications (Aleem, 1990). Adverse selection is a consequence of information asymmetry. It occurs when lenders struggle to distinguish between the good borrowers and the bad when advancing credit (Akerlof, 1970; Moro et al., 2015; Yan et al., 2015).

The economics of information theory that will be discussed in this section has three concepts, namely adverse selection (the lemons principle), signalling effects, and screening costs of the lenders (see Figure 1 below). According to the economics of information theory, a lender will seek to acquire additional information where the marginal benefit of acquiring additional information exceeds the marginal cost (Goldman & Johansson, 1978; Stiglitz, 2000; Yan et al., 2015).

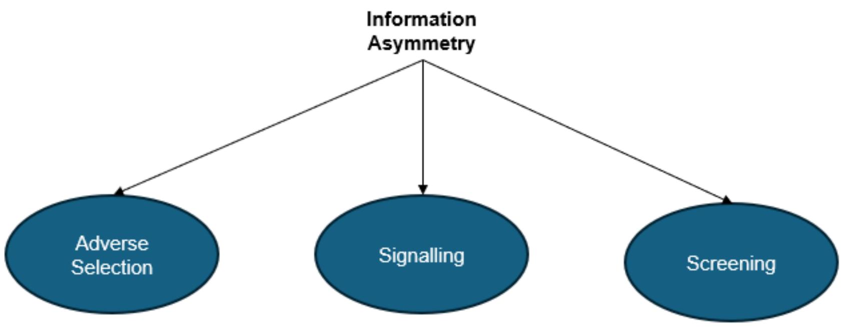

Figure 2: Economics of information in the lending business

Adapted from: Yan et al. (2015)

2.5.2 Adverse selection (the lemons principle in credit)

2.5.2.1 Concept definition

Akerlof (1970) in his Nobel prize winning paper used the automobile market as an example, positing that in the car market, good and bad cars will be sold at the same price due to information asymmetry as buyers find it impossible to tell the difference. Akerlof (1970) coined the lemons principle which states that the market price affects the quality of the goods offered and consequently the demand (Stiglitz, 2000). Traditional lenders using conventional credit processes struggle in terms of lending to SMEs with limited information. Differentiating between the good SMEs and the bad becomes challenging, consequently increasing lending rates and driving out good borrowers. Lenders lack visibility of the risk properties of each SME, resulting in adverse selection. Accordingly, adverse selection suggest that low-quality firms become more aggressive in the market to access credit in the hope that they are passed off as high-quality firms (Ono et al., 2014).

2.5.2.2 The traditional lenders' approach

The SME market is heterogeneous and is characterised by a high level of opacity. Given the thin SME credit files, traditional credit evaluation may result in false positive and false negative approvals. Everett (2015) posit that lenders use privately collected soft information to gain an economic advantage over their clients and competitors. The author further describes a situation that he terms the "hold-up" issue, which means borrowers then find it difficult to switch lenders since the private information proving that the borrower is not a lemon is kept confidential by the lender.

Yan et al. (2015) describes false positive as traditional lenders incorrectly approving an application from a borrower with high credit risk while false negative is declining a borrower with low credit risk given the limited information at application. A common assumption regarding credit ratings or credit scores is that firms within the same credit score or rating class have the same default rate (Crouhy et al., 2000). The SMEs know more about their business than the information that they can provide in their applications (OECD, 2015; Rosser, 2003). Some businesses operate informally and possess limited verifiable information, which consequently leads traditional banks to perceive such SMEs as high-risk borrowers.

As a result, Mills (2018) noted that traditional banks tend to make a binary decision, either approving or declining credit applications for SMEs. Small businesses are left with little choice but to approach smaller banks, non-bank financial institutions and micro lender institutions that primarily rely on conventional credit assessment methods and tend to approve credit for SMEs at significantly higher interest rates. This state of affairs results in adverse selection as good borrowers drop out of the market (Rosser, 2003; Stiglitz, 2000) and banks struggle to distinguish SMEs that will repay from those that will not (Moro et al., 2015). Only low-quality businesses remain in the market. In his paper, Akerlof (1970) further discusses the issue of adverse selection theory with reference to the medical insurance industry. He states that at certain price levels in medical insurance, the ones that use it have sub-standard life and are certain they will need insurance. Similarly in credit markets, if lenders could identify borrower risk perfectly, each borrower would be charged an appropriate risk premium instead of treating all borrowers as homogenous (Stiglitz, 2000).

In conventional credit, willingness to pay is a key component in assessing an SME's probability of default. This measure is subjective in that traditional banks rely on the SME's banking relationships. OECD (2015) in its report states that bank lending relies on relationship gathered through direct interaction with SMEs. The cost of dishonesty in the market is an inhibitor to banks advancing credit to SMEs that do not maintain a transactional relationship with them. Akerlof (1970) maintains that dishonest players in the market negatively impact all stakeholders. He further asserts that the cost of dishonesty impacts not only lenders and borrowers, but also forces legitimate businesses out of existence. Consequently, this situation leads to viable SMEs lacking access to the necessary cash for their sustainability.

Without clearly determinable willingness to pay and detailed information submitted by SMEs, traditional lenders rely on collateral to stimulate repayment behaviour and reduce losses given a default (Lehmann & Neuberger, 2001). In other words, a lack of information results in lenders rationing lending to all borrowers (Stiglitz, 2000).

2.5.2.3 The fintech approach

Alternative data are used by fintechs for their ability to more accurately and reliably separate trustworthy borrowers from untrustworthy ones (Yan et al., 2015). The authors further state that alternative data proactively gathered are more objective and cannot be easily manipulated.

Sheng (2021) found that in China, access to alternative data allows fintechs to turn soft information into hard information that can be relied on to offer credit to SMEs. By utilising alternative data to analyse SMEs, fintech can reduce issues of information asymmetry that causes SMEs to be treated as homogeneous entities. The added advantage of alternative data for SMEs is that they do not have to incur ex ante costs on time spent producing information to satisfy the lender’s request (Moro et al., 2015).

Alternative lenders rely on rich transactional data from other fintech partners that hold a deeper relationship with the SMEs to advance credit. From Day 1 of a lender-borrower relationship, the alternative lender can ingest data that provides years of transactional history. Generally, transaction level data used in SME credit scoring should exhibit regular and consistent trade that becomes difficult to manipulate. These data are typically monitored in real time with some fintechs collecting the money at source which takes the onus away from an SME having to prove willingness to pay. Moro et al. (2015) in their research paper found that reduction in information asymmetry has an economically relevant impact on the amount of credit provided. With the advancement in collecting and interpreting soft information for superior profiling of SME loan applicants, access to or price of credit is likely to improve (A. Morse, 2015).

In his paper, Akerlof (1970) states that credit is granted where the lenders have easy means of enforcing their contract. The ability to have visibility of cashflows influences the appetite of most fintechs to provide unsecured products.

2.5.3 Signalling in credit

2.5.3.1 Concept definition

Information relevant to advancing credit is conveyed through a number of variables (Stiglitz, 2000). Credit advancement requires certainty that funds will be repaid. Lenders therefore rely on signalling from SMEs in their subjective analyses of future trajectories. Kawai et al. (2022) argue that a business's approach to credit access is an important measure of creditworthiness. The authors posit that less creditworthy borrowers tend to be less sensitive to price or interest rate and place a higher value on getting funded, while creditworthy borrowers tend to be sensitive to pricing. Soft information which is collected in person and recorded in text generally provides unobservable signals (Liberti & Petersen, 2019; Zou & Wang, 2022), while hard information from alternative data is more reliable.

In his study on signalling theory Spence (1973) argues that signalling occurs when individuals communicate their qualities or abilities to prospective employers when there is information scarcity. He further argues that it takes place where there are many signallers (SMEs) tapping into the markets to position themselves in the best possible way. According to Tala (2021), SME owners often have more information about their businesses but have difficulty positioning them with traditional lenders. This predicament is in line with the position taken by Lehmann and Neuberger (2001) that SMEs provide less information to lenders than large firms.

Spence (1973, 2002) studied signalling in the job market. He found that recruitment is rife with inefficiencies through information asymmetry. In his example, an employer is unable to accurately decide on the skills of a candidate and therefore relies on signals such as investment in relevant education. Similarities were identified between this concept and its application in extending credit by traditional lenders, as well as in examining how fintechs employ signaling to strengthen their processes.

2.5.3.2 The traditional lenders' approach

At inception of a lending relationship, banks look for SMEs exhibiting business characteristics signalling sustainability such as an established management track record, well documented financial statements, a longer-term history, high turnover thresholds, solid credit scores and collateral (Asah et al., 2020). Collateral for example is a strong signalling variable for the borrower (Voordeckers & Steijvers, 2006). These authors found that the stronger the collateral traditional lenders have, the better the credit terms and ability for SMEs to access credit. Ability to provide collateral reflects signalling effects in traditional bank lending (Lehmann & Neuberger, 2001). It sends a message to the lender that the borrower puts up collateral for projects that will not require the collateral to be perfected (taken over by the bank). In fact, collateral induces borrowers to reveal risks that are unknown to the lender.

Irwin and Scott (2010) investigated whether SME managers' personal characteristics such as ethnicity, gender and education are signalling variables. The evidence from their research revealed a strong reliance by traditional lenders on such observable information. Irwin and Scott (2010) in their research conducted in the UK, found that ethnic minorities and women tend to struggle in accessing finance. The authors state that bank processes have a gender bias, are racialised and class based. This finding illustrates that there is a preponderance of traditional bank processes being heavily weighted towards attributes that Spence (1973) refers to as observable and unalterable signals.

The SME alterable signals include management's execution capability and strategic plans going forward. These alterable characteristics become an actual signal if they are negatively correlated with the company's sustainability and thus ability to repay (Spence, 1973). SME managers are equally concerned about sharing too much information which could possibly lead to traditional banks overreacting or misinterpreting the news and inferring negative signals about future performance (Moro et al., 2015). These authors argue that negative signals from banks lead to an increase in the interest rate, a reduction in the amount of credit made available, declines or additional guarantees being required.

2.5.3.3 The fintech approach

Fintechs are following the principle recommended by Spence (1973) to use other means to gather information and confirm its veracity where signalling is weak. Fintechs are using alternative data points to get comfort on the observable and unalterable signals discussed above. Jagtiani and Lemieux (2019) argue that when alternative data are included in credit risk analysis, it paints a fuller and more accurate picture of a borrower's creditworthiness.

Fintechs are collecting data proactively and assessing SMEs for consistency. This practice results in SME applicants being treated based only on behaviour. Some of the credit scoring engines the fintechs are using are agnostic to gender, race and ethnicity. They rely on transactional level data with such demographics released post loan approval. One such product is the merchant cash advance product offered by fintechs which relies on anonymised mobile point of sale transactional / cashflow data to make lending decisions. Furthermore, fintechs using alternative data have simplified the point at which SMEs can become eligible for funding.

2.5.4 Screening in credit

2.5.4.1 Concept definition

Information scarcity increases the costs of screening in the credit market. Rothschild and Stiglitz (1976) posit that if individuals were willing to reveal all their information, everyone would be better off. Given that information gathering is costly, banks will seek additional information until the expected marginal benefit of search equals zero (Lehmann & Neuberger, 2001).

Employers want to know the productivity of their workers, investors want to know the return on investments, insurance companies want to know the likelihood that various people they insure might have an accident or get sick (Stiglitz, 2000) and SME lenders want to know the likelihood that they will be repaid for credit provided. In the SME landscape, if SME managers were transparent with the full account of their businesses, traditional lenders would be able to price them appropriately and they (the former) would be able to access credit.

Currently lenders expend resources and time to screen applicants and pass on the costs to the borrowers (Aleem, 1990). Two elements are at play with this concept, the cost of resources invested in search activities and the efficiency with which the screening activities are conducted (Goldman & Johansson, 1978). In deciding whether to grant credit, a lender considers the costs of screening the borrower and the costs of servicing the borrower, capacity to repay and loyalty of the borrower (Arráiz et al., 2021).

2.5.4.2 The traditional lenders' approach

The time and resources invested are considered the most important cost element of screening (Goldman & Johansson, 1978). Traditional SME lenders must grapple with how much time is invested in gathering documents, underwriting and disbursement.

The screening process typically followed by traditional lenders was covered in a paper by Aleem (1990). Firstly, traditional bank lenders have less risk appetite for SMEs with no previous history of dealing with them. Lehmann and Neuberger (2001) point out that the cost of gathering information about a borrower is prohibitively high when a bank has no previous dealings with the SME. The natural inclination for traditional lenders for first time relationship SMEs is to access financial statements. They are used by the lenders to screen SMEs on their payment capacity and willingness (Tala, 2021). Therefore, a lack of this information results in SMEs failing the screening process. Lehmann and Neuberger (2001) found that bank lending to SMEs is not only influenced by firm characteristics and credit risk variables but also by social interactions between the loan officer and the bank manager. Traditional lenders have a preference to be more open to providing first time credit lines to large firms with detailed information (Mbedzi & Simatele, 2020)

Secondly, traditional lenders rely on existing credit relationships with other financial institutions. The second feature in the screening process makes it difficult for start-up SMEs to access funding with thin files. Aleem (1990) postulates that subject to the SME passing the first two stages of the screening process, the final stage involves approving the SME for an initial credit line (i.e., an introductory credit line).

In their paper, Rothschild and Stiglitz (1976) studied insurance markets. They found that insurance companies do not discriminate among their potential customers based on similar characteristics. SME lending is different in that traditional lenders treat businesses differently based on sector, size, age and profitability. Credit rationing is employed by lenders based on customers' characteristics.

2.5.4.3 The fintech approach

The common aspect in the fintech approach is the development of web-based platforms (Fasano & Cappa, 2022). These digital platforms allow fintechs to carry out their operations from a distance, reducing geographical, physical and social restrictions (Fasano & Cappa, 2022). The authors further highlight the fact that digital platforms give fintechs access to huge amounts of data, which enables them to optimise even more in terms of how they service their customers. The use of alternative data reduces the cost of making credit decisions, ongoing monitoring and lowers the operating costs for the lenders (Jagtiani & Lemieux, 2019).

Screening costs for collection of credit information retrieval has shifted from a passive to a more proactive information retrieval approach (Yan et al., 2015). Relationships and soft data provide advantages in borrower screening (Jagtiani & Lemieux, 2018). Through access to big data, fintechs have reduced information asymmetry as search costs for credit and the collection of data has become proactive (Yan et al., 2015). The fixed costs of acquiring these alternative data with technological improvements has reduced dramatically (Stiglitz, 2000). Fintechs work through APIs and can proactively access anonymised data regarding sector, size, age and profitability.

Fintech models are anchored in driving behavioural change by clearly articulating the benefits of disclosing information. SMEs are encouraged to proactively share their information by offering incentives for data enrichment. The consequence of fintech lenders offering unsecured credit products addresses the finding by Voordeckers and Steijvers (2006) that in SME lending, traditional banks tend to do minimal screening and rely excessively on collateral.

2.6 Conclusion of literature review

The apparent gap in the literature that has been reviewed is the critical role fintechs play in transforming the SME funding landscape. Fintechs play a critical role in alleviating the problem of information asymmetry by offering credit to SMEs in South Africa. Fintech have created access for SMEs that are typically excluded from traditional funding for low credit scores (Bank et al., 2022). Literature referenced across this report highlights the rise in the number of fintech models providing SMEs with a digital profile usable in credit decisioning (Abbasi et al., 2021; Charaia et al., 2021). Globally, the proliferation of fintechs is driving the digitisation of SMEs, particularly after the global pandemic (Charaia et al., 2021). Łasak (2022) posits that fintechs enable the flow of SME data among industry players and provide platforms that connect SMEs with alternative lenders and customers. This service is critical to enabling fintech lenders in South Africa to access data and make funding available for SMEs. This report will cover examples of companies that are generating alternative data that are being used by fintech lenders.

The ability of fintech lenders to assess alternative data and determine creditworthiness of SMEs while making capital more accessible ensures the sustainability of these businesses. Abbasi et al. (2021) found that the use of big data enables fintech lenders to accurately ascertain the credit risk of SMEs. The struggles of small business with accessing funding are well documented.

2.6.1 Proposition 1

Alternative data are used by fintechs to effectively augment SME credit analyses where there are gaps in the traditional data points.

Unlike traditional banks, new lending technologies allow fintechs to provide SMEs with credit without banking relationships while bypassing constraints such as information opacity and lack of collateral (Mbedzi & Simatele, 2020). Fintechs have improved the availability and accuracy of information, increased the number of information channels and sources, and reduced information friction between lenders and SMEs thereby enabling access to credit (Sheng, 2021)

2.6.2 Proposition 2

Alternative data allow fintechs to depend on hard information derived from alternative data sources to make informed credit decisions.

Zou and Wang (2022) conclude from their work that soft information is critical for lower levels of credit default. Sheng (2021) argues that fintechs have enabled the collection of information-rich alternative soft and hard datasets from other lenders and reduced the cost of screening and monitoring. They have transformed soft alternative data into hard information while eliminating friction in data collection and enabling real-time decision making.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Introduction

This qualitative report using the interpretivist paradigm explored the use of alternative data by fintechs to provide access to credit for SMEs in South Africa.

The chapter starts with the research paradigm and approach giving the rationale behind adopting an interpretivist-constructivist world view. It covers the assumptions made in the research and their appropriateness to the research study. Furthermore, it highlights the benefits inherent in this methodological choice.

Thereafter there is a discussion on the suitability of adopting a standard qualitative approach as a research methodology. This section leads to a discussion of the research demographics, encompassing the identified population, the sampling strategy, and the specifics of the sample itself. Emphasis is placed on the rationale for employing purposive sampling in this study.

Following this aspect, the section details the data collection methods including a description of the research instruments utilised in the study. In order to ensure the integrity of the research, a quality assurance sub-section is included, with a focus on establishing the credibility and dependability of the findings.

Finally, the chapter concludes with an overview of the ethical considerations.

3.2 Research paradigm

The choice of paradigm influences the research approach and design adopted for the study. It affects the manner in which research questions are asked and investigated (Kelly et al., 2018). The dominant research paradigms, according to Kelly et al. (2018), include positivism, post positivism, interpretivism, constructivism and pragmatism. This study adopted the interpretivism-constructivism paradigm. Interpretivism-constructivism holds the notion that individuals develop subjective meanings of their experiences, meanings that are varied and represent a complexity of views (Creswell, 2014). According to Meissner and Meissner (2016), action is contained in the lived experiences and points of view of those who experience it. The suitability of this paradigm comes from its ability to recognise multiple perspectives and versions of the truth (Kelly et al., 2018; Thanh & Thanh, 2015). Interpretivism reflects a recognition of subjective understanding and the need to interpret it through the researcher (Kelly et al., 2018). Given the focus of this research paper on how fintechs are using alternative data to provide access to credit for SMEs, this paradigm allows for diverse and unique views from the participants to be expressed, and it provides an opportunity for the researcher to interpret the findings.

3.3 Research approach

This study adopted a standard qualitative research approach, which was considered an appropriate choice for this study due to its qualitative nature and its focus on understanding processes, actions, or interactions from the viewpoint of participants (Creswell, 2014). According to Creswell (2014) qualitative research is an approach that explores and seeks understanding of the meaning individuals or groups ascribe to a social or human problem. Hammarberg et al. (2016) assert that qualitative research methods are appropriate for answering questions related to experiences, meanings, and perspectives from the participants' point of view. This process involves the researcher collecting data in the participants' settings and interpreting the data through inductive analysis and thematic development (Creswell, 2014).

Open-ended questions were designed to enable participants to express their unique worldviews, thereby contributing to a diverse range of perspectives on the topic. The context in which participants' views are formed is considered critical in this qualitative research approach, as it results in gaining an understanding of the whole picture. Furthermore, there is an awareness of the researcher's role in interpreting and analysing the data and potential biases.

There are several advantages identified in following this qualitative methodology. It offers flexibility in choosing methods that align with the research objectives and build trust with participants (Susilo et al., 2021). The selected sample, chosen based on participants’ professional experience and background, was best placed to contribute to a better understanding of the theoretical framework (Etikan et al., 2016). This methodology aligned with the exploratory nature of the research questions, focusing on understanding and interpreting the participants' experiences and viewpoints (Turner & Astin, 2021). These viewpoints can allow researchers to explore unexpected patterns and themes in the data. It is particularly relevant given the dynamic and nascent nature of fintech lending in South Africa (Asah et al., 2020).

Moreover, standard qualitative research provides a systematic and rigorous process for data collection and analysis. It involves collecting data, analysing, organising and contextualising patterns to gain comprehensive insights into the research topic (Nowell et al., 2017). Thematic analysis, as noted by the authors, is particularly effective in examining varying perspectives, identifying commonalities, differences, and unexpected insights.

While this approach offers numerous advantages, it is not without limitations. One potential disadvantage is that conducting detailed coding is time consuming, tiring and laborious (Hussein et al., 2014). In addition, the researcher's role as an active instrument in the study, analysing and providing meaning to the interviews can introduce subjectivity (Legard et al., 2003). In conducting this study, the researcher was aware of the argument by Peters and Halcomb (2015) that many novice researchers, find it challenging to separate their own preconceptions about the study topic and associated issues. This awareness necessitated the need for transparency in the research process to mitigate any potential bias.

Despite these challenges, standard qualitative research, with its focus on in-depth understanding and systematic analysis, remains a robust and well-established approach. Its focus on in-depth understanding and interpretation, coupled with a systematic approach to data analysis, makes it the most suitable choice for addressing the research questions and objectives.

3.4 Research methodology

The overall study design for this research employed a standard qualitative research approach. The primary methodology used was semi-structured interviews.

Semi-structured interviews were considered particularly suitable for this research due to their flexibility and depth. They allow for an in-depth exploration of participants' views and the meanings they attribute to the use of alternative data in credit assessment (Tong et al., 2007). Semi-structured interviews are the most frequently used interview technique in qualitative research (Kallio et al., 2016). The authors further note that the main advantage of semi-structured interviews is that they enable reciprocity between the interviewer and participant, enabling the interviewer to improvise and ask follow up questions based on the participants' responses.

The study predominantly relied on primary data collected through these interviews. Primary data provides firsthand insights directly related to the research question, ensuring the relevance and specificity of the information gathered. While secondary data from existing literature, reports, and studies on fintech, SME credit assessment, and alternative data usage were reviewed for background and contextual understanding, the core insights and conclusions of this research were drawn from the primary data.

A cross-sectional study design was chosen over a longitudinal approach. Cross-sectional research involves interviews conducted with participants at a single point in time (Rindfleisch et al., 2008), providing a snapshot view of the phenomena under study.

In the following sections, we outline the specific steps and procedures that were employed in conducting the standard qualitative research study, including data collection, thematic analysis, and interpretation.

3.5 Population and sample

3.5.1 Population

The research population for this study comprised diverse participants involved in the fintech-driven SME credit assessment landscape in South Africa, each of whom contributed unique perspectives and insights. Participants came from various sectors and industries within this landscape, all having exposure to alternative data and their role in enhancing credit accessibility for SMEs. The total population size of the fintech market in South Africa is unknown due to the scarcity of publicly available data on the market. However, the key populations of interest included in this study included:

- Fintech Lenders: This category included fintech companies and lenders operating in South Africa. These entities leverage alternative data sources and advanced scoring models to assess and extend credit to SMEs.
- Data-Generating Businesses: This category included businesses that provide services to SMEs and generate a wealth of data points that can be classified as alternative data. These enterprises offer embedded funding or credit solutions utilising the alternative data they generate as a foundation to SMEs, all in partnership with fintech lenders.
- Credit Bureaus: This category comprised globally recognised traditional credit bureaus with a presence in South Africa. Credit bureaus play a pivotal role in the credit ecosystem, working with both alternative and traditional data to construct comprehensive and commercially viable scoring models for SMEs.
- SME owners and operators: These individuals or entities serve as consultants specialising in SME advisory services. They collaborate closely with fintech and traditional credit providers across Africa and Europe, offering guidance on the strategic use of data to enhance credit accessibility for SMEs.

The inclusion of these diverse participant categories ensured a rich exploration of the use of alternative data by fintechs in extending credit to SMEs, shedding light on the intricate dynamics within this evolving landscape.

3.5.2 Sample and sampling method

The sampling method selected for this research was purposive sampling, which is a non-probability sampling technique. This sampling technique entails a deliberate choice to include a participant in the research due to the qualities that they possess (Tongco, 2007). This approach was regarded as suitable for this research as it seeks to gain in-depth insights from participants who have undergone the required experience or have the knowledge of the topic (Coyne, 1997). The participants were therefore purposefully selected based on their experience and direct involvement in the fintech-driven SME landscape where alternative data are used in the credit assessment landscape in South Africa.

In the current study there were 13 participants. The number was deemed sufficient as it provided rich and detailed data. Going beyond this point would have become repetitive and no new codes likely to emerge from the interviews. Most of the emergent themes were identified after 8 interviews, and to reach a total of 13, only minor refinements to the themes emerged. Purposive sampling focuses more on quality and richness of data rather than the number of participants (Hennink et al., 2017). Hennink et al. (2017) describes saturation as the point in data collection when no additional issues or insights emerge from the data. Creswell (2014) states that it is the point when the researcher stops collecting data because fresh data no longer spark new insights or reveals new properties.

The contacts with the participants were established through professional networks. The participants were initially contacted through WhatsApp, where they received a brief summary of the study and an invitation to discuss their potential participation in the academic research project. This method was chosen for its convenience and widespread use within the professional community. Following this initial contact, a formal email was sent to each participant, including a comprehensive Participant Information Document outlining the study's objectives, procedures, and their rights as participants. This email also contained a consent form, which participants were asked to review and sign electronically. To confirm their participation, participants either replied to the email with their consent or indicated their agreement at the beginning of each recorded interview session. Participants' privacy and anonymity was upheld with the utmost integrity. It is important to note that none of the participants received any financial compensation for their involvement in this study.

The participants were carefully selected with a view to providing depth and insights to this research study (Hammarberg et al., 2016). The selection criteria for the participants were based on specific criteria to ensure relevance to the research questions given their experience in the use of alternative data for credit in the SME landscape in South Africa. The criteria included:

- Alternative data experience: The participants needed to have had direct involvement/experience in working with alternative data and to have observed various use cases including SME credit.
- Role diversity: Participants represented various roles and perspectives within the fintech ecosystem, from C-suite executives, founders and consultants.
- Year of experience: Participants were required to have had at least 5 years industry experience.
- Established company: The companies represented had to have a presence in South Africa of at least 5 years and they needed to offer credit scoring or lending products directly or indirectly.

The identified sample provided both geographical diversity and experience diversity. Demographic profiling was not necessarily required for this study as insights sought had no bearing on demographic outcomes for this research project.

3.6 The research instrument

This research employed a semi-structured interview guide as its research instrument. The interview guide comprised of a set of 14 open-ended questions.

The guide was structured into sections, each addressing specific aspects of the research study. These sections included:

- Participant introduction: Incorporating introductory remarks to initiate the interview and give an overview.
- Benefits of Alternative Data for SMEs: Investigating the observed advantages and implications of utilising alternative data for SMEs in South Africa. These benefits were from a borrower and lender perspective.
- Types of Alternative Data Used: Exploring the various categories and sources of alternative data employed by fintech lenders and related businesses.
- Information Asymmetry: Probing the concept of information asymmetry and its relevance to credit assessment in the SME sector.
- Subjectivity in Credit Decisioning: Investigating the role of subjectivity in fintech credit decision-making and its implications.
- Signalling and screening: Examining the signalling and screening mechanisms used by lenders when assessing SMEs for credit using alternative data.

These interviews, lasted between 45 minutes to 120 minutes, with the average interview time being 90 minutes. Having a standardised guide ensured consistency across all interviews. Semi-structured interviews were chosen due to their flexibility, enabling the researcher to probe further as the participants responded and delved deeper into the participants' experiences, perceptions and opinions (Peters & Halcomb, 2015). The implication was that the structured interview guide was not strictly adhered to, allowing for open-ended questions with the participants.

From the literature and theoretical framework, this study addressed three key concepts derived from information asymmetry. These theoretical concepts were – the lemons market, signalling and screening. There were two research questions explored in this study.

The first question: How are fintechs transforming the SME credit landscape in South Africa by using alternative data?

Questions 2 to 7 of the interview guide (Appendix C) addressed this research question. The goal was to unpack the current uses of alternative data in the SME credit market. These questions looked at benefits for the fintech as well as benefits for the SME. Considering that the participants were industry experts, a common understanding of the common alternative data points used in South Africa was established. It was felt that this information could help SMEs understand the value of alternative data and tap into the credit market offered by fintechs.

The second question: How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?

Questions 8 to 14 of the interview guide (Appendix C) addressed this research question. The questions were primarily related to signalling by SMEs and screening. From the literature review, soft information qualifies SME for higher limits. The goal was to unpack whether soft information was regarded as essential for the subjectivity components on credit decisioning. Since the costs of screening SMEs increases as you gather soft information, it was considered essential to understand how fintechs are screening their businesses to obtain the signals they need to make funding lines available to these SMEs.

It is important to note that fintechs do not generate alternative data themselves. The study aimed to understand how they access and analyse such data. In cases where SMEs have thin files and lack traditional data for analysis, the research sought to determine if alternative data were considered a substitute or complement.

3.7 Data collection process

Participants for the interviews were purposefully selected based on their experience, industry role diversity, years of experience and age of the companies they led. The details of potential participants were obtained from the researcher's professional network. Interviews were conducted either face-to-face or through virtual platforms, including Microsoft Teams, Zoom, and Google Meets, depending on participant preferences and logistical feasibility. Prior to commencing each interview, participants were asked whether they were comfortable with the session being recorded. For in-person meetings, Microsoft Teams was used to facilitate the recording, while virtual meetings utilised the respective platform's recording features.

Each interview followed a structured yet flexible approach. It commenced with an overview of the research topic and an introduction to the theoretical framework underpinning the paper. Participants were then invited to introduce themselves, emphasising their professional experience in the field and academic background. Subsequently, the interview guide was used, but its sequence was adapted to the natural flow of conversation with each participant, allowing for a more natural and insightful discussion.

In order to ensure comprehensive data capture, all interviews were transcribed within seven days of their completion. During the transcription process, attention was paid to accuracy, and any discrepancies between the participants' spoken words and written transcripts were rectified to maintain data integrity. Transcriptions were then carefully anonymised to safeguard the identities of the participants.

At the beginning and closing of the interview, participants were guaranteed confidentiality (Susilo et al., 2021). This ethical principle was upheld throughout the data collection process.

Data collection occurred over several months to accommodate the schedules and availability of the participants. All collected data, including interview recordings and transcripts, were securely stored in OneDrive (a cloud-based storage service) and a password protected computer.

3.8 Data analysis strategies and interpretation

A thematic data analysis approach was followed as it is useful to identify patterns in relation to the participants' experience, views and perspectives (Clarke & Braun, 2017). Data collection and preliminary analysis occurred concurrently (Peel, 2020). The framework for data analysis as laid out by Peel (2020) included: (1) data collection, (2) engaging with the data; (3) coding the data; (4) generating the code categories; (5) conceptualising the themes from the data; and (6) contextualising and representing the findings.

- *Data Collection:* A few days after each interview, the audio and video recordings were transcribed word-for-word. This process was crucial to ensure that participants' responses were documented with accuracy and completeness. This process was important for maintaining data integrity and providing a rich dataset for subsequent analysis.
- *Engaging the data:* All transcribed materials were sorted and incorporated within a consolidated master document. The answers were then organised by question to identify themes. All transcripts were read at least four times and throughout the data collection, there was extensive cross referencing.
- *Coding the data:* Themes were then identified and coded using Atlas.ti. The coding was an iterative process which included clear descriptions to maintain the consistency as the data were being gathered.
- *Generating the code categories:* After coding the data, code categories were generated. This process involved grouping the recurring themes and grouping them into categories that gave insights into the research questions.
- *Conceptualising the themes from the data:* This step involved examining the categories generated and identifying common or interconnected elements of the participants' answers to build concepts.
- *Contextualising and representing the findings:* The apparent findings from the data were presented and supported by respondent extracts.

3.9 Possible limitations and challenges of the study

Sample size: The fintech lending industry in South Africa is still in its infancy and is dominated by a few industry players.

Intellectual property risk: Some participants were hesitant to disclose detailed information about their business models, due to concerns about potentially providing competitors with a competitive advantage. Other potential participants declined the interview requests for the same reasons.

Researcher bias: As an active participant in the industry through professional experience, the researcher may have had personal biases that influenced interpretation of the data. Through some of the work the researcher has done, some of the interviewed participants were colleagues in the industry. This shared experience in alternative data between the researcher and the participants might have led to an inclination to place undue weight on certain themes. The need for researchers to be objective is highlighted as critical during analysis. Hence, a clear trail of the evidence that emerges from research conducted needs to be available for validity and interrater reliability (Xu & Zammit, 2020).

Potential Bias with participants known to the researcher: The researcher had prior knowledge of some participant details; however, measures were taken to mitigate any potential bias. These measures included ensuring that participation was voluntary and obtaining ethical approval before commencement of the interviews. The researcher continuously reflected on their objectivity to maintain the integrity of the findings.

Qualitative analysis: This study did not delve deeply into the interest rates charged by fintech lenders and the impact of the credit access on the sustainability of the SMEs. In addition, the success of the fintechs in terms of credit loss ratios was not analysed, as these data are not publicly available.

Traditional lenders: This study focused on fintechs and their use of alternative data. Traditional lenders that have pivoted to fintech strategies were not interviewed for this study. Moreover, traditional lenders were not interviewed to understand their credit assessment methods, as sufficient information was obtained from the literature review.

3.10 Trustworthiness and quality assurance

3.10.1 Credibility

Credibility refers to the trustworthiness and plausibility of the research findings (Tracy, 2010). In order to further illustrate the credibility of the research, when the overarching perspectives of different participants were compared, the conclusions they provided, converged. This can be described as qualitative triangulation. Qualitative triangulation assumes that if two or more sources of data converge on the same conclusion, that makes the conclusion more credible (Tracy, 2010). In this study, purposive sampling was chosen which resulted in the selection of experienced and active participants from the fintech industry. This sample provided various perspectives and contributed to more in-depth insights from the study (Graneheim & Lundman, 2004) as the participants were from different fields, ages and experience levels. Multivocality takes place when differences in experience, class or age can be the basis for very different meanings in the field, and credibility is enhanced (Tracy, 2010).

3.10.2 Dependability

Dependability in research refers to the ability to obtain consistent results if the study were to be repeated, and is achieved through the use of consistent methods and processes during data collection (J. M. Morse, 2015; Singh et al., 2021). Every step, including the recording process, transcription, and the strategies for data analysis, have been meticulously documented. Moreover, the utilization of Atlas.ti as the coding tool not only streamlined the process but also ensured a comprehensive cloud-based audit trail. Archiving all raw data records, regardless of the data collection method used, provided an essential audit trail and served as a benchmark to assess the adequacy of subsequent data analysis and interpretations (Nowell et al., 2017).

3.10.3 Transferability

Transferability in research describes the extent to which sufficient information is provided for readers to evaluate the relevance of findings to other contexts (Stalmeijer et al., 2024). In this study, the research gave detailed descriptions of the research context and participants background to enable other researchers to evaluate the applicability of the findings in other data rich and technology driven contexts. Alternative data are generated in various contexts, and this study highlights the power of data analytics to make informed decisions, which would resonate with professionals across different fields. The participants in the study provided different perspectives on how they see alternative data being accessed and used. The participants discussed in detail what reliable alternative data is which bodes true across contexts. The study also discussed theoretical frameworks, such as the lemon market principle, which provide a broader conceptual basis for understanding the dynamics of any market where there is a buyer and a seller.

3.10.4 Confirmability

Confirmability is established when credibility, transferability, and dependability are all achieved (Nowell et al., 2017). It involves reflexive analysis, where a researcher is aware of and discussing one’s own bias and influence on research process and outcomes (Singh et al., 2021). In this study, the researcher documented every step of the data collection process thereby creating an essential audit trail. The researcher also discussed personal biases and potential participant biases as limitations to the study. This reflexive exercise was done to enhance the confirmability of the study.

3.11 Ethical considerations

This study received approval from the Wits Business School Ethics Committee, and an ethics clearance certificate was obtained before any data collection commenced. For reference, please see the certificate in Appendix D. Participants were advised that participation in the study was entirely voluntary and had the opportunity to withdraw at any point. None of the participants interviewed withdrew from the study.

Importantly, levels of confidentiality and anonymity of the participants were fully disclosed to each respondent.

As far as confidentiality is concerned:

- During the data interpretation process, no participant responses were identifiable and disclosed to other participants.
- If any participant disclosed information that they explicitly stated should be off the record, the information was excluded.

As far as anonymity is concerned:

- The interviews were conducted face-to-face or using video conferencing facilities. The identities of the participants were known to the researcher and are not disclosed in the final report.

CHAPTER 4. RESEARCH FINDINGS

4.1 Introduction

This chapter presents the findings of the study, obtained through analysing the data collected through the interviews. The persons who participated in the interviews comprised a diverse group of professionals, each offering a unique perspective based on their expertise in either the fintech or traditional lending landscape, with a particular focus on the SME sector.

The chapter commences by providing background information about the participants, ensuring that their identities are protected while offering context to their valuable insights. This is followed by an overview of the coding process employed during data analysis and an explanation of how the themes were developed from these codes.

In the third section, we delve into each identified theme, elaborating on its implications for the research questions. The themes are presented with supporting evidence from the interviews, grounding the findings in the participants' first-hand experiences and insights.

4.2 Background information on participants

Table 4 shows a high-level profile of the participants. All participants were assigned unique identifiers in the order in which their interviews were conducted (P001, P002, etc). The unique identifier ensured that anonymity and confidentiality were maintained, in accordance with the consent provided by the participants.

Table 4 shows that the sample participants are made up of senior executives at 62% and founders of companies at 38%. The average age of the participants was 46 years with an average experience of 18 years working with alternative data across diverse industries.

Table 4: Participant profiles

Participant ID	Gender	Alternative Data Experience	Position	Type of Organisation
P001	Male	20+ years	Head of Digital Product	Fintech
P002	Male	20+ years	Chief Risk Officer	Fintech
P003	Male	10+ years	Managing Director	Fintech
P004	Male	20+ years	Advisor (Financial Services)	Traditional
P005	Female	5+ years	Chief Product Officer	Fintech
P006	Male	10+ years	Head of Africa	Fintech
P007	Male	20+ years	Director	Traditional
P008	Female	5+ years	Co-founder	Fintech
P009	Male	10+ years	General Manager	Fintech
P010	Female	5+ years	Co-founder and CEO	Fintech
P011	Male	20+ years	Head of Credit Analytics	Fintech
P012	Male	20+ years	Chief Innovation Officer	Traditional
P013	Male	20+ years	Chief of Decision Analytics	Traditional

The industries to which the various participants belonged were distributed as follows:

- 31% of the participants represented fintech lenders that utilise alternative data sets to provide credit to SMEs in South Africa.
- 31% were affiliated with credit bureaus that leverage both alternative and/or traditional data to construct comprehensive and commercial scoring models for SMEs and consumers.
- 23% worked in businesses that service SMEs and generate a plethora of data points that can be classified as alternative. These businesses offer embedded funding or credit options to SMEs utilising the alternative data they generated.
- The remaining 15% were consultants who collaborate with fintech and traditional providers across Africa and Europe, advising on the use of data to enhance credit accessibility for SMEs.

This diversity of the participants offered comprehensive and multi-faceted insights into the research questions, even though the perspectives were not exhaustive. The participants for this research were predominantly male, which explains the unequal representation of females in the study. Although a more balanced gender distribution might have provided a wider range of perspectives, the lack of gender parity does not affect the validity of the findings.

4.3 Thematic coding analysis

The transcripts of the interviews were analysed using a thematic analysis/coding process and a deductive coding method was followed. In a deductive approach, the theoretical framework and the research questions are applied to develop the codes and then subsequently attach the codes to the text (Xu & Zammit, 2020). Microsoft Word was used for manual and first level coding, with broad topics identified. Second level analysis was digital using Atlas ti.

Each new idea or concept was assigned a unique code, and similar ideas or concepts were grouped under the same code. Initially, 25 anchor codes were identified. These initial codes are the smallest unit of analysis that capture interesting features of the data (Clarke & Braun, 2017) and the participant views (Xu & Zammit, 2020). From these codes, categories were developed to cluster and group the various codes together. This led to the identification of key themes emerging, which captured the core insights from the data. The themes shed light on the research questions and offered meaningful patterns to interpret the findings.

The themes that emerged from the data were analysed and juxtaposed against the literature review and theoretical framework (Xu & Zammit, 2020).

Table 5 lists the codes identified, categories developed and the respective themes that emerged.

Table 5: Theme identified

Count	Codes	Categories	Theme Identified
1	Data analysis techniques	Data Analysis Techniques and Verifiability	Fintech Innovation (Theme 1)
2	Verifiability, Reliability, Accuracy of Data		
3	Alternative Data (defined)	Data-Related Concepts	
4	Alternative Data Type		
5	Data acquisition and access		
6	Data types and sources		

7	Hard Data		
8	Soft Data		
9	Traditional Data (defined)		
10	Access to Credit	Fintech and Lending Practices	
11	Advantages: Fintech and Alternative Data		
12	Fintech business models		
13	Traditional Lenders vs Fintechs		
14	Adverse selection	Credit Assessment Concepts	Data driven transformation (Theme 2) Objective-Subjective Relationship (Theme 3) Mitigating Asymmetry (Theme 4) Signalling Creditworthiness (Theme 5) Screening Enhancement (Theme 6)
15	Credit assessment		
16	Creditworthiness		
17	Screening		
18	Signalling		
19	Substitute or complement		
20	Information asymmetry		
21	Interest Rates and Pricing		
22	Subjectivity		

4.4 Presentation of findings

4.4.1 Findings based on RQ1

RQ1: How are fintechs transforming the SME credit landscape in South Africa by using alternative data?

The findings in this qualitative research paper reveal that the fintech SME landscape in South Africa is evolving. It highlights the pioneering innovative approaches of fintechs to using alternative data as well as the profound impact of alternative data in reshaping credit assessments.

4.4.1.1 Innovative approaches and advantages of fintechs in SME Lending

The findings highlight that given the scarcity of funding for SMEs, fintechs have innovated in the credit space. They have developed products that exhibit a more nuanced understanding of the heterogeneity of SMEs. They offer tailored funding solutions through innovative ways of gathering alternative data.

Participant P003 emphasised the following aspect of alternative data:

> *"Alternative data as a concept is about using information in a way that improves the entire credit experience of SMEs". – P003*

Participant P012 was of the view that:

> *"From a lender perspective, the big role of alternative data is to assist lenders make credit decisions where there is very little data to start with. This could be due to the bureau infrastructure being limited or generally because there is not much traditional data lying around". – P012*

A similar viewpoint was expressed by P013.

Alternative data allow fintech credit providers to service markets that have traditionally been underserved by conventional credit methodologies such as micro enterprises. This sentiment was echoed by participants P002, P003, P007 and P011.

Below are a few comments from participants on the innovative uses of alternative data:

> *"We call it embedded finance. It becomes everywhere, if you need a bit of money for a short space of time, somebody will have scored you and will [have] an offer for you, using whatever data they found that works for their particular predictive model*." – P001

The participants described how SMEs are provided with credit offers without actively getting involved in the application process. This scenario is due to big data being used to score large numbers of merchants:

> *"You want something that happens without SMEs getting involved, you just get the offer. So that is possible when you start talking to the guys who are providing services to the SMEs and you agree that you have developed a mechanism to score headless data…the headless data that you've been given access to, allows you to*

> *score tens of thousands of merchants…there's no application process, it's just something that's available to your merchants as and when they need it."* – P001

Participant P001 discussed the improved application processes, emphasising the seamless and convenient nature of application processes. Participant P004 was of the view that fintechs using alternative data have fundamentally transformed the uncertainty in the customer's mind regarding the journey of accessing credit.

This viewpoint was echoed by participants P001, P003, P010 and P012:

> *"One of the things that we changed when we redeveloped the digital version of the merchant cash advance product was, there was no application process. We managed to get information which was completely and utterly anonymous but represented a business in its entirety…" – P001*

> *"The process has changed to using your trade data. Based on that [trade data], here's your offer, which means you make the decision when you want to take the funding, how you want to use it. So power moves to the small business" – P003*

> *"The secondary value of alternative scoring for SME, its certainly is a big pain point, most of them just don't know how to have the business information in the one place. Being able to help them aggregate all the stuff in one place and find it easier when it's in one place is also becoming an interesting value for these SMEs, so that essentially means that they now have one place to apply for financing" – P010*

> *"it's using alternative data [as a lender] to pre-empt what you're looking for versus you going out there and looking for something" – P003*

> *"SMEs do not trust credit providers, because they get turned down and embarrassed when they ask for money, whether they are a consumer or they are a business" – P012*

Participants P012 and P013 stated that SMEs are often reluctant to apply for funding due to negative experiences associated with the process. The idea of using alternative data changed the application process from a pull process to a push process which fundamentally transformed credit from a bad actor experience to a good actor experience, as articulated by participant P001.

Participant P008 highlighted the importance of innovating while keeping certain things constant. P008 spoke of integrating into an existing service in the markets in which they operate.

> *"Our solution is end-to-end on WhatsApp, so there is no app, there's nothing of that sort…. it's just a WhatsApp end-to-end solution, used to collect data which is verifiable by credit records and information that's available online and then some of it is verifiable against aggregated domain knowledge, some against social media. So we kind of get the soft data and then look at ways in which we can verify it ourselves without having to make the person upskill themselves to be able to meet that." – P008*

> *"The worst thing you'd want to do from a distribution perspective is to take people away from an ecosystem that they are familiar with and try to build something for them." – P008*

This sentiment was echoed by Participant P009 who highlighted how fintechs evolve to adapt to the changing needs of the SMEs:

> *"With the rise of digital commerce or ecommerce, there was a need to borrow particularly at the point of sale, as an example, and with digital commerce blossoming as it has, the various fintech players were able to fill that gap" – P009*

The participants further highlighted fintech innovative capabilities in leveraging real time information in predicting propensity to pay.

> *"…Now that you have readily available real time information that can be passed on to the lender, that changes the game versus waiting for information or collecting information [from the merchant]." – P002*

> *"We are a fintech that enables small businesses that cannot be scored with traditional credit scores to access financing better. How we do that is we essentially have built and alternative credit scoring infrastructure that allows us to collect cash flow sources directly from … multiple filters that SMEs use." – P010*

> *"Because it's real time data that's been provided, you want to take the funding. That fear of having to apply for funding and then being rejected has been removed…." – P003*

Participants P005 and P006, who were part of businesses that generate substantial amounts of data, highlighted the availability of real-time data on their respective platforms, which fintech lenders utilise for immediate scoring:

"We track actual stock, so we're a stock management solution and we can see which stock is moving fast, which stock is not moving. Ultimately we want to make the data available to multiple lenders with different kinds of products or lending products because the different traders are at different phases of their business and so their requirement is very different" – P005

"For the optionality of getting immediate payment on your invoices… SMEs can implement a supply chain finance programme interfacing with buyers ERP system,… and every single time a supplier delivers and invoices the corporate, their invoice comes into the platform, so they also get a working capital injection because now they can get their money almost immediately, like within five days of the invoice being approved" – P006

P005 echoed the idea that the same alternative data can be used for different types of credit.

In addition, the research revealed the significant benefits of utilising real-time data obtained from channels outside of the SME owners. Participants P001, P006 and P009 notably highlighted the inherent difficulty in manipulating alternative data:

"When using alternative data, you have to actually set up a bank account and simulate activity for 180 to 360 days before you can even start to get credit. That makes it a lot more difficult than just filling in a paper form with some dodgy details that can't be checked anyway" – P001

"In supply chain finance, [real time] alternative data, derisks the entire ecosystem, and removes the chance of fraud risk" – P006

"What alternative data does allow, is capability to rely on multiple sources of data. The chances of all of them being incorrect is lower than when you're relying on a single source. If there is an error in that or whatever the case might be it becomes difficult to create that" – P009

The real time nature of the data was discussed at lenghty by Participant P009. Essentially, P010, P012 and P013 emphasised the near real-time quality of these data. This characteristic makes the data less stale than the information that traditional bureaus use, and that aspect is quite useful.

"I think alternative data is going to enhance models, is it takes a view of you right now, it looks at your transaction behaviour over time, so there's velocity of changes in transaction behaviour, changes in spending behaviour, changes in purchase behaviour, changes in web data insights on, are there more complaints or less complaints in the last two days, which a traditonal model will never pick up" – P012

The participants unanimously emphasised the progress achieved by leveraging alternative data for working capital solutions. However, some believed that the adoption of alternative data remains in its early stages within the South African market:

"In general, very nascent usage of non-traditional data. But in the POS space actually if we're specific to that use case, yes, I'm excited that we're advancing credit, so that whole Square, Yoco, iKhokha, advancing, which is really working capital funding, I think that's innovative and that starts us moving in the right direction but, yes, in general though it's still very nascent." – P009

Participant P009 further pointed out that while point of sale lending has been a success, it has been limited to working capital credit and not other types of credit.

"..the POS device lending solutions, that's a start, and it helps a lot in terms of very viable SMEs accessing credit, but really that's working capital. So that's one type of capital. There's other types of capital that a lot of these SMEs actually may warrant, so growth capital and all other types, to fund their business, that's where they may not necessarily have financial statements but you may be able to consider other attributes, holistically. So we haven't seen that happening. We're seeing movement but we haven't reached critical mass…" – P009

There has been significant growth of unsecured credit, which has now become a critical source of funding for SMEs. The use of alternative data helps mitigate structural biases such as race, gender and location, within the SME population. This benefit was also strongly expressed by participant P003.

"The number one key differentiator or key important factor…with alternative data, particularly how we use it, which is anonymised, there's no bias in the method, there

is no bias in the population…We use only trade data of that person. So you remove any bias from the performance and through these patterns of looking only at trade data, you can then give funding to merchants that never qualified before" – P003

Participants underscored the substantial influence that the regulatory landscape plays in shaping the fintech industry and its ability to leverage alternative data. One participant noted:

"In the UK, its obviously a regulatory lend model, they rolled out open banking regulation which then meant essentially organisations are compelled to share customer data if the customer consents. So that obviously then created a huge opportunity for alternative lenders, other fintech players to access data and therefore innovative in terms of propositions that weren't being offered by the incumbents. So I think that's what also triggered that market to essentially be a lot more open to sharing data via open API standards. So in South Africa it's been more an incumbent led model rather than a regulatory lend model because regulations are still behind and so then data hoarding, unwillingness to share data and therefore information asymmetry still persist. So the organisation with the most data then feels like they have the upper hand but actually, really the whole point of the regulation is, that's not your data, that's customers' data, you shouldn't really be holding it. So that's where I think we are still moving into. We are at a slower pace and partly because there's no regulation to serve as a burning platform." – P009

Participants highlighted the increased financial inclusion from alternative data uses –

"The good thing about alternative data, it brings underserved SMEs into the more formal data world, not formal business world but formal data world… even though the company is still an informal business, it operates as a formal business, the data you have on the business is similar to the data you have on a medium sized business." – P012

"We know there's a huge proportion of people that we lend, it's their first real journey into credit…so I do not think they would really be any debate around that. The service that we provide and our competitors provide is incredibly necessary and clearly the mainstream lenders are not in that space at all, are not interested in it. I don't think there's any debate about that" – P002

“Because of alternative lenders, because of access to funding, we are actually formalising informal markets by default” – P003

“For financial inclusion it’s really important because with alternative data you get information about the company which the company cannot provide through traditional means...So that’s where alternative data tells how financial inclusion can really open up the spectrum of lending to a great extent” – P007

“Alternative data allows you to open up opportunities that SMEs haven't had before, and also gives the marginalised access to credit” – P011

4.4.1.2 Types of alternative data points for credit scoring (RQ1.1)

RQ1.1: What types of alternative data points do fintechs commonly rely on when credit scoring SMEs?

Definition of alternative data

The definitions of alternative data provided by the participants varied broadly. However, a consistent element across all definitions was the concept that these data were derived from sources other than the SME itself:

“It is data that you wouldn’t typically collect from a customer when you're onboarding them, and/or data that typically goes beyond what you would collect from the customer when onboarding them” – P009

“So, what we would deem as alternative is anything that is over and above this so called publicly available data that we can access from a credit bureau.” – P010

“Alternative data in the banking sense are the things like your turnover and the regularity at which you're able to do that, as well as now, which is very interesting, all the other non-financial data.” – P004

“Alternative data is anything that is non-transactional” – P005.

“It is soft data, that you can verify through social media scraping, in one way, but then also in the way that you collect the data and building in checks, building in truthfulness reports and understanding that data in the context of sectorial knowledge” – P008

In order to contextualise the topic, a few of the participants gave definitions of traditional data:

> *"What we deem traditional credit data sets are what credit bureaus have access to, so your history of credit, mostly from formalised sources like high street lenders like banks." – P010*

> *"Traditional data is obviously all the usual stability kind of questions that you normally get asked on an application form for credit, plus anything that you could derive from a bank statement or an income statement, and the credit bureau." – P002*

Participant P005 offered a unique perspective on the distinction between traditional and alternative data, introducing a viewpoint that diverged notably from the rest:

> *"Traditional data would be transactional, so if it's looking at financial behaviour...However, if you're looking on invoices but you're not looking at the rand value and you're looking at the stock and how often you order the stock, so you can deduce how quickly you sell it, that becomes alternative because it's not specifically on money. Its other factors that can feed a credit model. Traditional data focuses too much on the money side, but we're not focused on the stock and services that you're using that money for" – P005*

Participant P001 in defining alternative data as a data source, highlighted the fact that it has always been available. The methods of harnessing these data, however, have been archaic:

> *"I think that non-traditional data precedes traditional data in the sense that we weren't able to capture the stuff previously, it's always existed. So, the number of friends you have, that's always existed, it's only now that we've got social media that I can actually get the number of friends that you have via WhatsApp, Facebook and Instagram. Lenders are now able to pull information out of your life which has always been there, it just hasn't been recorded, and now it's recorded." – P001*

Types of alternative data (RQ2.3)

RQ2.3: What kinds of alternative data are fintechs utilising to limit issues of information asymmetry in the credit assessment process?

In South Africa, access to finance remains a significant hurdle for small businesses seeking credit, with low approval rates from conventional funding mechanisms (Mbedzi & Simatele, 2020). This gap in the market provides a commercial opportunity for fintechs to leverage alternative data to enhance financial access. A key finding of this research report is the enhancement of financial inclusion and a notably higher approval rate as a result of fintech initiatives.

> *"Fintechs don't tend to talk to the 10% to 15% of the total population that the banking sector tends to talk to. So the first thing is that we tend to talk to a much broader swathe of the population and that's possible because we're using data that the banking sector doesn't have access to." – P001*

Fintechs are employing a wide range of alternative data types, each playing a unique role in assessing an SME's creditworthiness. The participants gave examples of alternative data types:

> *"Behavioural scoring and bureau score is traditional data and demographic is also traditional data. Psychometric, the mobile, the online data and geospatial, those are all alternative data" – P011*

> *"Any sort of purchase, stock inventory purchase data could be useful, in theory, phone, telco data could be useful, obviously the point of sale." – P002*

> *"Stokvel data is alternative data. Stokvel market, is by far the biggest credit product that we have and is run completely outside of the traditional institutions." – P001*

> *"Currently common in the industry is payment data. This is well established actually becoming quite mature and established" – P002*

> *"I look at small businesses and so from a business perspective I think you can draw a lot of other indicators, like location and foot floor around the location, rather than purely from your bank account statement, especially when you know it's a high cash flow area." – P005*

Other examples of alternative data for which participants have started to see applications in the South African market, include tracking safe drops, supplier purchases, social media reviews, marketing engagement, accounting platforms, insurance data and cell tower information.

As fintechs evolve in SME lending, there has been a shift from traditional sources of data like bank statements into more agile and innovative sourcing strategies:

> *"We pivoted towards not only not using bank statements as much but using readily available information that was available online, always available…" – P002*

> *"We got a list of every single transaction that business was doing from a card perspective, and that was sufficient data for us to be able to score that business…" – P001*

Role of alternative data

Participant P004 challenged the notion that the use of alternative data is the exclusive domain of fintechs. He argued that traditional businesses are also starting to utilise alternative data:

> *"In defining alternative data, it is standard operating procedure for anybody who has an extended supply chain to proactively offer credit to that supply chain using alternative data, and they don't have to be a fintech." – P004*

This notion was echoed by participant P005, referencing how a traditional corporate is using alternative data:

> *"What South African Breweries does differently though, I think, and this is really alternative, is that they have sales representatives that go out to look at stores and based on location, store setup, they sponsor fridges. I think that is alternative data because you're actually looking at aesthetics of your store and interior design [for potential]" – P005*

The key driver for fintechs using alternative data was captured by P002 and P011:

> *"The key driver for using alternative data is either to get better risk models or specifically, just a general evolutionary thing, how do we get better discrimination in terms of being able to make risk decisions, but probably more importantly and more pertinently, it's to open up markets that are not traditionally served or not served by the traditional means" – P002*

“The big role of alternative data is around trying to assist you in making a decision where there is very little data to start with, that might mean because bureau infrastructure is limited… it’s general because there’s not much traditional data lying around.” – P011

Participant P011 discussed the success of leveraging psychometric scores as alternative data in SME credit assessments. The participant emphasised that psychometric tests can be used to tilt decisions:

"You can put in some buffers with the high risk customers based on psychometrics, generally, remember, if we split that risk up where we stated at the beginning, the risk of the customer, the risk of the company, the risk of the industry, the risk of the region, and then affordability, yes, you can combine them a little bit but generally your affordability decision must be fairly strict around your standard affordability rules, you can tweak a little bit with risk, just open up basically your exposure to the low risk people and close your exposure down to the high risk people” – P011

Online data and geospatial data were also mentioned as types of alternative data:

“Geospatial score is below twenty and then the online data is now above mid-teens. Mobile data scores also can be quite weak depending on how much you use your phone and what you are scraping from the phone but can be as high as sort of late twenties, maybe early thirties.” – P011

P007 highlighted the fact that some alternative data points are particularly useful for assessing short-term loans:

“…you just need to look at the bank statement to see income or the mobile phone statements, what money comes in, what money goes out, and then you know that within three months the chances of your loan being repaid are pretty high…’’ – P007

Collection of alternative data

The incumbent fintech providers pioneered the process of obtaining alternative data through channels other than the SMEs themselves. In addition, this collection of data by SMEs is important for signalling and screening of SME data:

“We were the first to partner with service providers within the SME industry and use their data as a credit risk instrument. We partnered with corporates that actually

process POS transactions … in some instances we are literally dealing with a machine that you are pushing the credit cards through on your counter top" – P001

"As a data company, we are generating lots of data. Any data is always good data, depending on how you collect it. So, it's the method in collecting and passing clean data. …I think for us it's very easy to align with people that need to use the data and it's easy for us to find alignment on how we get that data passed." – P005

"We use geospatial with non-SACRRA members and non-credit clients that we work with, and basically you can have models that say, listen, I've got no credit record of you but I know that you are like your neighbours, and there's some sort of correlation there that you can use which can be done at a very fine level" – P011

Participant P005 further reiterated that only anonymous data is shared with third party credit providers.

P001, P009 and P010 covered how fintechs have become smart in collecting data from multiple sources.

"By partnering with SME service providers, you've got real time access to the data, potentially, and that is game changing, because certainly in some businesses that credit decision needs to be made within the space of a half an hour." – P001

"Partnerships are valid insofar as it relates to open APIs and data sharing …there's a lot of data sharing going on via open APIs, and so that then allows a lot of these companies access to data at scale" – P009

The participants drew attention to new data verticals that could be mined in time and would be revealing of the stability of the business:

"So WhatsApp business side is becoming interesting and potentially means that they're going to be seeing the data that we're all interacting with, with all of these companies that have now sat themselves on WhatsApp, it won't be financial, so payments are not yet a thing, but they can certainly start gauging level of customer complaints, good versus bad reviews, how much interaction there is, so again useful data for a company." – P001

"We are scraping web data, so internet data on businesses. The data is not financial data, it is data such as do they have a website, do they have a Facebook page, how

many complaints are there on Hellopeter on this business, how many kudos on there on Hellopeter, are there any press releases or press clippings on this company, has that press release been positive or negative, as an example, and seeing what the predictive value is of that data…. It's a bit of a different approach and we have tested this in developed markets" – P013

The distinct aspect of mobile data collection merits individual discussion:

"You can launch a mobile app and a person's mobile phone is incredibly rich from an information perspective, and a lot of players are now looking at using the data on your mobile phone to build up a credit model for you which allows lenders to extend the appropriate amount of credit based on the interactions happening on your phone" – P001

"Tracking of mobile behaviour has become a common trend by fintechs. For example, Mr Price has a cash customer app were one can get specials. You download the app, a year later apply for credit, they could use your behaviour on that phone by having code on the app that actually tracks what other apps you have, how much airtime you use, how much data you use, and various other behavioural information around how you use your phone" – P011

On the validity of the alternative data, the participants believed that it matters where the data is coming from and whether it is susceptible to manipulation:

"I think where the data or the data is coming from or how it's collected doesn't necessarily affect the fact that it is alternative. What it does though, is it just affects the validity of it. So, you can have all these different data sources or data sets, but one would weigh more because of how verifiable it is" – P010

"There will always be an element of trust and reliance on the data being accurate. I think the key data that we would rely on from a partner in terms of making risk decisions, would also be the bread and butter of the partner and SME. So, they wouldn't be in business if that data was distorted or in any way inaccurate." – P002

"Banks themselves do due diligence on all of the companies that they partner with as a value-added services layer, to make sure that the bad actors get weeded out very quickly. So, the guy's you are dealing with are kind of accredited by the banking system, and that also promotes trust in the system." – P001

All of the aforementioned points were echoed by both Participants P003 and P010.

Participant P006 emphasised that as an aggregator of alternative data within the supply chain industry, they play a critical role as a third party in the ecosystem of confirming the veracity of the information that fintechs can rely on.

An interesting perspective of the validity of alternative data and how it is generated was provided by participant P002:

> *"It's not because its alternative data, that gives you the benefit, it's just the fact that it is ongoing [continuously] available. Ease of access to the data is what's giving the benefit there." – P002*

4.4.1.3 Complement or substitute for traditional data (RQ1.2)

RQ1.2: How does alternative data complement or substitute traditional credit data in evaluating an SME's creditworthiness?

The opinions of the participants varied on whether alternative data should be viewed as a complement to or a substitute for traditional credit data. The majority seemed to lean more towards the view of alternative data as a complement.

Participant P002, offered a distinctive perspective, questioning the pace at which alternative data is evolving –

> *"Now, which is the better thing, to have something that's a little bit old fashioned but stable and lasts longer or something that is just constantly iterating." – P002*

P003 noted that in some cases, fintechs have not replaced traditional data but are instead using the same data more efficiently –

> *"We took bank statement information only to assess the customer, and then of course everything improved. Turnaround time improved, approval rate improved, customer experience improved, but it didn't leapfrog from where it was, it became more efficient."* – P003

Participant P003 further noted that fintech lenders have streamlined the loan application process, reducing the turnaround time from weeks or months to merely hours or days.

Participants expressed views on alternative data being a substitute:

> *"I believe it's a substitute… it becomes difficult to mix the models together because by doing that you're then taking advantages and disadvantages on both sides away and that's what we actually find practically." – P003*

> *"Alternative data can be a substitute. I would hope that most fintechs who do this well rely on more than one source. When you are relying on multiple sources the chances of all them being incorrect is lower than when you're relying on a single source" – P010*

> *"We think on companies that we've got no data on we can build a score on alternative data only, we don't think … we are currently proving it, we are building those scores. Yes, you will always, especially on the smaller businesses, the individual, the owner actually, that data also comes into play, but we are definitely building a score based on alternative data." – P012*

P008 and P011 acknowledged that traditional credit scores from bureaus do provide some value (*"I don't think they're totally rubbish"*), especially for individuals who already have credit histories. However, they also pointed out the limitations of these traditional scores, suggesting they do not cater for certain personal factors or situations and in instances where there are no credit scores. In those instances, it can be a substitute.

Other participants argued that alternative data is a complement:

> *"In consumer credit, still relatively low value ticket lending, alternative data is a substitute. However, in your traditional SME lending, in terms of volume and size, you would go back to more of a complementary type of setup." – P004*

> *"I don't think alternative data can be a substitute, we still need to overlay something with something and … the more data you have, the better. I don't think one thing can substitute the other, I think it just can simply enhance it, like it can enhance and give you ever greater comfort. So, if you have 50% certainty, you can get to 80% because you've got a more enriched overview" – P006*

“So the use of alternative data replaces and supports traditional data. But there’s going to be a long tail in moving the banks to use alternative data, not because they don't want to, but because from a regulatory environment they're going to have to see how do they incorporate it into this strict regulatory world, whereas from a credit bureau perspective, easier to work with fintech funders that would jump at this opportunity.” – P012

“It’s a complement, I don’t think the market is ready.” – P012

Participant P007 distinguished between the uses of alternative data for short-term versus long-term credit assessments:

“If you want to give a three year loan you’d better have enough data which goes long enough back in order to feed the algorithms to get a credit decision...Three months is something which you can assess with alternative data, where you just need to look at the bank statement to see income or the mobile phone statements, what money comes in, what money goes out, and then you know that within three months the chances of your loan being repaid are pretty high, but you won't have the possibility to assess a three year loan. So inventory finance is possible but real investment or for machinery or whatever else is needed, building, that won't be possible.” – P007

Participants P012 and P013 argued that the South African market is not ready for a full transition from conventional assessment methods to fully alternative. Upon a more thorough examination of this question, it became evident that a one-size-fits-all approach is insufficient. The views of participants, initially varying, began to show notable similarities with cross referencing:

“The smaller and the shorter the term, the more it is a substitute” – P002

“You almost need to pick which one is your core decision model and the other one can be overlaid as a checking mechanism or a method to maybe talk up or down the offer, but it mustn’t affect the entire decision.” – P003

“It’s a hybrid type approach. You would add the weighting of each data set differently. For example, you would weight different data point in the same model, alternative and traditional” – P005

Participant P011 suggested a hybrid approach, particularly when making decisions about applicants who are borderline cases, stating –

“Particularly around margin calls, so those are the sort of middle ground where close to the cut off. You could add this extra level of insight and push it over the line, one way or another, and we’ll therefore accept or decline the person based on the margin call we’re making with the psychometric test.”

Participants P012 and P013 pointed out that adoption of alternative data will vary across different sectors of the lending market. For example, P013 reflected:

"If you think about it in the retail world and in the unsecured lending world, they’ll be adopting this a lot more where the formal banking world will test this but they won't aggressively adopt it."

Participant P010 suggested enriching traditional data with alternative data sources to create a more inclusive and effective credit scoring approach to the extent that the targeted sector is credit scorable. This view provides a balanced perspective for SMEs in the informal sectors where alternative data can act as a substitute.

Theme 1: Fintech innovation

This theme focused on the novel business models employed by fintech companies leveraging alternative data to provide access to credit for SMEs in South Africa. It is evident from the above sections that innovative business models are reshaping the SME landscape. It provides an understanding of how fintechs view alternative data and how they have consumed it to transform credit access for SMEs. The theme illustrates the distinct advantages SMEs experience when securing credit through fintech platforms, as opposed to traditional credit providers.

Participants highlighted advantages such as enhanced financial inclusion, frictionless and seamless lending processes, increased convenience, higher approval rates, more tailored financial products, and the competitive edge gained from these practices. Overall, the theme emphasised the differing paradigms in how fintechs approach credit for SMEs, which has in turn become the competitive edge over conventional finance providers.

The broad codes that came through from this theme when interviewing participants are set out in Table 6.

Table 6: Theme codes

Access to Credit (frequency = 24)
Advantages: Fintech and Alternative Data (frequency = 103)
Fintech business models (frequency = 55)
Traditional Lenders vs Fintechs (frequency = 38)

It is of interest that "advantages" emerged as the most frequent code across the data from the 13 participants interviewed, appearing 103 times.

This finding solidifies its significance in the conversation surrounding fintechs and SME lending. From the co-occurrence of the codes within this theme, a strong link can be seen around how fintech business models leverage the advantages of alternative data to enhance access to credit and differentiate themselves from traditional lenders.

Table 7: Code co-occurrences

	Access to Credit	Advantages: Fintech and Alternative Data	Fintech business models	Traditional Lenders vs Fintechs
Access to Credit	0	2	3	1
Advantages: Fintech and Alternative Data	2	0	14	9
Fintech business models	3	14	0	6
Traditional Lenders vs Fintechs	1	9	6	0

4.4.2 Findings based on RQ2

RQ2: How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?

The questions posed to the participants under RQ2 explored various topics aimed at capturing the fintech landscape in as far as information asymmetry is concerned. There were various interconnected sub-questions that sought to elicit the required insights.

RQ2.1 investigated how subjective measures are incorporated into credit algorithms. RQ2.2 delves into the sustainability indicators that fintechs rely on when using alternative data to make decisions about propensity to pay into the future. RQ2.3 explores the strategies that fintechs employ to limit information asymmetry. In order to understand information asymmetry, the research report looks at adverse selection, signalling and screening in greater detail.

4.4.2.1 Application of alternative data in credit assessment

The view emerged that the application of alternative data by fintechs has been transformative:

> *"Fintechs have fundamentally transform credit from something which is almost like a bit of a bad actor kind of experience" – P001*

The application of alternative data in analysing SMEs represents a needed shift. It fundamentally challenges the core credit principles typically embraced by traditional providers.

> *"The base assumption is that you are in business, and you are living off that business, and to continue to have your livelihood the business needs to continue to work. So, it's a completely different base equation. You're not actually assessing the ability and the willingness; you're assuming that the willingness will be there to pay because it is based on that your business will continue to trade. It is just a different equation and therefore it's a different risk assessment." – P002*

Despite the above, from a South African perspective, several participants pointed out that the application of alternative data is still nascent:

> *"I don't think that we've adopted alternative data fully, I know there's a few POCs that run but in terms of commercialising it, I think it's still a bit in the infancy stage." – P005*

> *"In general, very nascent usage of non-traditional data" – P009*

A factor considered when deciding which types of alternative data to use, is the measure of consistency:

> *"One of the key things in risk management is consistency in process, so if you've got an alternative data source, let's say Facebook, and you do a whole lot of analysis that tells you certain aspects of your Facebook profile, whatever, will be predictive. Now, what if significant portion of your population are not on Facebook?" – P002*

Questions like this one enhance understanding of issues of information asymmetry to be discussed in a section that follows.

In an insightful commentary on integration of alternative data in credit assessment, the participants highlighted how segmentation of SMEs is considered:

> *"When you have real time data of SMEs, you can start looking at various businesses that are similar together and you can do that without knowing anything about the business…You can then start predicting trends on those clusters" – P001*

> *"Access to real time allows you to have more targeted treatment, meaning that you are treating your customers more appropriately." – P002*

> *"We segment similarly risky people together so that ratios should be the same across all the cohorts. For example, Cohort A, 5% liars, cohort B, 5%, cohort C, 5% liars. When using traditional data, there is a lot of randomness" – P001*

> *"You have to understand the industry in the alternative data that you are looking at to ensure your product's suited." – P011*

Participant P002 pointed out the trade off in the application of alternative data. The more data points, the better:

> *"It's a question of probability but it's also a numbers game. If you are dealing with, one SME, your probability is much harder to put a number to than if you're dealing with 100,000 SMEs with similar data that is available on an ongoing basis." – P002*

Real-time monitoring was perceived to allow for more dynamic credit scoring models that are reviewed all the time:

"The inferences are also factoring in real life behaviour data of the collection on the active loans that we are able to monitor." – P011

Participant P011 emphasised how behavioural scores derived from alternative data over time should give a more robust view of creditworthiness than traditional data:

"Behavioural scores are generally your strongest models, they relate to how customers have paid you before. You only have behavioural scores if a customer has been a customer for a significant amount of time, all new customers would not have a behavioural score or repeat customers from long ago wouldn't have the behavioural…" – P011

The participants emphasised the importance of large data quantities to continuously feed their data warehouse and continuously revive credit models:

"A good model is built on data and that data and the Gini will tell you whether this thing separates good and bad well" – P013

Participants pointed out that it is important to test the alternative data credit models against the traditional models to build trust in the ecosystem:

"Alternative data should be validated against the traditional way of doing credit. This way, you can show the strong correlation and predictive power of alternative data" – P013

While traditional data are being used to build trust, there are some alternative data points that are independent of traditional data validations. This aspect was echoed by P004 and P001:

"Drawing from the consumer credit market, Jumo has seemingly not found any use case for traditional bank data in their credit scoring capabilities for the types of customers they are servicing" – P004

"The whole premise of psychometric testing is that psychometric profiling allows you to identify character traits like honesty. One can argue that the predictability of honesty in credit scoring models is a leading indicator in the long term." – P001

The higher the Gini Coefficient, the better the model is at distinguishing between defaulters and non-defaulters (Frunza, 2013). Bureau scores generally give a Gini Coefficient of 45%. Participant P011 discussed the benefit of leveraging psychometric scores as alternative data in SME credit assessments:

> *"Psychometrics score gives a gini coefficient in the mid-30%. We have seen quite wide ranges, from 20% to 40%, but generally the average is about 30% in terms of predictability. The great thing is everybody who takes a psychometric test gets a psychometric score. Under psychometric score, you also get an entrepreneurial score, so we can actually rate the customer's entrepreneurial flare and taking into account various pieces of business acumen. It's like an aptitude test for running your business and everybody who does that test gets a score." – P011*

The strength of alternative data in improving scoring models was discussed by the participants:

> *"Not all businesses with a credit bureau score of 720 are the same. Traditional scoring model are built on historic data, and it doesn't take into consideration a business' life stage right now. Where alternative data is going to enhance models, is it takes a view of the business right now, it looks transaction behaviour over time, velocity of changes in transaction behaviour, changes in spending behaviour, changes in purchase behaviour, changes in web data insights on, are there more complaints or less complaints in the last two days. A traditional model will never pick this up" – P013*

> *"...enable small businesses that cannot be scored with traditional credit scores to access financing better… We essentially have built alternative credit scoring infrastructure that allows us to collect cash flow sources directly from multiple filters that SMEs use. An example would be point of sale devices, their bank fee, mobile money, etc. as a core of the scoring, and then second to that is we basically also collect other alternative data sources which may be a bit more behavioural, so how they maybe interacted with credit before, social scores, how they obviously engage with their customers, if we do have that kind of data, and then also historical behaviour around their credit application behaviour, etc. and that is then, the data that we analyse to assess what their liquidity looks like, what their liquidity risk, credibility risk and default probability would essentially look like for some sort of credit." – P010*

Overall, the participants pointed out that they have observed vastly similar bad debt ratios between traditional and alternative data fintech players. This factor is important considering that alternative data have been opened to previously underserved and higher perceived risk SMEs. Participant P011 mentioned that the heterogeneity of SMEs makes it more challenging to apply credit models to SMEs because of business types, sizes and structures which make it difficult to apply a one-size-fits-all model.

Theme 2: Data driven transformation

The participants provided a diverse view of how fintechs are utilising alternative data in assessing creditworthiness of SMEs. Although, adoption of alternative data is transformative, it is not without its own challenges.

The participants discussed how the transformative nature of fintechs is observed in their approach to credit, by challenging certain core credit principles that are common to conventional credit. Despite these advances, participants emphasised that the integration of alternative data is still in its infancy in South Africa.

The participants discussed the importance of a consistent data approach in processing and sourcing. If the data processing and sourcing is inconsistent, it raises question on information asymmetry. The integration of behavioural and psychometric scores offers an enriched perspective on an SME's creditworthiness, though these too are tied to their own sets of challenges and limitations.

4.4.2.2 Incorporating subjective measures (RQ2.1)

RQ2.1: How do fintechs incorporate subjective measures into their credit algorithms/models, transforming soft data into decision-making criteria?

Participants P012 and P013 emphasised the robustness of data driven models. They pointed out that subjectivity is not superior to alternative data models that are more objective. The objectivity of these models is based on the models being trained on vast amounts of data that have been validated against traditional models:

"A good model is built on data and that data and the Gini will tell you whether this thing separates good and bad well. No person's subjectivity is going to improve on what that model does for you at scale and at pace. So, it's never, ever, ever going to outperform it." – P013

The participants expanded on how a hybrid of traditional and alternative data gives a robust scoring model that requires no subjectivity:

"With alternative data, there's going to be a lesser need for subjectivity, because we're taking historic and real time attributes into consideration immediately." – P012

This idea was echoed by P003 and P011:

"Why scoring works is its objective and generally better than fully subjective views." – P011

"You always get outliers. You need to be, the main aim is majority, 90 percent going through the objective assessment and the 10% that just falls outside of that you need subjectivity." – P003

Participant P008 cautioned against the inherent biases of traditional systems with regard to demography and location, "making things culturally relevant". It is important to ensure that fintechs using alternative data are vigilant in ensuring that their credit models do not reproduce or perpetuate existing biases.

A thread that came through from participant P005, P006 and P007 is that although subjectivity should not be the default, there are certain things that the data do not capture fully such as the operator risk and the structure of the loan:

"If I'm looking only at alternative data, I would not be able to know that your business seeks to expand" – P005

"You visit the site, see how they operate, ivory tower operation or lean operation." – P007

"...with alternative data only, you won't have the possibility to assess a three-year loan." – P006

Theme 3: Objective-subjective relationship

This emergent theme from the interview findings delved into the ways in which fintechs oscillate between the objectivity afforded by data-driven models and the nuanced role that subjectivity continues to play in the credit assessment process. Participants emphasised the superiority of objective data-driven decision making for reliability and scale. There was an acknowledgment of the potential biases associated with traditional systems. The participants noted the need for fintechs to be vigilant, ensuring their models do not replicate or exacerbate existing biases. Despite the emphasis on objectivity, there was a consensus that a degree of subjectivity remains essential. Certain aspects, like operator risk and loan structure, are not fully captured by data alone.

4.4.2.3 Information asymmetry (RQ2.2 and RQ2.3)

RQ2.2: What data points do fintechs depend on as indicators of SMEs' future sustainability?

RQ2.3: What kinds of alternative data are fintechs utilising to limit issues of information asymmetry in the credit assessment process?

Adverse selection

According to the participants, with alternative data the assessment becomes more objective that subjective:

> *"So, there will never be a lemon market because… [of] cohort management…A lot of the lemon problem is you not telling me stuff about stuff that I should know in order to make credit decisions, and given that that is the case, and I can't fix it, I'm going to ignore that stuff. So, I'm not even going to ask you to tell me the stuff that I should know because I know you're going to lie to me if you're a bad actor so I'm not going to ask you. So, I'm going to treat everybody the same and not ask them." – P001*

Participant P001 opted for equality of information in credit assessments, steering clear of potentially untruthful information self-reported by SMEs that could negatively impact lending decisions.

The participants highlighted a pertinent issue between banks and fintech lenders as far as information asymmetry is concerned. Although banks have access to cheaper funding, they do not have the appetite to lend to smaller SMEs that are data thin, while fintech lenders have the capability to address these funding challenges albeit at a higher funding cost.

> *"The banks have access to cheap money and the problem with a lot of fintech models is that, yes, they work but not at the price points that they're able to access money at. So, the cost of getting money for a fintech who isn't a deposit-taking institution is very expensive." – P001*

The participants discussed the fact that when access to funding is expensive, borrowers who are more likely to default are the ones inclined to take out expensive funding products. Alternative data and its integration in credit assessment enables fintech lenders to become efficient in accessing comprehensive data to enhance their credit assessment and limit adverse selection.

> *"I think the other one, going back, to your introductory point about lemons and selection bias, is that if your cost to serve is so low you now change that anti-selection payoff dramatically." – P002*

The participants emphasised that in using alternative data, the importance of obtaining consent from consumers and small businesses is critical to the success of these fintech models.

> *"You're never going to be able to build a score to present to a credit provider like yourself if I don't have the consent of the consumer to use the data or the small business, and similarly, we can have consent of them and build great models, but you don't want to use it, and we don't bring two parties together." – P012*

The above statement suggests that lenders can bridge the information gap by sourcing alternative data from sources that generate this SME data.

The participants pointed out that there is a reluctance by fintech lenders to share data which leads to continued adverse selection.

"We know there's a need to create this data, we also know it's not going to be easy to convince Retail Capital, the telcos, FinFind and all of them to pull their data, they're all going to go, never ever, I invested millions in building this up, I'm sharing it with nobody, and we know it's going to be difficult." – P012

"They [fintech lenders] are creating information. They live on information asymmetry." – P007

Theme 4: Mitigating asymmetry

This emergent theme was covered by the following codes from the interviews – adverse selection, creditworthiness, and credit assessment.

Overall, participants affirmed that lender visibility increases significantly when using alternative data which consequently reduces the issue of adverse selection. The move away from relying primarily on self-reported data has mitigated the risk of "lemons" which ensures a more informed and increasingly visible assessment of SMEs' creditworthiness. Some participants attributed the ability of fintechs to go deeper in their lending segments, to more efficient data utilisation. In the absence of efficiencies, thin file segments will remain underserved by fintechs. The synergy between businesses that generate SME data (e.g. payment acquiring aggregators, payment gateways and telcos) with alternative data fintech lenders was emphasised, showing the necessity of collaboration for comprehensive credit assessments.

Signalling

Participant P007 recognised that with alternative data, there was less reliance on information that was self-reported by the SMEs.

"We do use cash flow data but it's self-reported cash flow data, so we weight it not as highly but even that we benchmark against what we know to be more generalised data" – P007

The challenges of self-reported information were discussed by another participant:

"Is there ever a role for human response? No. Because we can even do a lie detector tests... Anytime I rely on [SMEs] to tell me stuff, if you're really good at it, even I put [SMEs managements or owners] on a lie detector and say you have to tell me the truth, they can beat it. So, we don't have any mechanism whereby we can ascertain that one is absolutely telling the truth. Why? Because there is no truth with human beings. Human beings make up their own truth, that's the nature of their brain." – P001

The participants highlighted the fact that alternative data can be used to look at signals such as entrepreneurship and business context signals:

"Some [alternative data] is more predictive than others but by and large what you're looking for is successful entrepreneurship." – P001

"With alternative data you get information about the company which the company cannot provide through traditional means" – P007

"I think you can draw a lot of other indicators [analysing a retail business], like location and foot floor around the location, rather than purely from your bank account statement, especially when you know it's a high cash flow area." – P005

This idea was echoed by P003.

P011 pointed out that the positive signals for creditworthiness can be observed from SMEs that have a good history of credit interaction:

"...also collect[ing] other alternative data sources which may be a bit more behavioural, so how they maybe interacted with credit before..." – P011

"...historical behaviour around their credit application behaviour..." – P011

P011 further asserted that an absence of a behavioural score for new and dormant customers shows a lack of information which signals higher risk in assessing business creditworthiness.

P010 suggested that personal credit profiles of the owners of SMEs still provide a valuable data point in scoring the business:

"A lot of alternative lenders still consider the personal credit profiles of the owner as a way to score these unscorable SMEs or informal SMEs." – P010

Fintech lenders look for indicators in alternative data that signal the SMEs future trajectory.

> *"Using non-traditional data, you can have a view of the SME's business through a digital funnel, how many of your transactions are dodgy, anomalous, outliers? It's always about statistics and what normal looks like and where you sit in terms of normal. The last one was, have you managed to get a bank account? Because it's hard to lie to the banking system because they've got so much regulation and so much governance that they do a really good job of KYC." – P001*

> *"I think [with alternative data] you can draw a lot of other indicators, like location and foot floor around the location, rather than purely from your bank account statement." – P005*

Participants P012 and P013 both argued that businesses may not consciously realise that by maintaining a positive online presence, responding to customer complaints effectively, or regularly updating their website, are signalling characteristics of the health of a business. P012 articulated his views as follows:

> *"The simple fact of some of those things exist is already a positive, the fact that the company has gone to the effort of actually setting up a website, ability to, in some companies, to order products on that website, even though it's very low level, the fact that they've gone through that effort and made that investment is already an indication of the chances more likely this company will stay in business for longer. A company that's got a website and a Facebook page, because the company only advertises via Facebook, the one with both, again, the chance is more like that company is more professional, it's likely to be longer in business. So those are the types of things we're looking at. On Hellopeter it's more a case of what are the positive versus the negative comments from their customers, because, again, very logically the more complaints there are against your business the more likely it is that that business is not going to make it, especially small businesses." – P012*

Understanding a borrower's willingness to pay is a fundamental aspect of credit assessment, which is why traditional credit providers foster relationships with SMEs to gauge this metric. However, when leveraging alternative data, this willingness to pay is derived through the unique data collection mechanisms:

“By virtue of not asking you to pay me back, by asking your provider to pay me back rather than you, we took the honesty [willingness] equation out of it totally.” – P001

“I think there is no logic behind it [relationships between borrower and lender], they [traditional banks] are just stuck with, that’s now how we lend to commercial … you have to look at their accounts [alternative data] and we have to know the business and it’s a mindset.” – P002

“We then infer the probability of them paying back, and so we took the approach of analysing their transaction behaviour and identify high risk things that could prevent collection. So, as a SME you're doing high withdrawals and gambling or, you know, those kinds of high-risk activities, we would weigh the behaviour as either good or bad and that then determines the collection power.” – P010

Participant P003 also stated that a business’s willingness to pay is higher for an SME than that of a consumer.

It is of interest that several participants highlighted social signalling. This approach predicts an SME's creditworthiness by examining the owner's social circle:

"The biggest characteristics that I found for successful payment of debt was your friends successfully paying debt. So, there’s an old saying, birds of a feather flock together, and the people you hang around with are incredibly predictive of your own credit behaviour." – P001

“I will categorically prove to you that your friends on Facebook and their behaviour is a direct correlation to you. We've proven it. We've proven it with the data. From a regulatory perspective we’re not allowed to use it, we’re not allowed to use it as a scoring mechanism, so you can't, but I can categorically prove to you that there’s a strong correlation between your credit behaviour and your friends’ credit behaviour, the people that you associate with.” – P012

P013 expressed similar sentiments.

The mindset shift of scoring a commercial life is something that was highlighted by the participants as a signalling consideration:

> *“if it’s a business the theory is, its their livelihood. The purpose of the loan is for something material, to keep their livelihood going, not a consumer lending, it’s not consumption” – P002*

Although most fintechs that use alternative data do not take collateral, they admit that the value of property ownership or a long-term rental is a strong signalling tool as discussed by P001.

The value of a having a bank account tended to legitimise a business according to some participants:

> *“The last one was [a bank account]? Because it’s hard to lie to the banking system because they've got so much regulation and so much governance that they actually do a really good job of KYC.” – P001*

Theme 5: Signalling creditworthiness

This emergent theme was covered by the following codes from the interviews – adverse selection, creditworthiness, screening and signalling. From the co-occurrence of the codes within this theme, it was possible to see a strong link between signalling and screening. There were 101 quotations from the participants that spoke directly to signalling.

In general, capacity and willingness to pay are critical tenets of a robust credit assessment criteria. Lenders ultimately look for signals that demonstrate trustworthiness and creditworthiness. The participants were all aligned on the unreliability of relying on self-reported data provided by the SMEs. They only reveal the parts that make lending to them attractive. The participants emphasised that an objective truth is what makes alternative data assessment distinct. It allows fintechs to group lenders into cohorts that have similar characteristics with them making it easier to identify anomalies in an SME’s data profile. The innovation around signalling for SMEs in credit assessment extends to social signals (social media analysis) and operational signals (customer feedback and online presence).

Screening

Participant P011 highlighted the heterogeneity of SMEs and the various factors that must be taken into consideration in analysing them:

> *"SME are heterogeneous, there are many different factors affecting small, medium enterprise loans, and we start with proprietor, the risk of the proprietor, then you go to the risk of the business itself, then you go to the risk of the industry and then potentially the geographic risk as well, so all of those can have a part. Then you go to the affordability and then you also go to the loan amount as well." – P011*

P012 and P013 made it clear that credit screening of SMEs is not limited to financial data, and emphasised the predictive value of web data. As P013 put it:

> *"We are scraping web data, so internet data on businesses. The data is not financial data, it is data such as do they have a website, do they have a Facebook page, how many complaints are there on Hellopeter on this business, how many kudos are there on Hellopeter, are there any press releases or press clippings on this company, has that press release been positive or negative, as an example, and seeing what the predictive value is of that data." – P013*

A lack of collateral, poor information transparency and limited financial information are significant factors contributing to the low approval rates of credit to SMEs from banks (Abbasi et al., 2021; Angilella & Mazzù, 2015; Asah et al., 2020; Mbedzi & Simatele, 2020; Sheng, 2021). SMEs need to provide lots of documentation to access credit. Participant P001 mentioned:

> *"Although you can't go and get a bureau score for a guy selling cigarettes on the side of the road but one can definitely credit score him and present him with a facility, and why should he be excluded from the credit landscape purely because he doesn't have an income statement or a balance sheet? …At the end of the day I'm scoring his commercial life." – P001*

Alternative data in the hands of fintechs have enabled lenders to assess and filter SME borrowers through an expanded range of data. By analysing patterns among similar customers or businesses, lenders have discriminatory power:

“They [SMEs] do leave breadcrumbs, they leave data trails, just not traditional ones.” – P001

"The likelihood of someone coming in with a similar profile, similar [input], etc. is likely to behave similar to that kind of customer." – P011

“Anything that you can show that will add depth to your discriminatory power is got to be looked on...” – P002

The participant suggested that alternative data provides a more granular view of the SME borrowers which improves screening capabilities:

“We've never found a need for the Sharpe ratio or the macroeconomic stuff, because the closer you get to the person generating the data, the less you need those upstream proxies.” – P001

“We benchmark against what we know to be more generalised data, like domain knowledge of the space, so a lot of things outside of cash flows, we looked at behaviour, intentions, volitions of traders, ambition, relationships.” – P007

This granular view, as other participants have highlighted, requires proactive gathering of information to reduce information asymmetry and improve screening as discussed by P011 and P010:

"We essentially have built alternative credit scoring infrastructure that allows us to collect cash flow sources directly from multiple filters that SMEs use. An example would be point of sale devices, their bank fee, mobile money, etc. as a core of the scoring..." – P011

"...the data that we analyse to assess what their liquidity looks like, what their liquidity risk, credibility risk and default probability would essentially look like for some sort of credit." – P011

“When you are relying on multiple sources the chances of all them being incorrect is lower than when you're relying on a single source” – P010

In the absence of a formal profile created by government or credit bureaus, lenders often resort to analysing the individual financial profiles of business owners. As one participant stated:

"Because they're not able to get a profile of an SME or a sole proprietor, they [lenders] would basically traditionally score the owner of the business...you can't really separate the finances of the business from you as an individual". – P010

"…SACCRA has tried to have [SME lenders] to start sharing business information, I think the project kicked off…in 2013. It's been ten years in the making, nothing has happened". – P013

This situation necessitates the examination of the individual's bank statements and financial activities.

In conventional credit, only certain deal sizes can be executed due to the diminishing returns from seeking additional documentation, rendering thin file businesses unprofitable. Participant P002 and P004 emphasised that with alternative data, if they are readily and easily accessible, you should persist in seeking them:

"Of course, there's always a trade-off between the difficulty and the cost associated with gathering the data to the benefit… If that data is readily available and easily available and cheap then it would warrant the analysis, it's almost certain to provide some additional insight" – P002

"Ability to use alternative data to risk score a customer, just by lowering the size of the loan actually improved the probability that the customer would pay back." – P004

Alternative data are typically provided more frequently than traditional data, which has enhanced fraud screening mechanisms for fintech lenders. They can observe an extended track record of these data, thereby reducing the risk of manipulated information. While this scenario is true for various data points, Participant P001 made the following observation specifically in the context of transactional level data:

"They've [fraudsters] now got to actually set up a bank account and simulate activity for 180 to 360 days before they can even start to get credit. That makes it a lot more difficult than just filling in a paper form with some dodgy details that can't be checked anyway." – P001

Participant P006 maintained how important it is in screening to have data verified by multiple sources.

Participants P001 and P002 mentioned consistency of observed data as crucial in screening to assess financial health and stability:

> *"Your turnovers were not going down too badly, so if we saw significant decreases in turnover, we assume that you are lying to us on the basis of you'd come to us for a credit application." – P001*

> *"[alternative data enables] there's no doubt that ongoing scoring, if you like to call it that, but an ongoing calibration of the probability of the customer paying on time and according to expectations is possible." – P002*

The use of alternative data has resulted in the reduction in the cost to acquire, cost to make credit decisions, ongoing monitoring and lowered the operating costs for the lenders thereby giving fintech borrowers the ability to scale and serve a wide array of businesses.

> *"I think that [alternative data] is truly transformational is this ability to reach millions of people at close to zero cost" – P004*

> *"We've scored everything from cigarette sellers on the side of the road using an MPOS device to build up a track record as a business, all the way through to optometrists putting prescriptions through their software every day to again build up a profile on that business and offer them money using that data." – P001*

The supplementary relationship between traditional and alternative data was addressed by certain participants:

> *"Lend our own money and then credit scores become a column rather than the yes or no" – P008*

Participants acknowledged that the use of alternative data, while beneficial, has certain limitations regarding the types of credit it can effectively screen:

> *"Of course, that would be true, once you get into larger advances and more secured advances and all those sorts of things, there's just more that needs to be done in order to make bigger advances, and if you're taking security then there's a whole*

different arrangement. But for short term, unsecured lending, I don't see the difference." – P002

This idea was echoed by participant P007.

An approach that combines alternative data and other credit scoring models is valuable. Participants discussed screening:

"…The great thing is everybody who takes a psychometric test gets a psychometric score. Under psychometric score, you also get an entrepreneurial score, so we can actually rate the customer's entrepreneurial flare and taking into account various pieces of business acumen. It's like an aptitude test for running your business and everybody who does that test gets a score." – P011

"…particularly around margin calls, so those are the sort of middle ground where close to the cut off. You could add this extra level of insight and push it over the line, one way or another, and we'll therefore accept or decline the person based on the margin call we're making with the psychometric test." – P011

Theme 6: Screening enhancement

The responses to interview questions in section 4.4.2.4 revealed the diverse roles that alternative data play in screening SMEs. The emergent theme is enhanced screen capabilities through alternative data integration. Alternative data offers a comprehensive approach that has shifted the limitations of traditional financial data, enabling a more informed and holistic screening of SMEs' creditworthiness. Real-time and granular data attributes offer a robust foundation for objective credit decisions, reducing reliance on subjective judgments.

4.5 Summary of findings

Table 6 below is included for ease of reference and aligns the emergent themes with the research questions that form the core of this research report. The themes have been identified from interviewing 13 participants and analysing the transcripts and video recordings.

The overarching commentary from the participants is that there is definite value in using alternative data to provide credit access for SMEs. The differences in insights drawn from the participants is the role alternative data play in conjunction with traditional data points.

Table 8: Research questions and Theme alignment

Emergent Theme	Alignment to Research Questions
Theme 1: Fintech Innovation	RQ1 & RQ1.1 & RQ1.2
Theme 2: Data driven transformation	RQ2
Theme 3: Objective-Subjective Relationship	RQ2 & RQ2.1
Theme 4: Mitigating Asymmetry	RQ2 & RQ2.2 & RQ2.3
Theme 5: Signalling Creditworthiness	RQ2 & RQ2.2 & RQ2.3
Theme 6: Screening Enhancement	RQ2 & RQ2.2 & RQ2.3

The next chapter fleshes out the findings within the context of the literature on the topic.

CHAPTER 5. DISCUSSION OF FINDINGS

5.1 Introduction

This chapter presents a discussion on the findings of the study, as outlined in Chapter 4. The primary research questions that this research report sought to answer were: RQ1: *How are fintechs transforming the SME credit landscape in South Africa by using alternative data?* and *RQ2: How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?*

The study was structured to cover analysis between the emergent themes from the participant interviews and the theoretical insights from the literature review. The report clearly shows the alignment and points of divergence between the study's empirical results and the existing literature.

A notable observation emerging from the analysis is the role of alternative data in enhancing credit access for SMEs. The participants' feedback revealed parallels with traditional credit assessment methods, yet with distinctive nuances in the application of alternative data. This chapter explores how the core principles of credit assessment are maintained, albeit executed differently when alternative data are incorporated.

Moreover, the discussion highlights the complementary nature of alternative data as emphasised by the majority of the participants. This aspect resonates strongly with the perspectives presented in the literature review, suggesting a synergistic relationship between traditional and data-driven credit assessment approaches.

5.2 Revisiting the demographic profile of participants

While the detailed demographic profiles of the participants were comprehensively presented in Chapter 4 (Table 4), it is important to briefly revisit these demographics in order to provide context to the discussion of the findings. This section does not repeat the information but rather focuses on how the demographics of the participants may have influenced the research outcomes and the interpretation of these outcomes.

The predominance of senior executives (62%) and founders (38%) in the sample provides a strategic-level perspective on the adoption and implications of alternative data in credit systems. The participants also enjoyed significant experience (average of 18 years) in working with alternative data across different industries, and not limited to credit, which is crucial. More importantly, all current projects the participants are working on have a focus on SMEs. This depth of experience offered valuable insights into the practical application of alternative data in credit assessment, particularly for SMEs.

The representation across fintech lenders, credit bureaus, data generating businesses (from servicing SMEs), and consultancy roles allowed for a holistic understanding of the alternative data landscape. Each participant brought unique viewpoints on how alternative data are used and perceived in the credit assessment process. Fintech lenders provide insights into the practical challenges and opportunities of utilising alternative data from a lending perspective, highlighting innovative approaches to credit assessment. Credit bureaus shed light on the evolution of credit scoring models, emphasising the incorporation of alternative data and the lessons learned through this process. Data generating businesses offer a unique viewpoint on the willingness and challenges associated with sharing data with fintech lenders. This emphasises the importance of data sharing within the SME ecosystem and its impact on reducing information asymmetry. Lastly, consultants bridge the gap between traditional and alternative markets, offering a broader industry perspective that combines regulatory, technological, and market trends.

The participants' experience was across geographies, particularly Africa and Europe and this diverse industry backgrounds offered insights into how South Africa uses alternative data in credit assessment. This consolidation of views ensures a well-rounded examination of how alternative data is reshaping credit provision to SMEs.

5.3 Discussion pertaining to proposition 1

Proposition 1 stated that alternative data are used by Fintechs to effectively augment SME credit analysis where there are gaps in the traditional data points.

Theme 1: Fintech innovation

Based on the findings from the study, alternative data represent a comprehensive and diverse range of non-traditional information sources. This type of data augments traditional credit assessment methods by providing additional insights and bridging gaps where traditional data sources lack full visibility. Beaumont et al. (2022) found that alternative data have enabled fintechs to offer SMEs real time digital underwriting and increase the propensity of SMEs to be funded. The participants in the study highlighted the fact that while non-traditional data sources have always existed, there have been technological advancements in data capture. This evolution in data capture has made SMEs have a digital footprint that can be leveraged in credit assessments. From the literature studied while doing this research, Sheng (2021) confirms that there has been an advancement in data collection to be more proactive and efficient.

The study confirmed that fintechs are incorporating alternative data in their credit processes to reach underserved SMEs and address the scarcity of funding. This practice involves formalising informal business through credit provision and developing products that recognise the heterogeneous nature of SMEs and offer tailored funding solutions. In his paper, Stiglitz (2000) stresses that credit lenders should reach a state where they can identify borrower risk with greater certainty instead of treating all borrowers the same, which is exactly what alternative data provide to fintech lenders. A recurring theme globally is talk of embedded finance. From the study, the participants acknowledged that big data and the need for personalisation is driving the change currently seen in credit. This change is in line with the findings by Yan et al. (2015) that big data have reduced information asymmetry with search costs for credit and the collection of business specific data becoming less costly.

Another point that comes through strongly in the study is the value of partnerships. There are many businesses in the SME ecosystem and all playing different roles. In order to effectively service SMEs, fintechs are forming synergistic partnerships to have more comprehensive insights into SMEs and deliver the funding at the point of need. An example of such partnerships is fintech lenders (balance sheet providers) partnering with point-of-sale providers (data and distribution providers). The study found that such partnerships result in a more nuanced understanding of each business, leading to more informed credit decisions and ultimately, enhanced credit assessment processes. The literature reviewed was silent on the value of partnership; however there was an acknowledgment by Zou et al. (2020) that SMEs interact with different business for specific reasons and there is a lack of knowledge among SMEs and corporates that the data they generate can be used for credit decisions.

Innovation is also addressing the concept of trust in the SME credit market. The study found that trust is low between traditional lenders and SMEs. This finding is consistent with the finding of the Oecd (2020a) that SMEs generally do not prefer to borrow from traditional financial institutions. However, with an aggregation of information into big data; and using that alternative data for credit decisions, lending transitions from pull credit (customer initiated) to push credit (lender initiated). In the context of SME lending, the push model is particularly transformative. Yan et al. (2015) found that alternative data have enabled lenders to be proactive in their credit assessment instead of reactive. They allow for greater financial inclusion by enabling access to credit for businesses that might not qualify through traditional methods or lack the resources to navigate the conventional application process.

The findings on collections protocols play a crucial role. The methods are important in as far as demonstrating the robustness and credibility of the alternative data. Fintechs are obtaining the alternative data through channels other than the SMEs themselves, which builds trust and ensures the accuracy and reliability of the data. The use of APIs and machine learning was found by Jagtiani and Lemieux (2019); Sheng (2021) to improve the availability and accuracy of the information. The study found that understanding the source and collection method of data is crucial for fintechs looking to use data to place their balance sheets at risk.

The study found that the real-time nature of alternative data provides an edge over traditional data, which often becomes outdated quickly. This timely data allows fintechs to respond more dynamically to SMEs' funding needs. This finding is consistent with the findings from Agarwal et al. (2020). Importantly, this particular study found that the same data points can be used by different fintechs to offer different products. There has been no evidence from the literature that different data points are being leveraged for different credit products.

Although the innovation has been there, it appears that adoption by fintechs has so far been limited to working capital solutions. None of the participants interviewed could give examples of longer-term financing uses for SMEs from alternative data. There was also no evidence from the literature on solely using alternative data for longer term financing. Traditional scoring models seem to be more stable in assessing longer term credit needs for SMEs. From the literature, Chavis et al. (2011) suggested that alternative data financing is more suited for younger firms. This statement is inconsistent with the findings of this study.

The participants in the study highlighted the pivotal role of alternative data in augmenting traditional data in credit analysis. This trend is not confined to fintechs alone; traditional businesses, such as FMCGs, are also increasingly incorporating alternative data into their credit decision-making processes which came as a surprise from the findings. The literature reviewed illustrates that traditional lenders are restricted to leveraging only traditional data such as financial statements and relying on existing credit relationships with other financial institutions and collateral (Aleem, 1990). There is insufficient evidence found in the literature on traditional banks leveraging both alternative data and traditional data. However, the literature covers how fintechs have made a shift from predominantly soft information to hard data-driven approaches (Fasano & Cappa, 2022).

On the face of it, the study findings were divided on whether alternative data represent a complement or substitute. While some participants' first inclination was for alternative data to be viewed as a substitute, particularly in contexts where traditional data are insufficient or unreliable, others saw it as a complement. This finding is confirmed by a recent study conducted by the International Finance Corporation (2021) which demonstrated that traditional credit assessment can be complemented with alternative data to promote access to credit. The participants that were in favour of substitution argued that alternative data offer robust insights, especially in short term lending and small ticket values. There was nothing found in the literature that contradicted this finding. They further argued that the credibility of alternative data in credit was derived from the fact that the data used in the underwriting would have been collected from different sources. There was an observable strong bias in that all of the participants that held this viewpoint were in short-term credit lending or scoring using alternative data.

Proponents of a complementary view believed that alternative data must and should be integrated with traditional data. They pointed out that in cases of larger loans and longer tenor, a hybrid of both data types provides a more comprehensive and accurate assessment of creditworthiness. This perspective was reinforced by concerns around regulatory constraints and lack of market readiness. The literature reviewed was clear on the need to regulate the incorporation of alternative data within the bureau environment to achieve the hybrid model (International Finance Corporation, 2021). On probing the participants further, there was consensus that alternative data are complementary. This consensus was based on tenure, the selection of a primary decision-making model supplemented by alternative data and a hybrid approach for borderline calls. There was no evidence found in the literature reviewed that pointed to the assessment for alternative data as a complement or substitute to include considerations such as tenure or size of loans.

Theme 2: Data-driven transformation

The application of alternative data has resulted in paradigm shifts in the credit assessment process. The International Finance Corporation (2021, p. 15) aptly puts it, “For many banks that are still struggling to serve South Africans in an impactful (as well as profitable) manner, the newfound ability to unlock alternative data securely and efficiently can be nothing short of transformative". In further explorations of the literature, Jagtiani and Lemieux (2019) argue that alternative data in credit assessment paints a fuller and more accurate picture of financial lives and creditworthiness.

A key distinction coming from the study is the differing mindsets of fintechs versus traditional lenders, particularly in their approaches to credit assessment using alternative data. The data suggest that fintechs adopt a unique perspective in engaging with their client base – which is the belief that SMEs seeking credit from fintechs are inherently good actors. This viewpoint is supported by Fasano and Cappa (2022) who concluded that fintechs have a strong impact on the traditional providers of financial services, as they change the way interactions with customers occur and consequently what information is gathered. Traditional providers have a more cautious stance to credit underwriting. The approach by traditional providers is confirmed by the International Finance Corporation (2018) which concluded that traditional institutions have typically perceived SMEs as risky or expensive to serve.

The study found that the real time nature of alternative data is transformative on credit markets. Individual credit scores, even if the same, may not fully reflect each SME's financial behaviour. This finding is consistent with the findings by Crouhy et al. (2000) who argued that the common assumption is that businesses with the same rating class have the same default probabilities. This aspect is precisely where the value of alternative data has been transformative in complementing traditional bureau data as found in the study. Real time nature of alternative data provides most recent transaction behaviour over time, velocity of changes in transaction behaviour, changes in spending behaviour, changes in purchase behaviour, changes in web data insights on, and whether there are more complaints or less complaints in the last two days. This value cannot be captured by bureau or traditional data points which are updated in a discrete fashion (Crouhy et al., 2000).

The study found that the default in the market for model acceptance is for an alternative data model to be validated against a traditional model. In a similar paper by Zou et al. (2020), the authors conducted head-to-head comparisons between traditional data and corresponding alternative data in assessing whether alternative data were useful in credit evaluation and fraud detection. As stated by the authors, the traditional data provided the largest sub-sample of observations with no missing values in order to maximise the power and external validity of the test. This finding emphasises the importance of validation against existing models to build trust.

An important metric used in credit to distinguish defaulters from non-defaulters is the Gini Coefficient. Frunza (2013) states that the Gini Coefficient is used for measuring the discriminative power of a credit model. Credit bureau data have a high Gini Coefficient of 45%. However, as confirmed by the study, and in line with findings in literature, information from bureaus is patchy on SMEs. Where information is available, it is generally associated with the individual owner and not the business entity (International Finance Corporation, 2021). The study observed that the effectiveness of alternative data is improving the Gini Coefficient which validates the model reliability. The progression to behavioural scoring over time makes the credit underwriting process more transformative. The data from the behavioural scores become a data point that is fed into the data warehouse and used to revive the credit models.

The findings of this study further revealed that there are some segments of the market that remain underserved due to high servicing costs for traditional lenders. As confirmed in the literature, the natural inclination for traditional lenders is to access financial statements and if that information is not available, it becomes prohibitively high for banks to seek more information (Lehmann & Neuberger, 2001). This is a gap in the credit markets that is now being filled by alternative data. Example of such segments include servicing of sole proprietary businesses.

The study's findings based on participant insights, that could not be verified quantitatively as they were beyond the scope of this study, suggest that the bad debt ratios for fintech lenders using alternative data is similar to the levels observed through traditional underwriting, although with a more seamless underwriting process. There was no evidence found in the reviewed literature to support this observation, making it a potential topic for future research studies.

5.4 Discussion pertaining to proposition 2

Proposition 2 stated that alternative data allow fintechs to depend on hard information derived from alternative data sources to make informed credit decisions.

Theme 3: Objective-subjective relationship

Based on the findings of the study, there is a strong reliance on objective, data driven credit models. The findings indicate that objective models will consistently outperform subjective models. From the literature, McEvoy (2014) found that credit assessment was often subjective, time consuming and expensive. Traditional banks rely on the SME's banking relationships. OECD (2015) in its report states that bank lending relies on relationship gathered through direct interaction with SMEs. As discussed under Theme 2, the application of alternative data has resulted in paradigm shifts in the credit assessment process, which is another example of this shift to less reliance on banking relationships.

The study further acknowledges that the topic of subjectivity has its nuances, positive and negative. While alternative data in credit assessment offer the advantage of anonymity, they do not completely eliminate inherent structural biases. The challenge lies in distinguishing between overt and covert biases. While alternative data effectively address overt biases (e.g. demographic biases), covert biases may still persist if they are structural (e.g. geographic and cultural biases). The literature reviewed supports this perspective, while it may not be possible to completely eliminate bias, the use of alternative data has the potential to reduce discrimination (Ryan, 2020).

The study found that relying only on alternative data does present trade-offs. This finding did not feature in any of the existing literature studied. The study found that alternative data do not fully capture operator risk and the structure of the loan. This finding explains why most products in the market that rely on alternative data are regarded as vanilla in as far as product construct goes.

The findings under this theme speak directly to the themes already covered above, namely, that a hybrid model that integrates both alternative and traditional data is the most effective approach to credit assessment.

Theme 4: Mitigating asymmetry

The findings of the study indicate that self-reported data by SMEs is likely to impact the credit assessment models negatively. This idea is the concept of dishonesty as covered by Akerlof (1970). The author states that the cost of dishonesty lies not only in the amount by which the purchaser (in this case borrower) is cheated; the cost also must include the loss incurred from driving legitimate business out of existence.

The study findings highlight the fact that self-reported data are likely to be tainted by misinformation from the SMEs, if they believe the information that they might share would have an adverse effect on the status of their applications. The focus is to proactively collect hard data. This notion is consistent with the findings from Yan et al. (2015) that proactively gathered alternative data are more objective and cannot be easily manipulated. This supposition is further supported by Sheng (2021) who argues that hard information can be relied on to offer credit to SMEs. The findings indicate that for successful alternative data use cases in credit assessment, the data gathered should be proactive and there should be less reliance on asking businesses to provide any information. Furthermore, treating all applicants with some degree of uniformity based on the cohort they fall under reduces information asymmetry.

Insights from the study highlight the fact that banks with access to cheaper funding lines exhibit a lower appetite for risk compared to fintech lenders. Traditional bank lenders have less risk appetite for SMEs with no previous dealing with them as confirmed by Aleem (1990) in the existing literature. While fintechs have high funding costs, they have a bigger appetite due to their ability to use alternative data in credit assessments. Consequently, fintechs often pass the high costs onto the SMEs. The study draws attention to the minimal price differentiation between good borrowers and bad borrowers, at least for the initial loans. With repeat loans, there is more information about a borrower’s risk profile.

However, as Akerlof (1970) points out, in such instances an asymmetry in available information has developed.

Furthermore, the study indicates that despite SMEs being charged a premium for their initial loans, there is an appetite to take funding at these high interest costs given the convenience advantage (speed and flexibility) of accessing fintech funding. The study also reveals that because of the higher cost of funding, the riskier borrowers are inclined to approach lenders for finance. This finding further implies that there is a higher likelihood of fraud with alternative lenders than there is with traditional data. However, the proactive data collection alleviates that risk. In his paper, Akerlof (1970) argues that bad money drives out the good money until updated information is developed. There is a risk of adverse selection by the fintechs as their best borrowers only use them once due to high interest rates or the fintechs reduce the amount of credit they extend to curb risk.

The findings suggest that successful implementation of alternative data models has a key dependency on customer consent. This consent is crucial for ensuring that from a legal perspective, all privacy regulations have been complied with and the data are being used ethically. The fine balance that fintechs have to navigate is bringing together data providers in a way that does not compromise the data subjects. In the literature reviewed, Jagtiani and Lemieux (2019) acknowledged the importance of borrower consent when processing private information.

The study found that the challenges of information asymmetry persist as the commercial advantages of each fintech lender trump the need to improve the system holistically. There has been slow progress made by the fintechs on sharing positive or negative data. As discussed in the paper by Akerlof (1970), in a market such as this one it becomes harder for lenders to discern good borrowers from bad. This situation is precisely what alternative data seek to bridge, namely, the ability to more accurately assess a borrower's credit worthiness in a market that suffers from information asymmetry.

While alternative data offer fintechs the ability to make more informed decisions, it is clear from the findings that the challenges remain – and alternative data have yet to reach their potential in reducing information asymmetry.

Theme 5: Signalling creditworthiness

Signalling in credit assessment is an important and pivotal concept to unpacking a borrower's behaviour and creditworthiness. This concept refers to the observable characteristics in the alternative data that the lender can rely on to make an informed, yet predictive decision about an SME's creditworthiness (Spence, 1973). The findings in the study suggest that there is a signalling effect depending on how the fintech receives the data, either directly from the SME or from third party sources. This process is in alignment with the discussion of findings under Theme 4 that self-reported data are less valuable than objective data (Sheng, 2021; Yan et al., 2015)

The signals that alternative data provide are predictive of an SME's behaviour as revealed in the study. These predictors include entrepreneurship success, historical credit behaviour, fraud risk, trade frequency, transaction analysis, geography and industry analysis. The findings of the study suggest that the multiplicity of data sources that are used to construct an SME's digital profile, makes the data more reliable. It is improbable for an SME to provide data that would be more insightful or to game the system. The literature review supports the concept of big data, mining that data for anomaly detection, association analysis, clustering and predictive modelling for risk underwriting (Yap et al., 2011). However, given the nascency of the alternative space in the market, there were no specific examples of data sources provided by the fintechs. They still viewed their data sources as proprietary and were not willing to share detailed information on all of them and how each data source completes the puzzle.

The findings of the study suggest that there is a lack of awareness among SMEs regarding the signalling impacts that having a digital profile presents to them. The lack of awareness by SMEs may impact their creditworthiness as they do not understand the importance of being digital and keeping an active profile. The study also highlights a controversial concept of social signalling, where an individual's creditworthiness is assessed based on their social circle. The utility of such a scoring model in the SME space is still untested; however, it has been successfully used in consumer lending. The participants confirmed that there had been quantitative proof of this utility but the incorporation of this model as a standard model will take time.

The findings of the study were inconclusive with regards to the role of the personal credit profiles of SME owners in credit assessment. The findings suggest that in a market where there is information scarcity, the owners' credit profiles can be a valuable starting point, although this approach could be punitive for SMEs. The literature reviewed has arguments that support both side. Kozubíková et al. (2015) argue that the SME owner’s personal characteristics are very important in assessing the level of credit in the SME. In contrast, the International Finance Corporation (2021) argues that research from other countries indicates that personal profiles of SME owners have limited predictive power for business credit.

Theme 6: Screening enhancement

Screening costs is a key component for lenders in determining which sectors of the SME market to provide access to credit. When lenders decide whether to service a market, they consider the costs of screening the borrower (Arráiz et al., 2021). The study found that alternative data have resulted in reduced costs for acquiring and processing credit applications. This finding is consistent with the findings by Mbedzi and Simatele (2020) as well as Sheng (2021). Fintechs can underwrite SMEs at speed and at scale. The reduction in the cost to screen has also extended down to the cost to service. Fintechs are able to provide low value loans due to their underwriting capabilities with big data. This ability is corroborated in the existing literature by Jagtiani and Lemieux (2019) who posit that alternative data had reduced the cost of making credit decisions and/or credit monitoring which has lowered the operating costs for lenders.

From existing literature, individuals are hardly ever willing to reveal all their information (Rothschild & Stiglitz, 1976). Nearly five decades after the authors made this statement it still holds true. The findings of the study confirm that fintechs are proactively gathering information to reduce information asymmetry. Fintechs are relying on the everyday digital trails that businesses are leaving behind to make credit decisions. Previous studies confirmed that information retrieval has shifted to a more proactive approach with technological improvements (Stiglitz, 2000; Yan et al., 2015).

The study emphasises that the uninterrupted flow of observed data is critical. The more consistent the data are, the more reliable the screening processes would be in determining a business's creditworthiness. These results are consistent with our earlier findings that collection protocols are important to build trust and ensure accuracy and reliability of data.

The study found that outside of creating a regulatory lend framework, there is little willingness to innovate in a way that benefits the ecosystem at large. Incumbents tend to be protective of their data, which is one of the reasons that initiatives like credit bureau data sharing through SACCRA have failed.

A finding that emerged from the study is that when fintechs are adopting a screening approach, there is a need to select a hybrid approach, which combines models and consequently alternative data points and traditional data. The results are consistent with our earlier findings.

5.5 Conclusion

The chapter started with a recap of the demographic profile of the participants, the purpose of which was to provide context to the findings and the credibility of the participants. The chapter analysed the findings of the study against the literature that was reviewed as part of this research paper.

The study reveals clearly, and in alignment with the literature that alternative data augment traditional data in building credit assessment models. A hybrid approach is the one most likely to deliver the best result. However, the study also finds that, there are certain considerations such as traditional data being insufficient or unreliable coupled with loan tenure, size of the loan and market readiness that justify reliance on only alternative data.

The value of partnerships between ecosystem players was highlighted in the findings. There is no one player in the ecosystem that has full visibility of the digital footprint generated by SMEs. The findings indicate that for successful alternative data use cases in credit assessment, the data gathered should be proactive and there should be less reliance on asking businesses to provide any information. Partnerships drive an increase in trust with the data used – and over time they reduce signalling and screening costs. Partnerships further mitigate that information asymmetry that is rife in traditional credit markets. Reduced signalling and screening costs with increasing information availability transitions lending from a bad actor to a good actor proposition, and from pull credit to push credit. Privacy risk (POPIA) with such partnerships should be guarded.

A further finding from the study is that the real time nature of alternative data and its ability to be accessible from third parties with little to zero reliance on self-reported data, is essential. This factor increases the objectivity of the models. Moreover, the study found that the more data sources a fintech relies on, the better the model outcome and its ability to identify fraud. Although objective, the trade-off not captured by incorporating subjective elements is that alternative data will not capture operator risk and the structure of the loan. Consequently, the products are vanilla and do not necessarily capture the heterogeneity of SMEs, at least with the first loan.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This concluding chapter of this research report integrates the findings from Chapter 4 and the discussion of the findings in Chapter 5 with the research questions outlined in Chapter 2. This research study had two primary research questions: (1) How are fintechs transforming the SME credit landscape in South Africa by using alternative data? and (2) How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?

Section 6.2 addresses the first research question, highlighting the benefits fintechs derive from the use of alternative data. Section 6.3 unpacks the second research question and discusses the ways fintechs are mitigating information asymmetry. There are three sub-components to information asymmetry that are covered in this section: adverse selection, signalling costs and screening enhancement.

Section 6.4 offers recommendations to various stakeholders, including business owners, credit professionals (in fintechs and conventional credit), CEOs, policy makers and future researchers. The section culminates with gaps identified in the study that present opportunities for future research.

6.2 Conclusions regarding research question 1

The research question was: How are fintechs transforming the SME credit landscape in South Africa by using alternative data?

The study found that alternative data have always existed. However, with technological advancements, companies have become more proficient in capturing and recording that data. Fintechs use alternative data points to augment traditional credit analysis, particularly where gaps exist in traditional data.

Fintechs are playing a critical role in the South African credit landscape by leveraging alternative data to broaden credit access for SMEs. SMEs continue to struggle to obtain credit due to their thin credit files, and innovations such as those led by fintechs are essential. This study highlights the emerging nature of the fintech industry in the South African context. It reveals that fintechs are shifting the thinking around formalisation. Business formalisation is not just regulatory but includes a recognition of a business's digital footprint as a step towards formalisation.

Financial inclusivity emerged as one of the most significant achievements of fintechs in the SME credit landscape. By harnessing alternative data, fintechs have expanded their reach, even to remote markets, offering business loans starting as low as R500. This capability is dependent on a business having a digital footprint and marks a significant departure from traditional funding methods. Conventional lenders are constrained in serving micro and start-up businesses, partly due to perceived risk, geographical area constraints and the absence of on-the-ground relationship managers. As highlighted earlier in this report, traditional lenders also consider banks relationships prior to providing lending limits. In addition, for traditional creditors, processing smaller loan sizes is often prohibitively expensive.

Another key finding in this paper is the use of headless data, which means, only anonymous data are being used in the credit underwriting process by fintechs. The implication is that financial inclusivity extends to race and gender. This approach addresses one of the enduring debates in the realm of credit – the disparity in approval rates between male and female applicants. The study suggests that alternative data may play a role in mitigating such disparities. The study reveals that alternative data mitigate these differences. However, this area warrants further research, comparing approval rates when using alternative data between females and males versus conventional data sources. Such research could provide meaningful insights into the effectiveness of alternative data in creating a non-discriminatory credit landscape.

Alternative data are laced with signals that enabled fintechs to shift away from reactionary "pull" credit offerings to more proactive "push" credit offerings. The data that are used in the underwriting process are collected from third party providers rather than being directly sourced from SMEs. There is less reliance in self-reporting by fintechs, fostering a more objective assessment process. The integration of multiple data sources provides increased reliability signals regarding an SME’s creditworthiness. The ability to push credit offering has notably increased the approval rates for SMEs significantly, as the underwriting process is often completed even before a business owner formally applies. The application processes have certainly become more seamless. As highlighted in the existing literature, SMEs have historically been hesitant to approach conventional lenders due to high rejection rates. Fintechs, by leveraging alternative data, are creating a more accessible and efficient path to credit for SMEs.

The use of real-time data from alternative sources has enabled fintechs to expedite the credit assessment process, reducing the turnaround time from loan application to disbursement. The availability of these data around the clock offers several advantages. Primarily, the data are more current compared to traditional credit data – and as part of the credit underwriting process, fintechs have the most updated view of the business’s operations. This vantage point gives fintechs enhanced credit assessment capabilities to differentiate businesses. Furthermore, the multiple data combined with traditional data, present an advantage that make it inherently difficult for the data to be manipulated. This integration not only significantly mitigates fraud risk but also does not entirely eliminate it. In scenarios where traditional data form the basis of credit assessment, alternative data can be leveraged to make decisions in borderline cases. This use of alternative data is a good example of the complementary role alternative data provide in enhancing credit decision making.

Alternative data have had the most influence in the development of unsecured working capital solutions for SMEs. This study highlights the fintech sector's recognition of the heterogeneity of SMEs, and their distinct needs. Products such as the merchant cash advance, bridging facilities, invoice discounting and purchase order finance is where the greatest impact has been observed of fintechs using alternative data. Fintechs are leveraging the same data points but are creating differentiated funding products. Notably, all these products have a primary term of less than 12 months. The study found limited evidence suggesting that alternative data could extend funding terms beyond the 12-month period, indicating a potential area for further exploration and development.

While alternative data offer fintechs the ability to make more informed credit decisions, several challenges remain. These challenges include the need for a balanced approach between objective data analysis and acceptable levels of subjectivity, the high cost of funds for fintechs which leads to high finance charges to SMEs, the protection of SME’s privacy within the ecosystem and the critical role of consent in data usage, and the industry's reluctance to share data with credit bureaus. Addressing these challenges is crucial for the effective use of alternative data in credit decisioning.

6.3 Conclusions regarding research question 2

The research question was: How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?

Adverse selection is an issue that is prevalent with traditional lenders and is being addressed differently by fintechs. The findings of the study reveal that fintechs are effectively separating good actors from bad actors. They depend on the extensive big data formed by the collection of alternative data.

SMEs tend to know more information about their businesses than the information they are willing to share with lenders. The study found that there is less reliance by fintechs on self-reported data. The findings indicate that fintechs proactively gather information through the business lifecycle, not only at points of funding needs. This proactive data collection results in more objective assessment procedures. Notably, some fintechs, as part of their credit application processes have eliminated self-reporting requirements completely. These same data are used by fintechs to look for signals that convey a message to fintechs that the borrowers are good payers. The ability to proactively collect these data as found in this study has reduced the cost of screening resources invested in search of this information and has significantly increased the efficiency with which the data are collected via technological advancements.

As lenders continually collect and analyse big data on SMEs, they refine their models. With access to real time data, fintechs are able to intervene earlier with businesses that show signs of distress. The evidence from the study suggests that lenders rely on alternative data to categorise businesses into risk-based cohorts, reducing the impact of adverse selection in their models. This access to real time data provides screening benefits as the borrower risk assessment is in real time.

The insights from the study suggest that fintechs tend to charge high interest rates, particularly with initial loans. This high charge is due to the high costs of funding. However, for subsequent loans, fintechs integrate behavioural data from previous loans which leads to more favourable pricing. The study notes a high drop off rate after initial loans due to mispricing. The evidence presented in the study suggest that fintechs attempt to offset this attrition by passing on the digital acquisition and service costs savings to the borrowers.

One significant finding is that fintechs have not substantially addressed the "thin file" issue, where SMEs lack extensive credit histories. As a participant put it, “fintechs live on information asymmetry”, suggesting that there has been a reluctance to report positive and negative data on SMEs, which perpetuates challenges of information asymmetry. There is no adequate incentive for fintechs to share the data. The implication is that there is a requirement for a regulatory lend model, where fintechs and traditional SME providers are compelled to share data in line with open banking principles and to compel the sharing of data with credit bureaus.

Fintech lenders look for signals in the alternative data for an SME’s creditworthiness. The study found that relying on hard data gives unalterable signals. The study also revealed that fintechs tend to rely on multiple sources when integrating SME alternative data into credit assessment processes, which makes the SME’s digital profile more reliable and harder to manipulate.

Findings from the study suggest that transaction patterns, an online presence, effective customer complaint resolution, regularly updated websites, social media accounts and a long history with service providers all send signals to fintechs of the financial health of the business.

A key finding is the weight placed by fintechs on SMEs having an active bank account which is a reliable indicator of the savviness of the business owner and the legitimacy of the business. The processes to open business bank accounts are subject to stringent regulatory requirements. Fintechs require a bank account to offer credit lines, as it serves as a crucial collection mechanism. The ability of fintechs to collect their funds directly from bank accounts or from source (invoice discounting or net settlement) is a key indicator of a business's overall risk.

Social signalling was highlighted in the study. The study found that examining an owner's social circle to predict an SME's creditworthiness has merit. This practice has been tested and proven in consumer credit; however, its application in SME credit assessment is yet to be established. However, the findings were inconclusive regarding the efficacy of using SME owner credit profiles in assessing business creditworthiness.

The study highlights the significant reduction in acquiring, processing and servicing costs associated with fintechs assessing SMEs. The use of APIs and secure file sharing has taken away the need for on-the-ground relationship managers to gather information and reduced the underwriting costs as this process is also automated.

The study highlights the fact that fintechs are extending their analyses beyond financial data to include for example, web data analytics and movement scoring on location as predictive indicators. By analysing patterns among similar borrowers, fintechs are able to more accurately assess risk. This ability not only aids in identifying potential defaulters but also in recognizing viable credit opportunities among SMEs traditionally considered as 'thin files.'

The study highlights the importance of a hybrid approach in screening. There is no one-size-fits-all approach. The study reveals that it is beneficial to combine models that utilise both alternative and traditional data points to the extent possible.

6.4 Recommendations

Recommendations to the SMEs (business owners)
Awareness: Business owners should recognise that they can now access unsecured credit products through fintech lenders, who typically have higher approval rates, do not require collateral, and rely on hard data for information transparency. In order to access funding from fintechs, SMEs should be open to providing data sharing consent to businesses that service them, which may lead to "push credit" offers. **Digital footprint:** SMEs looking to increase their funding chances should embark on digital transformation journeys. An active and positive digital presence, including digital payment solutions, transaction recording, online customer interactions, and social media engagement, can be crucial.
Recommendations to fintech lenders and traditional institutions
Explore partnerships: The SME ecosystem encompasses a variety of stakeholders, each playing a distinct function. In order to mitigate information asymmetry, fintech lenders and other key players in the ecosystem should seek synergistic relationships that facilitate information sharing. This approach results in effective sharing of alternative data with fintechs, enabling improved funding access for SMEs. In addition, traditional lenders should consider forming partnerships with fintech companies. Such collaborations can accelerate progress for traditional lenders in leveraging alternative data for credit assessments. **Invest in advanced data analytics:** Fintechs should invest in sophisticated analytics, particularly machine learning, to effectively handle diverse alternative data sources. This approach is likely to increase financial inclusivity for SMEs and result in the development of new financial products. Furthermore, it will enhance risk assessment and fraud detection. **Invest in secure platforms:** Technological platforms have enhanced and made efficient the flow of information between parties. Allowing data to be

shared via API integration facilitates real-time and secure data sharing. The ultimate benefit of this approach from a fintech lender perspective is reduced screening and signalling costs. As costs to acquire and underwrite these businesses reduce, scale markets open up and allow expansion into underserved segments. Data security becomes critical to instil confidence in market participants. **Adopt hybrid models:** The study is clear on the finding that the most effective credit models combine both alternative data and traditional data. Fintechs should view alternative data not as a substitute but rather as a complement to traditional data. That said, there is an appreciation that in certain markets were files are thin, alternative data alone would suffice. Alternative data when combined with traditional data can be used for longer term funding options.
Recommendations to CEOs of companies that generate data servicing SMEs
Explore partnerships: The SME ecosystem encompasses a variety of stakeholders, each playing a distinct function. In order to mitigate information asymmetry, CEOs should seek partnerships with fintech lenders to facilitate data sharing. This collaboration can enhance the sustainability of the SMEs they serve through value-added services and enable better funding access through alternative data.
Recommendations to policymakers
Regulate data sharing: The regulators should implement policies that encourage the open sharing of SME data. The implementation of open banking should be accelerated in South Africa to help alleviate information asymmetry problems. Institutions often possess extensive data, yet there is a reluctance to share this information, even when such sharing could benefit the data subjects themselves. This hesitancy hinders the growth of SMEs. **Support Innovation:** Encourage innovation in the fintech sector through supportive initiatives like allowing a sandbox to enable alternative data to be shared and credit models developed in a safe environment. The regulator should be more open to new products that come about as a result of leveraging alternative data to get funding to SMEs that have previously been marginalised.
Recommendations for credit bureaus

Integrate alternative data sources: Integrate alternative data into credit reports to provide a more comprehensive view of an SME's creditworthiness.
Collaborate with fintechs: Work closely with fintech lenders to understand the type of alternative data being used and explore ways to incorporate it into traditional credit scoring models. Collaborate in terms of sharing positive and negative data points to alleviate problems of information asymmetry.

6.5 Suggestions for further research

The scope of this study was limited to unpacking insights on the role of alternative data in providing access to credit for SMEs. Further research is required to understand whether businesses that have incorporated a hybrid model have a better bad debt experience than lenders that use only one of the two methods.

The study highlights the fact that fintech lenders generally charge a premium compared to traditional lenders. There is evidence provided in this study that SMEs at times value access over price; however, it was beyond the scope of this study to delve further into this aspect. Further research looking at the pricing dynamics for fintechs when providing funding to SMEs would be valuable. Such a study would look at the role that pricing and delayed access have on SME sustainability.

The Gini Coefficient is used to distinguish defaulters from non-defaulters. The higher the Gini Coefficient, the better the model. A study that investigates a combination of traditional data, social data, behavioural data and transactional data in a credit model to determine the Gini would be valuable. A key component of such as study is access to data.

Furthermore, a study that focuses on the predictive power of consumer credit scores on business credit would be beneficial. In a market like South Africa where lenders have resorted to relying on individual credit scores to assess SMEs, it becomes critical to understand its impact on access to funding.

From the literature review it emerged that the UK has a regulatory lending model which enabled them to roll out open banking. South Africa has a more incumbent led model. This model puts the regulators and incumbents at odds, specifically around data sharing. Further research on how open banking would work in South Africa’s SME sector would be of great value. This research should include the impact of regulations such as the Protection of Personal Information Act (2013) (PoPIA) and the Promotion of Access to Information Act (2000) (PAIA).

6.6 Concluding comments

The findings of the research highlight the significant gap in credit access affecting SMEs in South Africa. The research paper further highlights the transformative role of fintech innovation as they leverage alternative data to provide credit products to SMEs. South Africa is ripe for innovation as the SME sector increasingly embraces digital tools. The ability for fintechs to access SME-generated data is central to unlocking more inclusive and tailored products. A regulatory framework that ensures that SME own their data is critical. This necessitates a collaborative effort among all SME market stakeholders to create an ecosystem conducive to data sharing and innovation. The study envisions a shift towards 'push' credit offerings, characterized by reduced screening costs and an abundance of digital signals indicating SME health. Crucially, the research highlights that alternative data stands as a pivotal solution to the challenges of information asymmetry, heralding a new era of credit provision for SMEs in South Africa.

REFERENCES

Abbasi, K., Alam, A., Brohi, N. A., Brohi, I. A., & Nasim, S. (2021). P2P Lending Fintechs and SMEs' Access to Finance. *Economics letters*, *204*, 109890. https://doi.org/10.1016/j.econlet.2021.109890

Agarwal, S., Alok, S., Ghosh, P., & Gupta, S. (2020). Financial Inclusion and Alternate Credit Scoring for the Millennials: The Role of Big Data and Machine Learning in Fintech. *Business School, National University of Singapore Working Paper, SSRN*, *3507827*.

Akerlof, G. A. (1970). The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, *84*(3), 488-500. https://doi.org/https://academic.oup.com/qje/issue

Aleem, I. (1990). Imperfect Information, Screening, and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan. *The World Bank Economic Review*, *4*(3), 329-349. https://doi.org/10.1093/wber/4.3.329

Angilella, S., & Mazzù, S. (2015). The Financing of Innovative SMEs: A Multicriteria Credit Rating Model. *European Journal of Operational Research*, *244*(2), 540-554. https://doi.org/10.1016/j.ejor.2015.01.033

Arráiz, I., Bruhn, M., Roth, B. N., Ruiz-Ortega, C., & Stucchi, R. (2021). Borrower Leakage From Costly Screening: Evidence From SME Lending in Peru. *Journal of Development Economics*, *153*, 102719. https://doi.org/10.1016/j.jdeveco.2021.102719

Asah, F., Louw, L., & Williams, J. (2020). The Availability of Credit From the Formal Financial Sector to Small and Medium Enterprises in South Africa. *Journal of Economic and Financial Sciences*, *13*(1), e1-e10. https://doi.org/10.4102/jef.v13i1.510

Bank, E. C., Eça, A., Ferreira, M., Porras Prado, M., & Rizzo, A. (2022). *The Real Effects of FinTech Lending on SMEs: Evidence From Loan Applications*. https://doi.org/doi/10.2866/992620

Beaumont, P., Tang, H., & Vansteenberghe, E. (2022). The Role of FinTech in Small Business Lending. *SSRN Electronic Journal*.

Błach, J., Wieczorek-Kosmala, M., & Trzęsiok, J. (2020). Innovation in SMEs and Financing Mix. *Journal of Risk and Financial Management*, *13*(9), 206. https://doi.org/10.3390/jrfm13090206

Charaia, V., Chochia, A., & Lashkhi, M. (2021). Promoting Fintech Financing for SMEs in the South Caucasian and Baltic States During the COVID-19 Global Pandemic. *Business, Management & Education / Verslas, Vadyba ir Studijos*, *19*(2), 358-372. https://doi.org/10.3846/bmee.2021.14755

Chavis, L. W., Klapper, L. F., & Love, I. (2011). The Impact of the Business Environment on Young Firm Financing. *The World Bank Economic Review*, *25*(3), 486-507. https://doi.org/10.1093/wber/lhr045

Clarke, V., & Braun, V. (2017). Thematic Analysis. *The Journal of Positive Psychology*, *12*(3), 297-298. https://doi.org/10.1080/17439760.2016.1262613

Constable, S. (2019). What Is Alternative Data? . *Journal Report*(Generic). https://go.exlibris.link/Kf41lCjv

Coyne, I. T. (1997). Sampling in Qualitative Research: Purposeful and Theoretical Sampling; Merging or Clear Boundaries? *Journal of Advanced Nursing*, *26*(3), 623-630. https://doi.org/10.1046/j.1365-2648.1997.t01-25-00999.x

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE. https://wits.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwdV1LS8QwEB6kevAiPtn12ZPoYbtps-kmJw_aIoKexOuSNCkWpRUirP57J222VWGPw4SQhMnMl2S-CQBNIjL55xPKgiax1nNitECgxDHQaUlFKVShTZE6cnL2RLMXlj-yhyFVR9q3SFXKftuobpw07bL9p_a1Wd58uDdUNEgMueTr0rG7sO171TiHTIn7ziC_64_u2Mphk944E4FAgg6x3cu0JX8hKEGcL-JBQBiyKhDllX_l2P0qs9WN9VeAynchcKSFPdio1T6MO85t6PetDa98cenrA9hZpdqFus3dOISLPHu-vZ90nS78dc6in3FyBEHd1GYEIUN3lzLNTYrHmyKW3Kg5k5wLTWZUEjKG0bpejterTmAb4cGsu3A4hc0SN4Q5g2BZfdrzdnF_AGWlfi0

Crouhy, M., Galai, D., & Mark, R. (2000). A Comparative Analysis of Current Credit Risk Models. *Journal of Banking & Finance*, *24*(1), 59-117. https://doi.org/10.1016/S0378-4266(99)00053-9

De la Torre, A., Martínez Pería, M. S., & Schmukler, S. L. (2010). Bank involvement with SMEs: Beyond relationship lending. *Journal of Banking & Finance*, *34*(9), 2280-2293. https://doi.org/10.1016/j.jbankfin.2010.02.014

Del Gaudio, B. L., Sampagnaro, G., Porzio, C., & Verdoliva, V. (2022). The Signaling Role of Trade Credit in Bank Lending Decisions: Evidence From Small and Medium-Sized Enterprises. *Journal of Business Finance & Accounting*, *49*(1-2), 327-354. https://doi.org/10.1111/jbfa.12554

Einav, L., Jenkins, M., & Levin, J. (2013). The Impact of Credit Scoring on Consumer Lending. *The Rand Journal of Economics*, *44*(2), 249-274. https://doi.org/10.1111/1756-2171.12019

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1-4.

Everett, C. R. (2015). Group Membership, Relationship Banking and Loan Default Risk: The Case of Online Social Lending. *Banking and Finance Review*, *7*(2), 15-54. https://go.exlibris.link/TtjCxnl0

Fasano, F., & Cappa, F. (2022). How Do Banking Fintech Services Affect SME Debt? *Journal of Economics and Business*, *121*, 106070. https://doi.org/10.1016/j.jeconbus.2022.106070

Frunza, M.-C. (2013). Computing a Standard Error for the Gini Coefficient: An Application to Credit Risk Model Validation. *Journal of Risk Model Validation*, *7*(1), 61-82. https://doi.org/10.21314/JRMV.2013.099

Genesis Analytics. (2019). *Fintech Scoping in South Africa*. https://www.resbank.co.za/content/dam/sarb/quick-links/fintech/WB081_Fintech%20Scoping%20in%20SA_20191127_final%20(002).pdf

Goldman, A., & Johansson, J. K. (1978). Determinants for Search of Lower Prices: An Empirical Assessment of the Economics of Information Theory. *The Journal of Consumer Research*, *5*(3), 176-186. https://doi.org/10.1086/208729

Graneheim, U. H., & Lundman, B. (2004). Qualitative Content Analysis in Nursing Research: Concepts, Procedures, and Measures to Achieve Trustworthiness. *Nurse Education Today*, *24*(2), 105-112. https://doi.org/10.1016/j.nedt.2003.10.001

Hammarberg, K., Kirkman, M., & de Lacey, S. (2016). Qualitative Research Methods: When to Use Them and How to Judge Them. *Human Reproduction (Oxford)*, *31*(3), 498-501. https://doi.org/10.1093/humrep/dev334

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qualitative health research*, *27*(4), 591-608. https://doi.org/10.1177/1049732316665344

Hussein, M., Hirst, S., Salyers, V., & Osuji, J. (2014). Using Grounded Theory as a Method of Inquiry: Advantages and Disadvantages. *Qualitative Report*. https://doi.org/10.46743/2160-3715/2014.1209

International Finance Corporation. (2018). *The Unseen Sector: A Report on the MSME Opportunity in South Africa*. www.ifc.org

International Finance Corporation. (2021). *Diagnostic: Alternative Data Landscape in South Africa*. https://www.ncr.org.za/documents/Alternative%20Data%20Landscape%20in%20South%20Africa%2024.6.21.pdf

International Monetary Fund. (2018). *The Bali Fintech Agenda: IMF and World Bank Encourage a Principled Growth of Fintech*.

Irwin, D., & Scott, J. M. (2010). Barriers Faced by SMEs in Raising Bank Finance. *International Journal of Entrepreneurial Behaviour & Research*, *16*(3), 245-259. https://doi.org/10.1108/13552551011042816

Jagtiani, J., & Lemieux, C. (2019). The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence From the LendingClub Consumer Platform. *Financial Management*, *48*(4), 1009-1029. https://doi.org/10.1111/fima.12295

Kalitanyi, V. (2019). Enterprise Propellers (EP) and Identity of SMMEs, Informal Business, and Cooperatives in Gauteng Province of South Africa. *Acta Universitatis Danubius. Œconomica*, *15*(1), 53-80. https://go.exlibris.link/qksV2dMz

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic Methodological Review: Developing a Framework for a Qualitative Semi-Structured Interview Guide. *Journal of Advanced Nursing*, *72*(12), 2954-2965. https://doi.org/10.1111/jan.13031

Kawai, K., Onishi, K., & Uetake, K. (2022). Signaling in Online Credit Markets. *The Journal of Political Economy*, *130*(6), 1585-1629. https://doi.org/10.1086/718984

Kelly, M., Dowling, M., & Millar, M. (2018). The search for understanding: the role of paradigms. *Nurse Researcher (2014+)*, *25*(4), 9. https://doi.org/https://doi.org/10.7748/nr.2018.e1499

Kozubíková, L., Belás, J., Bilan, Y., & Bartoš, P. (2015). Personal Characteristics of Entrepreneurs in the Context of Perception and Management of Business Risk in the SME Segment. *Economics & Sociology*, *8*(1), 41-54. https://doi.org/10.14254/2071-789X.2015/8-1/4

Łasak, P. (2022). The Role of Financial Technology and Entrepreneurial Finance Practices in Funding Small and Medium-Sized Enterprises. *Journal of Entrepreneurship, Management and Innovation*, *18*(1), 7-34. https://doi.org/10.7341/20221811

Legard, R., Keegan, J., & Ward, K. (2003). In-Depth Interviews. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, *6*(1), 138-169.

Lehmann, E., & Neuberger, D. (2001). Do Lending Relationships Matter?: Evidence From Bank Survey Data in Germany. *Journal of Economic Behavior & Organization*, *45*(4), 339-359. https://doi.org/10.1016/S0167-2681(01)00151-2

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and Soft. *The Review of Corporate Finance Studies*, *8*(1), 1-41. https://doi.org/10.1093/rcfs/cfy009

Mbedzi, E., & Simatele, M. (2020). Small, Micro, and Medium Enterprises Financing: Costs and Benefits of Lending Technologies in the Eastern

Cape Province of South Africa. *Journal of Economic and Financial Sciences*, *13*(1), e1-e10. https://doi.org/10.4102/jef.v13i1.477

McEvoy, M. J. (2014). Enabling Financial Inclusion Through “Alternative Data”. *Mastercard Advisors: Bentonville, AR, USA*.

Meissner, R., & Meissner, R. (2016). Paradigms and theories in water governance : the case of South Africa's National Water Resource Strategy, Second Edition. *Water S. A.*, *42*(1), 1-10. https://doi.org/10.4314/wsa.v42i1.01

Mills, K. G. (2018). *Fintech, Small Business, and the American Dream: How Technology Is Transforming Lending and Shaping a New Era of Small Business Opportunity*. Palgrave Macmillan Cham. https://doi.org/https://doi.org/10.1007/978-3-030-03620-1

Moro, A., Fink, M., & Maresch, D. (2015). Reduction in Information Asymmetry and Credit Access for Small and Medium-Sized Enterprises. *The Journal of financial research*, *38*(1), 121-143. https://doi.org/10.1111/jfir.12054

Morse, A. (2015). *Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending* (0898-2937). https://go.exlibris.link/PRdrR6fv

Morse, J. M. (2015). Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry. *Qualitative Health Research*, *25*(9), 1212-1222. https://doi.org/10.1177/1049732315588501

Mpofu, O., & Sibindi, A. B. (2022). Informal Finance: A Boon or Bane for African SMEs? *Journal of Risk and Financial Management*, *15*(6), 270. https://doi.org/10.3390/jrfm15060270

Msomi, T., & Olarewaju, O. (2021). Evaluation of Access to Finance, Market, and Viability of Small and Medium-Sized Enterprises in South Africa. *Problems and Perspectives in Management*, *19*(1), 281-289. https://doi.org/10.21511/ppm.19(1).2021.24

Nanziri, L. E., & Wamalwa, P. S. (2021). Finance for SMEs and its Effect on Growth and Inequality: Evidence from South Africa. *Transnational Corporations Review*, *13*(4), 450-466. https://doi.org/10.1080/19186444.2021.1925044

National Small Business Act 102 of 1996, (1996). https://www.gov.za/sites/default/files/gcis_document/201409/act102of1996.pdf

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*(1), 1-13. https://doi.org/10.1177/1609406917733847

OECD. (2015). *New Approaches to SME and Entrepreneurship Financing*. https://doi.org/doi:https://doi.org/10.1787/9789264240957-en

Oecd. (2020a). *Financing SMEs and Entrepreneurs 2020 An OECD Scoreboard*. OECD Publishing. https://go.exlibris.link/6GmXcBK5

OECD. (2020b). *Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard*. OECD Publishing. https://go.exlibris.link/V9mBBLxm

Ono, A., Hasumi, R., & Hirata, H. (2014). Differentiated Use of Small Business Credit Scoring by Relationship Lenders and Transactional Lenders: Evidence From Firm–Bank Matched Data in Japan. *Journal of Banking & Finance*, *42*, 371-380.

Owens, J. V., & Wilhelm, L. (2017). Alternative Data Transforming SME Finance. http://documents.worldbank.org/curated/en/701331497329509915/Alternative-data-transforming-SME-finance

Peel, K. L. (2020). A Beginner's Guide to Applied Educational Research Using Thematic Analysis. *Practical Assessment, Research, and Evaluation*, *25*(1), 2.

Peters, K., & Halcomb, E. (2015). Interviews in Qualitative Research. *Nurse researcher*, *22*(4), 6-7. https://doi.org/10.7748/nr.22.4.6.s2

Rindfleisch, A., Malter, A. J., Ganesan, S., & Moorman, C. (2008). Cross-Sectional versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *Journal of Marketing Research*, *45*(3), 261-279. https://doi.org/10.1509/jmkr.45.3.261

Rosser, J. B. (2003). A Nobel Prize for Asymmetric Information: The Economic Contributions of George Akerlof, Michael Spence, and Joseph Stiglitz. *Review of Political Economy*, *15*(1), 3-21. https://doi.org/10.1080/09538250308445

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, *90*(4), 629-649. https://doi.org/10.2307/1885326

Ryan, K. (2020). The Big Brain in the Black Box: AI, Machine Learning, and Alternative Data Are Helping Banks and Nonbanks Alike Make Faster Decisions and Expand Access to Credit. While Fair Lending Concerns About "Black Boxes" Have Impeded Wider Adoption of These Technologies, the Regulatory Environment Is Shifting. *ABA Banking Journal*, *112*(3), 36. https://go.exlibris.link/45vgY2d2

Schueffel, P. (2016). Taming the Beast: A Scientific Definition of Fintech. *Journal of Innovation Management*, *4*(4), 32-54.

Sheng, T. (2021). The Effect of Fintech on Banks' Credit Provision to SMEs: Evidence From China. *Finance Research Letters*, *39*, 101558. https://doi.org/10.1016/j.frl.2020.101558

Singh, N., Benmamoun, M., Meyr, E., & Arikan, R. H. (2021). Verifying Rigor: Analyzing Qualitative Research in International Marketing. *International Marketing Review*, *38*(6), 1289-1307. https://doi.org/10.1108/IMR-03-2020-0040

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374. https://doi.org/10.2307/1882010

Spence, M. (2002). Signaling in Retrospect and the Informational Structure of Markets. *The American economic review*, *92*(3), 434-459. https://doi.org/10.1257/00028280260136200

Stalmeijer, R. E., Brown, M. E. L., & O'Brien, B. C. (2024). How to Discuss Transferability of Qualitative Research in Health Professions Education. *The Clinical Teacher*, e13762-e13762. https://doi.org/10.1111/tct.13762

Stefan, S., & Visser, R. (2021). What Role Can Small and Micro Businesses Play in Achieving Inclusive Growth? Questions Requiring Answers. https://www.cde.org.za/wp-content/uploads/2022/08/Small-Business-report-2021.pdf

Stiglitz, J. E. (2000). The Contributions of the Economics of Information to Twentieth Century Economics. *The Quarterly Journal of Economics, 115*(4), 1441-1478. https://doi.org/10.1162/003355300555015

Susilo, A., Sri Ramalu, S., Shahril, M., & Razimi, A. (2021). Generic Qualitative Research in Management Studies *7*, 1-13. https://doi.org/10.38204/jrak.v7i1.523

Tala, L. (2021). South Africa's Lending Infrastructure: Does it Facilitate or Constrain Access to Credit Finance by Small and Medium Enterprises (SMEs)? *Journal of Public Administration*, *56*(2), 276-287. https://doi.org/10.10520/ejc-jpad-v56-n2-a9

Thanh, N. C., & Thanh, T. (2015). The Interconnection Between the Interpretivist Paradigm and Qualitative Methods in Education. *American Journal of Educational Science*, *1*(2), 24-27.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-Item Checklist for Interviews and Focus Groups. *International journal for Quality in Health Care*, *19*(6), 349-357. https://doi.org/10.1093/intqhc/mzm042

Tongco, M. D. C. (2007). Purposive Sampling as a Tool for Informant Selection. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/5bdc92a7-99fe-4134-aaeb-6d0970562559/content

Tracy, S. J. (2010). Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. *Qualitative inquiry*, *16*(10), 837-851. https://doi.org/10.1177/1077800410383121

Turner, C., & Astin, F. (2021). Grounded Theory: What Makes a Grounded Theory Study? *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *20*(3), 285-289. https://doi.org/10.1093/eurjcn/zvaa034

Voordeckers, W., & Steijvers, T. (2006). Business Collateral and Personal Commitments in SME Lending. *Journal of Banking & Finance*, *30*(11), 3067-3086. https://doi.org/10.1016/j.jbankfin.2006.05.003

Wanke, P., Maredza, A., & Gupta, R. (2017). Merger and acquisitions in South African banking: A network DEA model. *Research in International Business and Finance*, *41*, 362-376.

World Bank. (2019). *Improving SMEs’ Access to Finance and Finding Innovative Solutions to Unlock Sources of Capital*. Retrieved 14 May from https://www.worldbank.org/en/topic/smefinance#:~:text=SMEs%20account%20for%20the%20majority,(GDP)%20in%20emerging%20economies.

Xu, W., & Zammit, K. (2020). Applying Thematic Analysis to Education: A Hybrid Approach to Interpreting Data in Practitioner Research. *International Journal of Qualitative Methods*, *19*, 160940692091881. https://doi.org/10.1177/1609406920918810

Yan, J., Yu, W., & Zhao, J. L. (2015). How Signaling and Search Costs Affect Information Asymmetry in P2P Lending: The Economics of Big Data. *Financial Innovation (Heidelberg)*, *1*(1). https://doi.org/10.1186/s40854-015-0018-1

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using Data Mining to Improve Assessment of Creditworthiness via Credit Scoring Models. *Expert Systems with Applications*, *38*(10), 13274-13283. https://doi.org/10.1016/j.eswa.2011.04.147

Zou, W., Vance, A., & Yan, J. K. (2020). The Differential Role of Alternative Data in SME-Focused Fintech Lending. https://core.ac.uk/download/pdf/326836124.pdf

Zou, Y., & Wang, X. (2022). Distance, Information and Bank Lending in China. *Pacific-Basin Finance Journal*, *74*, 101793. https://doi.org/10.1016/j.pacfin.2022.101793

APPENDIX A: INTERVIEW REQUEST COVER LETTER

Dear Participant,

I am writing to extend an invitation for you to participate in a research study that I am conducting as part of my pursuit of a Master of Management in Digital Business. My name is Mandla Khupe, and my research is focused on the use of alternative data by fintechs to provide access to capital for SMEs in South Africa.

I have carefully selected a limited number of participants for this study, and I would like to invite you to take part in an in-person or virtual interview session, whichever is more convenient for you. **Each interview session will last for approximately one to two hours in one sitting**. However, if you prefer, we can break down the interview into two separate sittings, whichever option is more comfortable and convenient for you.

The interviews will be conducted over an 8-week period between April and May, followed by data analysis and interpretation. **Please note that you will only be required to participate in one interview session (2 hours) and are not expected to make yourself available for the full 8-week period**. The outcomes of the research paper will be made available in H1 2024. If there are any questions during the interview that you would prefer not to answer, please feel free to decline to respond. I want to assure you that this research will not be circulated outside of my academic environment without first obtaining permission from all participants.

Should you have any questions or concerns, please do not hesitate to contact me at 2410076@students.wits.ac.za. I will be working under the supervision of Dr. Jabulile Msimango-Galawe, who can be contacted at jabulile.msimango-galawe@wits.ac.za should you have further inquiries and feedback.

I would appreciate it if you could kindly confirm in writing whether you are willing to participate in this study. Thank you for considering my invitation.

Sincerely,

Mandlenkosi Khupe

APPENDIX B: PARTICIPANT INFORMATION SHEET

Study title: The use of alternative data by fintechs to provide access to credit for SMEs in South Africa

Dear Participant,

I am studying towards my Master of Management in Digital Business at the University of the Witwatersrand, Johannesburg. My supervisor is Dr. Jabulile Msimango-Galawe.

Introduction:

The purpose of this qualitative research report is to explore the use of alternative data by fintechs sources to provide access to credit for SMEs in South Africa. The study aims to explore the benefits of using alternative data to assess creditworthiness and how it complements traditional credit data. Furthermore, the study will seek to provide insights into how alternative data is leveraged to address information asymmetry and how fintechs are using alternative data to signal business sustainability.

Invitation to Participate: You are invited to take part in an in-person or virtual interview session, whichever is more convenient for you. Each interview session will last for approximately one to two hours in one sitting. However, if you prefer, we can break down the interview into two separate sittings, whichever option is more comfortable and convenient for you. The interviews will be conducted over an 8-week period between April and May, followed by data analysis and interpretation. Please note that you will only be required to participate in one interview session (2 hours) and are not expected to make yourself available for the full 8-week period.

With your permission, I would like to record the interview. This data will be stored for 5 years and deleted. Only the researcher will have access to the data.

During the research activity, I will need to ask for some personal information about you such as name and surname, academic background, experience, and current role in the industry.

To maintain privacy, all Individual identities will be coded, and confidentiality guaranteed. Pseudo names will be given to participants to maintain their anonymity; however, this cannot be completely guaranteed as they are known to the researcher.

I understand that some questions may be sensitive, particularly those related to the intellectual property of associated companies. Please note that you may choose not to answer these questions if you prefer.

Once the research study is complete, I will write up a research report and make it available to my academic superiors. If you would like a summary of the report, I would be happy to send it to you.

If you have any questions or concerns about this research study, please feel free to contact me or my supervisor, Dr. Jabulile Msimango-Galawe. If you have any ethical concerns or complaints, you may contact the University Human Research Ethics Committee (Non-Medical) at telephone (+27) 11 717 1408 or email hrecnon-medical@wits.ac.za

APPENDIX C: THE RESEARCH INSTRUMENT

1. Can you please give an introduction about yourself and how long you have been in the fintech space (covering academic background, experience)? 2. Would you say South African fintechs are successfully using alternative data to advance credit to SMEs? If you have examples of this, that would be great?
3. How are fintechs and SMEs benefitting from the use of alternative data to assess creditworthiness to SMEs?
4. What are the common types of alternative data points that fintechs rely on for credit scoring of SMEs?
5. How do fintechs access and collect alternative data sources? And how do they ensure accuracy and reliability of the alternative data sources?
6. Would you say alternative data is more a complement to traditional data than it is a substitute or vice versa? 7. Where alternative data only is used, would you say its sufficient to provide a complete picture of SME creditworthiness? Are there particular industries where you need either or both?
8. Is alternative data addressing problems of information asymmetry? How? 9. What data points do fintechs use to signal sustainability of SMEs going into the future? 10. Are collection mechanisms in alternative data credit assessments the best measure of willingness to pay for SMEs? We know willingness to pay is a key component in assessing an SME's probability of default in traditional credit. In traditional lending, there is a heavy reliance on the relationship side to verify this. If the verification is not done, what levers are there to ensure that willingness to pay is there?
11. What subjective inputs are fintechs bringing into their credit processes or rather how do they balance this with the need to make data-driven decisions?
12. What types of alternative data are fintechs using to address information asymmetry?
13. How do fintechs ensure that the alternative data they use is accurate and reliable, and what challenges do they face in doing so? 14. What proactive screening processes do fintech lenders rely on, and how do they ensure the accuracy and reliability of the alternative data they use to assess SMEs for credit – while reducing the costs to gather this data?

APPENDIX D: ETHICS APPROVAL

Graduate School of Business Administration
University of the Witwatersrand, Johannesburg

Wits Business School Ethics Committee
Constituted under the University Human Research Ethics Committee (Non-Medical)

Ethics Clearance Certificate

Ethics protocol number: WBS/DB2410076/985

This certificate is only valid with a legitimate ethics protocol number and signed by the Researcher (belo

Project title	The use of alternative data by fintechs to provide access to credit for SMEs in South Africa
Investigator / Researcher	Mr Mandlenkosi Khupe
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed confidentialty.
Issue Date of Certificate	2023-02-21
Expiry date	Date of submission of the project / research report
Chairperson	Dr Pius Oba ☎ +27 11 717 3976 +27 82 733 6587 pius.oba@wits.ac.za

Declaration by Researcher

One copy must be signed by the Researcher and returned to the Chairperson of the Wits Business School Ethics Committee.

I fully understand the conditions under which I am authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I undertake to resubmit the protocol to the Committee.

Signature

10 March 2023
Date:

APPENDIX E: CONSISTENCY MATRIX

Research Question	Sub Questions	Proposition	Data Collection detail	Research Design	Data Type and Analysis Method
RQ1) How are fintechs transforming the SME credit landscape in South Africa by using alternative data?	RQ1.1) What types of alternative data points do fintechs commonly rely on when credit scoring SMEs?	Proposition 1: Alternative data is used by fintechs to effectively augment SME credit analysis where there are gaps in the traditional data points.	Interview guide questions: 3, 4, 5	Standard Qualitative Analysis	Qualitative / Thematic Analysis
	RQ1.2) How does alternative data complement or substitute traditional credit data in evaluating an SME's creditworthiness?		Interview guide questions: 6, 7		
RQ2) How are fintech lenders addressing information asymmetry in credit assessment for SMEs in South Africa?	RQ2.1) How do fintechs incorporate subjective measures into their credit algorithms/models, transforming soft data into decision-making criteria?	Proposition 2: Alternative data allows fintechs to depend on hard information derived from alternative data sources to make informed credit decisions.	Interview guide questions: 11	Standard Qualitative Analysis	Qualitative / Thematic Analysis
	RQ2.2) What data points do fintechs depend on as indicators of SMEs' future sustainability?		Interview guide questions: 7, 9, 12, 13		
	RQ2.3) What kinds of alternative data are fintechs utilising to limit issues of information asymmetry in the credit assessment process?		Interview guide questions: 8, 14		