Firm Productivity, International Trade and Competition:

Using micro data to examine the dynamics of South African firms

by

Tasha Lynn Naughtin

Dissertation presented for the degree of
Doctor of Philosophy (Economics)
in the Faculty of Economic and Management Sciences
at Stellenbosch University

Proefskrif ingelewer vir die graad
Doktor in die Wysbegeerte (Ekonomiese)
in die Fakulteit Ekonomiese en Bestuurswetenskappe
aan die Universiteit Stellenbosch

Supervisor: Professor Neil Rankin

December, 2016

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2016

With regard to chapter 2, "The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?", part of this chapter has been incorporated into the UNU-WIDER working paper entitled "The South African manufacturing exporter story" (Matthee, Rankin, Naughtin & Bezuidenhout, 2016) subsequent to the chapter being completed. The nature and scope of my contribution in Chapter 2 were as follows:

Nature of Contribution	Extent of Contribution
1. Data cleaning and all Stata coding except that mentioned explicitly below 2. Estimation and data analysis 3. Write up of literature review 4. Write up of results 5. All additional manuscript text 6. All tables and graphs	75%

The following co-authors have contributed to chapter 2, "The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?":

Name	e-mail address	Nature of Contribution	Extent of Contribution
Neil Rankin (Thesis Supervisor)	neilrankin@sun.ac.za	1. Conceptualising methodology 2. Assisting focus of chapter 3. Review and comments on chapter content	10%
Marianne Matthee	marianne.matthee@nwu.ac.za	1. Conceptualising research question 2. Review and comments on export-destination literature of chapter	10%
Carli Bezuidenhout	carli.bezuidenhout@nwu.ac.za	1. Initial coding of country-destination dummy	5%

		variables in SARS Customs data	

Signature of candidate: Tasha Naughtin *[Declaration with signature in possession of candidate and supervisor]*

Date: 26/07/2016

Declaration by co-authors:

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to chapter 2 "The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?",
2. no other authors contributed to chapter 2 "The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?" besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in chapter 2 "The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?" of this dissertation.

Signature	Institutional affiliation	Date
Neil Rankin *Declaration with signature in possession of candidate and supervisor*	Stellenbosch University	02/08/2016
Marianne Matthee *Declaration with signature in possession of candidate and supervisor*	North-West University	25/07/2016
Carli Bezuidenhout *Declaration with signature in possession of candidate and supervisor*	North-West University	25/07/2016

Abstract

Exports matter for economic growth. Exporting is associated with higher levels of employment, innovation, and investment. The South African government recognises the role of exports in stimulating the economy as evident in the New Growth Path, the National Exporter Development Programme, and the Medium-Term Strategic Framework 2014-2019. Despite this, relatively little is known about the dynamics of actual exporting firms in South Africa. Existing South African literature is limited due to the lack of access to comprehensive firm-level panel data. This thesis overcomes this by analysing two unique sources of substantial, detailed data on South African firms over time obtained from official government sources. This is one of the first instances in which data of this kind has been available for analysis in South Africa, and therefore it enables this thesis to study the South African exporting environment at the level of detail seen in the international literature.

Firstly, this thesis re-examines the 'stylised facts' of exporting in the case of South Africa in more detail. In contrast to the international literature, existing South African research concludes that exporters are, in general, no more productive than non-exporters. A number of possible explanations for this missing productivity premium have been suggested in the literature, however given the previous lack of sufficient firm-level data over time, few of these explanations have been adequately tested. This thesis is now able to test some of these explanations by making use of the two official firm-level datasets. It finds that both the nature of the data used in previous studies, as well as the homogeneous treatment of exporters, play a significant role in hiding South African exporters' productivity premium.

Secondly, this thesis employs a relatively novel unsupervised machine learning technique to test the robustness of the traditional classification of firms and exporters. Research using firm-level data usually classifies firms, and exporters, based on *a priori* assumptions. Firms are generally grouped by size, export participation, destination and products and correlations are reported based on these classifications. This study reverses the process through letting the data identify clusters. It uses cluster analysis techniques to identify classifications of South African manufacturing firms *a posteriori*. The findings highlight, among other things, the usefulness of exploratory techniques such as clustering for identifying potential heterogeneity among firms, particularly within large firm-level datasets.

Finally, the importance of identifying firm- and exporter-heterogeneity for policy purposes is illustrated. In particular, this thesis makes use of the substantial firm-level data, in conjunction with a natural experiment inherent in the South African tax legislature, to assess the impact of a specific tax incentive on small business investment and growth. The findings suggest that the incentive on small businesses did not have

the desired effect on capital accumulation in general. However, there were unintended benefits for small exporters, a result that is important for export-growth policy and one that would have been missed had all small firms been treated as homogenous in the analysis.

Opsomming

Uitvoere is van belang vir ekonomiese groei. Uitvoer word met hoër vlakke van werksgeleenthede, innovasie en belegging geassosieer. Die Suid-Afrikaanse regering erken die rol van uitvoere vir die stimulasie van die ekonomie, soos gesien kan word in die Nuwe Groeipad, die Nasionale Uitvoerontwikkelingsprogram en die Mediumtermyn Strategiese Raamwerk 2014-2019. Ten spyte hiervan bestaan relatief min kennis oor die dinamiek van werklike uitvoerfirmas in Suid-Afrika. Bestaande Suid-Afrikaanse literatuur is beperk weens ŉ tekort aan toegang tot omvattende firma-vlak-paneeldata. Hierdie proefskrif oorkom hierdie probleem deur twee unieke bronne van substansiële, gedetailleerde data oor Suid-Afrikaanse firmas, oor tyd, bekom vanaf amptelike regeringsbronne, te analiseer. Hierdie is een van die eerste gevalle waartydens data van hierdie aard beskikbaar was vir analise in Suid-Afrika, en dus maak hierdie dit vir dié proefskrif moontlik om die Suid-Afrikaanse uitvoeromgewing op dieselfde vlak van detail, soos in internasionale literatuur gevind word, te bestudeer.

Eerstens herbesin hierdie proefskrif oor die 'gestileerde feite' van uitvoer in die geval van Suid-Afrika in meer besonderhede. Teenstrydig met internasionale literatuur kom bestaande Suid-Afrikaanse navorsing tot die gevolgtrekking dat uitvoerders, oor die algemeen, nie meer produktief as nie-uitvoerders is nie. ŉ Aantal moontlike verduidelikings vir hierdie ontbrekende produktiwiteitspremie is in die literatuur voorgestel; gegewe die vorige tekort aan voldoende firma-vlak-data, oor tyd, is min van hierdie verduidelikings egter na behore getoets. Hierdie proefskrif kan nou sommige van hierdie verduidelikings toets deur gebruik te maak van dié twee amptelike firma-vlak-datastelle. Daar word bevind dat beide die aard van die data in vorige studies gebruik, sowel as die homogene hantering van uitvoerders ŉ beduidende rol in die versteking van Suid-Afrikaanse uitvoerders se produktiwiteitspremie speel.

Tweedens maak hierdie proefskrif gebruik van ŉ relatief nuwe, sonder toesig masjienleer-tegniek om die robuustheid van die tradisionele klassifikasie van firmas en uitvoerder te toets. Navorsing wat van firma-vlak-data gebruik maak, klassifiseer firmas, en uitvoerders, normaalweg op *a priori*-aannames. Firmas word normaalweg volgens grootte, uitvoerdeelname, bestemming en produkte gegroepeer, en korrelasies word gerapporteer gebaseer op hierdie klassifikasies. Hierdie studie keer hierdie proses om, deur die data die bondels te laat identifiseer. Dit maak gebruik van bondel-analisetegnieke om klassifikasies van Suid-Afrikaanse vervaardigingsfirmas *a posteriori* te identifiseer. Die bevindings lig, onder andere, die bruikbaarheid van eksploratiewe tegnieke, soos bondeling vir die identifisering van potensiële heterogeniteit tussen firmas, veral binne groot firma-vlak-datastelle, uit.

Laastens word die belangrikheid om firma- en uitvoerder-heterogeniteit vir beleidsdoeleindes te identifiseer, geïllustreer. Spesifiek maak hierdie proefskrif gebruik van substansiële firma-vlak-data, tesame met ’n natuurlike eksperiment inherent tot die Suid-Afrikaanse belastingwetgewing, om die impak van ’n spesifieke belastinginsentief op kleinsakebelegging en –groei te assesseer. Die bevindinge suggereer dat die insentief op kleinsakeondernemings nie die gewenste effek op kapitaalakkumulering, oor die algemeen, gehad het nie. Daar was egter onbedoelde voordele vir klein uitvoerders; ’n resultaat wat belangrik is vir uitvoer-groeibeleid, en een wat gemis sou word indien alle kleinsakeondernemings as homogeen in die analise hanteer sou word.

Acknowledgements

This thesis would not have been possible without the encouragement and support of my supervisor and mentor, Dr Neil Rankin, who has been a continued presence of inspiration and knowledge in my life since my undergraduate studies. Neil has patiently guided me through the daunting and often challenging doctoral process, providing clear and constructive advice throughout. I aim to take all that he has taught me and implement it in my future career as an economist, as well as in any future studies I may embark upon.

I am also grateful to my colleagues at the African Micro-economic Research Unit (AMERU) at the University of the Witwatersrand, which include Dr Gareth Roberts, Volker Schöer, and Dr Debra Shepherd (now at the University of Stellenbosch) who all provided a constant stream of debates and advice on multiple econometric issues as well as much needed emotional support and encouragement.

Particular thanks must go to the Economic Policy unit at the National Treasury, the South African Revenue Service (SARS) and the World Institute of Development Economic Research at United Nations University (UNU-WIDER) whose efforts to make available South African tax administrative data to select researchers were instrumental for the success of this thesis. I would further like to thank Dr Channing Arndt, Wian Boonzaier, Friedrich Kreuser and Dr Duncan Pieterse for the detailed discussions and assistance in getting this tax data into a useable format for the research. These individuals were always available for support and help and I have developed skills directly related to their inputs. In addition, thanks must be given to the other research groups who formed part of the pilot studies of this tax data who provided valuable feedback on certain parts of this research during various presentations and workshops.

Special thanks to Dr Marianne Matthee and Carli Bezuidenhout, my friends and colleagues, who provided valuable insights and constructive feedback on my research, particularly Chapter 2 of this thesis. I am grateful for the time spent discussing many economic issues, PhD woes and life lessons. I hope to continue both my professional, as well as personal, relationships that developed with these two outstanding researchers during these past years of study.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. It should be noted that opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF. I am thankful for this financial support of the NRF Innovation Doctoral scholarship, without which my studies would have been infinitely harder to complete.

Finally, my friends and family have provided years of support and encouragement. My parents, my sister and my grandmother have been pillars of strength throughout the process financially, physically and emotionally. I am grateful to be loved so completely by such wonderful people. My friends have been a great help, providing both inspiration and much needed distractions, bringing me back down to earth when I became overwhelmed. My parents-in-law have provided additional support in many ways for which am I ever grateful. I would, finally, like to thank my husband for putting up with all my ups and downs during this thesis. For bringing me comfort food and pushing me through many episodes writer's block. His belief and pride in my abilities kept me inspired and encouraged me to do the best that I can for my future.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

Chapter 1

Introduction

1.1. Motivation and Context

Access to individual firm data has significantly improved the understanding of how firms behave and how exporting works at the firm level. The empirical literature on exporting and firm performance, which has grown exponentially in the recent past, was initiated by the seminal work of Bernard and Jensen (1995 & 1999). These authors were the first to make use of substantial and official administrative data on firms to examine the differences between exporter and non-exporter performance. Their research found that exporting was rare: few firms export, and those that do export relatively little. In addition the authors found that exporters were different to non-exporters across a number of dimensions, including productivity.

This research inspired a stream of empirical work on the relationship between firm performance and international trade at a micro-level as researchers globally started to make use of the increasing access to rich, administrative firm-level datasets. See, for example, Bernard and Jensen (1995, 1999) for US firms; Bernard and Wagner (1997) for Germany; Aw and Hwang (1995) for Taiwan; Clerides, Lach and Tybout (1998) for Columbia, Morocco and Mexico; Muuls and Pisu (2009) for Belgium firms; and Kox and Rojas-Romagosa (2010) for the Netherlands.

This growing literature, in turn, inspired the development of 'new new', or micro, trade theories, such as Melitz (2003), Bernard, Eaton and Kortum (2003), and Bernard, Redding and Schott (2007). These theoretical papers incorporate these findings of firm-level heterogeneity, and other firm-level results discovered through the empirical work, to develop models of international trade that focus on the export-productivity relationship among heterogeneous firms. The Melitz (2003) model, for example, which is based on a Krugman-style (1980) model with product differentiation and increasing returns, incorporates firm heterogeneity into the framework by allowing productivity to differ between firms. Firm heterogeneity is also acknowledged in the model of Bernard, Eaton and Kortum (2003), which extends the Eaton and Kortum (2002) model of comparative advantage (which itself is an extension of the two-country Dornbusch *et al* (1977) Ricardian model of international trade) by introducing imperfect competition with variable mark-ups. Finally, the model of Bernard, Redding and Schott (2007) extends upon Melitz's (2003) monopolistic competition model by not only allowing for heterogeneity in firms' productivity levels, but also allowing for factor intensities to vary between industries and allowing countries to differ in their relative factor abundance.

It is clear that access to detailed and comprehensive firm-level data was the catalyst for a new wave of both the empirical and theoretical literature on exporting and firm performance internationally, and has substantially deepened the understanding of how firms operate in the international market. Development of the South African literature, in contrast, has been limited due to data availability. In particular, the majority of (if not all) South African studies, have been characterised by small samples based on ad-hoc surveys. Notwithstanding the data limitations, South African studies have found some results that are consistent with the international literature. Examples of these studies include Naude (2000), Rankin (2001), Rankin, Söderbom, and Teal (2006), Edwards, Rankin and Schöer (2008), and Matthee and Krugell (2012). These authors find that exporting in South Africa is similar to the stylised facts of the international export literature: exporting is rare and those firms that do export differ from their non-exporting counterparts across a number of dimensions such as size, wages, capital intensity and labour productivity.

An interesting anomaly in the South African literature, which contradicts international research, is that South African exporters are, in general, no more productive than non-exporters. The export-productivity relationship is studied in detail by Rankin (2013), Naughtin and Rankin (2014) and Matthee *et al* (2015). These authors conclude that it is only when export destinations are included in the analysis that the productivity premium becomes apparent. However, given the data limitations of these studies, it is difficult to determine whether these results are driven by the sample data, which generally oversample larger firms, or are actually a characteristic of the South African firm-level environment. Indeed, the lack of access to the type of official, substantial administrative data used in studies on other countries has significantly slowed the development of the South African export-productivity literature (Edwards *et al*, 2008).

This thesis extends the South African literature by being the first to use two substantial and official datasets to investigate export activity of South African firms. The first is a large, comprehensive data set collected every few years by the official statistics body of South Africa, Statistics South Africa (StatsSA). The sample of firms in this dataset is based on a sampling frame that is representative of the economy as a whole (and is thus more representative of the economy as a whole compared to the other datasets that previous research has used). The second set of data is based on the actual population of firms in South Africa. This rich, official dataset consists of official tax administrative data collected annually by the South African Revenue service (SARS).

The innovation in this thesis is therefore to make use of these two large, comprehensive datasets to investigate exporting at the firm-level in greater detail for South African manufacturing. There are a number of reasons for why this thesis focusses on manufacturing firms. Firstly, this thesis aims to replicate work from previous international and South African studies which use manufacturing firm-level data. Secondly, traditionally manufacturing exports have been an important target for stimulating growth in many countries (according to the Commission on Growth and Development (2008) report which further states that manufacturing growth also increases the incomes of the poor – an important consequence, particularly in developing economies). Finally, in order to make comparisons between the LSS dataset and SARS dataset, it was necessary to restrict the analysis of the SARS data to manufacturing firms since the available LSS dataset is on the manufacturing sector only.

Understanding the exporting environment is important in the South African context given that exports are a key mechanism through which countries can facilitate growth, and therefore improve overall wellbeing in the economy (Greenaway, Gullstrand, & Kneller, 2005). The crucial role that exporting plays in facilitating economic growth is further argued by the Commission on Growth and Development in their 2008 Growth Report which concludes that openness and exporting are one of the significant factors that characterised the high levels of sustained economic growth (for at least 25 years) of 13, mainly developing, economies. In addition, a recent study on three Sub-Saharan African countries (Botswana, Equatorial Guinea and Mauritius) which investigates the robust, sustained growth of these economies, finds significant evidence in support of export-led growth (Ee, 2016).

One way for government to stimulate exporting in the economy, and therefore encourage economic growth, is to intervene in such a way as to increase the real productivity of firms (Hausmann & Klinger, 2005). Aw and Hwang (1995: 314) believe that 'increases in productive inputs and/or greater average utilization of capacity play major roles in this higher output growth in many export-oriented developing countries'. The benefits of participating in international trade include improvements in allocative efficiency, increase in competition and investment and improvements in knowledge and technology over time (Roberts, 2000). Further, the increase in competition from trade can spur innovation and stimulate productivity growth in domestic firms (Edwards & Lawrence, 2012). Given these, and other, benefits of exporting, it is evident that understanding the underlying behavioural characteristics of South African exporters is an important and necessary task for policy makers.

The South African government acknowledges that it is good policy to increase export participation of domestic firms, see for example the Industrial Policy Action Plan (IPAP), the Medium-Term Strategic Framework (MTSF) 2014-2019, as well as the National Exporter Development Plan (NEDP), which was developed with the specific task of expanding the exporter base and fostering exporting in order to stimulate economic growth and generate employment within the economy (dti, 2013). However, despite recognising the importance of export growth for overall economic development, these government reports provide very little detail on which types of exporters need encouragement and through which mechanisms this encouragement can be facilitated. In order to adequately answer these questions and, consequently, in order to develop high-quality policy proposals, research must be done using high-quality firm level data (Falk, Hölzl & Oberhofer, 2015).

The studies in this thesis are some of the first in South Africa to take advantage of unique, quality firm-level datasets to examine firm performance and international trade in order to gain a more detailed understanding of the export environment. It does so in three ways. Firstly, while previous South African studies on exporter performance generally find results consistent with international studies, in that exporting firms exhibit a number of superior performance characteristics relative to non-exporters, little evidence exists to support the now stylised fact that exporters are more productive than non-exporters. According to Rankin (2001), exporters only exhibit significant productivity premiums when exporting outside of Africa. Chapter 2 therefore sets out to investigate whether this missing productivity premium is a characteristic of the small datasets previously used, or if it is indeed a South African exporter phenomenon. It does so by estimating productivity functions using the semi-parametric methodology proposed by Levinsohn and Petrin (2003), which given the limited nature of previous datasets had not been done previously. The use of these larger, comprehensive datasets, in conjunction with the Levinsohn-Petrin estimation, will enable a more detailed study of South African firms and their trade activities.

Chapter 3 uses these substantial datasets, along with a set of unsupervised machine learning techniques, to confirm the robustness of the groups, or typologies of exporters and non-exporters as identified in Chapter 2 and other South African studies. Instead of imposing *a priori* assumptions onto the data, as is done in Chapter 2 and previous research (both on South Africa and other countries), Chapter 3 makes use of unsupervised algorithms to identify correlations and groupings in the dataset without any external input. This serves as a robustness check of the *a priori* groupings used in Chapter 2 and an approach to uncover the broad 'segments' of firms in South Africa. The use of such techniques is increasing in many disciplines (Everitt *et al*, 2011), but remains rare in the broader field of economics (Murray, 2016).

Finally, this thesis makes use of the StatsSA dataset to examine how policy changes, which are not specifically focused on exporting firms, may have implications for exporting. In particular, Chapter 4 uses a natural experiment which changed the threshold for a specific tax exemption amongst small firms to investigate the response of small firms and small firm exporters. These type of impact assessments are important for policy-makers, but require detailed, representative data at the firm level over time, data which has only now become available. Chapter 4 thus provides a case study of how this type of data can be useful in evaluating specific policies which may affect export behaviour.

Despite the common theme of exporter heterogeneity and the use of substantial and official firm-level data, each of the three main chapters is presented in this thesis as a distinct research paper, as is common now for many PhDs. The remainder of this introduction will discuss in more detail the research questions and contributions of Chapters 2 to 4 of this thesis.

1.2. South African Exporter's Missing Productivity Premium

The understanding of export dynamics and firm behaviour has been improved by the increased availability of firm-level data. The international literature on this topic is growing exponentially (see Wagner 2007 and 2011 for an extensive literature review). While a number of South African studies have similarly set out to investigate exporting at a micro-level, they have been disadvantaged by a lack of access to good, detailed data that tracks the population of firms over time. Chapter 2 of this thesis overcomes this limitation through the use of newly available, highly disaggregated population data. This new dataset encompasses the population of firms from the tax administrative records of the South African Revenue Service (SARS) and allows for the detailed type of research found in the international literature that, until now, has been out of reach of South African researchers.

One of the major contributions of this chapter is the construction of a detailed dataset of the population of South African manufacturing firms matched to their export transactions. Substantial time and effort went in to the assembly of this dataset, which further contributed to the construction of the South African Revenue Service and National Treasury (SARS-NT) firm-level panel database: a substantial and unique source of data for the study of South African firm behaviour (Pieterse, Kreuser & Gavin, 2016).

Appendix A discusses the process of developing the manufacturing firm-level dataset in more detail, but in brief: data from three sources (company income tax records (CIT), employee income tax certificates (PIT) and customs transactions records) were merged together to form the workable manufacturing firm dataset. The CIT data, which contains information on approximately 700 000 unique firms (30 000 of which are identified as manufacturing firms), forms the base of the manufacturing firm dataset. This data comes from two overlapping tax forms (the IT14 and ITR14 tax returns), and therefore the first step was to combine the data from both these forms and clean the data where necessary. The fields in these two forms differ which created a challenge for combing the data, the details of which are given in Appendix A. This CIT data was then augmented with additional variables from the PIT and customs data. In particular, the total number of employees per unique firm from the PIT data was first calculated and then merged onto the manufacturing firm data by means of a link table (a table provided by SARS which contains all unique firm identifiers and corresponding PIT and customs identifiers). In addition, data on exporting activities of the firms was obtained from the customs transactions data, a dataset which contained approximately 10 million transactions for around 40 000 unique traders. This too was merged onto the manufacturing firm dataset to complete the manufacturing firm-level dataset on which the analysis in this and the next chapter is based.

Chapter 2 then makes use of this rich dataset, as well as another substantial dataset from StatsSA's Large Sample Survey on manufacturing firms, to investigate the South African exporter's missing productivity premium. The limited research on South African exporters has found similar results to international studies on firm-level exporting: few firms export, and those that do export very little (Rankin, 2013). In addition, exporters exhibit a number of export premia over non-exporters across a number of dimensions including wages, size, capital intensity and labour productivity. In contrast to international studies, however, little evidence exists that exporters are more productive than non-exporters. In fact the productivity premium seems to be linked to the destination of the exports (Rankin, 2001). Chapter 2 therefore attempts to answer the question of whether this missing productivity premium is a result of the small samples firms used in the previous South African research, which may not truly represent the population of exporters, or a true characteristic of the South African export environment.

The chapter contributes to the literature of South African firms by confirming that, in line with international studies, South African exporters are indeed more productive than non-exporters. One potential explanation for why the results in this chapter differ from previous South African research is that the previous studies were based on a non-representative, small sample of firms. The use of the population data in this chapter overcomes these potential sampling issues, and provides a more detailed analysis of exporting at the firm-level.

Chapter 2 also confirms the previous South African findings that when it comes to productivity premiums, export destination does matter. This chapter finds that higher productivity premium are associated with firms exporting outside of Africa relative to both non-exporters and Africa-only exporters. Further, the results in this paper suggest that firms exporting to multiple destinations are significantly more productive than firms which export to a single destination.

The results of this chapter are important from a policy perspective, particularly for policy that wants to encourage export growth. Chapter 2 confirms that not only are exporters different to non-exporters, but further that exporters also differ from each other. These different degrees of heterogeneity mean that export policy cannot provide a one-type-fits-all solution for exporters and expect to see significant export growth. Different firms and exporters have different characteristics and behaviours and are therefore likely to respond differently in certain situations. It is therefore important that policies to promote export growth first identify the 'type' of exporter they wish to support before implementing development strategies.

1.3. Clustering South Africa's Firms and Exporters

The identification of 'types', or groups, of firms has traditionally been based on *a priori* assumptions, often founded by past studies. In the exporting literature, for example, firms are generally classified into two predefined groups: exporters and non-exporters. However, as Chapter 2 suggests, there may be more heterogeneity within these groups. Methods which use *a priori* classifications of firms fall into the class of supervised machine learning techniques, most commonly known as discriminant analysis (Jain, Murty & Flyn, 1999). A limitation of supervised learning techniques becomes apparent in the presence of large, complex datasets: in practice, it becomes harder for supervised learning algorithms to learn models with deep hierarchies as the number of observations increase (Valpola, 2000).

On the other hand, unsupervised machine learning techniques try to identify the underlying, intrinsic structure of the data without imposing any external assumptions. In other words, these techniques let the data determine the groupings of firms. Unsupervised learning algorithms benefit from being able to identify patterns in the data that have not previously been considered and are further able to learn larger, complex models than supervised algorithms (Valpola, 2000).

One of the most commonly used unsupervised techniques is cluster analysis (Everitt *et al*, 2011). The object of cluster analysis is to group observations (such as firms), or in some cases variables, into clusters so that the observations in each cluster exhibit a high degree of "natural" similarity among themselves, but are relatively distinct from observations in other clusters (Anderberg, 1973).

Cluster analysis has been widely used in several disciplines, including statistics, software engineering, biology, psychology and other social sciences, in order to identify natural groups in large amounts of data (Tvaronaviciene, Razminiene & Piccinetti, 2015). The use of cluster analysis in economic fields is slowly gaining traction. For example, Vázquez & Sumner (2012) use clustering techniques to test the traditional *a priori* classification of developing countries into either low- or middle income groups, and find that a five-cluster solution is more appropriate and informative. In a similar fashion, Tiri, Peeters and Swinnen (2006) as well as Srholec and Verspagen (2008), use cluster analysis to identify innovation strategies and conclude that the traditional classifications of three or four strategies still contain significant heterogeneity within them and should therefore be segmented further.

When presented with large datasets, cluster analysis can be particularly useful for initial exploratory pattern analysis and decision making (Jain, Murty & Flyn, 1999). The results from the cluster analysis can then be used to inform additional discriminant, predictive analysis and therefore contribute towards less subjective based policy-making. Chapter 3 illustrates the usefulness of this type of analysis for the case of South African firm and exporter heterogeneity.

To the best knowledge of the author, unsupervised techniques such as cluster analysis have yet to be implemented on firm-level trade data in the South African context. It is here where chapter 3 makes its first contribution. The availability of the large, official datasets for use in this thesis allows this chapter to employ cluster analysis on the population of South African firms in order to identify the typology of South African firms in general, and exporters in particular. In this way the robustness of the *a priori* classifications of firms and exporters used in the traditional trade literature (and in Chapter 2) is tested.

Chapter 3 not only clusters firms, but further clusters key performance variables used in the dataset. Given the size of the substantial datasets used in this thesis, and given the standard problem of reducing dimensionality when presented with multivariate data (Mardia, Kent & Bibby, 1979), Chapter 3 applies a relatively recent algorithm ClustOfVar to identify groups of variables. Unlike the traditional methods used for dimension reduction, Principal Component Analysis (which is used only for quantitative data), or Multivariate Correspondence Analysis (used only for qualitative data), ClustOfVar is able to handle both types of data (qualitative and quantitative) data simultaneously. Following this, the principal components from each cluster of variables are then extracted in order to cluster firms. Finally, a classification and regression tree (CART) analysis is used to identify the variables that drove the formation of the clusters. There are few studies which employ techniques to cluster mixed datasets at the firm-level, particularly, to best of the author's knowledge, in South African trade studies and this is where Chapter 3 further contributes to the literature.

The results of the cluster analysis in Chapter 3 confirm that the existing *a priori* classifications of firms, and exporters, are broadly correct. Further, variables such as size, export status, productivity, and to some degree, export destinations are shown to be the main drivers of cluster formation. Therefore, without imposing any external assumptions on the structure of the data this chapter once again highlights the degree of heterogeneity among both firms and exporters but also the importance of firm size and productivity for export participation. The information obtained from clustering both firms and variables adds support to targeted policy making and cautions against the use of blanket policies, particularly for exporters, since the mechanisms or processes through which the various 'types' of exporters develop are different and therefore require different support/ facilitation strategies. The results also indicate that sectors are not an important driver for cluster formation – firms of similar sizes and productivity levels across sectors are more similar than firms within the same sector. This suggests that policies which aim to target similar firms need to use factors like size, rather than sector, to group these firms.

1.4. Tax Incentives and Small Exporters

Both Chapters 2 and 3 confirm that a certain degree of heterogeneity among firms, and exporters, is extant in South Africa. Identifying these different groups of firms and exporters is important for policy. As mentioned, the effect of policies may be muted if they are applied too generally: firms differ in terms of characteristics and behaviours, and consequently actions/ strategies to encourage growth and development of firms would benefit from being more targeted.

An example of a policy where the effects remain inconclusive is that of tax incentives for stimulating investment and growth of firms. Policies which aim to encourage small business investment and growth often do so through use of tax incentives (Guenther, 2012). However, despite its popularity, very little empirical research has been done on the impact of such incentives, particularly in the developing country setting (Klemm & Van Parys, 2009). Further, the current literature has yet to reach a consensus on the effectiveness of these policies (James, 2009). One potential reason for this may be due to the aggregate nature of the datasets used in majority of the tax literature.

Chapter 4 contributes to this sparse literature by analysing the impact of South Africa's small business tax incentive, Section 12E on the behaviour of firms using highly disaggregated firm-level data. This chapter recognises, based on the previous chapters' results, that firm heterogeneity plays a large role in the South African economy, particularly in terms of exporting. It therefore makes use of this knowledge to investigate the effect of the tax incentive at various levels of disaggregation. Very little work has been done on the heterogeneous impact of tax incentives for investment, particularly on exporting firms.

Section 12E allows for lower and progressive tax rates as well as accelerated depreciation for firms that qualified as small business corporations (SBCs). It was created for the purpose of stimulating capital investment, encouraging growth and generating employment in small businesses. Section 12E was implemented into the tax legislature in 2001, and over the years the threshold for firms to qualify as an SBC has changed. One such change occurred between 2005 and 2006. This change provides a natural experiment which Chapter 4 uses to examine the effectiveness of the small business tax incentive on the investment levels of small firms and small exporters. By taking advantage of the comprehensive StatsSA panel of South African firms between 2005 and 2008, Chapter 4 is able to assess the behavioural changes of firms that initially did not qualify for the tax incentive in 2005 but did qualify in 2006, relative to firms who never qualified.

The detailed analysis of firms in Chapter 4 again confirms the importance of firm heterogeneity. Firstly, while the results indicate that firms that qualified for the tax incentive in general exhibited higher levels of employment between 2005 and 2008, they also show a decrease in capital accumulation – a result that conflicts with the ultimate aim of the tax incentive policy, and a result that is most commonly found in the literature (Engelschalk, 2004; and Knittel, 2005). However, once heterogeneity is introduced, for example when firms are segmented into exporters and non-exporters, the results take on an additional dimension. Non-exporting SBCs saw an increase in employment and a reduction in capital, but exporting SCS exhibited the opposite: a decrease in employment and an increase in capital accumulation. These results alone caution against treating small firms as a homogeneous group when implementing growth policies since to do so would be likely to result in ambiguous outcomes.

Secondly, Chapter 4 further illustrates that by assuming homogeneity among small businesses it is easy to miss potential, unintended, impacts of policy interventions. One such unintended consequence of the SBC tax incentive, for example, was its association with survival among small exporters. This finding suggests that a tax incentive on small exporters may be useful in facilitating growth in South African exports, an important goal for the South African government. Without the ability to correctly identify heterogeneous groups of firms, these types of findings will remain hidden in the data.

1.5. Summary

The research contained in this thesis is based on two unique datasets: a large sample survey of South African manufacturing firms, conducted by the country's official statistical unit, StatsSA; and an extensive collection of population data from the South African Revenue Service's (SARS) tax administrative data. A major contribution of this thesis is in the construction of the latter. A considerable amount of time and effort was involved in setting up, cleaning and combining the substantial detail of the population of entities into a dataset of manufacturing firms to be analysed in this thesis, the efforts which further contributed to the construction of an even larger South African firm-level panel dataset developed by SARS and the National Treasury of South Africa.

South African studies on exporting at the firm level have been plagued by the absence of comprehensive, official data which tracks firms over time. An additional contribution of this thesis is therefore the use of these two substantial and official sources of firm level data to investigate the dynamics of South African exporters at a detailed, micro-level. It does so in three ways. Firstly, Chapter 2 uses this detailed data to replicate earlier South African studies on the exporter-productivity relationship. In particular it investigates whether South Africa's exporters' missing productivity premium is an anomaly of the South African economy, or if it is simply a characteristic of the small sample of data used in the previous research. Secondly, Chapter 3 takes advantage of these large, comprehensive datasets to assess the robustness of the *a priori* classifications of firms and exporters traditionally found in the literature. It does so by employing an exploratory technique, used often in the data science field, called cluster analysis which allows the data to intrinsically identify natural groupings of firms in the data. This type of analysis has not before been done on South African exporters at the firm-level and adds a degree of nuance to the way in which firms, and exporters are grouped. Finally, the findings in both Chapter 2 and Chapter 3 confirm the presence of significant heterogeneity among South African firms. Chapter 4 makes a further contribution to both South

African specific policy analysis but also more broadly to the understanding of the impact of small business tax incentives. It makes use of a natural experiment inherent in the South African tax legislature to highlight the importance of identifying these heterogeneous groups of firms from a policy perspective and shows that, at least in the South African context, a small business tax incentive may have a positive impact on small exporters.

In summary, this thesis highlights the importance and usefulness of comprehensive, detailed firm level panel data for analysing the dynamics of firms in South Africa. Further, the results of the three empirical chapters indicate that a close relationship exists between firm size, productivity and exporting. For example, Chapter 3 shows that firms are more similar by size, rather than sector or productivity. Therefore, classifying firms by sector is potentially the wrong variable to group firms by. In addition, Chapter 4 shows that the small exporters and small non-exporters (who were identified *a posteriori* in Chapter 3) exhibit heterogeneous responses to a tax incentive policy. One conclusion from these findings is that despite the importance of exporters and small business, particularly in terms of employment generation, small firms can only be made to export if they already have the underlying productivity characteristic that identifies other exporters.

On the whole, the results of this thesis suggest that exporting is an outcome and therefore policy research needs to be geared at understanding the underlying firm-level characteristics that influence this outcome. If government wishes to stimulate growth and generate employment it is imperative that a deep understanding of firm behaviour is obtained. The results of this thesis serve to caution against treating firms as homogeneous, particularly when implementing and analysing the impact of policy, since such broad classifications can result in misleading conclusions and potentially feeble future interventions.

Chapter 2

The South African Exporter's Missing Productivity Premium: A characteristic of the data or something more?

2.1. Introduction

It is well known that exporting encourages overall economic growth. South African policy makers' acknowledgement of the importance of export growth is evident in a number of policy documents, such as the National Development Plan (NDP) – 2030, the New Growth Path (2011) and the Industrial Policy Action Plan (IPAP) which all recognise the export sector as a generator of comprehensive, job-creating growth. In order to expand employment and bring about annual economic growth in excess of 5 percent the NDP suggests increasing exports, particularly in construction, mining and manufacturing. The New Growth Path (2011) further states that increasing exports particularly into the region and emerging economies will stimulate investment, productivity, employment and income.

This sentiment is echoed in a recent World Bank report which suggests three opportunities to stimulate export growth, and consequently employment, in South Africa: increased competition among firms; reduced infrastructure bottlenecks; and deeper regional trade integration (World Bank, 2014).

Export growth can ultimately lead to a higher standard of living. However, despite the emphasis on exports in government's growth strategies, very little research has been done on export dynamics at a micro-level in South Africa. Part of the reason for this is the limited access to good firm level data over time. A better understanding of the characteristics and behaviour of exporting firms, however, is crucial in order to design policies to increase exports.

Although there is a relatively large body of work on this topic, particularly in the developed setting, it has essentially been macroeconomic in direction. According to Bernard, Eaton and Kortum (2003, p. 1268) 'trade theory has been aimed at understanding aggregate evidence on such topics'. As highlighted by Edwards, Rankin and Schöer (2008), it is the sum of exports at the firm level which make up aggregate exports and it is therefore the actions and decisions of firms in which we should be interested. This is particularly the case in order to develop appropriate policies for improving South Africa's economic growth through exports.

International studies find that exporting firms are superior to non-exporting firms across a number of characteristics – most notably, exporters are more productive than non-exporters. It is therefore argued that increasing competition will allow 'good' firms to cross over some productivity threshold necessary for survival in the export market. South African literature, however, finds little evidence that exporters are more productive than non-exporters. It is only when firms export outside of Africa that productivity premiums are recognised.

This chapter adds to the literature on South African exporting firms through the use of two substantial firm-level datasets to analyse the export behaviour of firms in the manufacturing sector – a sector highlighted as one which is essential for low-skilled job creation (World Bank, 2014). It will attempt to replicate the findings of previous South African studies, which are based on small cross-sectional datasets, to determine whether South Africa indeed has a missing productivity premium, and, if so, examine how destination influences exporter behaviour.

The findings of this chapter confirm that firms exporting outside of Africa do perform better in terms of productivity than firms exporting only within the region, particularly among medium and large firms. Furthermore, there is evidence to suggest that Africa-only exporters are no more productive than non-exporters. In addition, exporting to multiple destinations is associated with higher productivity premiums relative to single destination exporters and non-exporters.

These results highlight the importance of recognising exporter heterogeneity. Policies which encourage regional export growth, such as those suggested in the World Bank report, might not be effective for medium and large firms who experience the biggest productivity jump when exporting outside of Africa and indeed exhibit no productivity premium over non-exporters if they export only to Africa. In addition, it is important to take cognisance of the lower productivity medium and large firms who may be competitive in the region but not internationally.

The following section presents a brief review of the trade theory. Section 2.3 briefly discusses the literature on exporting firms and export destinations. Section 2.4 presents an overview of the two datasets used in the analysis. Descriptive statistics are presented in section 2.5. Section 2.6 and 2.7 discuss the methodology used and results respectively. Section 2.8 concludes.

2.2. Theoretical Background

Traditional theories of international trade maintain that a country will export one set of products and import another. These 'old' theories explain the flow of goods between countries in terms of comparative advantage which arises either because of different relative productivities as in 'Ricardian' trade theory or because of a combination of differing factor abundance and factor intensity between countries as in the 'Heckscher-Ohlin' model of trade. Each of these is discussed below.

In the early nineteenth century British economist David Ricardo proposed an explanation of trade in terms of differences in labour productivity in different countries (Krugman & Obstfeld, 2006). The Ricardian model takes labour to be the only factor of production. Ricardian theory, which introduced the concept of comparative advantage, leads one to expect that when trade is free a country will, and should, export goods that their labour produces relatively efficiently. That is to say the country should specialise (Neary, 2009).

This degree of specialization, however, is extreme and although this model became, as Neary (2009, p.218) states, a 'cornerstone of international trade theory' it is lacking in that it does not take into account the effects of international trade on the distribution of income, nor does it allow any role for differences in resources among countries as a cause of trade (Krugman & Obstfeld, 2006). Ricardo took international productivity differences as given. However, over the years it became clear that traditional trade theories, based on comparative advantage, cannot account for observed intra-industry trade (Davis, 1995).

In the twentieth century Eli Heckscher and Bertil Ohlin developed an alternative theory which takes into account the effects of international differences in factor endowments as a major source of comparative advantage. This theory also discussed the effects of trade on the distribution of income. Both effects were ignored by the Ricardian model. The Heckscher- Ohlin theory of trade states that a country tends to produce and export the good that uses intensively its relatively abundant factor of production, however, in reality complete factor-price equalization is not observed because of wide differences in resources, barriers to trade, and international differences in technology (Krugman & Obstfeld, 2006).

Traditional trade theories, such as the ones discussed above, considered models in which trade is based on comparative advantage. In the Ricardian case comparative advantage arises because of productivity differences. In the Heckscher-Ohlin case comparative advantage arises because of a combination of cross-industry differences in factor intensity and cross-country differences in factor abundance (Bernard, Jensen, Redding, & Schott, 2007). However as the years progressed it became clear that international trade was not fully explained simply by factor endowments or even comparative advantage (Neary, 2009). Indeed a large share of international trade began to take place between relatively similar countries, a fact that could not be explained by the traditional theories of trade mentioned above (Krugman & Obstfeld, 2006).

This called for a new framework to be developed, the first of which was Edward Chamberlin's model of 1933. His principles were then broadened by Dixit and Stiglitz in their 1977 paper 'Monopolistic Competition and Optimum Product Diversity'. Their framework was one which contained 'all the ingredients needed to explain intra-industry trade' (Neary, 2009, p. 220). Specifically it incorporated a demand for differentiated products, implied that consumers prefer diversity and that returns to scale are increasing.

Paul Krugman was yet another author who developed trade theory to incorporate increasing returns to scale and intra-industry trade. He claimed in his widely celebrated paper of 1979, that international trade flows may simply be an effect of countries taking advantage of economies of scale, rather than differences in technology or endowments. He believed that while the role of scale economies in causing trade has been known for some time, it has been underemphasized in traditional trade theory (Krugman, 1979).

Krugman's model is consistent with the empirical evidence on intra-industry trade. Like traditional trade theorists, Krugman suggests both positive and normative predictions about trade, the difference being that the countries are identical so there is no role for comparative advantage (Neary, 2009).

New trade models explain how a combination of economies of scale and consumer preferences for variety leads relatively similar firms to specialize encouraging two-way trade between countries (Bernard *et al*, 2007). However, although new trade theory has attempted to address the shortcomings of traditional trade theories, Deraniyagala and Fine (2001, p.812) suggest that the new trade theory provides 'few, if any, unambiguous conclusions'.

Although their importance to international trade cannot be argued, both old and new trade theories are more useful at an aggregate level. Given the underlying assumption of homogeneous firms and products, a number of researchers argued that these theories were inconsistent with a number of firm-level facts, such as differences in productivity, capital intensity and wages between exporting and non-exporting firms (Bernard *et al*, 2003; Bernard *et al*, 2007; Edwards *et al*, 2008)

These empirical studies highlight the importance of understanding the role of firm heterogeneity in explaining trade patterns across countries. In order to better understand the intricacies of the export process, one needs first to identify the aspects of exporting at the firm level (Hallack & Sivadasan, 2006; Edwards *et al*, 2008).

Both old and new trade theory makes use of a representative firm in their models which assists in general equilibrium analysis. Melitz (2008) suggests that these models will predict that trade will affect all firms in a sector in similar ways and since this is not the case, as indicated by the empirical results above, the results cannot be explained by models based on representative firms. In response firm-level heterogeneity has been incorporated into what can be called 'new new' trade theory to account for the many new firm-level facts (Baldwin & Robert-Nicoud, 2008).

The most renowned of these 'new new' trade theories is that of Melitz (2003) and, later, Melitz and Ottaviano (2008). Melitz too held that previous theories of trade focussed too much on inter-industry reallocations and how trade affects all firms in similar ways and not enough on the intra-industry reallocations of trade. His theory builds upon Krugman's (1980) analysis of trade in the presence of product differentiation and increasing returns, by incorporating firm heterogeneity into a monopolistic competitive framework.

The Melitz (2003) framework shows how trade liberalisation leads to additional inter-firm reallocations toward more productive firms and thus explains how trade can stimulate aggregate productivity growth, without necessarily affecting intra-industry efficiency. Prior to entry into the international market, firms face uncertainty about their future productivity when making an irreversible investment decision, such as sunk costs of entry. Due to the nature of these sunk costs the more productive firms remain active whilst least productive firms face negative profits and as a result exit. This leads to the conclusion that exposure to trade will only allow for the more productive firms to enter the export market at the same time pushing the least productive firms to exit the market.

Melitz (2003, p.1714) argues that this theory illustrates how trade contributes to a 'Darwinian evolution within an industry' in which the least efficient firms are forced to contract or exit while the more efficient firms excel. This framework provides a significant contribution to the modelling of additional firm-level decisions in an open economy where heterogeneous firms self-select into different types of activities.

The influential contribution of the Melitz (2003) heterogeneous trade model to international trade theory cannot be argued. However, this model imposes a number of simplifying assumptions, for example: each firm produces one good, and sells it monopolistically; countries are symmetrical; the single factor of production is labour, which is homogenous; and firms export either to all countries, or no countries. Researchers were willing to accept these assumptions at the time, given the lack of firm-level data on export destinations, employee information and individual products.

An increase in access to transaction-, employee- and product data at the firm-level has more recently encouraged a number of extensions to heterogeneous firm models. Baldwin and Harrigan (2007), for example, develop a model which incorporates asymmetrical countries and heterogeneous quality among firms. In their quality heterogeneous-firms trade model, it is assumed that customer choices are based not solely on price, as in the Melitz model, but also on product quality. Higher quality is associated with higher prices, and therefore higher profits. Since higher quality firms are more likely to be competitive, they are more able to overcome distance-related trade costs and therefore more likely to succeed in exports markets.

Another modification to the heterogeneous firm model incorporates endogenous mark-ups that respond to the level of competition in the market along with heterogeneous firms. Melitz and Ottaviano (2008) find that market size brings about changes in the equilibrium distribution of firms and their performance measures: larger markets are generally more competitive. This tougher competition leads to lower average mark-ups and hence higher aggregate productivity. In their words this model ‘exhibits a link between bilateral trade liberalisation and trade reductions in mark-ups, thus highlighting the potential pro-competitive effects often associated with episodes of trade liberalisation’ (Melitz & Ottaviano, 2008, p. 296).

Further examples of extensions to the heterogeneous firm trade model include the introduction of relative factor abundance and comparative advantage (Bernard, Redding & Schott, 2007), multi-product firms (Bernard, Redding and Schott, 2011; and Mayer, Melitz & Ottaviano, 2011) and dynamics (Ghironi & Melitz, 2005; and Ottaviano, 2011)[1].

2.3. A Review of the Empirical Literature

The theoretical trade literature discussed previously has been greatly influenced by the findings of empirical, firm-level research. Much of this research has focussed on firm heterogeneity in international trade, in particular investigating the performance differences between firms that participate in trade and those that do not. Indeed, it is now a stylised fact that exporting firms exhibit performance premiums across a number of characteristics relative to their non-exporting counterparts.

Table 2.1 summarises the export premia found in a number of studies on firm heterogeneity in international trade[2]. The usual technique employed to extract the performance premium of exporters relative to non-exporters follows that of Bernard and Jensen (1995), who use a simple regression of firm characteristics, in natural logarithm, on an export status dummy and controls such as industry and size. This technique provides a useful way to discuss the differences in percentage terms rather than absolute values. In table 2.1, the studies have been separated into those which examine percentage differences and those which examine absolute differences.

The fact that exporters are superior to non-exporters holds regardless of the country and period of analysis. For example, using plant-level data spanning 1976 to 1987, Bernard and Jensen (1995) find that employment is around 94 percent higher among exporters relative to non-exporters, on average. In addition exporters pay 9 percent more wages, are 9 percent more capital-intensive and produce 15 percent more valued added per employee than non-exporters. Similarly, Isgut (2001) investigates the exporter premia for Columbian plant-level data over the period 1981 to 1991. Despite the developmental differences between Columbia and the US, the stylised facts remain: on average exporters employ more workers than non-exporters, pay higher wages, and are more capital intensive and more labour productive than non-exporters.

[1] See Redding (2010) and Melitz and Redding (2010) for a more in-depth look into a survey of trade theory.
[2] See Wagner (2007) and Wagner (2011) for a more comprehensive survey of international trade and firm performance.

In absolute terms, Aw and Hwang (1995), using firm-level data for 1986, find that the average Taiwan exporter employs 213 workers compared to the non-exporting firm average of 26. In addition, the authors find evidence that confirms that exporters are also more capital-intensive and have higher levels of output per employee than non-exporters. Similar results are found in Indonesia: exporters employ on average 252 workers compared to non-exporters who employ 65, employ higher levels of capital than non-exporters and are more labour productive than non-exporters (Blalock & Gertler, 2004).

These results are supported by a number of studies including: Bernard and Wagner (1997) for Germany; Muûls and Pisu (2009) for Belgium; Bernard *et al* (2007) for the US; Andersson, Johansson & Loof (2008) for Sweden; Castellani, Francesco and Tomasi (2010) for Italy and Kox and Rojas-Romagosa (2010) for the Netherlands.

The finding that exporters are more productive than non-exporters has gathered significant attention in the international trade literature. As previously discussed, exporting is associated with a number of sunk costs. Firms need to reach a certain productivity threshold before they can overcome these costs and make a profit in the international markets (Melitz, 2003). Exporters are therefore likely to have significant higher productivity relative to non-exporters.

A number of studies have tested this prediction of the productivity premium and found it to be true. For example, Muûls and Pisu (2009) for Belgium, show that exporting is associated with a 9 percent productivity premium. Anderson *et al* (2008) examine the relationship between exporting and productivity for Swedish firms and find, after controlling for various other factors which may affect productivity such as capital and firm size, a productivity premium of around 5 percent for firms that export. Kox and Rojas-Romagosa (2010) find similar results for the Netherlands: even after controlling for firm fixed effects, significant productivity premia are found for exporters compared to non-exporters (around 6 %).

South African research has been limited by the availability of good, firm-level data over time (Edwards, Rankin & Schoer, 2008). However, a handful of firm-level studies on Africa in general and South Africa in particular, have produced some similarities to the international literature. For example, in a study of nine sub-Saharan African countries, van Biesebroeck (2005) finds significant export premia. Relative to non-exporters, exporters are over 200 percent larger, 50 percent more capital-intensive and 56 percent more labour productive in terms of value-added per worker. In addition, exporters pay 34 percent higher wages than non-exporters. The superior characteristics found among African exporters are confirmed by Bigsten *et al* (2004) and Rankin, Soderbom and Teal (2006). Both studies find that exporters are larger and more capital-intensive and produce more per worker than firms which sell domestically only.

Table 2.1. Micro-economic studies on export premia

Author (s)	Country	Period	Data	Export Premia (averages)			
				Size (No. employees)	Average Labour Cost	Capital per worker	Value-added per worker
Percentage difference							
Andersson, Johansson & Loof (2008)	Sweden	1997-2007	Firm-level census data	69-152%	3-4%	23-33%	10-14%
Bernard & Jensen (1995)	U.S.	1976-87	Plant-level census data	94%	9%	9%	15%
Bernard & Jensen (1999)	U.S.	1984-92	Plant-level census data	51-102%	9-18%	7-22%	12-24%
Bernard & Wagner (1997)	Germany	1978-92	Plant-level survey data	51-72%	1-2%	5-12%	14-22%
Castellani, Francesco & Tomasi (2010)	Italy	1989-97	Firm-level census data	2-14%	-	4-43%	4-16%
Isgut (2001)	Columbia	1981-91	Plant-level survey data	123%	17%	48%	43%
Muûls & Pisu (2009)	Belgium	1996-2004	Firm-level census data	55%	-	3%	17%
Rankin (2001)	South Africa	1997-98	Firm-level survey data	-	20%	6%	34%
Van Biesebroeck (2005)	Sub-Saharan Africa[a]	1992-96	Plant-level survey data	213%	34%	50%	56%
Edwards, Rankin & Schoer (2008)	South Africa	1999	Firm-level survey data	0.54	-	0.18	42%
Absolute difference							
Aw & Hwang (1995)[b]	Taiwan	1986	Firm-level census data	187	-	161	392.3
Blalock & Gertler (2004)[c]	Indonesia	1990-96	Plant-level survey data	186.9	-	1.57	1.62
Bigsten *et al* (2004)	Cameroon, Kenya, Ghana & Zimbabwe	1992-95	Firm-level survey data	290	-	1.49	1.12
Kox & Rojas-Romagosa (2010)[d]	Netherlands	1997-2005	Firm- and plant-level survey data	-	1.5-1.7	-	10.7-12.4
Rankin, Soderbom & Teal (2006)	Sub-Saharan Africa[e]	1992-2003	Firm-Level survey data	206	-	1.76	1.06

[a] Includes: Ethiopia, Tanzania, Burundi, Zambia, Kenya, Ghana, Cote d'Ivoire, Cameroon, Zimbabwe

[b] Value in '000 NT$, capital per work & value-added per worker are not in natural logarithm (unlike the other studies)

[c] Value in '000 Rp, average of capital and value-added not per worker

[d] Total value in '000 €

[e] Includes: Ghana, Kenya, Tanzania, Nigeria & South Africa.

For South African firms, Rankin (2001) uses cross-sectional survey data to examine exporter characteristics. In line with the stylised facts, he finds that South African exporters, on average, pay higher wages, employ more capital per worker and produce more output per worker. These results are confirmed by Naude (2000) and Edwards, Rankin and Schoer (2008). In further support, a more recent study by Matthee and Krugell (2012) found significant differences between South African exporters and non-exporters in terms of age, size, foreign ownership and productivity. In addition the authors show that, after controlling for unobserved firm heterogeneity, firm size, productivity and access to finance affect the ability of a firm to export.

Overall, South African exporters seem to conform to the stylised facts of exporting: they are larger, pay higher wages and produce more per worker. However, unlike the international literature, very little work has been done on examining the productivity-exporting relationship any further. In one study, Rankin (2001) uses the Cobb-Douglas estimation of productivity to examine whether exporting is associated with increased efficiency and finds that, in general, exporters are no more productive than non-exporters. This presents a puzzle: South African exporters exhibit the same superior characteristics relative to non-exporters that international exporters do, with the exception that they exhibit no significant productivity premium as predicted by the heterogeneous-firm models of trade (such as Melitz, 2003) until such time as they export outside of the region.

Previous South African research, however, has used limited, cross-sectional survey data to examine the export-productivity relationship. The missing productivity premium may simply be a characteristic of the small samples used and not representative of South African exporters. Another possible explanation for the previous South African findings relates to the degree of exporter heterogeneity in terms of destinations served. This heterogeneity is evident throughout the literature in terms of accessibility of the export market, as well as the type and number of export markets served (high versus low income, single versus multiple destinations).

The ease of accessing the foreign market is determined by the ability of the firm to reach the productivity threshold. The literature here is divided into two hypothesis: namely self-selection and learning-by-exporting. The majority of empirical studies have found a strong self-selection effect: firms are more productive prior to entry into the export market (see Wagner (2011) for a detailed discussion on the empirical evidence of the self-selection hypothesis).

According to Bernard and Jensen (1999), firms that are looking to export in the future already have most of the advantageous performance characteristics some years before they enter the export market. This is because there are costs involved in entering the export market. These costs include transport costs, costs of developing relationships with foreign customers as well as production costs involved in modifying the domestic product for the foreign market. These costs provide a barrier to entry that less successful firms cannot overcome.

In the developing economies literature, evidence of learning-by-exporting is becoming more apparent. Blalock and Gertler (2004), for example, found that Indonesian firms became more productive only once they started exporting, not before, and that these productivity gains remained even after the firms exited the export market. Van Biesebroeck (2005), in a study of a number of African countries, finds that African exporters improve their relative performance after they enter the export market by reaping benefits such as scale economies or adopting a technology with higher productivity growth.

Other studies which find evidence of learning-by-exporting include Eaton *et al*. (2007) and Fernandes and Isgut (2005) for Columbia, Granér and Isaksson (2009) for Kenya, and Boermans (2013) for African firms. Evidence from these studies indicate that exporters learn from exporting when exporting to regional or neighbouring markets. In such cases access to the export market is easy: the destination is close (both in terms of distance, and likely technology), and the productivity threshold for entry is lower. This gives relatively low-productivity firms the opportunity to learn from exporting, thereby increasing their productivity. At a certain stage after productivity growth the firms can self-select into more developed markets, essentially using the regional market experience as a stepping stone into a higher productivity space.

It is clear, therefore, that a certain level of exporter heterogeneity is evident in terms of where the firm decides to export to. Indeed, Boermans (2013) shows that African firms exporting within the region exhibit poorer performance characteristics relative to those exporting out of the region in terms of capital, skilled labour and productivity. He further finds evidence of decreasing productivity among firms exporting within Africa.

Fernandes and Isgut (2005) find that learning-by-exporting effects are higher for firms that export to high-income countries. This introduces the second form of destination-heterogeneity, which relates to the type of market served. Consumers in high-income countries are assumed to have high-income tastes, i.e. these consumers expect a certain level of quality from the goods and services they purchase. This encourages competition among firms looking to export to such markets, which in turn drives firms to improve on their product, ultimately increasing productivity. A number of studies have found evidence of this positive relationship between high-income markets and productivity such as De Loecker (2007) for Slovenian firms, Tromifenko (2008) for Colombia, Pisu (2008) for Belgium firms, Park *et al*. (2010) for Chinese firms, Vacek (2010) for Czech firms, Bastos and Silva (2010) for Portuguese firms and Cebeci (2014) for Turkish firms.

A third strand of the exporter heterogeneity literature relates to the number of export markets served. While exporting to multiple destinations is likely dependant on certain firm characteristics, like productivity, it can also be argued that exporting to multiple destinations can have a different impact on a firm’s productivity relative to exporting to a single market since there is a greater opportunity for knowledge transfer when dealing with multiple export partners (Masso & Vahter, 2011). Evidence in support of this positive multi-destination-productivity relationship can be found in the case of US manufacturing firms (Bernard *et al*. 2011), Estonia's manufacturing firms (Masso & Vahter, 2011) and German manufacturing firms (Wagner, 2012).

Limited research has been done on firm performance and export destination in the South African context due to a lack of detailed firm-level panel data. As previously mentioned, Rankin (2001) shows that destination is an important consideration in the South African context. He shows that it is only when firms export to destinations outside of Africa that they exhibit positive productivity premiums. Another study by Rankin and Schöer (2013) also find positive performance indicators for firms exporting outside Africa in terms of wages and further find that that firms exporting in the region pay lower wages than non-exporters.

It is clear from this discussion that destination matters. South Africa is located far from the developed markets which makes exporting to these markets difficult in terms of accessibility. This is an important aspect to acknowledge since it can ultimately influence productivity gains from exports to these markets. In addition, although increasing exports to the region is encouraged, little evidence exists to indicate any positive export premium when exporting within Africa. It is recognised that exporting within the region has the potential for learning from exports, however there is evidence to suggest that exporting to developed countries is associated with exporter premiums and this should not be ignored.

Another element that matters is that of quality, particularly quality upgrading. Baldwin and Harrigan's (2007) quality-heterogeneous firm trade model assumes that customers' choices are based not only on price, but also on product quality. Higher quality firms are more likely to be competitive which has implications for exporting: firms with higher quality are better able to overcome trade costs and therefore more likely to succeed in the export market. This is an important factor to take account of since improving the quality of goods produced helps improve the existing comparative advantage thereby leading to an increase in export revenues and productivity (Henn et al, 2015). Further, increasing product quality, particularly for exporting, has also been linked to growth (Hausmann, Hwang & Rodrik, 2007). This in turn has stimulated policy discussions in developing economies on improving quality, thereby growing the workforce of higher quality (or skilled labour) and overall growth (Iacovone & Javorcik, 2012).

However, as discussed in Verhoogen (2008, p.489), more productive firms will produce higher quality goods, particularly for the more developed and "richer" customers. These firms will therefore pay higher wages for higher quality workers, which ultimately results in within-industry wage inequalities (as well as a change in the skills composition of those workers). A number of recent studies have investigated quality upgrading and trade at the firm-level (with majority focussing on the change in product mix during a period of trade liberalisation). These studies include Amiti and Khandelwal (2013), Castellares (2015), Iacovone and Javorcik (2012), and Lileeva and Trefler (2010).

Notwithstanding the importance of the impact of quality upgrading and exports, this chapter focuses on South Africa's missing productivity premium. This chapter contributes to the South African trade literature by addressing the issue of the missing exporter productivity premium and providing a detailed destination-oriented, firm-performance investigation. It will do so in two ways: first, by making use of two rich data sets: a large sample survey dataset collected by Statistics South Africa (Stats SA) and a substantial population dataset provided by the South African Revenue Service (SARS). This chapter attempts to replicate the previous work on South African exporters at the firm-level to determine whether the previous finding of the missing productivity premium still holds when more substantial datasets are used and what the relationship is between firm performance and export destination.

2.4. Data

The first dataset used for the analysis of South African exporters' productivity premium is that collected by Statistics South Africa (Stats SA) in its Large Sample Surveys of manufacturing firms (LSS). These surveys are used for the purposes of calculating the national accounts and although designed to be cross-sectional in nature many firms can be linked between the years to create a panel dataset. There are approximately 10 000 manufacturing firms in each of the rounds (2005 and 2008).

These surveys collect data on industrial classification, employment, imports and exports, income and expenditure, profit or loss, inventories, carrying value of assets as well as details of products manufactured. To compare 2005 data to 2008, the 2008 data was deflated using industry level deflators, except for wages which are deflated by the CPI.

While useful for estimating exporter characteristics and productivity premiums, no information on destinations is available in this dataset. In order to examine destinations as potential explanation for South Africa's missing productivity premium, this chapter turns to an additional set of data.

The second dataset used is generated from population tax return data provided by the South African Revenue Service (SARS). SARS provided a number of separate and substantial datasets. A more detailed discussion of the various datasets and how they were manipulated for the purpose of this, and the following, chapter can be found in Appendix A. A brief discussion is presented here.

Information from three separate datasets was combined into one dataset that was used to compare results from the LSS. These datasets are the Company Income Tax (CIT) return data, the Pay as You Earn, or Personal Income Tax employee data (PIT/PAYE) and customs transactions data.

The CIT data is available for the tax years 2009 to 2013, however the form used to generate the CIT return data was significantly changed in May of 2013 to accommodate the fact that larger firms require more detailed questioning relative to smaller firms. As such the CIT data is divided into two groups:

those firms which completed the old CIT form (known as the IT14), and those which completed the newer, more detailed form (known as the ITR14).

This dataset contains information on company balance sheet and income statement as well as extensive information on tax allowances and deductions. It does not include any information on total employment or exporting. This information is obtained from the PIT and customs transactions datasets respectively.

The IRP5 employee data contains information on unemployment insurance fund contributions, total employee tax amount, provident fund contributions, taxable income etc. It also allows for the calculation of number employees per firm. This information was merged to the CIT data for analysis.

The customs transactions data includes information on products, destinations, quantities, values and tariff rates applied for exports and imports. For this chapter, only exports are examined. From this data, the total customs value of exports per firm is calculated, along with the total number of destinations. This data is then merged to the CIT data in order to calculate the number of exporters and non-exporters in the dataset, as well as other firm-level descriptive statistics.

The next section provides a descriptive analysis on the final dataset, which comprised of around 700 000 unique firms per tax year. The focus will be on manufacturing firms, of which there were on average 29 000 per tax year.

2.5. Descriptive Statistics

This section presents a brief picture of exporting and non-exporting manufacturing firms in South Africa. Using data from both the LSS and SARS datasets it is found that, in terms of the propensity to export, exporting within the manufacturing sector is quite rare (figure 2.1). Among all manufacturing firms, less than a third export (26% in 2005, 30% in 2008 and 19% in 2013). This finding is not dissimilar to that of U.S. manufacturing firms, 27 percent of which export (Bernard *et al*, 2007). When the sample is restricted to only medium and large firms the proportion of exporting firms increases (33% in 2005, 42% in 2008 and 45% in 2013). It is anticipated that samples consisting of larger firms will have a higher share of exporters since, given the sunk costs associated with exporting, smaller firms are less likely to trade.

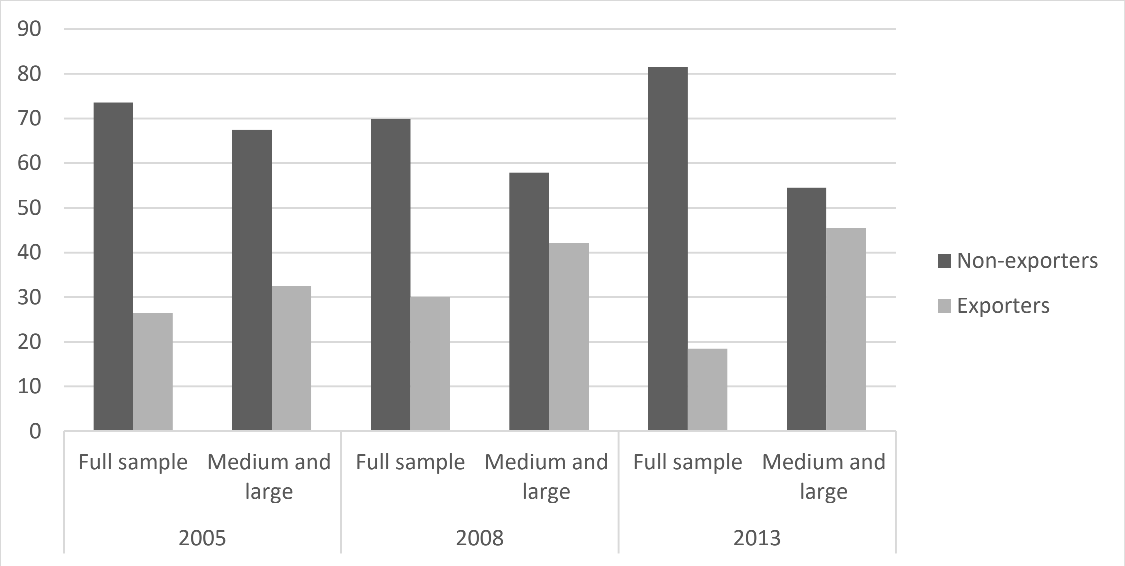


Figure 2.1. Proportion of Exporters
Source: Authors own calculations using LSS data for 2005 and 2008 and SARS data for 2013.

Not only is exporting rare among South African manufacturers, but as figure 2.2 indicates, very few South African manufacturing firms specialize in exporting. Given that a firm exports, few exporters export more than 50 percent of their total output (36% in 2005 and 27% in 2008). Further, the average (median) exporter exported around 18 (7.5) percent of total output in 2005 and 12 (5) percent in 2008. The SARS data shows similar findings (Figure 2.3): fewer than 10 percent of exporters export more than 50 percent of total sales with the average (median) exporter exported around only 21 (4) percent.

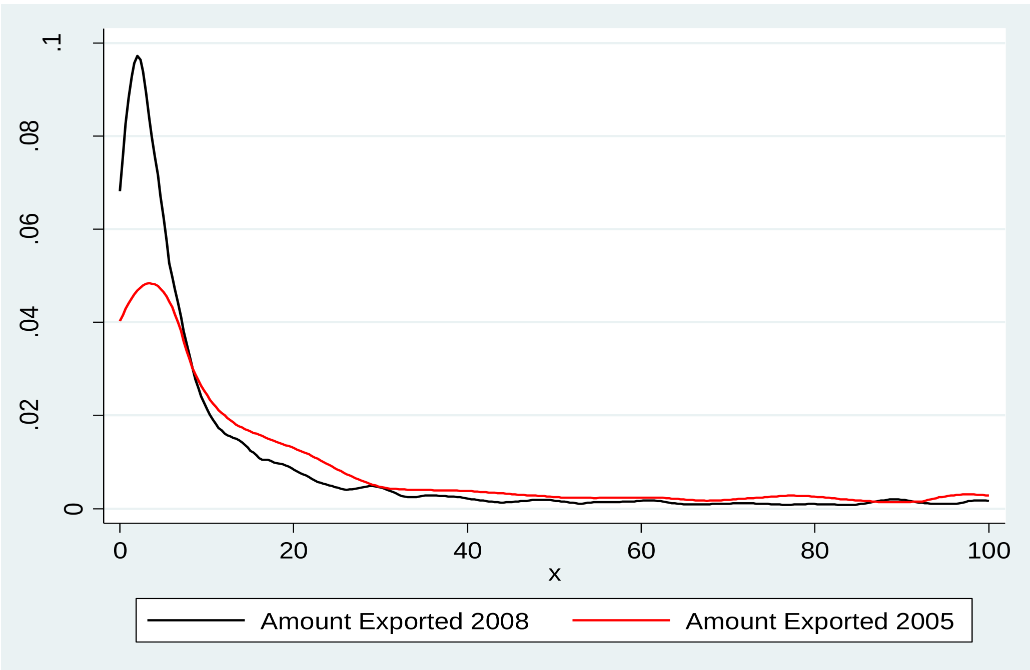


Figure 2.2. Proportion of Output Exported if a Firm Exports –LSS data

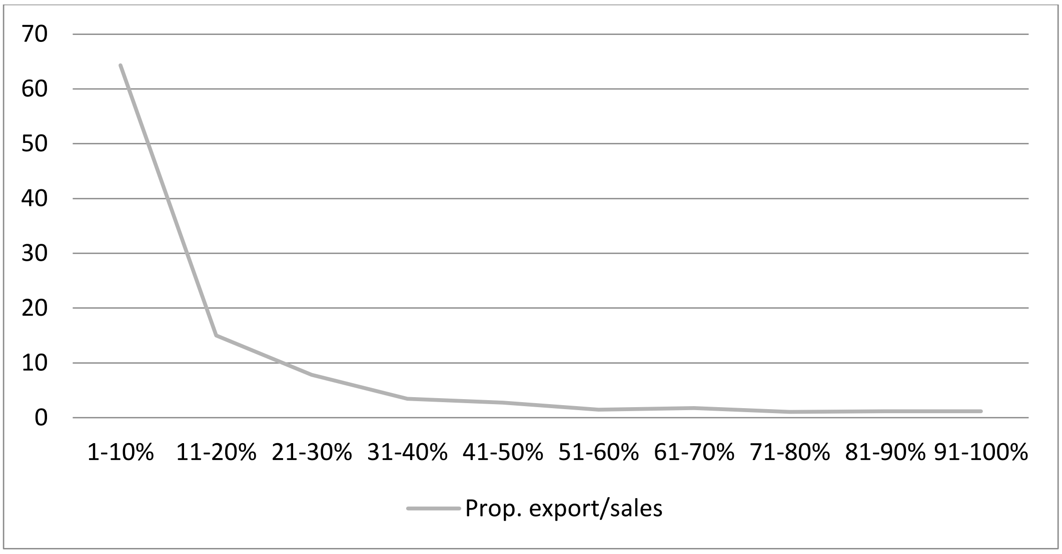


Figure 2.3. Proportion of Output Exported if a Firm Exports – SARS data
Source: Authors own calculations using SARS data.

Looking at destinations served by manufacturing exporters in South Africa, it is found that more than half of manufacturing exporters (57%) export only to destinations within Africa, yet contribute merely 7 percent to total export value (figure 2.4). A large proportion (73%) of South African exporters export to multiple destinations and these multiple-destination exporters contribute 98 percent to total export value. The proportion of multi-destination exporters within Africa relative to the proportion of multi-destination exporters outside of (and also potentially in) Africa is 34 and 39 percent respectively. However, multi-destination exporters within Africa account for only 6 percent of total export value compared to the 92 percent contribution of multi-destination outside (and within) Africa exporters, while single destination exporters make up the rest. These figures suggest that, as noted in previous studies, destination is a notable form of heterogeneity among exporters.

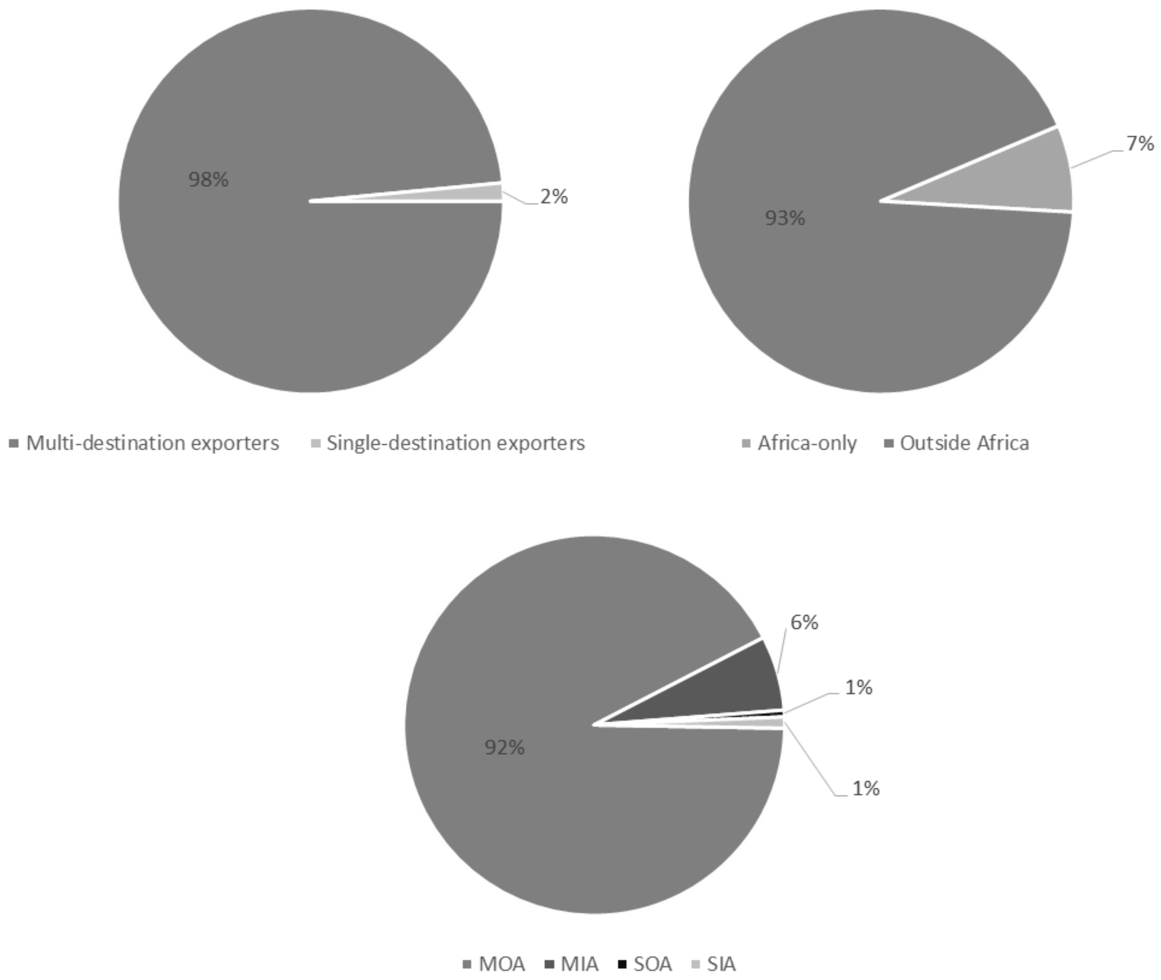


Figure 2.4. Proportion Contributed to Total Exports – Exporter Type
Source: Authors own calculations using SARS data.
Notes: MIA – Multiple destination exports within Africa only; MOA – Multiple destination exports outside Africa; SIA – Single destination exports within Africa only; SOA – Single destination exports outside Africa.

According to the international trade literature, exporters exhibit superior characteristics relative to their domestic counterparts. It is now a stylised fact that exporters are larger, have higher levels of labour productivity and pay higher wages than non-exporters. Following the methodology used by Bernard and Jensen (1995) export premia for a number of firm characteristics are estimated using regressions of the general form

$$\ln(X)_i = \alpha + \beta Exporter_i + \delta Industry_i + \mu_i \tag{2.1}$$

where $(X)_i$ is a vector of firm characteristics; $Exporter_i$ is a dummy variable that takes the value one if a firm exports and zero otherwise; $Industry_i$ is a control dummy for six digit SIC industry; and β represents the export premia which indicates the average percentage difference between exporters and non-exporters. For the LSS dataset the regression also includes a year dummy to control for year fixed effects between 2005 and 2008. A second set of regressions are run for the SARS dataset which control for firm size as measured by the (log) total number of employees. The analysis further restricts the regressions above to medium and large firms. The results for the medium and large firms are presented in Appendix B.

Table 2.2 reports the export premia for the general case of exporters relative to their domestic counterparts for both the LSS sample data and the SARS population data. The table indicates that for exporters relative to non-exporters the export premia are positive and significant for all characteristics. Exporters are notably larger, both in terms of output and number of employees, more labour productive, pay higher wages and are more capital and intermediate input intensive than non-exporters. Even after controlling for firm size, relative to non-exporters, exporters produce 46 percent more output per worker, pay 27 percent higher wages and are 12 and 57 percent more capital and intermediate input intensive respectively. Similarly for medium and large exporters (see table B1).

Table 2.2. Manufacturing Exporter Premia – Exporters versus Non-exporters

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
LSS data						
Exporter	1.161***	0.840***	0.417***	0.435***	0.292***	0.489***
	(0.0385)	(0.0294)	(0.0231)	(0.0176)	(0.0345)	(0.0245)
Exporter*2008	-0.00308	-0.0560	-0.0205	-0.132***	0.408***	0.0428
	(0.0500)	(0.0381)	(0.0300)	(0.0228)	(0.0448)	(0.0317)
2008	0.446***	0.386***	0.138***	0.356***	-0.00327	0.00224
	(0.0256)	(0.0187)	(0.0154)	(0.0112)	(0.0222)	(0.0156)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*21,237*	*23,758*	*21,152*	*23,747*	*22,995*	*23,751*
SARS data						
Exporter	2.118***	0.708***	0.343***	1.690***	0.151***	0.466***
	(0.0293)	(0.0283)	(0.0218)	(0.0281)	(0.0500)	(0.0263)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
SARS - Controlling for firm size						
Exporter			0.460***	0.273***	0.124**	0.569***
			(0.0220)	(0.0210)	(0.0514)	(0.0267)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes

Observations	*25,881*	*10,065*	*9,955*	*9,681*	*9,427*	*9,710*

Source: Authors own calculation using the LSS data and SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

As previous African studies have shown, firms that export to destinations within Africa only exhibit different characteristics to those that export outside of Africa. Table 2.3 contains the results of export premia regressions for firms which export to Africa only. These regressions were run for the SARS data only, since no information on destinations is available in the LSS data.

The results indicate that firms exporting to destinations within Africa only, while still significantly superior to non-exporters, exhibit poorer performance premia relative to firms which export to destinations outside of Africa across all characteristics. These findings hold after controlling for firm size as well as for medium and large exporters (see table B2).

Table 2.3. Manufacturing Exporter Premia – Africa-only exporters

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
Exporter	2.506***	1.028***	0.408***	2.093***	0.304***	0.515***
	(0.0417)	(0.0362)	(0.0282)	(0.0397)	(0.0644)	(0.0339)
Africa Only Dummy	-0.675***	-0.556***	-0.112***	-0.702***	-0.266***	-0.0862**
	(0.0518)	(0.0398)	(0.0310)	(0.0490)	(0.0706)	(0.0372)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Controlling for firm size						
Exporter			0.585***	0.431***	0.274***	0.672***
			(0.0286)	(0.0272)	(0.0668)	(0.0348)
Africa Only Dummy			-0.208***	-0.263***	-0.250***	-0.171***
			(0.0304)	(0.0289)	(0.0713)	(0.0370)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations	*25,881*	*10,065*	*9,955*	*9,681*	*9,427*	*9,710*

Source: Authors own calculations using SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Not only does the destination matter, but so too does the number of destinations served. As table 2.4 reports, multiple-destination exporters are significantly larger, more labour productive, pay higher wages and are significantly more capital and intermediate input intensive than single destination exporters, who, in turn, exhibit superior performance characteristics relative to non-exporters, even after controlling for firm size. Similar results are found after restricting the estimation to medium and large firms (table B3).

Table 2.4. Manufacturing Exporter Premia – Multiple Destinations

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
Exporter	1.310***	0.124***	0.179***	0.939***	-0.00528	0.277***
	(0.0508)	(0.0426)	(0.0333)	(0.0487)	(0.0765)	(0.0400)
Multiple Dest. Dummy	1.112***	0.798***	0.224***	1.028***	0.213***	0.258***
	(0.0574)	(0.0440)	(0.0344)	(0.0546)	(0.0790)	(0.0413)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Controlling for firm size						
Exporter			0.202***	0.0892***	-0.00826	0.296***
			(0.0324)	(0.0311)	(0.0766)	(0.0394)
Multiple Dest. Dummy			0.366***	0.261***	0.188**	0.389***
			(0.0340)	(0.0326)	(0.0803)	(0.0413)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations	*25,881*	*10,065*	*9,955*	*9,681*	*9,427*	*9,710*

Source: Authors own calculations using SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Given that exports outside of Africa as well as exports to multiple-destinations result in superior performance premia, it serves to examine the characteristics of firms who export to multiple destinations within Africa relative to those that export outside of Africa. These premia are reported in table 2.5, and for ease of comparison are presented in figure 2.5. Multiple-destination exporters are still shown to be the superior performers and, as expected, exporters to multiple destinations outside of Africa display superior characteristics relative to all other exporters, while firms which export to multiple destinations within Africa are larger, more labour-productive, pay higher wages and employ more intermediate inputs per worker than firms which export to single destinations (both inside and outside of Africa). Restricting the estimation to only medium and large firms yields similar results (table B4).

Interestingly, these results further indicate that firms which export to a single destination within Africa seem to perform better than firms which export to a single destination outside of Africa in terms of output, labour productivity, wages and intermediate-input intensity (and after controlling for firm size, labour productivity and intermediate-input intensity). However, once the estimation is restricted to medium and large firms, these differences are no longer significant.

Table 2.5. Manufacturing Exporter Premia – Multiple Destinations Within and Outside Africa

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
Exporter	1.043***	0.140	-0.0271	0.754***	-0.140	0.0166
	(0.122)	(0.0966)	(0.0758)	(0.116)	(0.176)	(0.0909)
Africa Only Dummy	0.319**	-0.0223	0.244***	0.222*	0.158	0.309***
	(0.132)	(0.103)	(0.0809)	(0.126)	(0.187)	(0.0969)
Multiple Dest. Dummy	1.634***	0.987***	0.482***	1.493***	0.491***	0.554***
	(0.127)	(0.0994)	(0.0780)	(0.121)	(0.180)	(0.0936)
Interaction:	-0.855***	-0.406***	-0.353***	-0.810***	-0.458**	-0.383***
Afri*Multi-dest	(0.145)	(0.112)	(0.0879)	(0.138)	(0.203)	(0.105)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Controlling for firm size						
Exporter			0.00304	0.116	-0.142	0.0404
			(0.0737)	(0.0707)	(0.176)	(0.0895)
Africa Only Dummy			0.236***	-0.0319	0.157	0.303***
			(0.0786)	(0.0753)	(0.187)	(0.0953)
Multiple Dest. Dummy			0.659***	0.357***	0.465**	0.716***
			(0.0762)	(0.0730)	(0.181)	(0.0925)
Interaction:			-0.423***	-0.214***	-0.446**	-0.447***
Afri*Multi-dest			(0.0854)	(0.0818)	(0.203)	(0.104)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations	*25,881*	*10,065*	*9,955*	*9,681*	*9,427*	*9,710*

Source: Authors own calculation using SARS data

Notes: ***p<0.01**p<0.05*p<0.1

(Is significant at the 1% level, 5% level and 10% level respectively)

Values are given in natural logarithms.

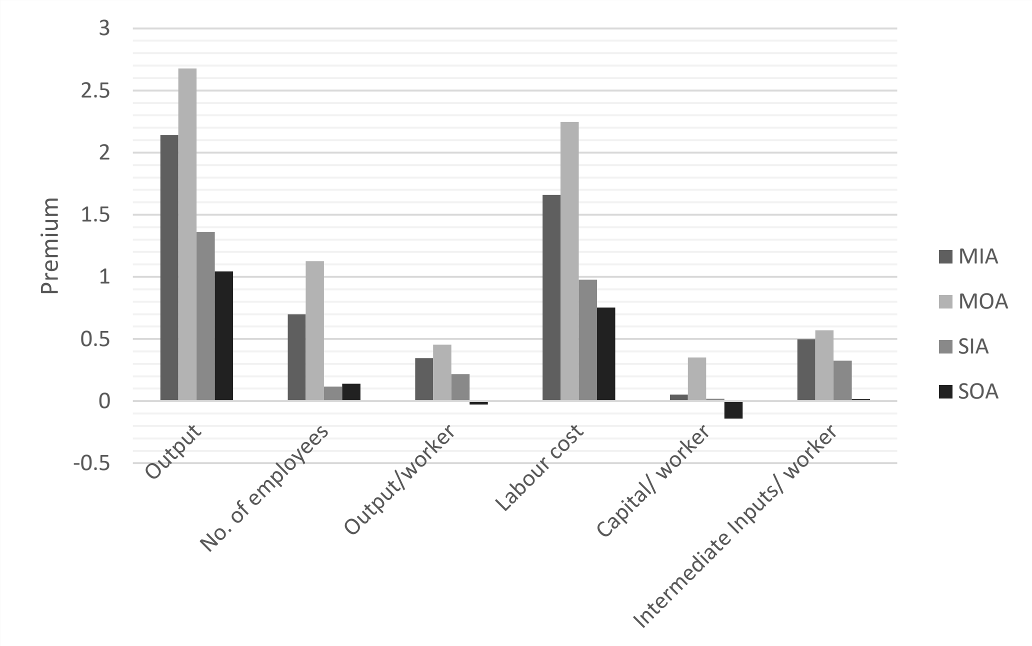


Figure 2.5. Manufacturing Exporter Premia - Multiple Destinations Within and Outside Africa
Source: Authors own calculation using SARS data
Notes: MIA – Multiple destination exports within Africa only; MOA - Multiple destination exports outside Africa; SIA – Single destination exports within Africa only; SOA – Single destination exports outside Africa; and non-exporters represented by the zero-line.

Overall, the results indicate that the stylised facts of exporting hold true for South African manufacturing exporters. In addition, destination matters both in terms of within and outside of Africa exports as well as the number of markets served. Africa-only exporters are inferior to international (outside Africa) exporters and multiple destination exporters are superior to single destination exporters. Further, exporting to multiple destinations outside of Africa results in the largest export premia. There is some evidence which indicates that when exporting to only one destination firms which export within Africa perform better than those which export to one destination internationally, however this does not hold true once the analysis is restricted to medium and large firms only. The next section goes a step further to examine how exporting and destinations relate to firm performance in terms of total factor productivity.

2.6. Econometric Methodology – The Productivity Premium

Production function analysis enables inference about the productivity difference between traders and non-traders. One can deduce these productivity differences from the estimated production functions because the coefficient of the expected trade status dummy variable gives the percentage difference between the productivity of traders and non-traders (Yasar, Nelson, & Rejesus, 2006).

South African manufacturing firms' total factor productivity (TFP) will be approximated by a Cobb-Douglas specification:

$$ln({}^{Y_{it}}/_{L_{it}}) = \alpha_{it} + \beta_1(EX_{it}) + \beta_2(ln\,{}^{K_{it}}/_{L_{it}}) + \beta_3\left(ln\,{}^{I_{it}}/_{L_{it}}\right) + \beta_4(\ln L_{it}) + \beta_5(Ind_{it}) + \mu_{it} \tag{2.2}$$

where i is the firm subscript, t is the time subscript, L is labour, ${}^{Y}/_{L}$ is real output per worker, ${}^{K}/_{L}$ represents real capital per worker, and ${}^{I}/_{L}$ represents real intermediate inputs per worker, Ind is a vector of industry characteristics (measured at the 5 digit SIC), and μ_i is the residual.

The variable EX_{it} is a dummy variable that will represent export status. It takes the form $EX_{it} = 1$ if the firm exports; $EX_{it} = 0$ otherwise.

For both the LSS sample data and the SARS population data, the coefficients will initially be estimated by Ordinary Least Squares (OLS) regressions. Specifically β_2 , β_3 and β_4 are the elasticity of output productivity with respect to capital per worker, intermediate inputs per worker and labour respectively. The coefficient of interest, β_1, signifies productivity differences between exporting firms and domestic traders. Industries are explicitly included in the regression equation to account for some of the heterogeneity in this sample.

An issue with OLS estimation lies with its key assumption that inputs of production are uncorrelated with omitted unobservables, i.e.: the error term in equation (2.2) is assumed to be uncorrelated with the factor inputs. In reality, however, the level of factor inputs is likely to be influenced by unobservables; for example, a positive productivity shock, which is observed by the firm but not by the analyst, will lead the firm to increase output levels and, consequently, input levels. By ignoring this relationship, OLS estimation is likely to give biased and inconsistent estimates.

A fixed effects regression is therefore run on the LSS and SARS panel datasets in order to obtain more consistent estimates. This methodology, however, is also problematic in that it assumes that the relationship between firm behaviour and productivity is time invariant. Since this may not necessarily be the case, another more robust methodology should be used to solve these issues such as the semi-parametric methodologies of Olley and Pakes (1996) and Levinsohn and Petrin (2003).

The Olley-Pakes (OP) estimation addresses the endogeneity problem by making use of investment as a proxy for shocks in the unobserved productivity levels, where the Levinsohn-Petrin (LP) methodology makes use of intermediate inputs. The success of the OP investment method in controlling for bias can be hampered if the data contains a large number of zero-investment observations. Since the monotonicity condition will not hold for these observations, they will be truncated from the dataset. The dataset used in this study contains a large number of observations where investment is zero. As such, the LP methodology is used to estimate productivity, making use of intermediate inputs as a proxy.

The first stage in estimating TFP is to estimate the log-linearised form of the Cobb-Douglas production function represented by equation (2.2) above. Unlike OLS, LP recognises that the error term, μ_{it}, is comprised of two components: transmitted productivity (observed by the firm, but not the analyst) and an error term that is uncorrelated with the choice of inputs, given by ω_{it} and ε_{it} respectively. The transmitted productivity component is the source of the simultaneity problem and the root of inconsistent estimates when using OLS.

Levinsohn and Petrin assume that intermediate inputs, I_{it}, can be written as a function of the two state variables K_{it} and ω_{it}. Under the condition of monotonicity, Levinsohn and Petrin show that the unobservable ω_{it} can be written as a function of intermediate inputs and capital:

$$\omega_{it} = \omega_{it}(K_{it}, I_{it}) \tag{2.3}$$

Using this expression, the production function in (2.2) can be expressed as:

$$ln(^{Y_{it}}/_{L_{it}}) = \beta_1(EX_{it}) + \vartheta_{it}(ln\,^{K_{it}}/_{L_{it}},\ ln\,^{I_{it}}/_{L_{it}}) + \beta_4(\ln L_{it}) + \beta_5(Ind_{it}) + \varepsilon_{it} \tag{2.4}$$

where $\vartheta_{it}\left(ln\,^{K_{it}}/_{L_{it}},\ ln\,^{I_{it}}/_{L_{it}}\right) = \alpha_{it} + \beta_2\left(ln\,^{K_{it}}/_{L_{it}}\right) + \beta_3\left(ln\,^{I_{it}}/_{L_{it}}\right) + \omega_{it}(K_{it}, I_{it})$ (2.5)

The estimation procedure involves two steps. First, a third-order polynomial approximation is used to estimate $\vartheta_{it}\left(ln\,^{K_{it}}/_{L_{it}},\ ln\,^{I_{it}}/_{L_{it}}\right)$. The second step involves solving the generalised methods of moments (GMM) minimisation problem in order to identify β_2 and β_3, the coefficients on capital and intermediate inputs. The LP technique will thereby give rise to consistent parameter estimates of the productivity equation (2.2).

Following the process in section 2.3, the productivity premia are estimated for exporters in general, Africa-only exporters, multiple destination exporters as well as the interaction between multiple destinations and Africa. The results are presented and discussed in the following section.

2.7. Estimation Results

In contrast to the international literature, previous South African studies have found that despite exhibiting superior export premia across a number of characteristics, exporters, in general, are no more productive than non-exporters and it is only when exporting outside of Africa that the productivity premium becomes significant (see Rankin (2001)).

As previously mentioned, South African studies have been limited by inaccessibility to good firm-level data. For the first time in South African literature, official population data will be used to reproduce the findings of the South African studies to determine whether or not South African exporters exhibit productivity premiums relative to non-exporters. The OLS results of equation (2.2) are presented in table 2.6 for the LSS sample data and in table 2.7 for the SARS population data. The more robust Levinsohn-Petrin (LP) estimations are contained in table 2.8.

According to the data contained in the LSS sample, no productivity premium exists for exporters in 2005. Indeed among medium to large firms, exporters are less productive than non-exporters, although this effect disappears in the fixed effects estimations. The OLS estimates indicate that exporters in 2008, are no different to non-exporters. However the fixed effects estimates premium of between 15 and 7 percent.

Table 2.6. OLS Estimation of Exporter Productivity Premium – LSS data

	Ordinary Least Squares		**Fixed Effects**	
VARIABLES	(1) All firms	(2) Medium-Large	(3) All firms	(4) Medium-Large
Exporter	0.00099	-0.0249*	-0.0247	-0.0302
	(0.0133)	(0.0139)	(0.0189)	(0.0196)
Exporter*2008	-0.0148	0.0105	0.0549**	0.0673***
	(0.0158)	(0.0165)	(0.0217)	(0.0232)
2008	0.0798***	0.0547***	0.0796***	0.0722***
	(0.00946)	(0.0107)	(0.0129)	(0.0145)
ll	-0.0131***	-0.0186***	-0.114***	-0.172***
	(0.00291)	(0.00351)	(0.0152)	(0.0202)
lkl	0.0378***	0.0420***	0.0164*	0.0170*
	(0.00316)	(0.00353)	(0.00847)	(0.00943)
lil	0.858***	0.865***	0.770***	0.745***
	(0.00421)	(0.00494)	(0.0155)	(0.0184)
Industry controls	Yes	Yes	Yes	Yes
Observations	*9,358*	*7,330*	*9,358*	*7,330*
R-squared	*0.880*	*0.886*	*0.751*	*0.777*
Number of nent_num			*7,246*	*5,590*

Source: Authors own calculation using LSS data
Notes: ***p<0.01**p<0.05 *p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)

The SARS data shows a more detailed story. The results of the OLS, Fixed effects and Levinsohn-Petrin estimations are presented in table 2.7[3]. These results are for all manufacturing firms: the micro-small and medium-large firm estimations are discussed here, but presented in tables B5 and B6 in Appendix B.

The Levinsohn-Petrin estimators are mostly similar to those of the OLS regressions. The Fixed effects estimates are mostly insignificant. In general the estimates indicate that manufacturing exporters are indeed more productive than non-exporting firms, around 4-7 percent, this holds regardless of firm size. The South African export premium of between 4 and 7 percent is in line with (albeit slightly lower than)

[3] It is noted that the input variable estimates (labour, capital and raw materials/intermediate inputs) differ from those found in other studies, such as Thomas & Narayanan (2012) and Mukherjee (2014), in two ways. Firstly in this chapter capital and intermediate inputs are divided by labour for a per worker contribution. Secondly, in previous studies labour is calculated/defined in terms of compensation, whereas in this study labour is defined as the actual number of workers employed.

international studies which find premia of around 6 to 9 percent (see Muûls and Pisu (2009), and Kox and Rojas-Romagosa (2010)).

Columns (4) to (6) show the productivity premium for firms exporting within Africa only relative to non-exporters and firms exporting outside of Africa (but might export to Africa too). Firms that export to destinations only within Africa are around 6 percent less productive than firms which export outside of Africa, who in turn are around 8 percent more productive than non-exporters. Interestingly, Africa-only exporters exhibit little, if any, productivity premium over non-exporting firms, particularly among medium to large firms. This is perhaps not surprising given the findings of previous African studies which show that the productivity threshold is low for firms exporting to the region.

Studies have also shown that the number of markets served impacts on productivity premiums. This relationship is represented in columns (7) to (9). The estimates indicate that productivity increases with the number of export destinations, but at a decreasing rate.

The LP estimation suggests that for the population of manufacturing firms, exporters become more productive than non-exporters when exporting to at least one destination and reach peak productivity at 56 destinations. Medium and large exporters also exhibit positive productivity premiums over non-exporters when exporting to at least one destination, but reach their maximum productivity at around 66 destinations. Smaller exporters' productivity starts to decrease after fewer destinations, around 40.

Columns (10) to (12) show that firms exporting to multiple destinations are significantly more productive than single-destination exporters (5-6%) as well as firms selling to the domestic market only (6-7%). Further, there is no evidence to suggest that single destination exporters are any more productive than non-exporters: indeed among all firms single destination exporters seem to be less productive than non-exporters. This holds true even among small exporters.

The combined Africa-only, multiple destination estimations are interesting (table 2.8) particularly when distinguishing between medium to large firms and smaller firms tables (see Appendix B, tables B7 and B8 respectively). As previously seen, exporting to multiple destinations is associated with a higher productivity premium relative to exporting to single destinations. The OLS estimates for medium and large firms indicate that exporting to a single destination outside of Africa results in a productivity premium of around 8 percent relative to non-exporters, 3 percent more than if the firm exported to multiple destinations including those outside of Africa. However, the less biased LP estimates indicate otherwise: medium to large exporters exporting to single destinations both within and outside of Africa are less productive than multiple destination exporters. Indeed, the LP estimates suggest than exporting to a single destination outside of Africa is not associated with a greater productivity premium relative to non-exporters.

Table 2.7. Exporter Productivity Premium by OLS, Fixed Effects (F.E.) and Levinsohn-Petrin (L.P.) estimation– SARS data

	Standard			Africa			No. of Destinations			Multi-destinations		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**
2013	0.0246**	0.0237**	0.0326***	0.0242**	0.0238**	0.0324***	0.0235**	0.0237**	0.0316***	0.0244**	0.0238**	0.0325**
	(0.0117)	(0.00992)	(0.0117)	(0.0117)	(0.00993)	(0.0105)	(0.0116)	(0.00992)	(0.0104)	(0.0117)	(0.00997)	(0.0130)
Exporter	0.0523***	0.00506	0.0448***	0.100***	0.00687	0.0814***	-0.0138	-0.00220	-0.00393	-0.0119	0.00266	-0.0157
	(0.0141)	(0.0192)	(0.0144)	(0.0172)	(0.0242)	(0.0158)	(0.0148)	(0.0205)	(0.0125)	(0.0217)	(0.0226)	(0.0184)
Exporter*2013	-0.0373**	-0.00446	-0.0227	-0.0424**	-0.00303	-0.0362**	-0.0258	-0.00507	-0.0148	-0.0249	-0.0119	-0.00125
	(0.0159)	(0.0136)	(0.0149)	(0.0195)	(0.0160)	(0.0160)	(0.0158)	(0.0136)	(0.0127)	(0.0248)	(0.0251)	(0.0229)
Africa_only				-0.0875***	-0.00182	-0.0677***						
				(0.0189)	(0.0215)	(0.0127)						
Africa_only*2013				0.0175	-0.00315	0.0305*						
				(0.0217)	(0.0188)	(0.0156)						
No. dest							0.0153***	0.00483	0.0114***			
							(0.00130)	(0.00484)	(0.00131)			
(No. dest)^2							-0.00013***	-2.69e-05	-0.0001***			
							(2.53e-05)	(9.40e-05)	(3.02e-05)			
Multi-dest.										0.0905***	0.0142	0.0847***
										(0.0221)	(0.0242)	(0.0195)
Multi-dest. *2013										-0.0153	0.00826	-0.0275
										(0.0252)	(0.0247)	(0.0221)
ll[4]	-0.0579***	-0.303***	-0.041***	-0.0611***	-0.303***	-0.0432***	-0.0707***	-0.304***	-0.0507***	-0.0614***	-0.303***	-0.0440***
	(0.00253)	(0.0158)	(0.00363)	(0.00256)	(0.0158)	(0.00384)	(0.00265)	(0.0158)	(0.00348)	(0.00257)	(0.0158)	(0.00331)
lkl	0.0237***	0.00412	0.00615	0.0234***	0.00417	0.00622	0.0231***	0.00413	0.00924	0.0238***	0.00412	0.00662
	(0.00151)	(0.00447)	(0.0153)	(0.00151)	(0.00448)	(0.0168)	(0.00150)	(0.00447)	(0.00753)	(0.00151)	(0.00447)	(0.0147)
lil	0.754***	0.578***	0.875***	0.753***	0.578***	0.872***	0.747***	0.577***	0.840***	0.752***	0.578***	0.871***
	(0.00284)	(0.0107)	(0.171)	(0.00284)	(0.0108)	(0.152)	(0.00286)	(0.0108)	(0.0283)	(0.00285)	(0.0108)	(0.161)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,609	12,609	12,609	12,609	12,609	12,609	12,609	12,609	12,609	12,609	12,609	12,609
R-squared	0.888	0.737		0.888	0.737		0.890	0.738		0.888	0.737	

[4] It is noted that the input variable estimates (labour, capital and raw materials/intermediate inputs) differ from those found in other studies, such as Thomas & Narayanan (2012) and Mukherjee (2014), in two ways. Firstly in this chapter capital and intermediate inputs are divided by labour for a per worker contribution. Secondly, in previous studies labour is calculated/defined in terms of compensation, whereas in this study labour is defined as the actual number of workers employed.

Number of id	9,940	9,940	9,940	9,940

Source: Authors own calculation using SARS data

Notes: ***p<0.01, **p<0.05, *p<0.1 is significant at the 1% level, 5% level and 10% level respectively.

Table 2.8. Multiple-destination/Africa Interaction Productivity Premium+

VARIABLES	(1) **OLS**	(2) **F.E.**	(3) **L.P.**
Non-exporter 2013	0.0242**	0.0237**	0.0323***
	(0.0117)	(0.00997)	(0.0110)
Single destination out of Africa 2013	-0.0324	0.0770	-0.0333
	(0.0292)	(0.0512)	(0.0514)
Single destination out of Africa 2012	0.0342	-0.0129	0.00602
	(0.0452)	(0.0448)	(0.0512)
Single destination within Africa 2013	-0.00907	0.00263	0.0241*
	(0.0155)	(0.0248)	(0.0143)
Single destination within Africa 2012	-0.0223	0.00650	-0.0205
	(0.0235)	(0.0246)	(0.0189)
Multiple destinations out of Africa 2013	0.0689***	-0.00936	0.0924***
	(0.0129)	(0.0236)	(0.0151)
Multiple destinations out of Africa 2012	0.0739***	0.00140	0.0936***
	(0.0220)	(0.0241)	(0.0154)
Multiple destinations within Africa 2013	0.0287**	0.0408	0.0537***
	(0.0140)	(0.0259)	(0.0145)
Multiple destinations within Africa 2012	0.0379*	0.0139	0.0382*
	(0.0198)	(0.0267)	(0.0199)
ll	-0.0633***	-0.302***	-0.0453***
	(0.00258)	(0.0158)	(0.00330)
lkl	0.0234***	0.00423	0.00594
	(0.00151)	(0.00448)	(0.0144)
lil	0.751***	0.578***	0.871***
	(0.00285)	(0.0108)	(0.154)
Industry controls	Yes	Yes	Yes
Observations	12,609	12,609	12,609
R-squared	0.889	0.738	
Number of id		9,940	

Source: Authors own calculation using SARS data
Notes: + Base category is non-exporters in 2012
***p<0.01, **p<0.05, *p<0.1 significance at the 1%, 5% and 10% level respectively.

For smaller firms, the productivity premium is greater for exporting to multiple destinations, and greatest for multiple destinations outside of Africa. Interestingly, small firms who export to a single destination outside of Africa are significantly less productive than domestic producers: indeed, if a small firm decides to export to one destination, the results suggest that exporting to an African country will result in a positive productivity premium relative to domestic sales and sales to outside the region.

Overall multiple destination exporters exhibit higher productivity premiums, with exports to multiple destinations outside of Africa displaying the greatest productivity premium for all firm sizes. Exporting to a single destination outside of Africa results in the lowest productivity for firms, and even negative productivity for micro to small firms.

2.8. Concluding Remarks

Current South African policy recognises the importance of export growth in contributing to overall economic growth and not only aims to encourage increased exports among and within firms, but also to encourage regional integration. In order to develop appropriate export-policies, it is first necessary to investigate the behaviour of exporters at the firm level.

Very little research has been done on this topic in the South African sense due to a lack of good, detailed panel data. Recently, access to a substantial set of population tax income data has been made available to a select few researchers for analysis. This has enabled a more detailed examination of South African firm behaviour. This chapter forms part of this research.

Using this official population data, this chapter was able to confirm the stylised facts of exporting for the case of South Africa: exporters are larger, more labour productive and capital intensive and pay higher wages than firms that sell domestically only.

In addition, this chapter finds evidence of exporter heterogeneity in terms of export destinations. Firms that export only to destinations within Africa significantly underperform in terms of output, labour productivity, and capital and intermediate input intensity, relative to firms that export internationally. Further, exporting to multiple destinations generates higher characteristic premia over single destination exporters. Firms exporting to multiple destinations outside of Africa, therefore, display superior characteristics relative to all other exporters as well as non-exporters.

In the sparse South African literature on exporters, there is little evidence to suggest that manufacturing exporters in general are more productive than non-exporters. The evidence in this chapter contributes to the literature by re-examining the South African case, using this considerable data, to determine possible explanations for this missing productivity premium among exporters. One such explanation involves the large degree of exporter heterogeneity in terms of the destination served as well as how many destinations are served.

In the first case, the study by Rankin (2001) found that no productivity premium existed for exporters relative to non-exporters, and it was only when firms exported outside of Africa that a premium became evident. This heterogeneity among export destinations is an important aspect to acknowledge and examine, particularly since policy makers in South Africa are encouraging firms to increase exporting within Africa.

This chapter too finds that there is a higher productivity premium associated with exporting outside of Africa relative to non-exporters and Africa-only exporters. However, in contrast to Rankin (2001), this chapter also finds that exporters are, in general, more productive than non-exporters. A possible explanation for these contrasting results may simply be that the exporting environment has changed substantially over the decade and a half between the studies. Alternatively, despite the identical specification of the Africa/non-Africa in both papers, the Rankin (2001) result for the exporters in general may have to do with the sample oversampling Africa-only exporters.

Heterogeneity is also evident in terms of the number of export markets served. The results in this study suggest that firms exporting to multiple destinations are significantly more productive than firms which export to a single destination. In addition, firms exporting to multiple destinations outside of Africa (and perhaps to some destinations within) exhibit the highest productivity premium relative to other exporters. Interestingly, small firms exporting to a single destination outside of Africa display a lower productivity than firms exporting to a single destination within-Africa as wells as firms selling for the domestic market only.

A potential explanation for why beyond-Africa exporters, as well as multi-destination, exhibit the productivity premiums can be drawn from the theory. Access to foreign markets is associated with a number of barriers, or sunk costs of investment (transport costs, barriers to trade, technology improvements, etc.). According to Melitz (2003) before entering international markets, firms face uncertainty concerning their future productivity when making these costly and often irreversible investment decisions. Due to the nature of sunk costs, only the more productive firms will be able to enter the export market, while the less productive firms will be pushed out. Beyond Africa exporters are more productive than in-Africa exporters since the technological distance, as well as transport costs, are lower within the region. Therefore the entry-level productivity requirement is lower when exporting within the region compared to beyond (see for example Granér and Isaksson (2009)). Indeed, some studies have found that African firms use region exporting, with the lower productivity requirements, as a steppingstone into more technologically distant international markets with higher productivity requirements (see Eaton et al. (2007)). This is based on the assumption that exporters 'learn-by-exporting' – a hypothesis that is becoming more relevant in developing economies (Boermans, 2013). As for why multi-destination exporters exhibit a productivity premium, if there is a fixed cost involved in entering a specific market (as a Melitz-style model suggests), then firms not only have a baseline level of productivity which they need to overcome, but an additional productivity threshold for each of the new markets (since each new market requires that the firm overcome a new set of sunk costs such as market research, trade permits, transport costs, etc.).

Overall the results suggest that it is important for policy makers to know that all exporters should not be treated as homogeneous. Regional exports are encouraged in policy: the productivity threshold is likely to be lower within the region, which will allow the less productive firms to enter the export market. This could provide a stepping stone for smaller firms who, as they learn and grow, become productive enough to enter the more competitive international markets.

However, there is little evidence to suggest that exporting within the region is associated with greater performance outcomes than selling domestically. Indeed, the highest productivity jump comes from exporting to multiple destinations, be it within Africa or internationally. In general, firms should be encouraged to export to multiple destinations (since single-destination exporters exhibit little, if any, premium over non-exporters). In particular, firms should be encouraged to export to multiple destinations outside of Africa which will likely improve efficiency, technological knowledge and ultimately overall economic growth. More research is required to understand the dynamics behind this exporter-productivity-destination relationship in order to better understand the potential impact of policy among heterogeneous exporters.

Chapter 3

Clustering South African Manufacturing Firms and Exporters: A hierarchical and ClustOfVar approach for segmenting firms using mixed data

3.1. Introduction

In the international trade literature, exporters are increasingly shown to differ from non-exporters across a number of dimensions. In particular they are shown to be more input-intensive, larger, pay higher wages and are generally more productive than non-exporters (Wagner, 2011).

In South Africa, comparatively less research has been done on manufacturing exporters and non-exporters. However, the limited South African research confirms the findings of the previous international studies. Rankin (2001), for example, finds that manufacturing exporters in South Africa are more labour-productive, capital-intensive and pay higher wages relative to firms selling domestically. These findings are confirmed by Matthee and Krugell (2011), and Matthee *et al* (2016) who show that South African manufacturing exporters are significantly different across a number of characteristics including, size, capital and intermediate input intensity, wages, age and productivity.

In chapter 2, additional support for differences across these performance characteristics between exporters and non-exporters were given. That chapter further highlighted the differences in terms of these characteristics among exporters. Specifically, it found that exporting outside of Africa was related to higher performance premia than exporting within Africa only. Further, firms exporting to multiple destinations (both within and outside of Africa) exhibit higher performance characteristics over both non-exporters as well as firms who export to a single destination.

These previous studies have mostly used discriminant function analysis to study the characteristics and behaviours of exporters and non-exporters. While admittedly well-founded, these techniques use *a priori* assumptions to categorise firms prior to analysing their differences. An alternative approach, which serves as a robustness check of these *a priori* assumptions, is to allow the data to determine the classification of firms and exporters. This chapter uses two sets of rich, official data on South African firms, namely: Stats SA's Large Sample Survey and The South African Revenue Services (SARS) population tax administration data to classify firms, and exporters, through the use of cluster analysis.

Cluster analysis is a powerful tool used to aggregate firms according to their characteristic similarities (Murray, 2016) and according to Beckstead (2002) is useful for developing typologies, generating hypotheses through data exploration, and testing whether groupings of observations as defined by other means in the literature are in fact present in the data. In this chapter, therefore, the data informs the discussion without imposing any *a priori* structure on it.

Cluster analysis techniques are widely used across many disciplines including market research (Chakrapani, 2004); astronomy (Faúndez-Abans, Ormeno, & de Oliveira-Abans, 1996); genetics (Cortese, 2000); as well as in business and sociology (Murray, 2016). Clustering is also referred to in the literature as numerical taxonomy, segmentation analysis, typology analysis and Q analysis (Everitt *et al*, 2011). Regardless of the name, all techniques aim to identify groups in data.

The growth of cluster studies is indicative of the increasing interest in the topic as well as the desire among analysts to deepen the understanding of it (Tvaronavičienė, Razminienė, & Piccinetti, 2015). Despite its growing popularity, it has only recently been applied more widely in the economic discipline (Murray, 2016).

This work therefore adds to the economic literature on cluster analysis by taking advantage of the rich datasets to classify South African manufacturing firms based on their key performance indicators. The aim is to determine whether the traditional classification of manufacturing firms, as determined *a priori* in the trade literature, hold true in the population data. In addition, finding and assessing homogeneous, distinct groups of manufacturing firms will assist in identifying segments in the manufacturing environment who could benefit from targeted policy interventions. This study is one of the first to analyse South African population data through cluster analysis and one of the first in the South African trade literature to make use of a relatively recent algorithm called ClustOfVar for dimensionality reduction on mixed datasets.

The results of the cluster analysis on both firms and variables confirm the findings of previous South African trade studies. Firms group into small and medium/large exporters and non-exporters. Exporters then further group according to the destination of exports: within African, or outside of Africa. The analysis shows that firm size, as well as export status and productivity play an important role in distinguishing firms. It also confirms that existing *a priori* classifications are broadly correct. The findings of this chapter points to heterogeneity not only among firms in general, but also among exporters themselves. This has important consequences for too-general policies and interventions, since aspects that are important to any one firm in a particular cluster will also be important to other firms in that cluster, but will be different to what is important to firms in other clusters.

The remainder of the chapter is set out as follows. Section 3.2 discusses the theory behind clustering techniques. Section 3.3 briefly presents an overview of the clustering literature. The two datasets used in this study are described in section 3.4, while section 3.5 layouts the methodology employed. Section 3.6 presents the results of the cluster analysis and section 3.7 concludes.

3.2. Cluster Analysis: The Theory

Cluster analysis is a numerical technique used to classify multivariate data into a limited number of homogenous, yet distinct groups (known as clusters) in such a way that maximises similarity (homogeneity) within each group and minimises the dissimilarity (heterogeneity) between the groups (Johnson & Wichern, 1998).

This technique originated more than seven decades ago in the early 1930's[5]. Driver and Kroeber (1932) first developed cluster analysis in the field of anthropology and a few years later were followed by Zubin (1938) and Tryon (1939) who introduced the technique to psychology (Bailey, 1992). Soon after, the social sciences took up the formulation of clustering techniques, which experienced an increase in development, alongside computerisation, between the 1950's and 1960's (Bailey, 1994). The use of cluster analysis in multiple disciplines has continued to see tremendous growth, particularly since 2000 (Murtagh & Kurtz, 2016).

A substantial number of methods have developed over the years for clustering data (see Kaufman and Rousseeuw (2005) for a review of the methods). Before outlining some of these methods, it is necessary to first discuss how similarity is measured between objects.

3.2.1. Proximity Measures

Central to cluster analysis is the estimation of similarity, or distance, between objects (or units, observations, entities, etc.). The closer the objects are in terms of their distance, the more similar they are declared to be (Mooi & Sarstedt, 2011). These similarities (or distances) are usually expressed in an *n* x *n similarity matrix*, generally referred to as *distance* or *proximity matrices* (Kaufman & Rousseeuw, 2005).

There are a number of methods for measuring proximity, but the most popular measure of distance between the observations $(X_{1i}, X_{2i}, \ldots, X_{pi})$ and $(X_{1j}, X_{2j}, \ldots, X_{pj})$, is the *Euclidean Distance* (Everitt *et al*, 2011)[6]:

$$d(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{pi} - X_{pj})^2} \tag{3.1}$$

Object i is considered to be closer (and therefore more similar) to object j than it is to some other object k if $d(i,j) < d(i,k)$.

[5] For an extensive history of cluster and classification techniques, see Bailey (1992 & 1994) and Murtagh and Kurtz (2016).

[6] Equally as popular is the use of Squared Euclidean Distances (Mooi & Sarstedt, 2011).

Another commonly used method is the *city block distance* (also known as the *Manhattan metric*) (Kaufman & Rousseeuw, 2005), which is measured as the absolute distance travelled between points i and j if these points were located at opposite corners of a city block:

$$d(i,j) = |X_{1i} - X_{1j}| + |X_{2i} - X_{2j}| + \cdots + |X_{pi} - X_{pj}| \tag{3.2}$$

Other distance metrics include; the *Minkowski distance* (a generalisation of the Euclidean and city block measures), the *Canberra distance*, *Angular separation* and *Mahalanobis distance* (Kaufman & Rousseeuw, 2005; Everitt *et al*, 2011; and Mooi & Sarstedt, 2011). These methods are useful when the data is made up entirely of continuous variables (Everitt *et al*, 2011).

When data is binary or categorical, similarity $s(i,j)$ between two objects, i and j, is measured using matching coefficients which are 'defined in terms of the entries in a cross-classification of the counts of matches and mismatches in the p variables for two individuals' (Everitt *et al*, 2011, p.46). In its simplest form, this classification (or association) table is given as:

Table 3.1. Association Table for Matching Coefficients

		Object i		
	Outcome	1	0	Total
Object j	1	a	b	*a+b*
	0	c	d	*c+d*
	Total	*a+c*	*b+d*	*p=a+b+c+d*

Adapted from Mooi and Sarstedt (2011)

The *simple matching* coefficient is the most well-known and is defined as the ratio of the number of matches (or equivalently mismatches) to the total number of attributes (Kaufman & Rousseeuw, 2005), given by

$$s(i,j) = \frac{a + d}{(a + b + c + d)} \tag{3.3}$$

This measurement has some limitations, for example in the treatment of zero-zero matches or uninformative co-absences (Everitt *et al*, 2011). A number of alternative matching coefficients are therefore available such as the *Jaccard*, *Rogers and Tanimoto*, and *Sokal and Sneath* coefficients (see Kaufman & Rousseeuw (2005) and Everitt *et al* (2011) for a more extensive discussion).

A challenge arises when data is mixed, which is quite often the case with large datasets. As discussed in Kaufman and Rouseeuw (2005), one option is to treat all variables as interval- scaled, i.e.: treat all variables as if they are measured on a linear scale. In many cases this will work well for binary variables but not so well for variables with more than two levels or categories since, for example, it assumes that the distance between (categories) 0 and 4 is twice the distance between (categories) 0 and 2 and attaches meaning to these differences in distances when it is very likely that there is none. An alternative option is to transform all variables into binary variables, however in some cases important information could be lost in the transformation. A number of more complicated techniques have therefore been proposed (Everitt *et al*, 2011). A widely used measure for mixed data is Gower's general similarity coefficient (Niakšu, 2013). Gower (1971) provides a similarity measure which takes care of mixed data (interval, nominal and binary). The similarity coefficient for objects i and$j$ is given by

$$s(i,j) = \frac{\sum_{k=1}^{p} w_{ijk} s_{ijk}}{\sum_{k=1}^{p} w_{ijk}}, \tag{3.4}$$

where s_{ijk} is the similarity between objects i and j on the kth variable, and w_{ijk} is the assigned weight function. If observations for the kth variable for either, or both, objects i and j are missing, then w_{ijk} is set to zero. In addition, for nominal variables (binary or categorical), s_{ijk} is equal to one when two objects have the same value for variable (x) and zero otherwise, i.e.:

$$s_{ijk} = 1\,, iff\ x_{ik} = x_{jk}, and$$

$$s_{ijk} = 0, when\ x_{ik} \neq x_{jk} \tag{3.5}$$

For continuous variables:

$$s_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{r_k}, \tag{3.6}$$

where r_k is the difference between the maximum and minimum values of the kth variable (i.e.: the range of the kth variable).

3.2.2. Clustering Algorithms

Once the similarities or distances have been estimated, objects can be clustered. Clustering methods can broadly be split into hierarchical and non-hierarchical (or partitioning) techniques (Kaufman & Rousseeuw, 2005). Each will be discussed in turn.

3.2.2.1. Hierarchical Clustering

Hierarchical clustering methods can either be agglomerative or divisive (Mooi & Sarstedt, 2011). The agglomerative technique is a bottom-up approach which begins by assigning each object to a cluster made up only of that one object. Subsequently, similar clusters are merged according to the smallest distance between them. The algorithm ends with the extreme of every object being in one large cluster. Divisive clustering is a top-down approach beginning with all objects in one large cluster and then dividing them according to the largest distance between them. This algorithm ends with each object being a cluster unto itself.

The most frequently used methods for combining (or dividing) clusters at each stage are as follows (Kaufman & Rousseeuw, 2005; Mooi & Sarsteadt, 2011):

Single linkage (nearest neighbour). The distance between two clusters is defined as the smallest distance between two objects in the different clusters. At every step, the distance between two clusters is taken to be the distance between their two closest members.

Complete linkage (furthest neighbour). The opposite approach to single linkage. The distance between two clusters is defined as the distance between the two furthest points.

UPGMA (unweighted pair-group method using arithmetic averages). The distance between two clusters is the average of the distances between all pairs of cases in which one member of the pair is from each of the clusters.

Average linkage within groups. This method differs from UPGMA in that it combines clusters so that the average distance between all cases in the resulting cluster is as small as possible. The distance between two clusters is, therefore, the average of the distances between all possible pairs of cases in the resulting cluster.

Centroid. The distance between the two clusters is the distance between the two cluster centroids (the geometric centre of each cluster). This algorithm is intended for use with Euclidean distances.

Ward. Clusters are merged based on the size of an error sum-of-squares criterion. The objective at each stage is to minimize the increase in the total within-cluster error sum of squares. Like centroid linkage, this method makes use of Euclidean distances and is intended for use with continuous data.

Given the various options of linkage methods available to the analyst, the natural question is how to decide which method to use? Sneath and Sokal (1973) claim that in order to make interpretation of the results easier, the simplest measure should be chosen. Kaufmann and Rousseeuw (2005) agree with this claim and argue that UPGMA is not only easy to use and explain, but compared to the other techniques discussed previously, performs well in practice and is not restricted to interval scaled data which requires Euclidean distances. Further, Everitt *et al* (2011, p. 79) conclude that the UPGMA linkage method is 'relatively robust'.[7]

3.2.2.2. Partitioning Methods

Another popular group of classification programs are non-hierarchical or partitioning methods. The most well-known, and commonly used, partitioning method is k-means clustering, due to its simplicity

[7] Kaufmann and Rousseeuw (2005) provide a detailed comparison of the various techniques in Chapter 5 of their book. In addition Everitt *et al* (2011) provide a summary of the techniques and some of their characteristics.

and speed (Venkatesan, 2007)[8]. K-means clustering differs from hierarchical clustering in that it does not need to estimate a matrix of distances. K-means clustering aims to minimise within-cluster variation (Mooi and Sarstedt, 2011). In addition, the number of clusters, *k*, are not defined by the algorithm, but instead should be known beforehand. Tryfos (1998) provides a simple summary of the k-means approach:

> *Step 1*. Specify the number of clusters and assign each object (either randomly or deliberately) to a cluster.
>
> *Step 2*. Calculate each cluster's centroid and the distances between each object and centroid. If an observation is nearer the centroid of a cluster other than the one to which it currently belongs, re-assign it to the nearer cluster.
>
> *Step 3*. Re-compute the new cluster centroid and repeat step 2 until some convergence criteria is met (usually that the re-assignment of objects to clusters remains unchanged over multiple iterations (Venkatesan (2007)).
>
> *Step 4*. If it is the case that the number of clusters in step 1 cannot be chosen with confidence, repeat steps 1 to 3 with a different number of clusters and evaluate the results.

Since there are no objective criterion for selecting the most appropriate clustering model, Vázquez and Sumner (2012) conclude that selection will depend on the interpretability of the final result as well as expert knowledge. As will be seen in the literature section which follows, there is little consistency in the techniques used both within and across various fields and disciplines. However, a number of validation and robustness checks have been suggested by Venkatesan (2007), Everitt *et al* (2011) and Mooi and Sarstedt (2011). For example, an analyst can test the stability of the solution by re-running the analysis on the same dataset using different distance measures, or clustering algorithms. The stability of the cluster solution can also be tested by splitting the dataset into two halves, by running on different datasets or by simply changing the order of the objects in the dataset. Further, validity of the solution can be assessed by a number of validity criterion (are the clusters parsimonious, accessible, actionable?). In this sense, therefore, Venkatesan (2007) concludes that the clustering process will ultimately require some level of expert intuition and careful judgement.

3.3. Cluster Analysis: A Literature Review

The literature on the empirical application of cluster analysis is vast and can increasingly be found within multiple disciplines including: Marketing (for example in the classification of supermarket shopping paths (Larson, Bradlow, & Fader, 2005) and product branding (Moroko & Uncles, 2009)),

[8] See Kaufmann and Rousseeuw (2005), Andrews and Currim (2003), and Ayramo and Karkkainen (2006) for discussions on various partitioning methods.

Medicine (for example Beckstead (2002) applies hierarchical cluster analysis to research questions in the field of nursing, while Farmer, McGuffin, and Spitznagel (1983) and Hay, Fairburn and Doll (1996) used cluster analysis to classify mental illness and eating disorders respectively), Astronomy (see Wagstaff & Laidler (2005) and Jang & Hendry (2007) for examples), Climatology (Unal, Kindap & Karaca (2003); Yokoi *et al* (2011); and Corporal-Lodangco & Leslie (2016)) and Archaeology (see Papageorgeou , Baxter & Cau (2001); Mucha, Bartel & Dolata (2005); and Baxter (2009)) to name just a few[9].

Despite its multidisciplinary popularity, the use of clustering techniques in the discipline of social sciences, particularly within the field of economics, has only recently started to gain traction (Murray, 2016). There is evidence of the use of cluster analysis in development economics (see Vázquez & Sumner (2012) who classify developing countries into five clusters, as opposed to only low income or middle income); insurance (see Alexandru, Strat & Gogonea (2012) who use principal component analysis (PCA) and hierarchical clustering to identify key performance indicators for Romanian Insurance companies) and agronomics (see Apili Ejupu, Makhura & Kirsten (1999) and Diaz-Bonilla *et al* (2000) who both use cluster analysis to evaluate food security in rural Uganda and WTO countries respectively).

A number of studies have also used clustering techniques to describe various socio-economic topics. Lombardo and Falcone (2011) make use of a non-hierarchical clustering technique known as partitioning around the medoids (PAM), a method related to k-means analysis, to classify Italian provinces into seven distinct clusters based on criminal, economic and other socio-demographic indicators. The authors' findings contradict the previous belief that crime is indistinguishably tied to geographic location. Other studies attempt to identify the typology of welfare regimes, and describe the characteristics of each identified typology, by making use of hierarchical, k-means and PAM clustering techniques (Saint-Arnaud & Bernard, 2003; and Minas *et al*, 2014).

Another relatively popular social science topic in the clustering literature relates to innovation or, more precisely, the classification of innovation strategies or modes. Srholec and Verspagen (2008) use firm-level data on a number of European countries and, through the combined use of factor and k-means cluster analysis, identify five clusters which differ in terms of their innovation strategies. They use this finding to caution against the traditional method of classifying countries at the sector or country level since a large amount of heterogeneity exists within clusters at these levels. These findings compliment the study by Tiri, Peeters and Swinnen (2006) who, using a combination of hierarchical and k-means clustering on Flemish firms, identify seven clusters (interpreted as innovation strategies) as opposed to the traditional 'narrow' classification of firms into three or four strategies. Additional analyses use

[9] See Murtagh (2014) and Tvaronaviciene, Razminiene & Piccinetti (2015) for a more detailed review of this extensive literature.

clustering techniques to investigate the behaviour and patterns of innovation, for example Hollenstein (2003) for firms in the Swiss service sector and Wziatek-Kubiak, Balcerowicz and Peczkowski (2009) for manufacturing firms in the Czech Republic, Hungary and Poland.

Trade literature has recently made use of clustering techniques to explore various research questions. In the macro-economic trade literature Kharlamova and Vertelieva (2013) classify 36 countries in terms of their relative competitiveness into five homogeneous and distinct clusters using a combination of hierarchical and k-means clustering. The aim of the paper was to create a relatively objective competitiveness index for the sample of countries which could serve to assist these countries in developing appropriate strategies with various trade partners. Similarly, Popa and Stefan (2015) employ a combination of factor and hierarchical cluster analysis in order to classify 35 European countries in terms of the main characteristics of competitiveness (as determined by the factor analysis). The authors conclude that the European countries can be classified into four groups, differing in terms of growth strategies and market development, which can be used to assess the objectives of European competition policy among the member states.

Another macro-economic trade study set out to construct a typology of traders based on value-added, economic and trade policy indicators (Escaith & Gaudin, 2014). Countries were clustered using hierarchical techniques and used to characterise the underlying patterns within the global value chain. Recently, a study by García, Grigonytė, & Oliver (2015) used cluster analysis to evaluate the commercial relationships between counties of the European Union. The authors found a two-cluster solution which classified country pairs into either standard trade relationships or intense trade relationships.

The use of cluster analysis on trade indicators at the micro level has also been growing. K-means clustering was used to classify provinces in China into four groups according to their levels of foreign direct investment (FDI), namely: highly developed provinces; low-middle provinces; middle to middle-upper provinces; and economically developing provinces (Yu & Zhang, 2007). The results of this study gave insight with regards to the ability of provinces within each cluster to attract FDI, an important classification for FDI decision makers in terms of future investment strategies. Likewise, Friedman *et al* (2011) used FDI (as well as other indicators of trade openness) to classify sectors within the Chilean economy. A combination of the hierarchical and k-means method of clustering identified three categories of openness: high, medium and low. Further analysis of these clusters by the authors highlighted a significant wage gap based on openness to trade: higher wages characterise more open sectors.

At the exporting level, clustering techniques have been used to explore export diversification and strategy. Shirotori, Tumurchudur and Cadot (2010), for example, classified products into six, broad clusters according to their factor intensities. The authors then describe the industry composition of each cluster, highlighting the differences both within and between industries and concluding that policies aimed at increasing export diversification should be implemented at a high degree of disaggregation. Onodi and Pecze (2014) used k-means cluster analysis to describe the characteristics of successful exporting firms in Hungary. After classifying exporters into two groups (successful export-oriented firms and stagnant firms), further analysis concluded that strategic management is likely to play a large and significant role in exporter success.

The economic literature using cluster techniques in the South African context spans various topics, but is ultimately sparse. Montalbano and Nenci (2011) employ cluster analysis on South African data at the macro-level to evaluate trade clusters within the WTO (World Trade Organisation) countries, with a particular focus on the China, India, Brazil and South African (CIBS) cluster. Hierarchical clustering results suggest no evidence of a CIBS cluster, at least in terms of trade patterns. In particular the authors find that China and India trade paths differ substantially, while Brazil and South Africa both follow patterns similar to that of other developing countries more so than that of China or India. In a later study, Montalbano and Nenci (2014) again use macro-economic data on the emerging CIBS countries to test the hypothesis that emerging economies pose a threat to industrialised countries in terms of their level of competitiveness. The authors use fuzzy cluster analysis to assess the level of trade competitiveness for the CIBS countries, and find little evidence in support of this hypothesis.

At a more disaggregated level, Tyler and Gopal (2010) explore data on South Africa, and other sub-Saharan African countries, to examine regional patterns of development. The authors make use of both the PAM technique as well as Kohonen's Self-Organising Map (SOM) to identify eight clusters which 'accurately represent the differences among countries within sub-Saharan Africa' (Tyler and Gopal, 2010, p.17). By further analysing the characteristics of each cluster, the authors find sufficient evidence to suggest that a direct relationship exists between governance and economic well-being, particularly in South Africa. Bunting *et al* (2010) made use of data on South Africa's higher education institutions to classify South African universities according to their purpose or functions (as opposed to the more formal classification on rankings). A k-means cluster analysis classified South African institutions into three significantly distinct groups which differ in terms of the students they serve and the type of qualifications produced.

By making use of household survey data, a number of studies have used clustering techniques to assess the South African economy at the micro-level. Some studies, for example, have employed clustering techniques in the area of agricultural economics where the authors evaluate small-holder farming in South Africa. Specifically, Van Averbeke and Mohamed (2006) used a visual clustering technique to classify three clusters of South African farmers in terms of their differing farming styles. The identification of these different farming strategies lent support to the author's argument that the impact of general agricultural policies for small-holder farmers are likely to be muted since the different groups of farmers are unlikely to respond to policies in the same way. In a recent paper, Pienaar and Traub (2015) used a more objective approach to analyse this topic, which included the use of PCA, hierarchical and k-means clustering methods to empirically assess South African smallholder farmers. Similarly to the study before them, the authors found a degree of segmentation among smallholder households (they identified seven clusters), again leading to the conclusion that the treatment of smallholder farmers as one homogeneous group could mislead development planning and policy goals.

Other studies use cluster analysis to assess the welfare of South African households. For example, Vella and Vichi (1997) clustered South African households by means of PCA and k-means techniques into poverty classes based on several socio-economic factors (such as water and sanitary conditions, asset ownerships, household debt, dependency ratio, and demographics). The authors were thus able to group households into clusters with similar standards of living and were further able to compute a composite index of poverty (or deprivation) for South Africa. In another welfare study, McParland *et al* (2014) developed a hybrid model based on impulse response theory and factor analysis to cluster their study population (households living in North-east South Africa) in terms of their asset status. Four distinct, homogeneous clusters were identified and interpreted as four socio-economic states (SES). The characteristics of each SES cluster, as stated by McParland *et al* (2014, p. 23) could, 'aid decision-making with regard to infrastructural development and other social policy'.

It should briefly be clarified here that the term 'cluster' has an alternative application in economic literature to that defined in the previous section and discussed above. In the economic (geography) literature, clusters are defined as, 'geographic concentrations of interconnected companies, specialised suppliers, service providers, firms in related industries and associated institutions in particular fields that compete but also co-operate' (OECD., 2009, p. 25). This definition is derived from Porter's economic cluster theory, a theory which hypothesised why particular industries become competitive in particular locations (Porter 1990 & 1998). This lead to the development of a key tool of analysis known as the "Diamond" model, which presents a broad view of the driving forces of competitiveness of clusters. The empirical literature on industrial geographic clusters generally involves the investigation of local supply chains and other characteristics associated with industry or regional agglomerations and international competitiveness and has been well researched (see for example; Lagendijk & Charles (1999); Porter (2007); Aylward & Glynn (2006); OECD (2009); Garanti & Zvirbule-Berzina (2013); and Howard *et al* (2014)).

This chapter, unlike industry clustering which uses *a priori* assumptions to identify agglomerations ('industrial/ regional clusters'), uses exploratory data techniques (such as hierarchical cluster analysis) to objectively identify homogeneous, yet distinct, groups of South African exporting firms and compares the findings to those of previous South African exporter studies (see Chapter 2). This chapter adds to the growing economic literature in two ways. Firstly, it is, to the best knowledge of the author, one of the first studies to use an unsupervised machine learning technique (such as cluster analysis) to analyse heterogeneity among South African exporters at the firm-level. Secondly, while most South African studies using cluster analysis are based on small survey data sets, this study is one of the first which has the opportunity to run such exploratory analysis on two substantial and official firm-level data sets, one of which is actual population data. The next section discusses the various unsupervised methods that will be used to explore such rich datasets.

3.4. Methodology[10]

3.4.1. Standard Hierarchical Clustering

This chapter adopts two alternative clustering methodologies developed for the analysis of mixed variables. The first method employs agglomerative hierarchical clustering to the selected variables described in Section 4. Hierarchical clustering is selected over the non-hierarchical k-means technique since k-means clustering traditionally uses Euclidean distances and is therefore unsuitable for mixed data (Ahmad & Dey, 2007). Distances between firms are calculated using Gower's general similarity coefficient (equation 3.4). Gower's measure is an option available in the daisy function of the R statistical software cluster package.

Kaufman and Rousseeuw (2005) provided a detailed discussion of the daisy function in Chapter 5 of their book. The main advantage of this function is its ability to handle many types of data such as ordinal, nominal and binary variables even if they appear in the same dataset.

Following Kaufmann and Rousseuw (2005) and Montalbano and Nenci (2011), hierarchical clustering is applied to the estimated distance matrix using the average linkage (UPGMA) method, since, according to these authors, this method of linking clusters together performs well in most situations. Additionally, the agglomerative hierarchical algorithm provides an agglomerative coefficient which quantifies whether or not there is a natural cluster structure in the data. The agglomerative coefficient ranges between 0 and 1 with higher values indicating a clear clustering structure (Kaufman & Rousseeuw, 2005).

Selecting the number of clusters from the analysis above is argued to be largely subjective, and based on expert judgement and theoretical sense (Saint-Arnard & Bernard, 2003 and Escaith & Gaudin, 2014). However, a number of methods are available, some more objective than others, which together can assist in deciding on the number of clusters.

In the first instance, a dendrogram is used to illustrate the hierarchical classifications produced by the agglomerative method. This diagram embodies a tree-like structure which graphically displays the

[10] The analysis in this chapter was carried out using the R Project for Statistical Computing, or simply R for short. R is a free (open source) available from https://www.r-project.org/.

distances at which firms (and clusters of firms) are merged: the length of the stems (vertical lines) represent the distances, while the nodes represent the clusters. The general rule of thumb regarding the choice of the number of clusters is to cut the tree where the distances are relatively large (see Everitt *et al* (2011) for a description of dendrograms and their properties in Chapter 4).

In many cases, particularly in large datasets, given the large number of mergers generated at each stage, a visual determination of the ideal place to cut the tree can be challenging. Therefore, a second method for determining the number of clusters makes use of the agglomeration schedule (Mooi & Sarstedt, 2011). In particular, a scree plot (which graphically represents the agglomeration schedule, i.e.: the estimated distances against the number of clusters) is plotted. A distinct break (or elbow) in the plot illustrates the point at which adding another cluster does not sufficiently improve information. This point generally indicates the cluster solution (the optimal number of clusters).

Finally, a more precise method for determining the number of clusters is utilised. Calinski and Harabasz (1974) proposed the variance ratio criterion (VRC), a cluster-stopping rule, to determine the optimal number of clusters. According to the survey of Milligan & Cooper (1985) this index appears to be the better performer among a number of alternative stopping rules. The Calinski and Harabas index is given by:

$$CH_k = \frac{SS_B/(k-1)}{SS_W/(n-k)} \tag{3.7}$$

Where k denotes the number of clusters, n the objects, SS_B the overall between-segment variation and SS_W the overall within segment variation with regard to all clustering variables. An optimal number of clusters is then defined as a value of k that maximizes CH_k[11].

Once the cluster solution has been determined, analysis of the characteristics of each cluster group is presented and discussed. Student t-tests are run on the coefficients between each cluster to determine whether any characteristics are significantly different or not. The output of the t-tests are presented in Appendix C.

3.4.2. ClustOfVar

Whereas the previous method involves clustering firms based on a number of key performance variables, the next method used in this chapter clusters the variables themselves. The method employed is useful for analysis of large, mixed datasets and follows that of Brida *et al* (2014).

The first stage involves reducing the dimensions of the data by transforming the space of the initial dataset into another space of lower dimension (ideally containing a set of uncorrelated principal

[11] The `NbClust` package in R provides the optimal cluster solution as determined by the Calinski and Harabas index (see Charrad *et al*, 2014).

variables), while maximising the quantity of information recovered from the initial space. This technique is more commonly known as Principal Component Analysis (PCA) or quantitative variables, or Multivariate Correspondence Analysis (MCA) for qualitative variables (Chavent *et al*, 2012). This type of technique is useful for when there are a large number of variables in the dataset and/or when there is a chance that a number of variables in the original dataset are correlated. However, PCA and MCA cannot be applied to datasets of mixed type (Brida *et al*, 2014).

This analysis therefore uses the relatively recent `ClustOfVar` algorithm, developed by Chavent *et al* (2012) for R statistical software, which allows for the clustering of variables regardless of their type and is based on the PCAMIX method of Kiers (1991). PCAMIX is a principal component method for a mixture of numeric and categorical variables.

The ClustOfVar approach clusters variables according to a homogeneity criterion which includes both the correlation ratio (between groups variance to total variance ratio) of each qualitative variable as well as the squared correlation coefficients used for the quantitative variable (Brida *et al*, 2014). ClustOfVar employs either a hierarchical ascendant algorithm or a k-means type partitioning algorithm. For the sake of consistency the analysis again uses the hierarchical algorithm. This algorithm is implemented in the R `ClustOfVar` package, available on the CRAN (Comprehensive R Archive Network) and is defined as follows:

Let $\{z_1, z_2, \dots, z_{v1}\}$ represent a set of $v1$ categorical (or qualitative) variables and $\{x, x_2, \dots, x_{v2}\}$ represent a set of $v2$ continuous (or quantitative) variables. Further, denote $\boldsymbol{Z}$ and $\boldsymbol{X}$ as the corresponding (standardised) qualitative and quantitative matrices, respectively, of dimension n x p, where n is the number of observational units. The hierarchical method can then be expressed in the following steps:

Step 1. Begin with $V = v1 + v2$ partitions (i.e.: begin with each variable in a cluster of its own).

Step 2. Merge the two clusters which have the smallest dissimilarity between them. For clusters K_1 andK_2 this dissimilarity condition is given by:

$$(K_1, K_2) = H(K_1) + H(K_2) - H(K_1 \cup K_2) \tag{3.8}$$

The homogeneity of the kth cluster of variables $C_k \in \{C_1, C_2, \dots, C_K\}$is estimated by:

$$H(C_k) = \sum_{x_j \in C_k} r^2_{x_j y_k} + \sum_{x_j \in C_k} \eta^2_{y_k | z_j} = \lambda_1^k \tag{3.9}$$

where:
y_k - is the central synthetic quantitative variable which represents the first principal component of PCAMIX applied to the standardised variables of C_k,
r^2 – is the squared Pearson correlation of x_j(for quantitative variables) and y_k,
r^2 – is the correlation ratio of z_j (for qualitative variables) and y_k, and
λ_1^k – represents the first eigenvalue of PCAMIX applied to C_k

The strategy, therefore, consists in merging the two clusters that result in the smallest decrease in H. Using this aggregation measure the new partition into $p - l$ clusters maximises H among all the partitions into $p - l$ clusters obtained by the merging of two clusters of the partition into $p - l + 1$ clusters.

Step 3. Stop when single cluster consisting of all variables is achieved.

The number of clusters are then determined in a similar fashion as before, and the composition of each cluster of variables is defined and discussed.

Once the clusters of variables have been identified, the second step turns to the use of Classification and Regression Trees (CART). CART analysis uses historical data to construct so-called decision trees which in turn are used to classify new data (Brieman *et al*, 1984). A usefulness of CART analysis is that of identifying important clustering variables, i.e.: the classification tree identifies which variables have the most significant effect on partitioning the data into clusters (Brida *et al*, 2014).

A simple breakdown of the CART methodology is as follows. Let t_p be a parent node and t_l and t_r the left and right child nodes (respectively) of parent node t_p. The classification tree is built based on a splitting rule - the rule that performs the splitting of the data sample into smaller parts. At each split the data are divided into two parts with maximum homogeneity which is defined by an impurity function, most commonly the Gini index:

$$\underset{x_j \leq x_j^R, j=1,\ldots,M}{\arg\max} \left[-\sum_{k=1}^{K} p^2 (k|t_p) + P_l \sum_{k=1}^{K} p^2(k|t_1) + P_r \sum_{k=1}^{K} p^2(k|t_r) \right] \tag{3.10}$$

The reader is referred to Yohannes and Webb (1998) for a review of CART theory and practical implementation. This study employs CART analysis in order to estimate the variables that significantly affect the partition in clusters of the original datasets based on the first principal components discussed above.

3.5. The Data and Descriptives

The results of this study are based on two substantial and official South African firm-level datasets, namely: Statistics South Africa's (Stats SA) Manufacturing Large Sample Survey (LSS) and the South African Revenue Service's (SARS) tax administrative data. These will be described below.

3.5.1. Stats SA's Large Sample Survey

The LSS is a periodic survey which collects information on the South African manufacturing industry's economic activities. This includes industrial classification, employment details, trading income, profit/loss, inventories, the book value of assets and the products manufactured. Both private and public enterprises operating in the manufacturing industry are surveyed. These firms are classified into the following industries: Food products and beverages; Textiles, clothing, leather and footwear; Wood and wood products, paper, publishing and printing; Petroleum, chemical products, rubber and plastic products; Glass and non-metallic mineral products; Basic iron and steel, non-ferrous metal products, metal products and machinery; Electrical machinery and apparatus; Radio, television and communication apparatus and professional equipment; Motor vehicles, parts and accessories and other transport equipment; Furniture and other manufacturing.

This study uses data from the 2005 and 2008 LSS surveys. Given that the data consists of variables measured in different scales, and therefore to avoid giving more weight to any one variable, the data is first standardised to have a mean of 0 and a standard deviation of 1. This requires that the records need to be complete, and thus missing data is removed. The 2005 data consists of 9 499 of unique firms, which reduces to 3 188 due to missing data. Likewise, the 2008 data consists of 10 736 of unique firms, which reduces to 6 080 due to missing data.

In general, selection of variables into a cluster analysis is non-trivial and will depend largely on subject-matter knowledge and common sense (Kaufman and Rousseeuw, 2005). In order to make the results of the cluster analysis comparable to previous South African studies based on the Bernard and Jensen (1995) methodology (described in Chapters 1 and 2), this dataset (as well as the SARS dataset) is limited to include only certain key firm-level performance variables (chosen based on the theoretical and empirical findings of Chapter 2). These variables are presented in table 3.2.

Table 3.2. Variables and Descriptive Statistics – Stats SA LSS dataset

Quantitative Variables		**LSS 2005**		**LSS 2008**	
Variable description	**Label**	**Mean**	**Median**	**Mean**	**Median**
Log(Output per worker)	lryl	6.157	6.106	6.15	6.05
Log(Number of employees)	ll	4.215	4.220	4.14	4.04
Log(Capital per worker)	lrkl	3.841	3.832	3.85	3.87
Log(Intermediate Inputs per worker)	lril	5.873	5.793	5.78	5.64
Log(Wages per worker)	lrwl	0.034	0.062	0.34	0.40
Log(Total factor productivity)	ln_TFP_OLS_all	0.005	0.014	0.002	-0.005
Exports as a % of total output[a]	exp_sales_perc	17.970	7.452	12.09	4.51

Qualitative Variables		**LSS 2005**	**LSS 2008**
Variable description	**Label**	**% of sample**	**% of sample**
Medium to large dummy[b]	med_large	70.51%	82.94%
Exporter dummy[c]	exporter2005	27.98%	51.89%
Industry SIC Classification	common_sic3[d]		
***Food** products and beverages (SIC30)*		13.08%	12.38%
***Textiles**, clothing, leather and footwear (SIC31)*		9.47%	7.63%

***Wood** and wood products, paper, publishing and printing (SIC32)*	11.95%	11.02%
***Petroleum**, chemical products, rubber and plastic products (SIC33)*	14.62%	15.12%
***Glass** and non-metallic mineral products (SIC34)*	5.43%	5.12%
***Basic iron** and steel, precious and non-ferrous metal products, fabricated metal products and machinery (SIC 35)*	24.31%	28.52%
***Electrical machinery** and apparatus (SIC36)*	3.70%	3.98%
***Radio**, television and communication apparatus and professional equipment (SIC37)*	3.73%	2.20%
***Motor vehicles**, parts and accessories and other transport equipment (SIC38)*	7.06%	6.78%
***Furniture** and other manufacturing divisions (SIC39)*	6.65%	7.25%

Notes: [a] Conditional on a firm exporting; [b] a dummy variable of 1 if a firm is recognised as a medium or large firm, 0 if a firm is micro or small (categorised using the SARS definition of firm size based on total assets and gross income); [c] a dummy variable equal to 1 if a firm exports, and 0 otherwise; [d] this variable is measured in the data at 3-digit SIC (illustrated in the table as 2-digit SIC for ease of visualisation)

The mean and median value for most of the quantitative variables are similar which suggests that the distribution of these variables is approximately normal. However, this is not the case for exports as a percentage of total output. This variable's mean is higher, in both years, than its median suggesting that there may be a handful of exporters in the sample exporting a large proportion of their total sales relative to the sample's average exporter. This may be something to take note of when interpreting the results of the cluster analysis.

A large proportion of the LSS data (with complete cases) is made up of medium to large firms: around 71 percent of the 2005 sample and 83 percent of the 2008 sample. In 2005, just over a quarter of the sample was classified as an exporter, whilst over half the sample in 2008 reported exporting activities[12]. In addition, the average exporter exports less than 20 percent of total output. Again, these statistics need to be remembered during the cluster analysis. In both years more than 50 percent of the sample was represented by firms engaged in either; basic iron and steel, precious and non-ferrous metal products,

[12] Chapter 2 reported different percentages to this chapter. This likely has to do with the incomplete cases that had to be dropped in order to standardise the data (for example in the LSS data approximately 55% of firms from Chapter 2 did not have productivity estimates due to missing values for the performance variables: lrkl/ lrwl/ ll/etc.). Further in the LSS data, 74% of incomplete cases in terms of missing productivity estimates were micro or small firms, therefore since majority of complete cases were more likely to be large, they were also more likely to be exporters (assumption based on trade literature and the relationship between firm size and export status) and thus likely to push up the proportion of exporters in this chapters dataset relative to Chapter 2.

fabricated metal products and machinery (SIC 35); petroleum, chemical products, rubber and plastic products (SIC33); or food products and beverages (SIC30).

3.5.2. SARS Tax Administrative Data

The second dataset used in the cluster analysis was provided by the South African Revenue Service (SARS) in partnership with The National Treasury of South Africa as part of UNU-WIDER's (United Nations University and World Institute for Development Economic Research) Firm Level Analysis project. The release of this data through this project has, for the first time in the South African context, allowed for the analysis of highly disaggregated tax administrative data, which by definition is based on the population (as opposed to the more commonly used survey data sets).

A considerable amount of time and effort was put into the development of a subset of this extensive tax data base for the purpose of this study. As previously mentioned in Chapter 2, the method of generating this subset is discussed in detail in Appendix A. To recall here briefly, the three sources of data that are used in this, and the previous, chapter are (i) Corporate Income Tax data (CIT); (ii) Personal Income Tax data (PIT); and (iii) Customs transaction data.

The CIT records provide the primary data for this study. These records contain balance sheet, income statement and detailed tax information (including deductions and allowances) for businesses in South Africa. This information is extracted from two forms, the IT14 and the ITR14. The main difference between the two forms is that the former is a 'once size fits all' form whereas the updated, digital ITR14 form is customised based on the company type (micro, small, medium/large, dormant, etc.). In other words, the ITR14 reduces the administrative burden on smaller firms, who have fewer fields to complete, relative to medium and large firms who are required to complete significantly more, detailed fields.

The PIT records are obtained from employee tax certificates (IRP5 forms) contain information on various sources of income (both taxable and non-taxable), allowances, benefits and contributions (such as UIF, SDL and medical fund). These employee tax database records are used to estimate employment data which is then appended to the CIT data.

Finally, in order to obtain data on international trading activities of South African firms, customs transactions data is used. These records contain information on products imported and/ or exported, their value and volume, origin and destination, as well as tariffs and duties at the transaction level. These records are also appended to the CIT data.

As before, this chapter only retains the key performance and status variables used in Chapter 2 for consistency and comparability purposes (these variables are presented in table 3.3)[13]. In addition, only observations on manufacturing firms are kept (again for comparability). The SARS dataset for the financial year 2013 used for this chapter contains approximately 30 000 unique manufacturing firms. In order to standardise the data, observations with missing data for any of the key variables of interest are removed leaving a final SARS dataset of 8 879 manufacturing firms.

Table 3.3. Variables and Descriptive Statistics – SARS dataset

Quantitative Variables			
Variable description	**Label**	**Mean**	**Median**
Log(Output per worker)	lryl	13.50	13.62
Log(Number of employees)	ll	2.94	2.94
Log(Capital per worker)	lrkl	10.70	10.41
Log(Intermediate Inputs per worker)	lril	12.98	13.05
Log(Wages per worker)	lrwl	11.57	11.49
Log(Total factor productivity)	ln_TFP_OLS_all	-0.02	0.01
Exports as a % of total output [a]	exp_sales_perc	14.75	4.41

Qualitative Variables		
Variable description	**Label**	**% of population**
Medium to large dummy [b]	med_large	52.24%
Exporter dummy [c]	exporter	52.80%
Exporters exporting to a single destination that is out of Africa	sing_dest_out_afr_2013	3.95%
Exporters exporting to a single destination that is in Africa	sing_dest_in_afr_2013	21.78%
Exporters exporting to multiple destinations out of Africa	multi_dest_out_afr_2013	39.44%
Exporters exporting to multiple destinations within Africa	multi_dest_in_afr_2013	34.83%
Industry SIC Classification [d]	industry	
Manufacturing *of food products*		5.33%
Manufacturing *of beverages*		1.66%
Manufacturing *of tobacco products*		0.17%
Manufacturing *of textiles*		4.45%
Manufacturing *of wearing apparel*		3.18%
Manufacturing *of leather and related products*		1.33%
Manufacturing *of wood and of product of wood and cork, except furniture; manufacture of articles of straw and plaiting materials*		3.64%

[13] Import status has also been shown to relate strongly with export status (see for example Lawrence & Edwards (2008)).

Manufacturing *of paper and paper products*	2.84%
Printing *and reproduction of recorded media*	2.48%
Manufacture *of coke and refined petroleum products*	0.40%
Manufacture *of chemicals and chemical products*	5.33%
Manufacture *of pharmaceuticals, medicinal chemical and botanical Products*	1.86%
Manufacture *of rubber and plastic products*	8.19%
Manufacture *of other non-metallic mineral products*	1.87%
Manufacture *of basic metals*	4.38%
Manufacture *of fabricated metal products, except machinery and Equipment*	10.80%
Manufacture *of computer, electronic and optical products*	1.72%
Manufacture *of electrical equipment*	4.79%
Manufacture *of machinery and equipment n.e.c.*	8.41%
Manufacture *of motor vehicles, trailers and semi-trailers*	2.03%
Manufacture *of other transport equipment*	1.43%
Manufacture *of furniture*	2.99%
Other *manufacturing*	17.21%
Repair and installation *of machinery and equipment*	3.55%

Notes: [a] Conditional on a firm exporting; [b] a dummy variable of 1 if a firm is recognised as a medium or large firm, 0 if a firm is micro or small (categorised using the SARS definition of firm size based on total assets and gross income); [c] a dummy variable equal to 1 if a firm exports, and 0 otherwise; [d] this variable is measured in the data at 6-digit SIC (illustrated in the table as 2-digit SIC for ease of visualisation).

Just over half the population of manufacturing firms in this data set (of complete cases) are medium-sized or larger. In addition, a large number of manufacturing firms export some proportion of their sales (around 53%)[14], with the average exporter exporting just under 15 percent of their total output. Again, it is noted that, with the exception of the percentage-exported variable, there is little difference between the mean and median value for the quantitative variables (labour productivity, capital per worker, number of employees, etc.) suggesting an approximately normal distribution.

The SARS data further allows for the analysis of export destinations. A larger proportion of exporters in this study export to multiple destinations relative to those exporting to only a single destination, regardless of whether that destination is within or outside of Africa. However, if an exporter exports to a single destination only, they are more likely to be exporting within Africa (22% relative to the 4% exporting to only one destination out of Africa).

[14] This figure is higher than that expressed in Chapter 2 (less than 20% for the full sample of manufacturing firms), and more in line with the 45 percent of medium-large exporting firms from Chapter 2. As mentioned in footnote 8, this is likely to do with the incomplete cases that were dropped from the subset due to missing performance variables (which is more common among the smaller firms than the larger firms).

The next section implements the methodologies discussed in section 4 to cluster South African firms from both the LSS and SARS dataset according to their key performances variables in order to assess the different types of firms in South African manufacturing without imposing any *a priori* assumptions as to what those different types are.

3.6. Results

The purpose of the analysis in this chapter is to dissect the data in order to illustrate the usefulness of cluster analysis for grouping firms without imposing any prior assumptions about the groups. In essence, allowing the data to be clustered using the techniques discussed in the chapter, enables the confirmation of the *a priori* assumptions of firm groupings found in the literature. Further the process may also be able to pick out new groups of firms that were previously hidden within the data.

The first step of the analysis firms into clusters based on a set of variables. The next step describes these clusters of firms in terms of their characteristics (the average value of each variable in each cluster). The second step clusters the variables together in order to determine which variables are the most important in distinguishing these firm-groupings from the first step. This second step involves a sort of PCA/MCA mix reduction by grouping similar variables together creating synthetic variables. A tree analysis is then used to identify the importance of both the original variables as well as the synthetic variables. This technique is most useful when presented with many variables but remaina an interesting illustrative exercise in this chapter.

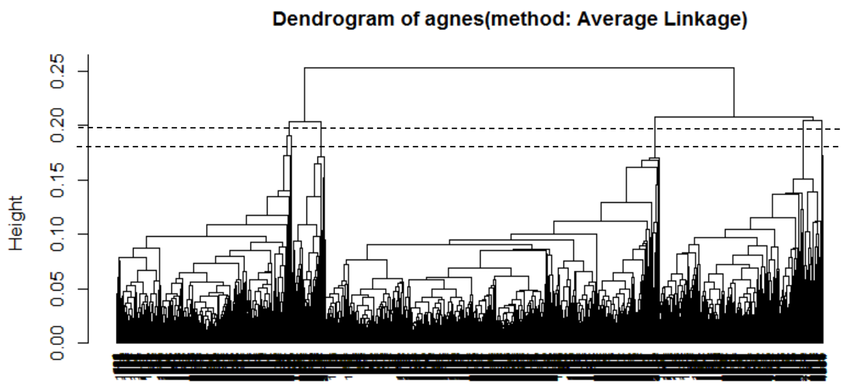
3.6.1. Standard Hierarchical Cluster Analysis using Gower Distance Measure

3.6.1.1.Clusters of Firms

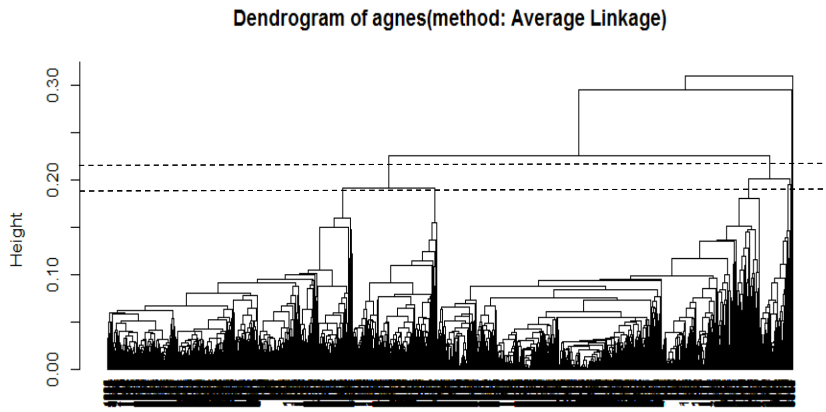
The analysis begins with a standard hierarchical (agglomerative) clustering procedure, where distances are estimated using the Gower measure and clusters are linked via average linkage (UPGMA). This will be performed for both the 2005 and 2008 LSS data as well as the SARS data discussed in section 5. Following the methodology set out previously, the decision on the number of firm clusters will be guided by the dendrogram, agglomeration scree plot, and the Calinski and Harabasz (1974) index.

Firstly, a visual inspection of the dendrograms for LSS 2005, LSS 2008 and SARS data (figure 3.1a, b, and c respectively) suggests that a number of natural clusters exist in all three datasets. Indeed the agglomerative coefficient 0.92, 0.96 and 0.96 respectively, suggest very clear structuring in the data. It should be noted here that this coefficient can be influenced by even one outlier (Kaufmann and Rousseeuw, 2005). However, inspection of the maps show that while there may be some indications of outliers, a tight clustering structure is evident.

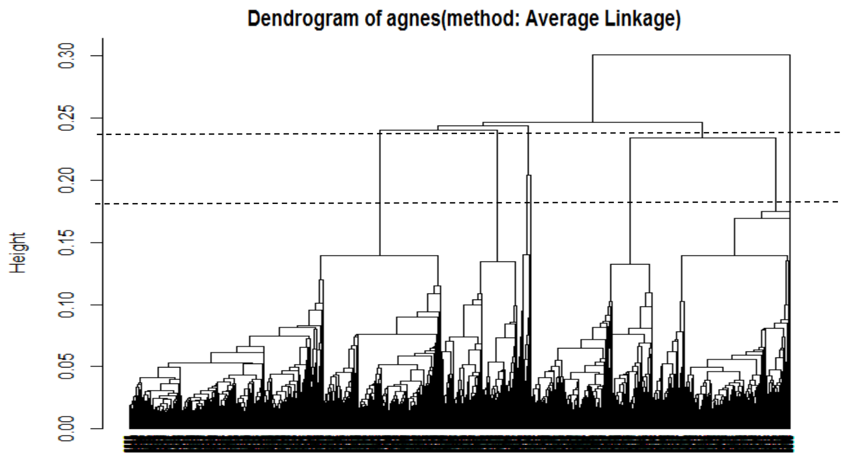
A rough guide as to how many clusters to retain can be determined by drawing a horizontal line through relatively large distances (measured by the vertical lines, or stems, in the dendrogram) and counting the number of stems it passes through. Figure 3.1(a) therefore suggests that either 5 or 6 clusters may be appropriate for the LSS 2005 data. Figure 3.1(b) suggests that a cluster solution may lie somewhere between 4 to 7 clusters for the LSS 2008 data. Finally, a solution of between 5 and 7 clusters is identified in figure 3.1(c) for the SARS data.



(a) LSS-2005



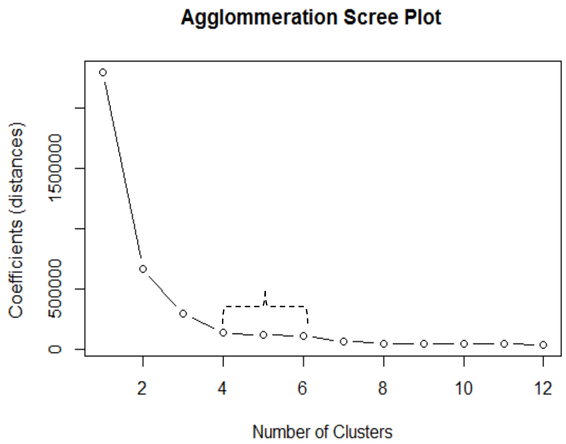
(b) LSS-2008



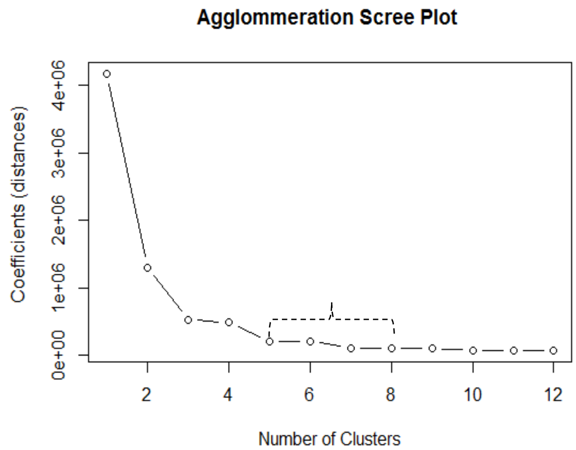
(c) SARS

Figure 3.1. Agnes dendrogram on LSS dataset. Panel (a), (b), and (c) report on LSS 2005, LSS 2008 and SARS data. The agglomerative coefficients are 0.92, 0.96 and 0.96 respectively.

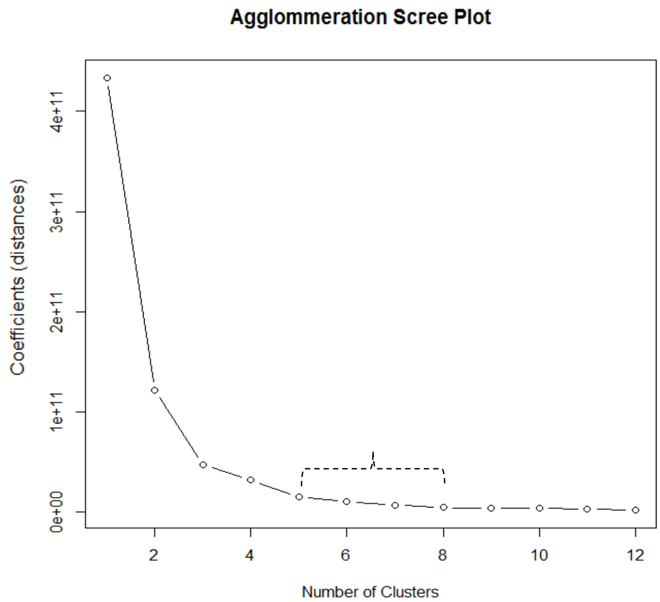
The agglomeration scree plot is another method used to advise the choice of cluster solutions (figure 3.2). Recall that the number of clusters is illustrated by a break (or "elbow") in the plot, or a flatting out of the curve, which indicates that an addition of more clusters after that point yields little additional information. For the LSS 2005 data this occurs between 4 and 6 clusters, while both the LSS 2008 and SARS data show a flatter curve between 5 and 8 clusters.



(a) LSS-2005



(b) LSS-2008



(c) SARS

Figure 3.2. Assessing the number of clusters with aggregation distances for the LSS 2005 (a), LSS 2008 (b), and SARS (c) data.

Both the dendrogram and agglomeration scree plot are somewhat subjective. Therefore, this study uses the NbClust[15] package available in R to extract the cluster solution based on the Calinski and Harabasz (1974) index. The NbClust package provides thirty indices for determining the number of clusters such as; Krzanowski and Lai (1988) (KL); Calinski and Harabasz (1974) (CH); Scott and Symons (1971)

[15] Charrad *et al* (2014) <available from: http://CRAN.R-project.org/package=NbClust>.

(SS); Milligan and Cooper (1985) (MC); and Friedman and Rubin (1967) (FR) to name just a few. However, as discussed in section 4, Calinski and Harabasz (1974) provide the preferred method. According to NbClust on the CH index the optimal number of clusters for LSS 2005, 2008 and SARS data are 6, 6, and 5 respectively. Given that these numbers fall within the ranges given by the dendrogram and agglomeration plot, for each dataset, it can be confirmed that (at least with some confidence) the optimal number of clusters in the three datasets are: 6 for the LSS 2005, 6 for the LSS 2008 and 5 for the SARS.

3.6.1.2.Characteristics of Clusters

The characteristics of the clusters for the LSS 2005 and 2008 datasets are presented in table 3.4. The 2005 (table 3.4a) dataset is clustered into 6 groups that can be classified as: small exporters (Cluster 1); medium/large non-exporters (Cluster 2); regular medium/large exporters (Cluster 3); high-performer medium/large exporters (Cluster 4); regular small non-exporters (Cluster 5); and high-performer small non-exporters (Cluster 6). In the following description and comparisons of the clusters, the high-performing outlier clusters 4 and 6 (who show significant[16] differences across most performance indicators relative to all other clusters) will initially be ignored in the description of cluster 1, 2, 3, and 5, and discussed individually. The idea of the cluster analysis is to identify comparable typologies among South African manufacturing firms. Outliers, while (arguably) important in their own right, can unnecessarily complicate the comparisons of the more 'regular' clusters. According to table 3.4a, the clusters can be described as follows:

Cluster 1. Small Exporters (4.7% of sample):

Relative to medium/large firms (regardless of export status), median small exporters exhibit significantly lower levels of labour productivity, capital- and intermediate input intensity and number of employees. Relative to the median small non-exporter, however, a median small exporter is significantly more labour productive, employs more workers and is more intermediate-input intensive. When it comes to the median level of total factor productivity, small exporting firms do not appear to differ from regular firms (regardless of size or export status). Interestingly, small exporters (as well as small non-exporters) exhibit a higher median wage per worker than the medium/large firm clusters, despite employing fewer workers. This may imply that small manufacturing firms are being forced by labour market issues to employ high skilled workers, who are potentially ill-suited to production (as indicated by the relatively lower levels of labour productivity).

[16] Refer to table C1 in Appendix C of this thesis for the p-value output from the Student t-test used to test for the significance of any differences in the variables.

Table 3.4. Median Values of Variables in each Cluster for LSS 2005 and 2008 Datasets

	Clusters	**Variables (Performance Indicators)**											
		LSS - 2005											
		med_large	exporter	lryl	ll	lrkl	lril	ln_TFP	lrwl	exp_sales_p	common_sic3	*Freq.*	*% sample*
Cluster 1	*Small Exporters*	0	1	5.59	3.14	3.28	5.21	0.02	1.12	6.48	356	*150*	*4.71*
Cluster 2	*Medium/large Non-exporters*	1	0	6.44	4.70	4.03	6.18	0.03	-0.34	0.00	338	*1506*	*47.24*
Cluster 3	*Regular Medium/Large Exporters*	1	1	6.34	4.83	4.18	6.06	0.02	-0.28	7.50	338	*734*	*23.02*
Cluster 4	*High-performing Medium/large Exporters ("super exporters")*	1	1	8.97	3.88	8.11	8.96	0.22	1.65	80.43	343	*8*	*0.25*
Cluster 5	*Regular Small Non-exporters*	0	0	5.36	2.82	3.25	5.03	-0.03	1.11	0.00	351	*785*	*24.62*
Cluster 6	*High-performing Small Non-exporters ("high technology firms")*	0	0	8.84	0.11	6.94	8.74	0.20	5.24	0.00	335	*5*	*0.16*
		LSS - 2008											
		med_large	exporter	lryl	ll	lrkl	lril	ln_TFPl	lrwl	exp_sales_p	common_sic3	*Freq.*	*% sample*
Cluster 1	*Small Exporters*	0	1	5.68	3.29	3.58	5.28	-0.02	1.13	4.41	354	*283*	*4.65*
Cluster 2	*Medium/large Non-exporters*	1	0	5.99	3.87	3.59	5.60	0.00	0.44	0.00	342	*2923*	*48.08*
Cluster 3	*Regular Medium/Large Exporter*	1	1	6.12	4.25	4.11	5.77	-0.01	0.31	4.44	351	*2862*	*47.07*
Cluster 4	*High-performing Medium/large Exporters ("super exporters")*	1	1	8.74	1.70	3.29	8.77	0.40	4.20	86.27	355	*8*	*0.13*
Cluster 5	*Small "super-exporters"*	1	0	0.90	7.40	-1.08	1.99	-1.78	-4.58	0.00	313	*2*	*0.03*
Cluster 6	*Low-performing Large Non-exporters*	0	1	8.99	0.86	7.01	9.02	-0.02	5.11	25.16	366	*2*	*0.03*

Cluster 2. Medium/large Non-exporters (47.24% of sample):

Relative to smaller firms, regardless of export status (Clusters 1 and 5), firms in this cluster produce significantly more output per worker, are significantly more capital- and intermediate input intensive, and pay lower wages despite employing more workers (in terms of median values). Firms in this cluster are further more productive in terms of TFP than small non-exporting firms (Cluster 5), although exhibit no significant productivity premium over smaller exporters (Cluster 1). Finally, the median medium/large non-exporter differs significantly from the median medium/large exporter only in terms of lower employment levels as well as higher levels of capital and intermediate inputs per workers.

Cluster 3. Regular Medium/Large Exporters (23.02% of sample):

As already mentioned, medium/large exporters exhibit higher (median) levels of labour productivity, number of workers, and capital- and intermediate input intensity, and lower wages than smaller exporting firms (Cluster 1). In addition, the median firm in this cluster shows significantly higher levels for all performance indicators relative to small, non-exporting firms (Cluster 5).

Cluster 4. High-performing Medium/large Exporters - "super exporters" (0.25% of sample):

These firms (who make up less than 1% of all exporters) export significantly more of their total output than other medium to large exporters (a median of 80% relative to 8%). These firms resemble the South African 'super exporters' as identified by the World Bank (2014) and Matthee *et al* (2016) who exhibit higher levels in terms of productivity (labour and TFP), capital and intermediate-input intensity, and wages. The number of workers employed appears to be lower among these firms relative to other medium/large firms, however the difference between groups is not significant.

Cluster 5. Regular Small Non-exporters (24.62% of sample):

Firms in this cluster have significantly lower median values across all performance variables, except for wages which, as seen before, are significantly higher relative to larger non-exporters as well as larger exporters. Comparing small non-exporters to small exporters, it is noted that the former are less labour productive, employ fewer workers and use less intermediate inputs per worker.

Cluster 6. High-performing Small Non-exporters - "high technology firms" (0.16% of sample):

Finally, small non-exporters are further split into a handful of high-technology, skill-intensive firms. Despite employing fewer workers than the other group of small non-exporters, firms in this cluster pay significantly higher wages. In fact, these firms exhibit the lowest median number of employees but highest wages on average than any other cluster. In addition, relative to other small non-exporters as well as small exporters, these firms are more capital and intermediate-input intensive, and are significantly more productive (both in terms of labour productivity and TFP). This suggests that these firms are using relatively advanced technology.

To summarise, the LSS 2005 dataset can been clustered into groups of medium to large exporters and medium to large non-exporters, as well as small non-exporters and small exporters. It can further be segmented into two groups of outliers: super exporters and high technology firms.

Similarly to the LSS 2005 data, the LSS 2008 dataset (refer now to table 3.4b) also identifies groups of: medium/large non-exporters (Cluster 2); medium/large exporters (Cluster 3); small exporters (Cluster 1); and the high-performing medium/large exporters ("super exporters" of Cluster 4). In addition, the 2008 data identifies a pair of small "super-exporters" (Cluster 5) who only differ significantly from the larger super-exporters in terms of lower TFP (indeed these small "super-exporters" are no more productive than firms in the other clusters). Further, a pair of low-performing large non-exporters are identified (Cluster 6). These firms employ the largest number of employees at the lowest wage relative to all other firms, exhibit the lowest capital- and intermediate input intensity, and are less labour and TFP productive than all other clusters. These firms can therefore be characterised as low technology non-exporting firms and appear to be in the textiles industry.

The 2008 dataset fails to identify small non-exporters (both regular and high-performing) as identified in LSS 2005. This may be due to the composition of the 2008 subset used in this study. Recall that, as seen in table 3.2, the 2008 subset of complete cases is made up of over 80[17] percent of medium/large firms. Further, over 50 percent of this subset is engaged in exporting. An inspection of the incomplete cases dropped from the subset in 2008 show that 58 percent were small non-exporters. This is something to take note of and something to address in future studies (i.e.: perhaps an extrapolation of the missing values/observations to prevent dropping cases). For now however, in this initial exploration of the data, the robustness of clusters will be determined by comparing the cluster results of the three data sets. So far both LSS datasets identify four consistent clusters: a group of medium/large non-exporters; medium/large exporters; small exporters; and a small group of super-exporters.

The analysis now turns to the examination of the 5 cluster solution identified in the SARS 2013 dataset. The cluster characteristics are presented in table 3.5. Similarly to both LSS datasets, the SARS dataset clusters firms into a group of small exporters (Clusters 1 and 2 in table 3.5), and large exporters (Cluster 3). However, the addition of destination variables gives a further dimension to the South African manufacturing firm environment. In particular, the SARS data can be clustered as follows:

Cluster 1. Small exporters exporting to a single destination within Africa (11.49%)

Relative to other exporters (i.e.: small exporters exporting to multiple destinations outside of Africa, Cluster 2), these firms exhibit significantly higher median values for labour productivity and intermediate inputs per worker, despite exporting significantly lower (median) values of total output. As we saw previously in the LSS clusters, relative to small non-exporters (Cluster 4), firms in this

[17] In the SARS population data manufacturing firms are made up of around 40 percent of med/large firms. The LSS 2008 subset of complete cases therefore over-represents larger firms.

cluster are more labour productive, employ more workers and are more intermediate input intensive. The median firm in this cluster is significantly less labour productive, employs significantly fewer workers, is less capital- and intermediate input intensive, and pays lower wages than the median medium/large exporter (Cluster 3).

Cluster 2. Small exporters exporting to a single destination outside of Africa (2.08% of sample)
Firms in this cluster not only exhibit lower median levels of labour productivity (and intermediate input intensity) relative to other small exporters (Cluster 1), but also relative to small non-exporters. Again, these small exporters have significantly lower median values relative to medium/large exporters across all characteristics (with the exception of TFP and wages).

Cluster 3. Medium/large exporters exporting to multiple destinations outside of Africa (39.22% of sample)
Firms in this cluster are characteristically similar to the LSS medium/large exporters in that they exhibit higher (median) levels of labour productivity, number of workers, and capital- and intermediate input intensity, than smaller exporting firms and significantly higher levels for all performance indicators relative to small, non-exporting firms.

Cluster 4. Small non-exporters (47.19% of total sample)
This cluster of firms exhibit similar characteristics to the small non-exporters identified in the LSS data.

Cluster 5. High-performing medium/large pair (0.02% of sample)
The final cluster identified is an outlier cluster, containing a pair of medium/large, high performance firms; one of which does not export, and the other who does (albeit a very small proportion of total sales).

Before summarising the findings of this section, reference is made to the findings of Chapter 2. Recall that in this chapter firms were (*a priori*) classified as either exporting to a single destination within Africa, a single destination outside of Africa, multiple destinations within Africa and multiple destinations outside of Africa. Out of interest, this analysis therefore cuts the SARS data into 6 clusters (which still falls within the range suggested by the dendrogram and agglomeration scree plot). These clusters and their characteristics are presented in table C1 in Appendix C.

Table 3.5. Median Values of Variables in each Cluster for SARS 2013 Dataset

	Variables (Performance Indicators)															
	SARS															
	med_large	exporter	lryl	ll	lrkl	lril	ln_TFP	lrwl	exp_sales_p	industry	SDIA	SDOA	MDIA	MDOA	*Freq.*	*%sample*
Cluster 1 *Small exporters exporting to a single destination within Africa*	0	1	13.50	2.75	10.62	12.99	-0.01	11.51	1.55	25111	1	0	0	0	1020	11.49
Cluster 2 *Small exporters exporting to a single destination outside of Africa*	0	1	13.24	2.63	10.53	12.63	0.01	11.66	2.38	25994	0	1	0	0	185	2.08
Cluster 3 *Medium/large exporters exporting to multiple destinations outside of Africa*	1	1	13.72	3.43	10.77	13.25	-0.01	11.72	5.89	25993	0	0	0	1	3482	39.22
Cluster 4. *Small non-exporters*	0	0	13.31	2.60	10.67	12.74	-0.04	11.45	0.00	25111	0	0	0	0	4190	47.19
Cluster 5. *High-performing medium/large firms*	1	0.5	19.48	2.41	16.77	19.01	1.16	16.70	0.64	10605.5	0.5	0	0	0	2	0.02

Notes: SDIA refers to "sing_dest_in_afr_2013"; SDIO refers to "sing_dest_out_afr_2013"; MDIA refers to "multi_dest_in_afr_2013"; and MDOA refers to "multi_dest_out_afr_2013" variables.

The clusters are as before, with the exception of a sixth cluster that is split from the medium/large exporter cluster (Cluster 3 in table 3.5). This sixth cluster can be labelled as “Medium/large exporters exporting to multiple destinations *within* Africa”. Relative to medium/large exporters who export to multiple destinations outside of Africa, these firms have lower median values across all performance variables (with the exception of TFP which does not differ). Similarly to Cluster 3, these firms exhibit higher values than small non-exporters over all characteristics (except TFP) as well as relative to small exporters (although only in terms of labour productivity, number of employees and intermediate-input intensity)[18].

In summary, all three datasets identify a group of small exporters and a group of large exporters. Both LSS datasets additionally identify a group of large non-exporters and a (small) group of “super exporters”. Further, both LSS 2005 and the SARS dataset identify a group of small non-exporters. Finally, with the added dimension of destinations, the SARS data is further classified into two groups of small single-destination exporters as well as two groups of medium/large multiple-destination exporters. These findings suggest that export status and firm size may be correlated.

In terms of the key performance characteristics (and consistently over the three datasets), medium/large exporters appear to show the highest levels relative to all other firms, particularly small firms, with the exception of TFP (and no labour productivity difference relative to medium/large non-exporters). In particular, it is the medium/large exporters exporting to multiple destinations outside of Africa that exhibit the highest levels of performance (at least in terms of labour productivity, number of employee and intermediate input intensity). Small non-exporters, on the other hand, show lower performance levels relative to all other firms. Finally, there is little evidence of productivity differentials in terms of TFP between clusters. Indeed it is mostly the outliers (the “super exporters” and “high performance” clusters) that indicate significantly higher TFP medians relative to all other clusters (although there is also evidence of higher TFP levels for all large firms relative to small non-exporters).

Overall this agglomerative hierarchical method for clustering South African manufacturing firms appears to differentiate firms based on their size as well and their export status (and then export destination). It is therefore of interest to identify which variables are actually important for classification of manufacturing firms in South Africa. This calls for the use of the second clustering methodology: ClustOfVar.

3.6.2. ClustOfVar Cluster Analysis

As discussed in section 4, the second stage in the analysis uses the ClustOfVar algorithm to cluster variables in order to examine which variables are important for distinguishing firms. The first step

[18] It is further noted that both the 5 and 6 cluster solutions in the SARS data do not identify a group of large non-exporters. As before this is likely due to the composition of the data subset and the incomplete cases.

involves identifying the groups (or clusters) of variables that carry the most similar information. Similarity is determined by the ClustOfVar algorithm, and clusters are chosen in as similar fashion as before (using the dendrogram and agglomeration plot) as well as by using a stability plot (an output of the ClustOfVar package). Once these groups of variables (synthetic variables) are identified, firms will again be clustered but now based on these new, synthetic variables. Finally, CART analysis will be used to identify which variables were important in the clustering of firms. The results of each step follows below.

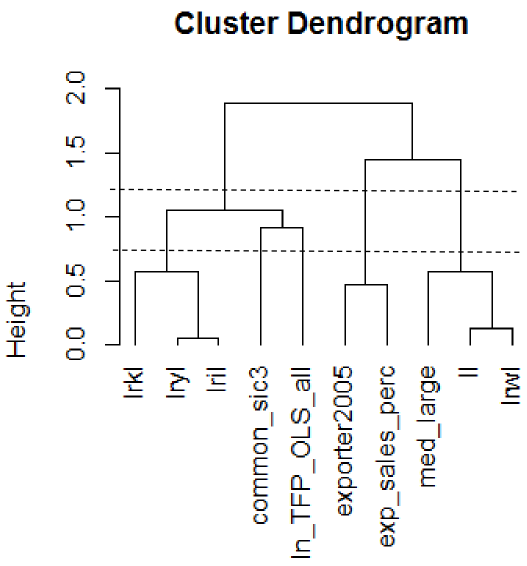
3.6.2.1.Clusters of Variables

The ClustOfVar dendrograms of the key performance variables for the LSS 2005, LSS 2008 and SARS datasets are shown in figure 3.3a, b, and c respectively. As before, using a horizontal line to cut the dendrograms at the (vertical) points where distance is relatively large, identifies potential cluster solutions for each dataset. Figure 3.3 thus indicates a cluster solution of between 3 and 5 for the LSS datasets and between 4 and 7 for the SARS dataset.

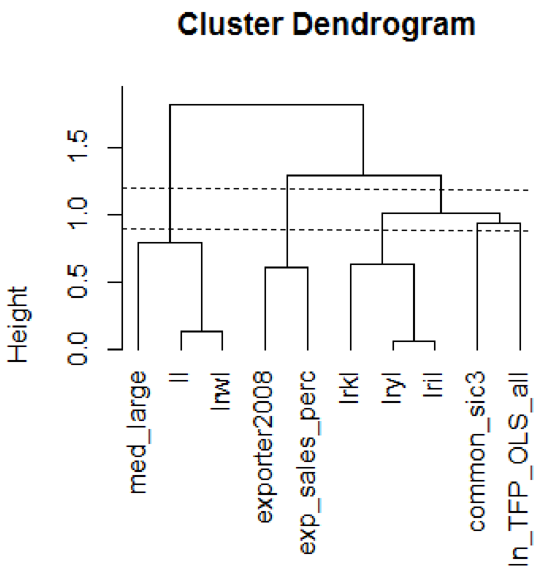
The agglomeration scree plots are shown in figure 3.4a, b and c for the respective datasets. The schedule appears to flatten out between 4 and 6 clusters for the LSS 2005 and SARS data. For the LSS 2008 data (figure 3.4b), the schedule suggests a cluster solution of either 4 or 5.

Finally, the ClustOfVar package in R does not provide a measure of the Calinski and Harabasz (1974) index, but rather outputs a stability plot which allows for identifying the number of clusters by analysing stability of the partitions via bootstrapped mean-adjusted Rand index (Chavent, 2012). These stability plots are given in figure 3.5 for the LSS 2005 and LSS 2008 data.

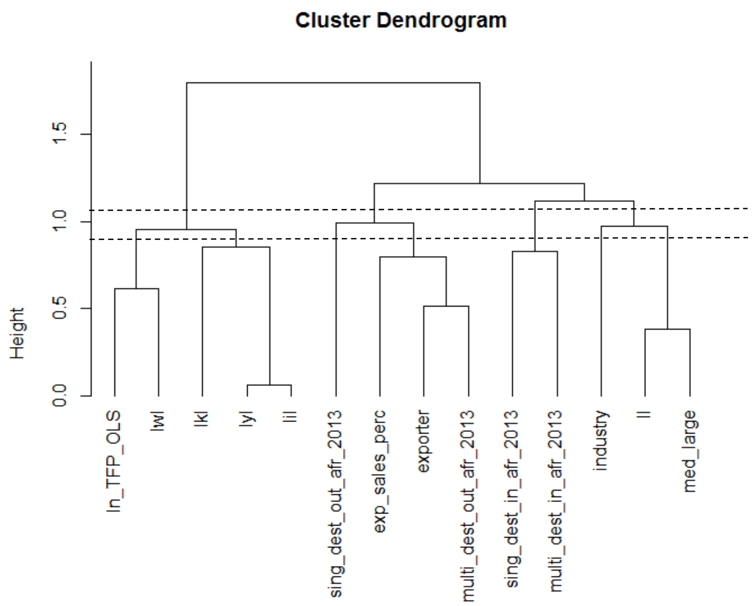
It is noted here that a stability plot for the SARS data is not presented since computing power and (terminal access) time limits the application of this coding to the SARS data. Figure 3.5 suggests that, at least for the LSS data, and adequate cluster solution lies between 4 and 6 clusters. There is no clearly objective choice of variable clusters, therefore both the 4 and 5 cluster solution will be presented since these seem to be consistent options in figures 3.3 to 3.5.



(a) LSS 2005

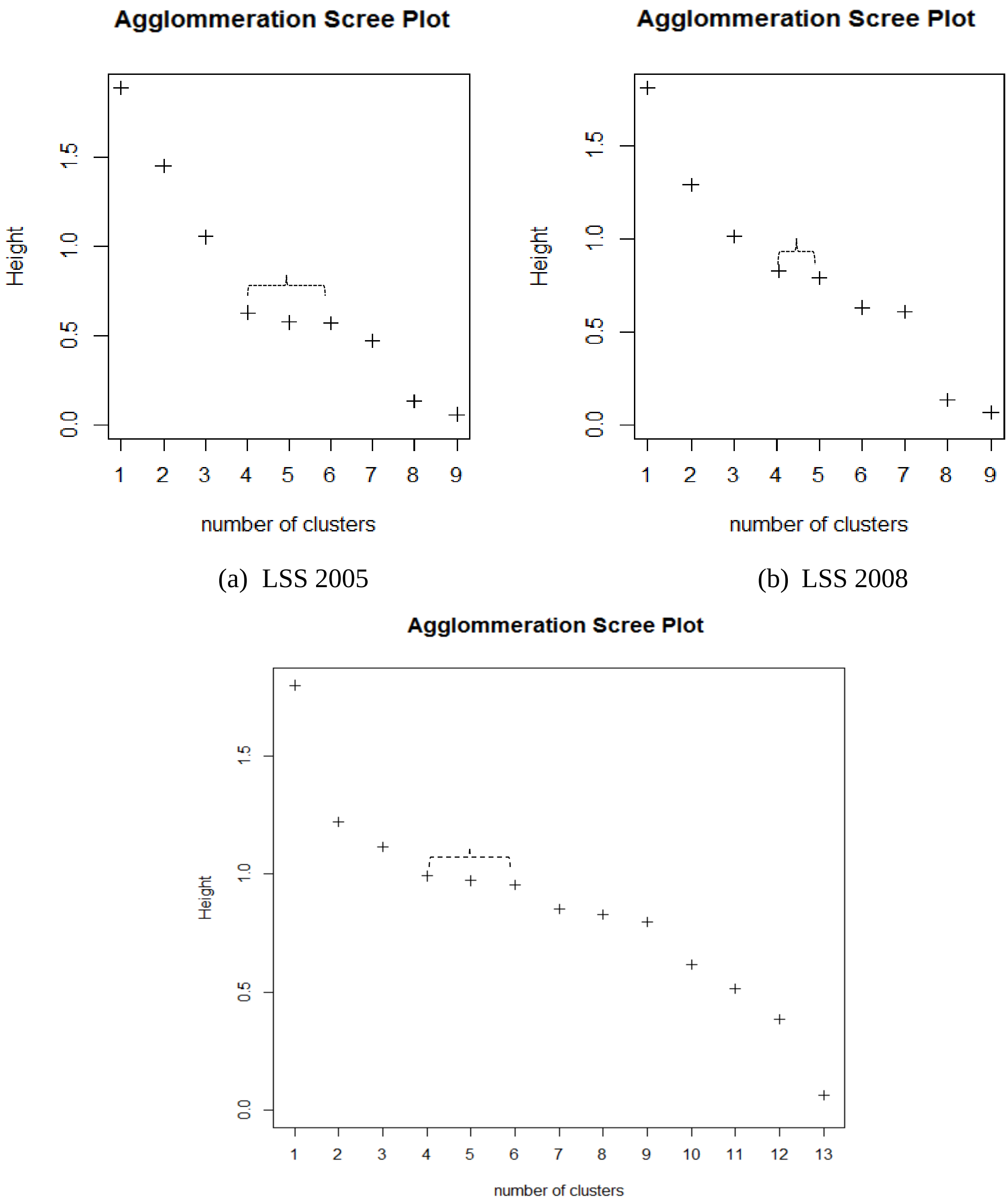


(b) LSS 2008



(c) SARS

Figure 3.3. ClustOfVar hierarchical cluster dendrogram on the variables of the three datasets

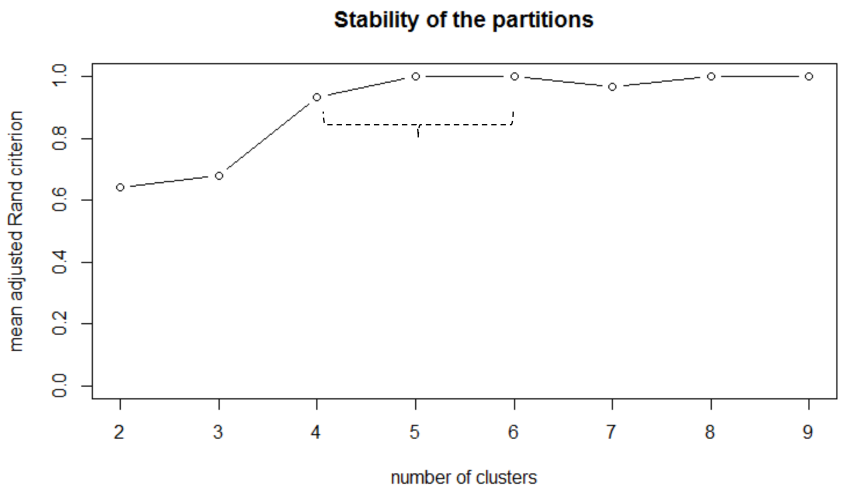


(a) LSS 2005

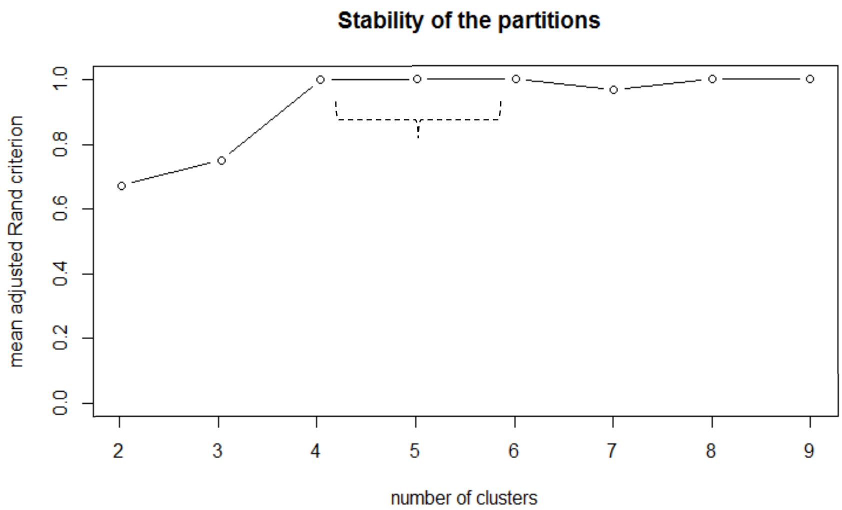
(b) LSS 2008

(c) SARS

Figure 3.4. ClustOfVar: Agglomerative scree plot for assessing the number of variable clusters in the LSS 2005, LSS 2008 and SARS datasets



(a) LSS 2005



(b) LSS 2008

Figure 3.5. ClustOfVar: assessing the number of variable clusters from stability of the partitions, mean adjusted Rand index from bootstrap.

Recall, ClustOfVar lumps together strongly related variables and separates them into clusters that can be scored as a single, synthetic variable for the purpose of dimension reduction and redundancy removal. Table 3.6 displays the composition of the clusters of variables ("cluster*n*") for both the 4 and 5 solution on the LSS data, together with the factor loadings (see table 3.7 for SARS data). In particular, it provides for each cluster the squared loadings with the central synthetic variable of the cluster (which is the first principal component of PCAMIX), which corresponds to the correlation ratio for qualitative variables, and the squared correlation ratio for quantitative variables, with the synthetic variable.

Table 3.6. ClustOfVar: (Synthetic) Variable Cluster Composition of LSS Datasets

LSS 2005				LSS 2008			
3 Cluster Solution		**4 Cluster Solution**		**3 Cluster Solution**		**4 Cluster Solution**	
Variable	**Squared Loading**	**Variable**	**Squared Loading**	**Variable**	**Squared Loading**	**Variable**	**Squared Loading**
Clust1		Clust1		Clust1		Clust1	
Exporting status		*Exporting status*		*Exporting status*		*Exporting status*	
exporter2005	0.76346810	exporter2005	0.76346810	exporter2008	0.69720310	exporter2005	0.69720310
exp_sales_perc	0.76346810	exp_sales_perc	0.76346810	exp_sales_perc	0.69720310	exp_sales_perc	0.69720310
Clust2		Clust2		Clust2		Clust2	
Production technology choice		*Production technology choice*		*Production technology choice*		*Production technology choice*	
lryl	0.93038980	lryl	0.90557540	lryl	0.92721195	lril	0.89942630
lril	0.86509021	lril	0.89973140	lril	0.86456828	lryl	0.89397580
lrkl	0.54994644	lrkl	0.56564010	lrkl	0.48117238	lrkl	0.50887090
ln_TFP_OLS_all	0.05492446			ln_TFP_OLS_all	0.06637754		
common_sic3	0.00446970	Clust3		common_sic3	0.00581154	Clust3	
		Employment and size				*Employment and size*	
Clust3		ll	0.92854970	Clust3		ll	0.91355810
Employment and size		lrwl	0.79183880	*Employment and size*		lrwl	0.82132290
ll	0.92854970	med_large	0.57076700	ll	0.91355810	med_large	0.34283310
lrwl	0.79183880			lrwl	0.82132290		
med_large	0.57076700	Clust4		med_large	0.34283310	Clust4	
		Industry productivity				*Industry productivity*	
		common_sic3	0.54391570			common_sic3	0.52872500
		ln_TFP_OLS_all	0.54391570			ln_TFP_OLS_all	0.52872500

A challenge with any type of principal component analysis is the interpretation of the synthetic variables. Examination of cluster1 suggests a variable that encompasses information on a firm's export strategy. This "export status" variable is consistently found in both the 3 and 4 cluster solution for both 2005 and 2008. A firm's production technology choice, factor inputs versus factor outputs, is represented in cluster2. Again, this synthetic variable is present in both 2005 and 2008 data, as well as in the 3 and 4 cluster solution. However, in the 4 cluster solution, total factor productivity and the industry variable are separated from the cluster of productivity variables into their own synthetic variable (with a larger factor loading), cluster3 (in the 4 cluster solution). The remaining 3 variables are then clustered together to form a variable that describes a firm's employment, and partly size, strategy (cluster3 in the 3 cluster solution and cluster4 in the 4 cluster solution). Based on this table, the analysis will continue using the 4 cluster solution (i.e.: the 10-variable datasets are reduced to 4 synthetics variables). The addition of the industry/TFP variable is interesting, and may add to the discussion without taking away from the general productivity variable.

Table 3.7. ClustOfVar: (Synthetic) Variable Cluster Composition of SARS Dataset

SARS 2013			
3 Cluster Solution		**4 Cluster Solution**	
Variable	**Squared Loading**	**Variable**	**Squared Loading**
Clust1		Clust1	
African exporters, by size		*Size*	
med_large	0.74659203	ll	0.79126582
ll	0.73988565	med_large	0.78144447
multi_dest_in_afr_2013	0.11396529	industry	0.06977505
sing_dest_in_afr_2013	0.04668393		
industry	0.05172673	Clust2	
		Out of Africa exporters	
Clust2		exporter	0.66518538
Out of Africa exporters		multi_dest_out_afr_2013	0.61962876
exporter	0.66518538	sing_dest_out_afr_2013	0.02144226
multi_dest_out_afr_2013	0.61962876	exp_sales_perc	0.39168642
sing_dest_out_afr_2013	0.02144226		
exp_sales_perc	0.39168642	Clust3	
		Production technology choice	
Clust3		lyl	0.91950270
Production technology choice		lil	0.74942140
lyl	0.91950270	lwl	0.53021760
lil	0.74942140	lkl	0.17648770
lwl	0.53021760	ln_TFP_OLS	0.13765990
lkl	0.17648770		
ln_TFP_OLS	0.13765990	Clust4	
		African exporters	
		sing_dest_in_afr_2013	0.58556000

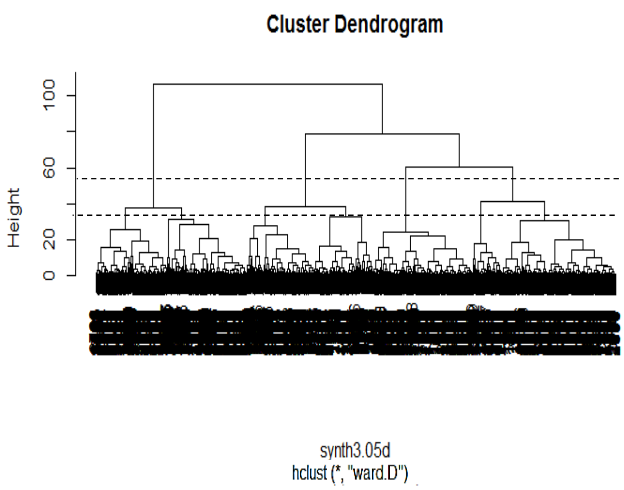
multi_dest_in_afr_2013	0.58556000

Given the added destination dimension, the SARS dataset is examined separately in table 3.7. The fourteen variables have been clustered as follows. In the 3 cluster solution, the three synthetic variables include: a variable which combines size and within Africa exporting strategies (cluster1); an out-of-Africa exporting variable (cluster2); and a productivity variable (cluster3). The 4 cluster solution provides a clearer classification of the variables. Cluster1 refers to a firm's size (both in terms of assets as well as number of employees). Exporting to destinations outside of Africa is the second synthetic variable (cluster2) while exporting within Africa forms a separate cluster (cluster4). Finally, production variables are once again clustered into their own synthetic variable (cluster3). Again the choice is made to retain the 4 cluster solution since the clustering is clearer, which will assist in making further analysis clearer.

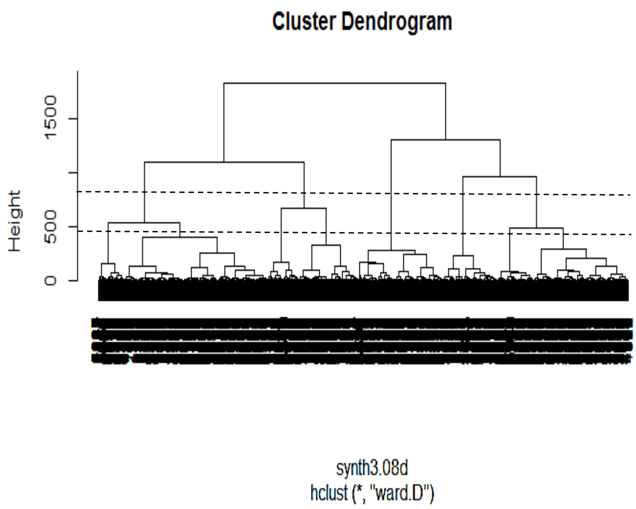
3.6.2.2.CART Analysis of Firms Based on Synthetic Variables

After identifying groups of variables, the second step in the analysis uses these new synthetic variables to segment the South African manufacturing firms. The principal components of the 4 synthetic variables identified in both the LSS datasets and SARS datasets were used to cluster the firms as illustrated in figure 3.6. As before, hierarchical clustering was used, but due now to the presence of the *numeric* synthetic variables, the analysis follows the example of Brida *et al* (2014) (among others, for example Saint-Arnaud & Bernard 2003) by using Ward's linkage technique. Based on these Principal components, an examination of the dendrograms and CH index suggests segmenting the firms into 6, 5 and 7 clusters for the LSS 2005, LSS 2008 and SARS datasets respectively.

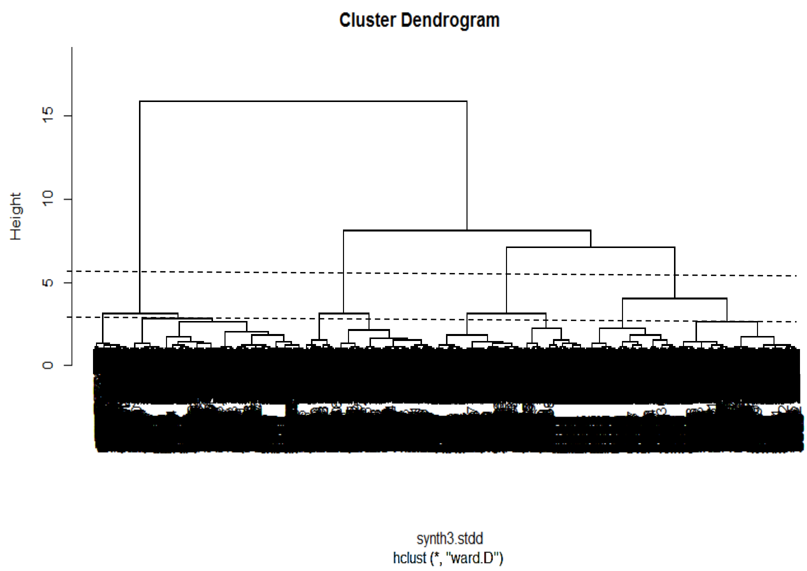
The third, and final, step employs CART analysis to assist in identifying the variables that significantly affect the partitioning of firms in the original datasets. The classification tree is run on both the raw variables (tables 3.2 and 3.3 respectively for LSS and SARS datasets) as well as on the 4 synthetic variables identified in the previous section.



(a) LSS 2005



(b) LSS 2008



(c) SARS

Figure 3.6. ClustOfVar: Hierarchical clustering of manufacturing firms on synthetic variables.

The CART analysis on the original and synthetic LSS 2005 dataset is presented in figure 3.7a and 3.7b respectively. Within the ten original variables, the classification tree identifies 6 important variables for grouping firms, namely: firm size ("med_large"), export status ("exporter"), amount exported ("exp_sale"), intermediate input intensity ("lril"), wages ("lrwl") and labour productivity ("lryl").

The first, and most important variable for classifying firms in the LSS 2005 dataset is firm size. For medium and large firms (the left branch), the amount exported is the second most important discriminating variable with those medium/large firms who export a very small proportion of output making up almost the entirety of cluster 1 (868 firms out of the 923 firms in cluster 1, or 94%) and those who export relatively more making up less than 10 percent of cluster 4. The complementary subset (right branch) is characterised by small firms. For these firms whether they export or not (regardless of the amount) is a further discriminating factor. Exporters are represented by the right hand side branch, which is split again by intermediate input intensity with majority of (small, exporting,) low intermediate input intensive firms making up a large proportion (72%) of cluster 4 and relatively high intermediate input intensive firms making up majority of cluster 6 (90%). Intermediate input intensity is also used to split up small non-exporters (left branch). Low intensity non-exporters are then categorised into low wage-paying firms (making up 72% of cluster 2) and higher wage-paying firms (who enter 20% of cluster 3). Small non-exporters, with higher levels of intermediate input intensity are also further discriminated in terms of wages, with those paying a relatively lower wage per worker being a large part of cluster 3 (62%). Those paying relatively higher wages are finally split in terms of their relative labour productivity where higher productivity firms represent a large proportion of cluster 5 (66%) and lower productivity firms representing only 4 percent of cluster 3.

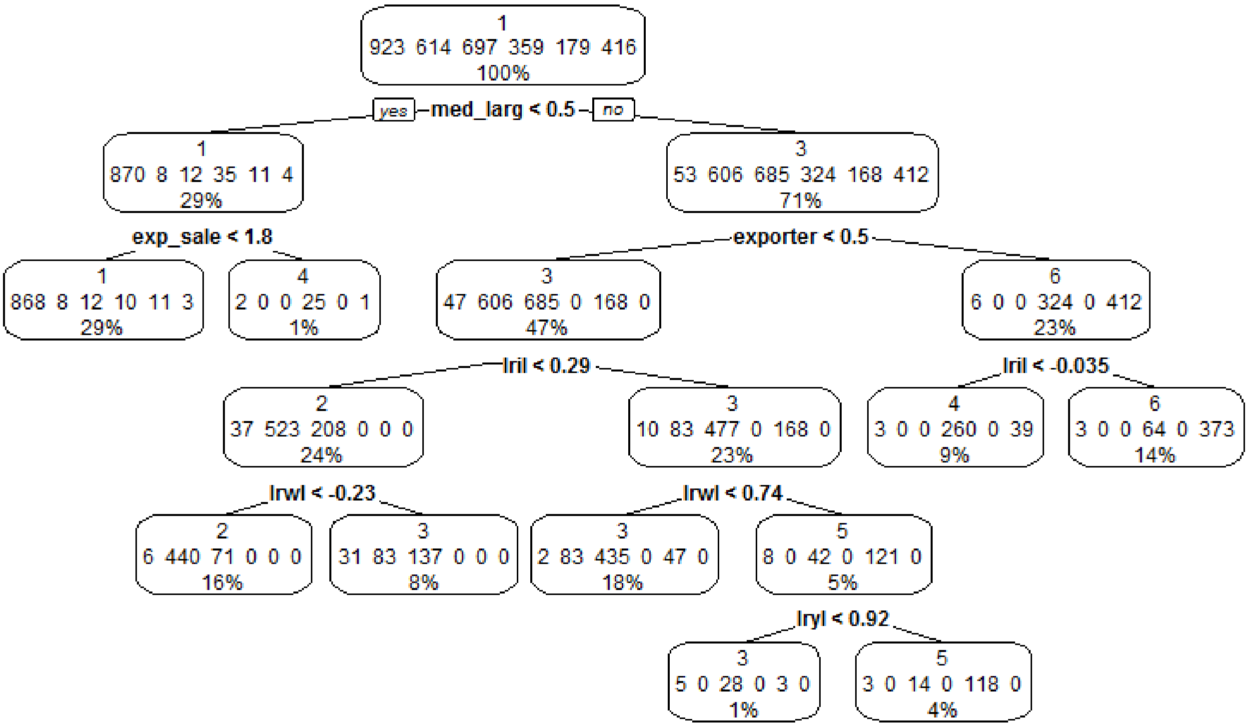


Figure 3.7a. CART: Decision tree on the raw (original) LSS 2005 variables. Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 6 clusters (based on principal components).

Interpreting the synthetic variables in figure 3.7b is slightly less clear since the latent dimensions of the variable clusters in table 3.6 (i.e.: "cluster*n*") are more complex than the single variables in table 3.2 and may thus represent more complicated concepts (Bridal *et al*, 2014). Recall, the four synthetic variables were identified as: export status (cluster1), production technology choice (cluster2), employment strategy (cluster3), and industry/TFP (cluster4).

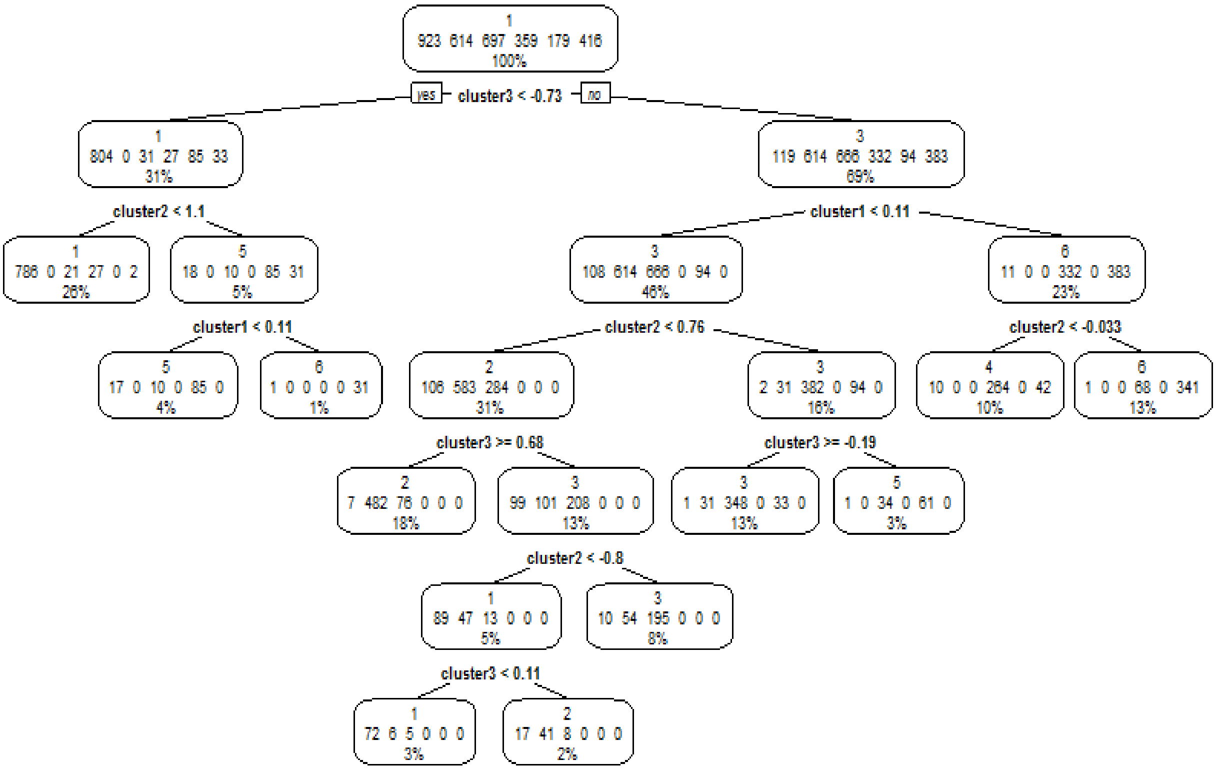


Figure 3.7b. CART: Decision tree on the synthetic LSS 2005 variables ("cluster*n*"). Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 6 clusters (based on principal components).

Figure 3.7b shows that of the 4 synthetic variables identified, a firm's employment (and size) strategy (cluster3) play the most important role in classifying firms. For those firms with lower values for this variable, the production technology choice plays an important distinguishing role where firms with lower capital and intermediate input intensity, as well as lower output produced per worker, enter into 85 percent of cluster 1. Firms with a higher value for the production strategy variable are then split by exporting status, where firms with lower values represent 47% of firm-cluster 5 and firms with higher values representing 7% of firm-cluster 6. Firms in the right branch as generated by variable cluster3, which are those with higher values employment and size strategy, are further divided by variable cluster1 which reports an exporting dimension. Exporters are further split according to their production choice into a group entering into 74 percent of firm-cluster 4 (lower production function) and another higher production function group making up 82 percent of firm-cluster 6. Non-exporters are also distinguished further by production strategies (cluster2) and then further by employment-size strategy (cluster3) where firms with higher values for both variables make up half of firm-cluster 3 and those with high production values but lower values for employment-size strategy enter into 34 percent of firm-cluster 5. On the other branch, lower production firms with high values for employment-size strategy form majority of cluster 2 (79%), while those with lower employment-size strategy variables make up 28 percent of cluster 3.

The CART analysis on LSS 2008 is given in figures 3.8a and 3.8b on the original and synthetic variables respectively. The variables firm size ("med_large"), export status ("exporter"), amount exported ("exp_sale"), intermediate input intensity ("lril"), wages ("lrwl") and labour productivity ("lryl") are also identified as important variables for clustering firms in the LSS 2008 dataset. In addition, industry ("common_s") is identified as important, but on a lesser level than the other variables.

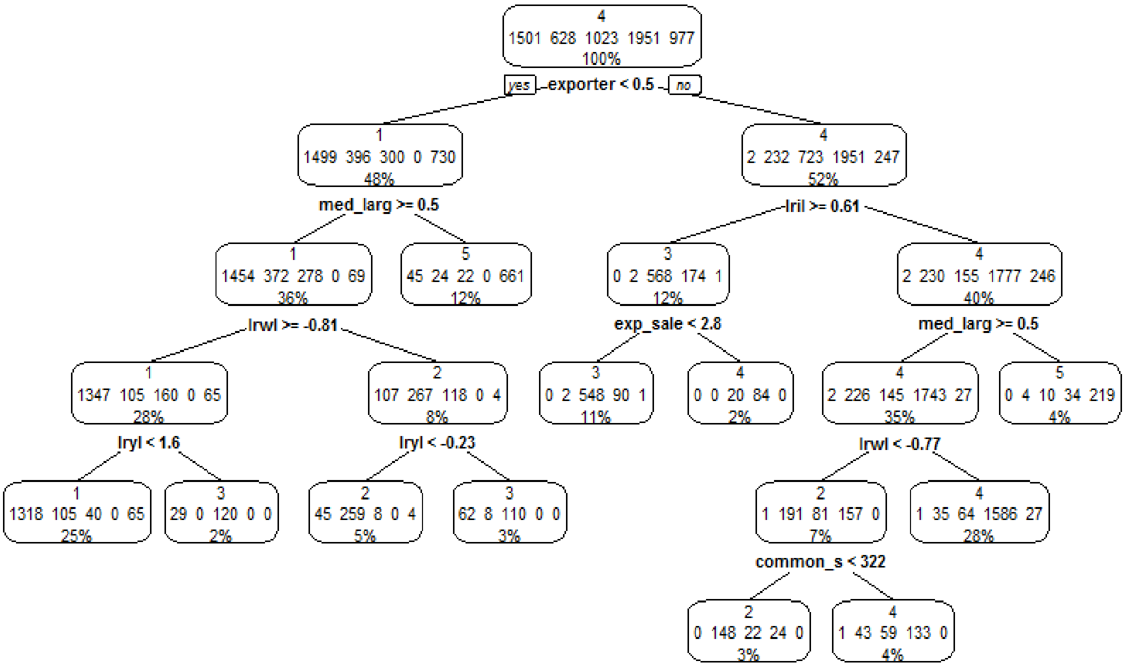


Figure 3.8a. CART: Decision tree on the raw (original) LSS 2008 variables. Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 5 clusters (based on principal components).

Whereas the LSS 2005 identifies firm size ("med_large") as the first and most important clustering variable, LSS 2008 first distinguishes firms on export status. Non-exporters are then split further by size, followed by wages and finally labour productivity. In particular, non-exporters who are relatively small make up 68 percent of firms in cluster 5. Larger non-exporters are split into higher-wage, lower-labour productivity firms forming majority of cluster 1 (88%) and lower-wage, lower-labour productivity firms entering 41 percent of cluster 2. On the right hand branch it can be seen that exporters are further distinguished by intermediate input intensity. Those with higher intensity are then classified by export intensity into a group of high export-intensive firms (making up only 4% of cluster 4 and 2% of cluster 3) and low export-intensive firms (entering 54% of cluster 3). Exporters with low-intensity can further be distinguished by firm size into a group of medium-large firms paying low wages (making up 30% of firms in cluster 2) and those paying higher wages (representing 81% of cluster 4).

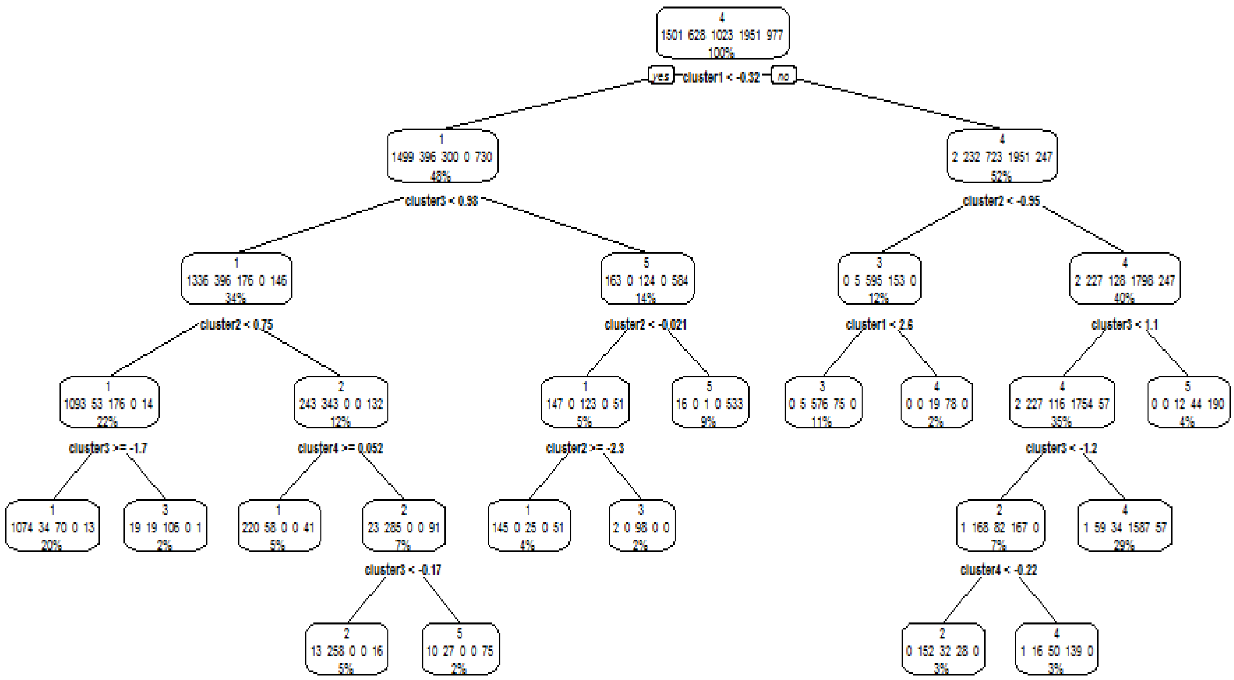


Figure 3.8b. CART: Decision tree on the synthetic LSS 2008 variables ("cluster*n*"). Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 5 clusters (based on principal components).

In terms of the synthetic variables in the LSS 2008 dataset, the export status variable (cluster1) plays the most important role in distinguishing firms. Non-exporters are then classified first by employment-size strategy and then by their production technology choice, and in some cases by the industry-TFP variable. Exporters are classified first by their choice of production technology followed by the employment-size strategy or, in some cases, export-intensity. In particular, 73 percent of firm-cluster 1 is classified as non-exporters with lower relative values on the employment-size variable, and lower production techonology. Non-exporters with lower relative values on the employment-size variable but higher levels of the production technology variable and higher levels of industry/TFP variable make up 45 percent of firm-cluster 2. Along another branch, non-exporters are classified into a group with higher employment-size strategy levels and higher production strategy levels (representing 55 percent of firm-cluster 5). Firm-cluster 3 and 4 represent the exporter clusters, with 56 percent of firms in the former cluster being classified firstly by their relatively lower production technology and secondly by their export intensity (relatively low). Finally, exporters with higher production technology values are distinguished by their lower employment-size variable representing 81 percent of firm-cluster 4.

Overall the CART analysis on both the original and synthetic variables on LSS 2005 and LSS 2008 identify the primarily important variables as firm size (both in terms of asset value and gross income, as well as number of employees) and export strategy, and to a lesser extent production technology and choice of inputs. This compliments the findings of the standard hierarchical method used in section 6.1, which also appears to classify firms in terms of size and export status.

The final step in this chapter regards the CART analysis on the SARS data (figures 3.9a and b). Similarly to the LSS datasets, SARS data is first distinguished by firm size. Labour productivity is the second most important distinguishing variable, while exporting to multiple destinations outside of Africa as well as export intensity are also highlighted as important variables.

Smaller firms are split by labour productivity where a group of lower labour productivity firms enter into 47 percent of cluster 3. Small, higher productivity firms are further classified in terms of whether they export to multiple destinations outside of Africa or not (including non-exporters). Those that do represent only 3 percent of cluster 1, 9 percent of cluster 3 and 26 percent of cluster 6. Firms that do not engage in exporting to multiple destinations outside of Africa make up 65 percent of cluster 1.

Larger firms in the right hand branch are also classified in terms of labour productivity. The higher productivity firms make up 84 percent of cluster 1 and 71 percent of cluster 5 (the later firms being more labour productive than the former). Firms with lower labour productivity are distinguished further by export intensity: lower export intensity firms making up 4 percent of cluster 1, 18 percent of cluster 3 and 12 percent of cluster 4, and higher export intensity firms representing 15 percent of cluster 2 and 11 percent in cluster 3. These higher export intensity firms are further distinguished from one another again by labour productivity.

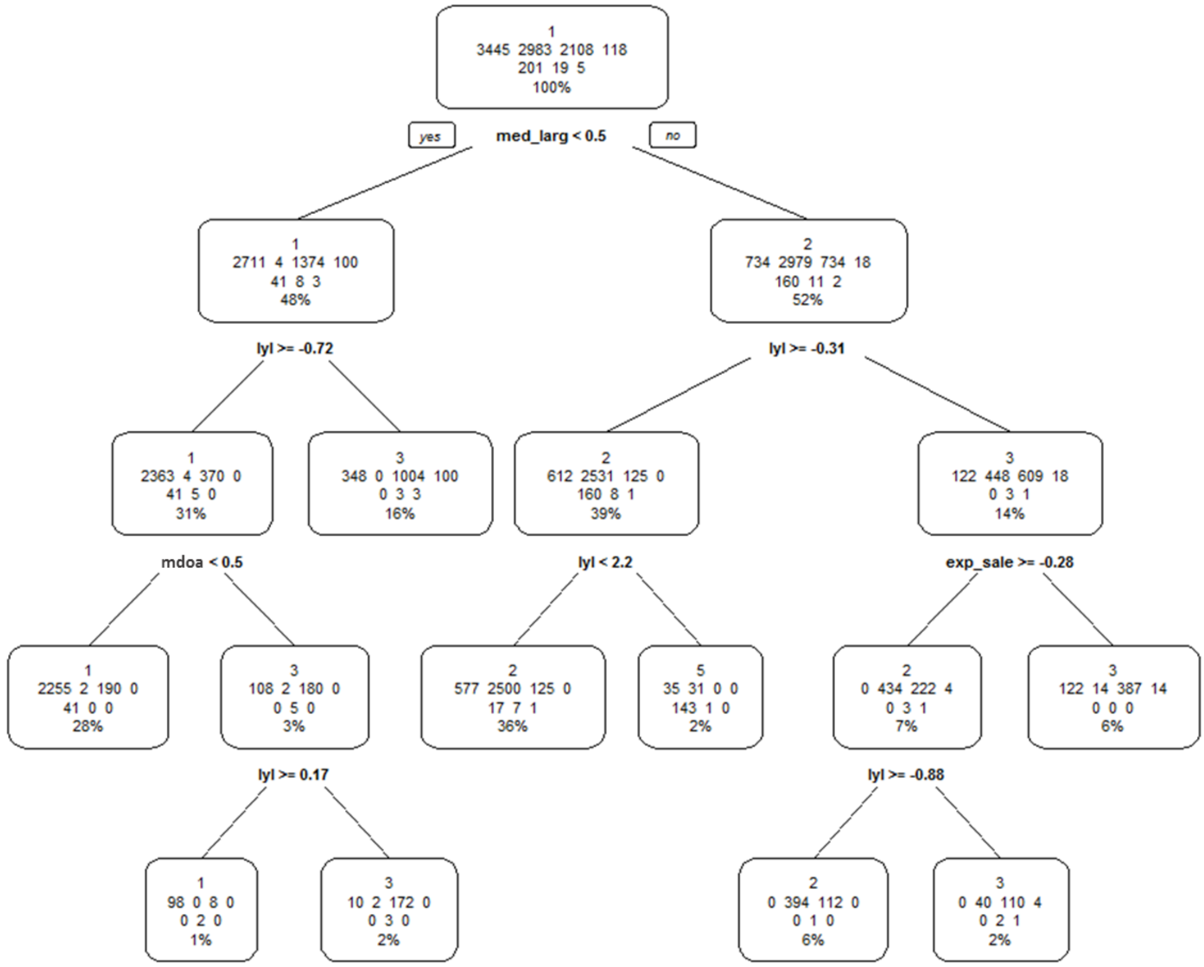


Figure 3.9a. CART: Decision tree on the raw (original) SARS variables. Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 7 clusters (based on principal components). Note: mdoa refers to the "multi_dest_out_afr_2013" variable.

Finally, figure 3.9b illustrates the classification tree for the synthetic variables identified previously, namely: firm size, in terms of assets as well as number of employees, (cluster1); exporting to destinations outside of Africa (cluster2); production technology choice (cluster3); and exporting within Africa (cluster4).

In figure 3.9b variable-cluster1, the highly related variable to firm size, is the most important discriminating variable. The firms with the lowest values of the variable, in the left branch, are further split by their production technology choice into a group of production technology firms and a group of lower production technology firms. The former entering into 47 percent of firm-cluster 3 and 73 percent of firm-cluster 5. The smaller, lower production technology firms are further distinguished by their decision to export outside of Africa. Those that do represent 6 percent of firm-cluster 1, 11 percent of firm-cluster 3 and firm-cluster 5, and 32 percent of firm-cluster 7 (with firms in firm-cluster 3 exhibiting higher levels of production technology than most small firms in firm-clusters 1 and 5). On the other branch, those firms that do not export outside of Africa (either exporters exporting only within Africa or non-exporters) are further classified again by production technology choice with the 44 percent entering firm-cluster 5 resembling those with higher production technology than the 71 percent entering firm-cluster 1.

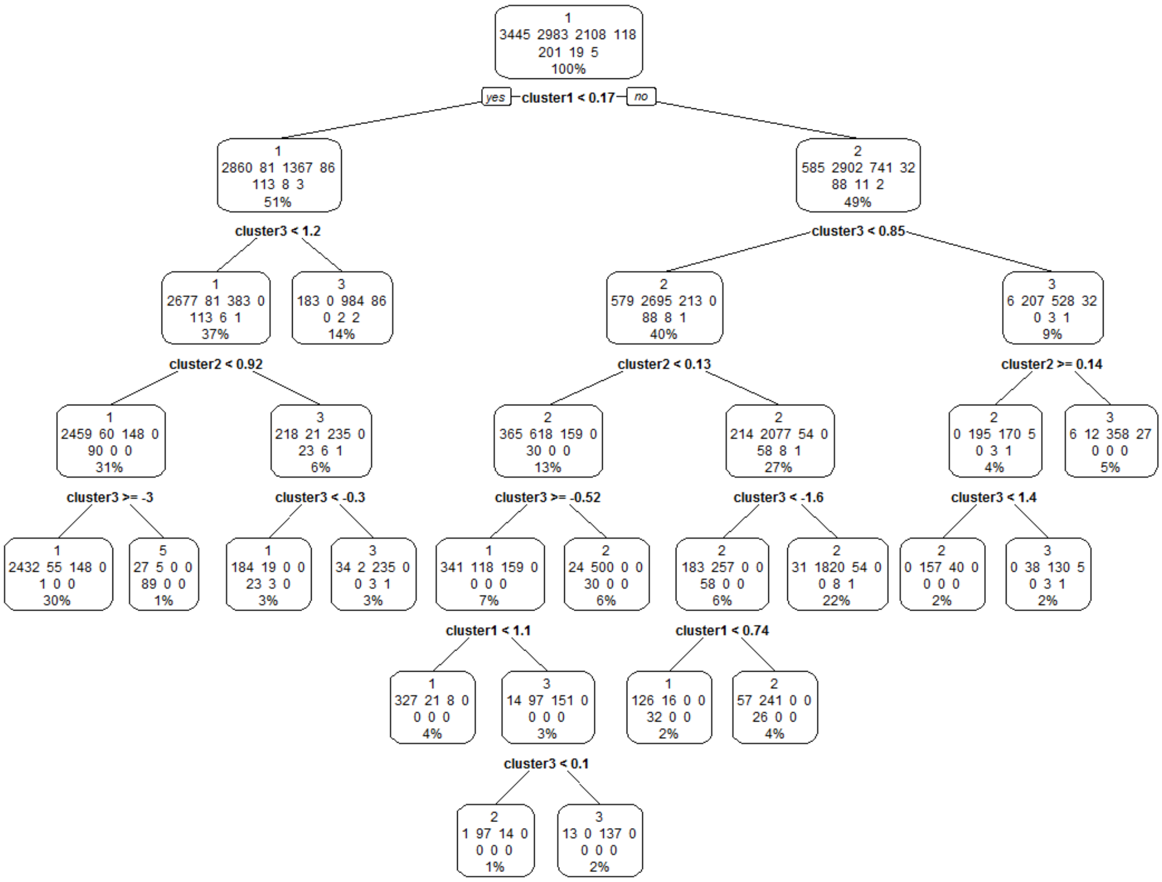


Figure 3.9b. CART: Decision tree on the synthetic SARS variables ("cluster*n*"). Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 7 clusters (based on principal components).

Firms in the right hand branch generated by variable-cluster1 (i.e.: larger firms), are also further divided by the synthetic production variable. Firms with higher levels of production technology are then classified into two groups according to whether the firm exports outside of Africa or not. Those that do not represent 17 percent of firm-cluster 3 and 23 percent of firm-cluster 4. Those that do export outside of Africa are divided again by production technology with lower level firms making up 7 percent of firm-cluster 2 and higher production technology firms 7 percent of firm-cluster 3.

The larger, lower production technology firms are also split in terms of the decision to export outside of Africa, followed again by the choice of production technology and then firm size. For example larger firms with lower levels of production technology exporting outside of Africa make up just over 61 percent of firm-cluster 2. These firms exhibiting higher levels of this synthetic production technology variable than those out-of-Africa exporters represented by 5 percent of firm-cluster1, 9 percent of firm-cluster 2 and almost 30 percent of firm-cluster 5. On the other branch, larger firms with lower levels of production technology *not* exporting outside of Africa are divided again into a group of lower production technology firms (making up 17% of firm-cluster 2) and higher production technology firms (entering into only 10% of cluster 1 and 8 % of cluster 3).

In brief, the SARS dataset with the additional destination dimension identifies firm size as the most important variable, followed by production technology choice and then export strategies (in particular the decision of whether to export outside of Africa). The LSS datasets also identified firm size and export status as important variables for classifying firms. This gives support to the categorisation of clusters identified in section 6.1, i.e.: South African manufacturing exporters can be classified into groups of small and medium/large non-exporters and exporters. Exporters can further be classified by their choice of destination. All groups of firms are further distinguishable in terms of their productive capacity.

3.7. Conclusions

Previous studies on South African manufacturing firms found that exporting firms differ to non-exporting firms across a number of performance indicators. These studies were largely based on small survey datasets. Chapter 2 of this thesis, by making use of substantial and official firm-level datasets, confirmed that South African manufacturing exporters and non-exporters do indeed differ from each other. Further, chapter 2 found that heterogeneity is also evident among exporters themselves, particularly in terms of firm size and the destination of exports.

These previous studies are examples of discriminant analysis, where the groups of firms (exporters, non-exporters, etc.) are decided *a priori*. This chapter takes advantage of the two substantial and official datasets on South African firms, one of which is population data, to perform cluster analysis. This technique, by allowing the data to speak for itself, can help identify optimal groups of firms *a posteriori*. By clustering observations based on a number of key performance attributes, this chapter therefore set out to describe manufacturing firms in South Africa in order to assess whether the classification of firms as determined by previous discriminant studies, in particular chapter 2, are in fact present in the data.

The first part of this study uses a standard hierarchical clustering technique to classify South African manufacturing firms according to a number of key performance indicators. The analysis grouped firms into small exporters, medium/large exporters, small non-exporters and medium/large non-exporters. In general terms, the analysis further found that the median exporter (particularly medium/large exporter) does exhibit significantly higher levels for a number of performance variables relative to the median non-exporting firm. These findings confirm the results of chapter 2 which found that exporters are larger, more capital-intensive and pay higher wages than non-exporters.

Chapter 2 further found that heterogeneity was evident among exporters themselves, in terms of export destination. Based on *a priori* assumptions, chapter 2 categorised exporters into four groups: firms that export to a single destination within Africa; firms exporting to a single destination outside of Africa; firms exporting to multiple destinations within Africa; and firms exporting to multiple destinations outside of Africa. The cluster analysis in this chapter also identifies these groupings of firms. In addition, it shows that medium/large exporters exporting to multiple destinations outside of Africa exhibit higher levels of employment, intermediate input intensity and labour productivity relative to all other groupings of firms (with the exception of handful of outlier clusters).

The cluster analysis further identifies a group of high-performing "super-exporters". This supports the relatively recent findings of Freund and Pierola (2012), Cebeci *et al* (2012) and Matthee *et al* (2016), who identify a handful of large exporters which dominate the export market. These "super-exporters" are found to be larger, more capital-intensive, and more productive than other exporters. The "super-exporters" identified in this chapter are also found to exhibit higher levels across all performance indicators, including TFP, relative to non-exporters and other exporters.

The second stage of this study adopted a three-step multivariate segmentation analysis which addressed the classic problem in reducing dimensionality by making use of ClustOfVar, a relatively recent algorithm that is able to handle mixed (qualitative-quantitative) datasets. The purpose of this stage was to identify the variables that significantly affected the classification of clusters as identified previously.

Firm size (both in terms of number of employees as well as total assets and income) was consistently identified as a main driver for cluster formation. To a lesser extent, export status and productivity played a role in distinguishing firms as well as the destination of exports. This again complements the findings of the previous South African studies on manufacturing traders as well as the classifications identified in the chapter.

Overall, the results of this chapter do confirm the results in previous chapters and literature, but they also add nuance. In this chapter, size is identified as the major factor (most important variable) for grouping firms, not sectors or choice of production inputs, etc.. Therefore, if policies are geared towards targeting specific groups of firms then size is the correct dimension to do this by. An explanation for the size-exporting link can be drawn from the Melitz (2003) model which predicts that exporters are larger than non-exporters and this difference exists prior to entry into exporting. There are at least two explanations for this, economies of scale and administrative burdens. Entry into exporting is accompanied by a number of fixed costs including market research, customs and trade permits, capital purchases, etc.. In the Melitz model therefore, a firm is able to export only if it is able to cover the fixed (sunk) costs associated with entering exporting. Due to economies of scale larger firms are better able to handle these costs since higher levels of production, and consequently employment, spread these fixed costs over the largest number of units thereby reducing their average unit costs (Greenaway & Kneller, 2007). In addition, exporting is administratively burdensome – a dedicated set of workers is needed to manage foreign clients, payments and tracking. Larger firms with a greater number of employees allows for specialisation, which increases efficiency and reduces the administrative burden associated with exporting (Rankin, 2011). Therefore the findings of this chapter, i.e.: size is an important determinant of firm clusters particularly for exporting, has important implications in that a firm needs to reach a certain size threshold before it has scale enough to become a successful exporter. The focus, therefore, should be on helping firms grow before helping them to access international markets.

After firm size, firms are segmented based on exporting activities and productivity. In general, this implies that if, for example, smaller firms are to become exporters they would either need to grow or become more productive. Finally, this chapter highlights that export destination, and number of destinations, matter. Exporting outside of Africa is generally different to exporting within Africa. This again confirms the previous chapter's findings, however the segmentation in this chapter suggests different processes by which exporters are distinguished. This in turn suggests that the process of converting a single-destination within-Africa exporter into a single-destination outside of Africa exporter will differ to that of previous studies.

The aim of this exploratory analysis is not to present casual mechanisms or identify predictive models but rather to illustrate the usefulness of techniques such as cluster analysis for compacting information from an entire population into information about specific groups. According to Escaith and Gaudin (2014) these techniques are helpful in organising empirical knowledge by identifying underlying structures within the data as well as stylised facts. Indeed, once groups have been identified through clustering, predictive discriminant techniques can be used to further analyse the behaviours of the objects in each group relative to those in other groups.

This study is not only one of the first in South Africa to make use of Stats SA's official large sample survey data, as well as the SARS population tax administration data to study the firm-level environment, but further it is one of the first to make use of cluster analysis in the international trade literature to group exporters and non-exporters *a posteriori* rather than *a priori*. Further it introduces a relatively new multivariate clustering algorithm for mixed datasets to the South African firm-level literature. As the availability of large, mixed variable datasets grows, this technique may prove useful in assisting analysts in investigating the underlying structure of the datasets, reducing dimensions, and organising or testing hypotheses without having to impose any *a prior* assumptions on the data.

A limitation of this study is given by the data. Despite the obvious advantages of using substantial, official datasets to perform the analysis, a large number of observations on key performance variables were missing for certain firms. It would therefore be of interest to estimate values for these missing cases in order to prevent dropping them from the analysis. The clustering could then be run again to determine whether or not the findings still hold. Nevertheless, the analysis in this chapter uses three datasets to test the robustness of the clusters. Further research should make use of the techniques described in this chapter on additional datasets in order to better describe the manufacturing environment and profile of exporters. This will ultimately help in the development of more targeted policies.

Chapter 4

Tax Incentives on Small Businesses: The unintended consequences for exporters

4.1. Introduction

Taxation is primarily a means to fund public spending, but it can also be a policy instrument to influence the behaviour of companies. Corporate tax policy has been used by governments over the decades to achieve various policy objectives including job creation, increased (or decreased) investment, export stimulation and the facilitation of new and small businesses (Bird, 1996).

The goals of job creation, increased investment and economic growth are high on the agenda of South African policy makers (National Planning Commission, 2013). One way to achieve these goals is through increased exports. According to the New Growth Path (2011), increasing exports has significant potential to stimulate investment, productivity, employment and income and it is for this reason that export stimulation is one of the key aims of South Africa's National Development Plan (NDP). However, despite the emphasis on exports in government's growth strategies, these strategies provide limited details on how this will happen. Part of the reason for this is that very little is known about export dynamics at a micro-level in South Africa (Rankin, 2013). A better understanding of the characteristics and behaviour of exporting firms can help inform policies designed to increase exports.

Exporting firms are not the only facilitators of employment and economic growth. Small and medium enterprises (SMEs) also play a recognisable role in the economy in terms of job creation and economic growth (Gumede, 2000; FIAS, 2007; SBP, 2013). The NDP argues that 90 percent of jobs will be created in small and medium firms. However, many barriers to growth exist for SMEs, including limited access to international markets, abundant and cumbersome rules and regulations and, importantly, the lack of access to debt and equity finance (Abor & Quartey, 2010). These barriers can significantly limit an SME's ability to increase investment and consequently its ability to expand. The NDP furthers states that much of this small firm employment creation will come about through small exporting firms. Therefore, policies geared towards employment creation and overall economic growth should include the facilitation and support of not only the small business sector in general, but small exporters too.

Tax incentives are commonly used to stimulate investment and support small business (see for example Hasset & Hubbard, 2002; Van Parys & James 2010; Booysens 1997; and Guenther, 2012). However, empirical evidence on the impact of these tax incentives is very scarce, particularly for developing countries (Klemm & Van Parys, 2009). In addition, not much attention has been paid to the level of heterogeneity among small firms. The previous chapters of this thesis highlighted this heterogeneity, not only among firms in general, but also among small firms. Indeed Chapter 3, for example, identified at least two groups within the small firms: exporters and non-exporters. The work in this chapter adds to the literature on tax incentives for small businesses by examining the effects of a small business tax incentive for small South African manufacturing firms in general, and small exporters in particular by making use of a dynamic firm-level dataset from StatsSA.

This tax incentive was passed into the tax legislature in 2001. Recognising the critical role played by SMEs in the growth of the economy, the National Treasury along with the South African Revenue Service (SARS) introduced Section 12E of the Income Tax Act No. 58 of 1962: a policy aimed at fostering employment, stimulating capital investment and promoting growth in the small business sector. Section 12E allowed for lower and progressive tax rates as well as accelerated depreciation for firms that qualified as small business corporations (SBCs).

The requirements necessary to qualify for the tax incentive have changed over the years. This provides a natural experiment in which research can be done to examine the changes in behaviour of firms that did not qualify for the tax incentive in 2005, but subsequently did in 2006, relative to firms that never qualified. By making use of a firm level panel data set for 2005 and 2008, this chapter attempts to investigate whether the tax incentive was successful in changing the investment and employment behaviour of small businesses as it intended.

In addition, survey evidence from a number of countries shows that exporters are likely to change their behaviour in response to tax incentives geared towards investment (James, 2009). Given the importance of exporters to employment, investment and economic growth, this chapter also examines if there were any unintended impacts on the behaviour of (small) exporters as a result of section 12E.

The rest of the chapter is organised as follows: the next section discusses a number of stylised facts on exporting. Section 4.3 follows with a discussion on the impact of tax incentives on small firms. Section 4.4 presents the data and a few descriptive statistics on South African exporters and small business corporations. The methodological approach is discussed in Section 4.5. Section 4.6 presents and discusses the empirical results and section 4.7 concludes.

4.2. Stylised Facts about Exporting

Exporting is good for the growth and development of the economy (Bernard & Jensen, 2004). In a study of 13 high-growth countries, the Commission on Growth and Development state that a key lesson to take away from these countries is their ability to successfully take advantage of the global economy (The Growth Report, 2008). The report argues that exploiting international markets is one of the biggest tools available for poverty reduction available. Given this, governments should advocate policies that support and foster exporting and exporters (Rankin, 2013). In order to do so, policy makers require information about the underlying characteristics of exporting firms, particularly at the firm-level.

The literature on exporting at the firm-level has produced several well-known ‘stylised facts’. In the vast majority of empirical studies it has been shown that exporters exhibit superior performance characteristics to non-exporters: they are larger, pay higher wages and are more productive and capital- and skill-intensive than non-exporting firms (Bernard, Jensen, Redding, & Schott, 2007). It has also been found that exporting is rare, and survival rates for the export market are low, particularly among small firms (Bernard *et al* 2007; Mayer & Ottaviano, 2008; Freund & Pierola, 2012). In what follows, the themes of exporting and productivity; exporting and size; and export survival will be discussed in more detail.

4.2.1. Exporting and Productivity

In order for firms to participate in the international markets, they need to be globally competitive. This requires them to be able to overcome certain productivity hurdles. The relationship between exporting and productivity has been explored by many authors commencing with the seminal contribution of Bernard and Jensen (1995, 1999) for U.S firms. Wagner (2007, 2011) provides an extensive survey of the recent literature on this topic.

Two hypotheses are present in the literature to explain why exporters differ to non-exporters in terms of productivity. The first hypothesis is that of self-selection: the Melitz (2003) model predicts that exporting firms are more productive than non-exporting firms even prior to entry into the export market. This is because there are fixed costs involved in entering the export market. These costs include transport costs, costs of developing relationships with foreign customers as well as production costs involved in modifying the domestic product for the foreign market. These costs provide a barrier to entry that less successful firms cannot overcome. Since competitive pressures in international markets are more severe than domestic ones, firms must improve their productivity prior to entry in order to survive.

The second hypothesis is that of learning-by-exporting: it is the entry of firms into the export market that raises their productivity. Participation in the export markets results in knowledge- and technology-spillovers from international competitors and buyers. In addition, competitive pressure in foreign markets encourages increased innovation among exporters relative to non-exporters (Clerides, Lach, & Tybout, 1998).

Although not mutually exclusive, empirical evidence overwhelmingly favours the self-selection hypothesis over learning-by-exporting (Singh, 2010). Bernard and Jensen (1999) empirically show that U.S manufacturing firms have many of the desirable performance characteristics several years before they actually enter the export market: exporters are 20% - 45% larger and 7% - 8% more labour productive than non-exporters before entry.

Self-selection is also evident in other countries. Bernard and Wagner (1997) found that German manufacturing firms which export already have many of the superior characteristics, such as size and productivity advantage, several years prior to exporting, and that these characteristics are accentuated in the run-up to exporting. For Columbia, Mexico and Morocco it was found that the positive export-productivity relationship is as a result of the self-selection of relatively more efficient plants into exporting (Clerides, Lach, & Tybout, 1998). Further support for this hypothesis has been found for Italy (Castellani, 2002; Serti & Tomasi, 2009), Belgium (Pisu, 2008), the Netherlands (Kox & Rojas-Ramagosa, 2010) and Portugal (Silva, Alfonso & Africano, 2010).

In contrast, these authors find little evidence of learning-by-exporting: ‘firms entering the export market are unlikely to substantially raise their productivity, even if they export continuously’ (Bernard & Jensen, 1999, p. 24). Bernard and Wagner (1997) and Clerides *et al* (1998) agree with the U.S studies. German plants showed no improvement after the start of exporting and evidence from Columbia and Mexico found that productivity improvements generally did not continue after entry into the foreign markets (Clerides *et al*, 1998).

There are, however, some authors who do find evidence of learning in some economies, particularly in developing ones. A possible explanation for this is given by Van Biesebroeck (2005) who believes that exporting firms in developing countries are relatively well positioned to absorb foreign knowledge. Blalock and Gertler (2004, p.398) support this and give a suggestion as to why their results differ from the studies discussed previously, ‘Whereas our work and that of Van Biesebroeck examine very poor countries, Indonesia and sub-Saharan Africa respectively, the majority of earlier work considers more developed economies... firms in the poorest countries may have much more to gain from exposure to international export markets’. Salomon and Shaver (2005) further find *ex post* benefits to exporting, measured by product innovation, and conclude that learning-by-exporting is present among Spanish exporters.

A possible explanation for why learning-by-exporting effects may not be evident is given by Westphal (2002) who suggests that it may be the case that technology spill-overs pass through from exporters to non-exports. If non-exporting firms benefit from the new knowledge of their exporting counterparts, they will likely follow similar productivity trajectories to these exporters and thus exporting benefits all domestic firms, not only those that participate in the export market.

4.2.2. Exporting and Firm Size

A second prediction of the Melitz (2003) framework is that exporters are bigger than non-exporters. Again, this difference exists prior to entry, since not only is it necessary for these firms to expand production for the international market, but also because they are more productive and therefore hold a higher market share even before they start exporting.

A number of authors have confirmed that exporting is positively correlated with size. For German manufacturing firms it was found that 84 percent of all large firms exported a portion of their output compared to only 30 percent of small firms (Wagner, 1995), while U.S. manufacturing exporters were found to be four times larger in terms of employment than non-exporters (Bernard & Jensen, 1995). Even within size categories, exporters are between 30 and 50 percent larger than non-exporters (Bernard & Wagner, 1997).

The positive correlation between exporting and firm size is not unique to developed economies. A number of studies in African economies have also found that firm size is robustly associated with export participation (Bigsten *et al*, 2004; van Biesebroeck, 2005; Rankin, Soderblom & Teal 2006; Edwards *et al*, 2008).

There are a number of possible explanations for why firm size and exporting are related. Firstly, exporting firms are more capital intensive than non-exporters (Bernard, Jensen, Redding, & Schott, 2007). Capital purchases are ‘lumpy’ and larger firms are better able to bear the cost of this investment. This is due to economies of scale: higher levels of production, and consequently employment, spread the fixed costs over a greater number of units and thus larger firms are better able to reduce their average unit costs (Greenaway & Kneller, 2007; Rankin 2013).

Exporting is not only associated with investment in capital. Entry into the export market is accompanied by fixed costs. These include costs associated with changing the products to meet international standards, market research, customs and trade permits and so on. Therefore, to keep unit costs at a minimum, firms need to be larger in order to spread these fixed costs across the greatest number of units (Wagner, 1995; Rankin, 2013).

A second explanation for the size-export relationship relates to the administrative burden of exporting. Exporting requires skilled individuals to connect with and manage foreign relationships and sales. Larger firms are able to fully utilise their employees and their skills by assigning them to such specialised roles. In addition, various departments such as advertising, sales, tracking etc. are more likely to exist in firms with a large number of employees (Bernard & Wagner, 1997; Westhead, Wright & Ucbasaran, 2001; Rankin, 2013).

In addition to economies of scale and administrative easing, the link between firm size and exporting may be explained by the size of the domestic market. In cases where the domestic demand is limited, exporting can provide firms with the opportunity to expand into international markets and generate additional profits (Edwards *et al*, 2008; Rankin 2013).

4.2.3. Export Survival

The final stylised fact emerging from the current literature relates to the low survival rates of firms in the export market. The export market is characterised by a large degree of experimentation. Alvarez and Lopez (2008), for example, find high rates of entry into and exit out of exporting for Chilean firms, as do Eaton, Eslava, Kugler and Tybout (2007) for Columbia and Freund and Pierola (2010) for Peru. This experimentation has been found to be associated with low rates of export survival (Cadot *et al*, 2013). Research has found that single year exporters are relatively common (Eaton *et al*, 2007; Besedes & Blyde, 2010; Cadot *et al*, 2013) and that these short lived export endeavours are not necessarily limited to developing economies. In a study of 47 developed and developing countries, Besedes and Prusa (2007) show that the median manufacturing exporter survives for only one and, in some cases, two years.

This common finding of low survival rates has encouraged researchers to examine what characterises survivors (see for example Besedes & Blyde, 2010; Cadot *et al*, 2013; Wagner, 2011; and Rankin, 2013). Export survival can be explained by a number of country- and industry- level factors including geography, exchange rates, transport costs, the local business environment, and other factors that may affect the level of competitiveness in international markets. At the firm-level, productivity and size seem to play a large role in the survival of exporters. In general 'weaker firms', that is firms that are smaller and less productive, are less likely to survive in the export market (Greenaway & Kneller, 2007, p. 150). Indeed, many authors have found that firms that exit exporting are small, particularly if they only export for one year (Eaton *et al*, 2007; Martincus & Carballo, 2010; Rankin, 2013).

One possible explanation for the correlation between firm size and export survival relates to the costs of exporting. As previously discussed, there are substantial sunk costs associated with entry into the export market. These costs are most pertinent for small exporters (Greenaway & Kneller, 2007). In light of this, weaker firms are likely to experiment with small trials into exporting in order to avoid high entry costs (Freund & Pierola, 2010), since the costs associated with 'testing the waters' are considerably less than those associated with significant exporting indentures (Eaton et al, 2007, p.17).

However, in most cases the sunk costs associated with exporting are substantial and in such cases high exit rates (and low survival rates) become inefficient (Cadot *et al*, 2013). Bernard and Wagner (1997) find that the consequences to firms of exiting export markets are disastrous: firms that stop participating in exporting experience negative growth in all performance measures including productivity and employment. These findings are supported by Rankin (2013) who finds that exiting firms decrease size and output per employee by around 60 and 45 percent respectively on exit.

Conversely, firms that do survive experience positive outcomes. Eaton *et al* (2007) find that in Columbia, new exporters who survive the first year of exporting are likely to continue to survive and experience rapid growth in the years that follow. This is also true of Latin American firms whose risk of failing decreases the longer they stay in the export market (Besedes & Blyde, 2010), as well as a number of African firms (namely Malawi, Mali, Senegal and Tanzania) who, not only are significantly less likely to fail if they survive the first year of exporting, but who also experience sizable growth in transaction volumes (Cadot *et al*, 2013).

In addition, although new exporters initially contribute very little to total exporters, some evidence suggests that if these new exporters survive the first year of exporting, their contributions to total exports will be substantial. Eaton *et al* (2007) find that new exporters together contributed close to 50 percent of the total expansion of trade in Columbia over a ten year period.

From a policy perspective, given that entry into exporting is associated with successful firms and given that the risk of failure is highest in the first year of exporting, but significantly reduces thereafter, policies aimed at supporting export expansion should focus not on helping firms to enter the export market, but rather on facilitating survival in the market and should be implemented during the early stages of exporting when the likelihood of survival is low (Rankin, 2013; Besedes & Blyde, 2010).

In summary, the literature suggests that certain productivity and size thresholds need to be overcome in order for firms to enter the export market. Size and productivity are also related to the continued success of these firms in the export market. Larger, more productive firms are more likely to succeed as exporters than smaller firms. However, the important role of small firms in the economy is well known (FIAS, 2007; SBP, 2013), and the relationship between exporting and the survival of small firms is becoming more apparent. According to Westhead *et al* (2001, p.334), the ability to participate in the export market is a 'necessary ingredient to ensure the survival and growth of new and small firms'. This is supported by Abor and Quartey (2010) who maintain that limited access to and participation in international markets constrains the development of small and medium enterprises. It is therefore proposed that targeted policies be implemented with the aim of promoting small business as well as small business exports (Gumede, 2000).

4.3. Tax Incentive Literature

Tax policy has for many years been aware of the need to assist the small business sector: tax policies designed to promote small business 'have always been popular with governments and will likely continue to be' (Bird, 1996, p.12). While tax incentives may be used in a variety of ways to stimulate business growth, one of the most common ways is through investment.

4.3.1. Tax Incentives or Investment

In many countries tax incentives are, or at least have been, used in an effort to stimulate investment and spur economic growth. In many cases, tax incentives are implemented when it is thought that the low levels of capital investment are as a result of either the current tax system or another obstacle for which the tax system can compensate (Klemm, 2010). The effectiveness of tax incentives on investments levels and economic growth has been the debate of much research and little consensus has been reached (James, 2009).

According to the neoclassical theory of capital accumulation, founded by Jorgenson (1963) and later extended by Hall and Jorgenson (1967), factors such as tax rates, depreciation, tax holidays and interest rates affect the levels of investment through the cost of capital. This theory assumes that, *cetris paribus*, a change in the cost of capital will lead to an automatic change in investment; in particular an increase in the cost of capital will reduce investment. Thus, in line with the neoclassical theory of capital accumulation, tax incentives which serve to reduce the cost of capital can be assumed to stimulate capital investment.

While admittedly simple, the theoretical impact of tax incentives is clear. Empirically, however, the relationship between tax incentives and levels of investment remains inconclusive. Tanzi and Shome (1992), for example, find that while tax incentives were successful in stimulating investment and economic growth in Taiwan, Korea and Singapore, they were unsuccessful at that time in the Philippines, Thailand, Malaysia and Indonesia.

Additional support for tax incentives in stimulating investment is found by Bernstein and Shah (1993) who make use of a dynamic production structure model to investigate the impact of tax policy in Turkey, Mexico and Pakistan. The authors find that tax incentives have a positive impact on levels of investment and production in these countries. Further, in an overview of developed country studies Hassett and Hubbard (2002) conclude that the empirical literature is consistent with the assumption made the neoclassical theory of capital accumulation: a reduction in the cost of capital will increase investment. The authors find that, on average, a one percent reduction in the cost of capital will increase investment by between 0.5 and one percent. An even larger effect is found in the developing country literature for Mexican manufacturing firms who exhibited an elasticity of investment with respect to the cost of capital of -2.0 (Verdugo, 2005).

More recently, however, little effect of tax incentives on investment behaviour has been found. For example, in a study of tax incentives in around 40 Latin American, Caribbean and African countries Klemm and Van Parys (2009) find little evidence to suggest that tax incentives encouraged private investment or economic growth. This result is supported by Van Parys and James (2010) who find no conclusive evidence of any impact of tax incentives on levels of investment in West and Central Africa.

A possible explanation for the lack of consensus among economies relates to the characteristics of the economy in question. It is argued that a country's economic characteristics play a larger role in the success of an industry than tax policy (Zee, Stotsky & Ley, 2002). Poor infrastructure, weak governance, and macroeconomic instability create an unattractive investment climate, which is unlikely to be effectively countered by tax policy (James, 2009). Indeed, Tanzi and Shome (1992; p.60) conclude that the positive experiences in Taiwan, Korea and Singapore are 'unusual' and are unlikely to be seen in economies that exhibit high levels of corruption and incompetence. Verdugo (2005) further cautions that investments are more sensitive to instability and uncertainty in the economy than fiscal incentives. The author finds additional results to suggest that the large positive impact of tax incentives in Mexico may simply be due to the small open economy characteristic of the country.

The inconclusive results found by these authors may also be as a result of the aggregate nature of the datasets used. With the exception of Verdugo (2005) majority of these studies use aggregate country data. The effect that is being estimated is an average across a large number of different types of firms with different characteristics and may be positive for some firms and negative for others, for example. By aggregating, this heterogeneity among firms is missed.

Other possible explanations for the inconsistent results include monopolistic competition, irreversibility, inter-asset reallocation, simultaneity, and measurement error (Goolsbee, 2000; Dixit and Pindyck, 1994; James 2009). Finally, others argue that the relationship between the cost of capital and investment is more complex than the neoclassical theory of capital accumulation predicts and due to the stickiness of capital at the core tax policy is unlikely to impact on investment decisions (Van Parys & James, 2010).

Regardless of the ambiguity of empirical results governments continue to use tax incentives. However, even if tax incentives are successful in stimulating investment, they are not always cost-effective (Zee *et al*, 2002). Bernstein and Shah (1993), for example, found that certain selective tax incentives, such as investment tax credits and accelerated depreciation, were more cost effective than general tax incentives in stimulating investment in Turkey and Pakistan. However, the authors found that while these selective tax credits were also successful in encouraging investment in Mexico, they were not cost-effective. Thus even if tax incentives encourage investment, the result can be inefficient and can result in the diversion of scarce resources into suboptimal uses (Bird, 2008). It is thus prudent to undertake cost-benefit analysis to establish not just the impact of the tax policy but also whether it is cost-effective or not.

The costs associated with tax incentives for investment include more than the direct loss of tax revenue. Indirectly, revenue may be lost if the incentivised investment crowds out other, more profitable investments (Klemm, 2010). In addition to the costs associated with the inefficient allocation of capital, tax incentives erode the tax base, increase administration costs and can create opportunities for corruption and rent-seeking activities (Zee *et al*, 2002; Bird 2008; James 2009).

The benefits of tax incentives include not only higher revenues from the increased investment but also an increase in social benefits such as job creation, signalling effects and overall economic growth (James, 2009). However, Klemm (2010) cautions that the benefits of tax incentive are difficult to measure since they are targeted at medium-term goals which are likely to be influenced by factors other than taxes and therefore the benefits of tax incentives should be measured in terms of higher levels of investment and economic growth.

Despite the limitations of using tax incentives to encouraging investment, they continue to be used by policy makers and governments in both developed and developing countries, particularly in underperforming sectors. An example of such would be the use of tax incentives within the small business sector.

4.3.2. Tax Incentives and Small and Medium Enterprises (SMEs)

Internationally it has been recognised that small and medium enterprises (SMEs) play an important role in the economy in terms of job creation and economic growth (FIAS, 2007). In the European Union, 55 percent of jobs are generated by small firms with fewer than 100 workers (Booysen, 1997). Similarly, small manufacturing firms in Ghana and South Africa provide for 85 percent and 61 percent of employment creation respectively and contribute around 70 percent and 57 percent to GDP respectively (Abor & Quartey, 2010). However, the development of small firms is hindered by a number of factors including tax-related issues and access to finance (OECD, 1997; Ojeka, 2011). It is thus important to evaluate the impact of current tax system on SMEs in order to establish whether it facilitates or hinders SME growth.

Many small and medium-sized firms lack access to short and long term capital markets since they are considered risky by banks and other financial institutions (Ojeka, 2011). As a result these firms rely heavily on the retained earnings and, in some cases, income of the proprietor of the business to fund future investments (Guenther, 2012). Thus a tax imposed on the SME directly reduces the main source of equity capital available to the business and therefore reduces its ability to invest and grow (Booysen, 1997). It is therefore in the best interest of the SME that enough profit remains after taxation to allow the business to expand.

Not only do taxes directly impact on the disposable income of the small business but also indirectly in the form of compliance costs. Tax compliance costs are regressive; the smaller the firm the larger the burden of compliance (Blažić, 2004). Smaller firms lack the management skills and training inherent in larger firms which is necessary to adequately comply with complicated and onerous tax regulation. As such smaller firms report large time and financial costs associated with tax compliance (Stern & Barbour, 2005). Evidence of the regressive nature of tax compliance costs has been found in many countries including Albania, Croatia, Zambia, South Africa, Rwanda and Nigeria (Engelschalk, 2004; Blažić, 2004; Stern & Barbour, 2005; Eragbhe & Modugu, 2014).

These high tax compliance costs may encourage some small firms to stay in the informal sector since the costs associated with participation in the tax net are perceived to be greater than the benefits (FIAS, 2006). However, despite high tax compliance costs it may prove worthwhile for some small firms to become part of the formal sector in the long run. Participating in the formal sector is a sign of legitimacy and as such allows small firms access to formal credit markets and government contracts (Stern & Barbour, 2005).

It is also beneficial to government if small firms join the tax net since it encourages a culture of compliance, which in turn facilitates the growth of the small firm and consequently substantial tax contributions from the now larger taxpayers (Shome, 2004; Stern & Barbour, 2005). However, the administrative costs involved in monitoring and collecting tax revenues from these small firms can in some cases outweigh the revenues collected and therefore may not be a cost-effective exercise (Engelschalk, 2004; Ojeka, 2011).

It is thus clear that the appropriate tax policy for SMEs should be one that is simplified and transparent and reduces the tax compliance burden thereby discouraging informality (Engelschalk, 2004). In addition it is essential that the tax policy reduces the actual tax burden of small business in order to allow the firm enough equity to reinvest and grow (FIAS, 2006). According to Ojeka (2011), if provision is made for an enabling environment for small firms through appropriate regulation, the SME sector has the potential to transform the economy.

Several studies have examined tax policy reforms for small businesses and the impact on their investment behaviour and growth. Booysen (1997) for example assessed the impact of tax incentives on SME development in a number of developed countries and two developing countries. The incentives took various forms including differentiated tax rates, tax holidays, and tax credits and accelerated depreciation allowances. The author concludes that many of these countries have recognised the benefits of providing tax incentives to SMEs such as increased investment and economic growth.

Other studies however find little evidence to suggest that SME investment and growth can be stimulated by tax incentives. In a study of policy reforms in Georgia, Ukraine, Russia and Albania it was found that SMEs did not significantly change their investment behaviour despite the introduction of simplified tax systems and generous tax reductions (Engelschalk, 2004). Similarly, Knittel (2005) conclude that the accelerated depreciation offered to small taxpayers in the United States, which was intended to stimulate capital investment (and consequently employment and overall economic growth), was unsuccessful. The author found that the benefits of the tax incentive were insufficient in persuading small firms to use the provision and consequently their investment behaviour remained relatively unchanged. However, considering that small firms lack access to information it may have simply been the case that many small taxpayers were unaware of this incentive or uniformed by their tax practitioner (Knittel, 2005).

This section discussed the potential impact of tax incentives on stimulating investment and economic growth. The conclusion remains unclear. This study attempts to add to the literature by examining the impact of a tax incentive on small business behaviour in the developing country of South Africa. The next section describes this tax incentive in more detail.

Section 12E of the Income Tax Act No. 58 of 1962

Section 12E was introduced in the South African government's Income Tax Act No.58 of 1962 (hereafter referred to as the Act) in 2001. The purpose of section 12E is to provide relief to small business corporations (SBCs) through the granting of preferential tax rates and accelerated depreciation allowances for the acquisition of any plant or machinery. These benefits are unavailable to non-SBC companies.

The Act defines a small business corporation (SBC) as:

> any close corporation, co-operative or private company in terms of the Companies Act 1973, all shareholders of which are at all times during the year of assessment natural persons where the gross income for the year of assessment does not exceed [*a predetermined threshold*]…not more than 20% of the receipts of the company consists of investment income [*in the form of interest, dividends and rental income*] and income from the rendering of a personal service and such company is not a personal service provider.

Since the passing of section 12E into the legislation in 2001, both the qualifying requirements as well as the tax benefits available to qualifying companies have been adjusted on numerous occasions. For example, the qualifying gross income threshold has been increased from R1 million in 2001 to R20 million in 2014. Progressive tax rates and brackets have been introduced and adjusted every other year, along with accelerated depreciation allowances and various administrative-easing programmes.

The motivation behind the introduction of these tax breaks follows a similar reasoning set out previously in the empirical literature; namely to support and facilitate the development of the country's small business sector.

Appreciating the somewhat critical role of small and medium business in job creation and overall growth of the South African economy, the National Treasury and South African Revenue Service (SARS) recognised a potential way to incentive these firms through the tax system.

The benefits provided for in section 12E, in particular the immediate expensing of investments, were expected to significantly reduce the cash flow constraint of growing small businesses and thus improve not only on the ability of such firms to create jobs, but also on the ability of the small business to reinvest and grow (National Treasury, 2001 & 2005). The expected benefits to the South African economy of the section 12E tax incentive was therefore an increase in both employment and investment in plant and machinery and eventually an improvement in growth of the small business sector and thus the economy as a whole.

Despite the fact that this incentive has been present for over a decade, very little research has been done to examine its effectiveness. Recently the Davis Tax Committee (DTC) released its interim report which examines the impact and compliance costs associated with various tax incentives that are available for small businesses in South Africa, including section 12E (Davis Tax Committee, 2014).

In terms of section 12E, the DTC finds that the main beneficiaries of the tax incentive are well-established, profitable niche SMEs. Small start-up businesses and those businesses struggling to survive experience very little, if any, benefit from section 12E. In addition, this incentive is costly and carries large administrative burdens for both business and government. For the 2012 tax year, the report estimates a cost of the tax incentive (in terms of forgone tax revenue) of about R1.261 billion. In light of the results, the report suggests that the current SBC incentives be eliminated and replaced by a system that rewards tax-compliant SBCs.

In turn, the National Treasury has given consideration to the DTC recommendations and is investigating the idea of a refundable tax compliance credit for SBCs in place of the graduated tax structure offered by section 12E (National Treasury, 2014).

While the DTC report provides a useful discussion on the potential benefits (or lack thereof) and estimated cost of the incentive, it does so by making use of a cross-sectional analysis of SARS tax data and the survey report of Smulders, Stiglingh, Franzsen and Fletcher (2012). It would be of additional interest to investigate the dynamic impact of the incentive on the levels of investment and employment at an empirical level. This chapter attempts to do just that. The next sections discuss the data and methodology that will be used to investigate the effect of the section 12E tax incentive on investment and employment of SBCs.

4.4. Data and Descriptive Statistics

This study makes use of official data collected by Statistics South Africa's (Stats SA) in its Large Sample Surveys of manufacturing (LSS). These surveys are used for the purposes of calculating the national accounts and although designed to be cross-sectional in nature many firms can be linked between the years to create a panel dataset. There are approximately 10 000 manufacturing firms in each of the rounds (2005 and 2008) used in this chapter.

These surveys collect data on industrial classification, employment, imports and exports, income and expenditure, profit or loss, inventories, carrying value of assets as well as details of products manufactured. To compare 2005 data to 2008, the 2008 data was deflated using industry level deflators derived from the PPI, except for wages which are deflated by the CPI.

As previously mentioned, section 12E has undergone many revisions over the past few years. However, given that the data used in this analysis is for the years 2005 and 2008, the discussion will be restricted to the changes that occurred between these years. It is noted here that no material changes were made to section 12E in the years 2007 and 2008.

By 2005 section 12E had expanded on the benefits provided to qualifying SBCs to include a number of administrative-easing programmes (in addition to those introduced in 2004) such as a small business help desk, community tax helpers and the provision of accounting packages for small business; addition progressivity of the tax rate structure and a third corporate income tax bracket; 100% immediate expensing of plant and machinery used in manufacturing as well as a three year accelerated depreciation for non-manufacturing capital in the ratio 50:30:20%.

By 2006 the benefits remained largely unchanged, with the exception of a widening of the tax brackets. However, the qualifying gross income threshold was increased from the 2005 qualifying threshold of R6 million to R14 million. This exceptional increase creates a natural experiment which enables this study to examine how firms, which initially did not qualify as an SBC in 2005 but did qualify in 2006, changed their behaviour in response to the tax incentive, relative to firms who never qualified as an SBC. The identifying assumption for this analysis is that the change in the tax incentive threshold was exogenous: firms were unlikely to know about the increase in the threshold and even if they did, the actual level of the new threshold was unanticipated.

In what follows, a number of descriptive statistics are presented for South African exporters as well as South Africa's small business sector.

4.4.1. Exporters in South Africa

Exporting in South Africa is rare, but does seem to be increasing over time. The data shows that less than a third of the South African manufacturing firms are exporters (26% in 2005 and 30% in 2008). In addition, those who export, export very little (Figure 4.1). Only 36 percent of exporters in 2005 exported more than 50 percent of their total output, the median exporter exporting a mere 7.5 percent. In 2008 less than a third of exporters exported more than 50 percent of total output with the median exporter exporting only 4.5 percent.

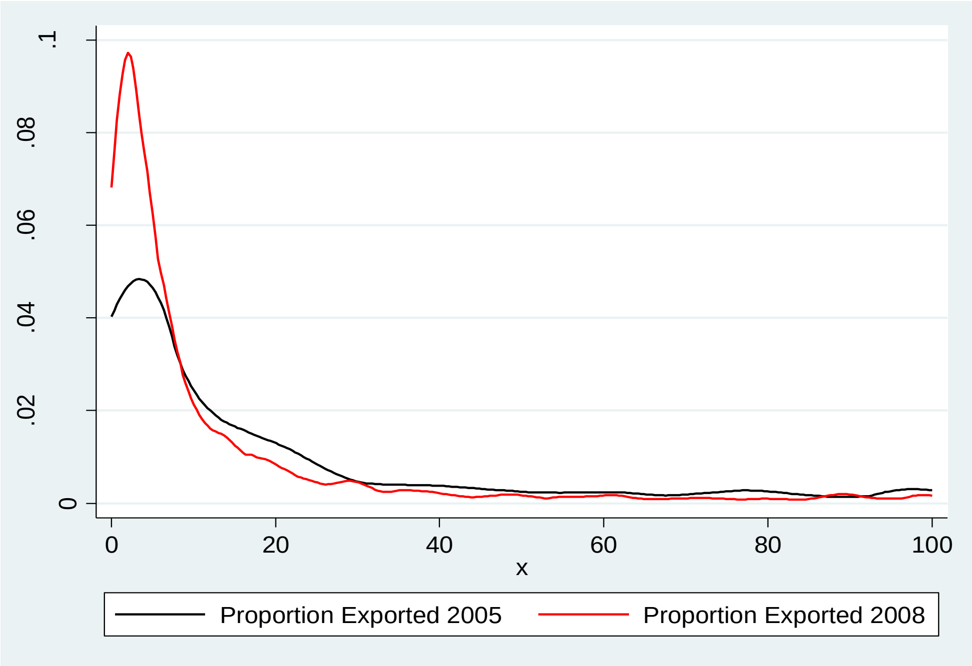


Figure 4.1. Amount Exported as a Proportion of Total Output

The data also shows that South African manufacturing exports are dominated by a small group of very large exporters (table 4.1). In order to rank exporters we calculate the total aggregate amount of exports in each year using the weights supplied by Statistics South Africa. Using the weights we calculate the total number of exporters in each year (1566 in 2005 and 2864 in 2008) and then rank exporters based on the amount their exporters contribute to total aggregate exports. The number of firms exporting almost doubles between the two years.

The degree of concentration is marked – the top 1% contribute between 65 and 72% of total export value and the top 5% contribute between 77 and 88% of total export value. Furthermore, the absolute number of these super-exporters is small – 104 in 2005 and 191 in 2008 (for the top 5%). Relative to these exporters, the contribution of the next tier (5-10%) is small. This group adds only between 4-10% to total export value. In total the top 10% of exporters produces approximately 90% of export value.

Table 4.1. Concentration of South African Manufacturing Exporters

	2005		2008		Pooled	
Proportion Exported	**No. of exporters**	**Contribution (%)**	**No. of exporters**	**Contribution (%)**	**No. of exporters**	**Contribution (%)**
Top 1%	*21*	***65.58***	*38*	***72.23***	*59*	***71.18***
Top 2%	42	72.12	76	80.27	118	78.03
Top 3%	63	75.88	115	84.51	177	80.12
Top 4%	84	76.03	153	86.87	236	82.69
Top 5%	*104*	***76.74***	*191*	***88.10***	*295*	***85.37***
Top 10%	*209*	***87.03***	*382*	***92.64***	*591*	***91.07***
Top 20%	418	94.81	764	96.51	1182	96.11
Top 25%	522	96.87	955	97.56	1477	97.20

Bottom 75%	1566	3.13	2864	2.44	4431	2.80

In terms of exporter characteristics, the data confirms the findings of the general literature that exporters exhibit superior characteristics relative to non-exporters. Table 4.2 below shows that exporters are larger than non-exporters both in terms of sales and employment and, although for most variables, this difference reduces in 2008 it remains significantly different from 0. They are also more labour productive, capital intensive and pay higher wages than their non-exporting counterparts even after controlling for industry and firm size.

Table 4.2. Characteristics of Exporting Firms

	Sales	**Total employment**	**Output per employee**	**Ave. wage**	**Capital per employee**	**Intermediate inputs per employee**
Controlling for Industry (SIC3)						
2008 dummy	-0.152***	-0.0667**	-0.0276	0.212***	-0.143***	-0.117***
Export dummy	0.738***	0.635***	0.164***	0.311***	0.358***	0.227***
2008×exp. dummy	-0.187***	-0.222***	-0.00796	-0.108***	0.167***	-0.00909
Observations	*9,561*	*9,950*	*9,514*	*9,945*	*9,770*	*9,945*
Controlling for Industry (SIC3) and Firm size						
2008 dummy			-0.0322	0.213***	-0.135***	-0.119***
Export dummy			0.186***	0.302***	0.291***	0.245***
2008×exp. dummy			-0.0150	-0.105***	0.190***	-0.0154
Observations			*9,514*	*9,945*	*9,770*	*9,945*

Notes: ***p<0.01, **p<0.05, *p<0.1 is significance at the 1% level, 5% level and 10% level respectively. Values for sales, employment, output per employee, average wage, capital per employee, and intermediate inputs per employee are given in natural logarithms

Exporters are also more likely to invest in capital than non-exporters over the years. This is unsurprising given that exporters are more capital intensive than non-exporters. The data shows that 72 percent of firms that entered into exporting between 2005 and 2008 increased their purchases of plant and machinery. This is compared to the 47 and 49 percent among firms that exited the export market and non-exporters respectively. Of the continuous exporters (those firms who have exported for at least three years), 67 percent increased their investment in plant and machinery. Thus it seems that investment in capital is an important part of the exporting process in South Africa.

The results in table 4.3 suggest that entry into the export market is associated with a marked increase in capital investment. Entering exporters are significantly more likely to increase their investment in capital than, not only non-exporters, but also the more experienced/continuous exporters. This seems to confirm the notion that successful entry into international markets requires a relatively large initial investment in more appropriate and internationally recognised technology in order to produce at a competitive level. Once the initial expense is overcome, capital investment only needs to be maintained at a competitive level. However, after this significant initial investment, maintaining these high levels may not always be feasible for the firm, particularly for small firms. Small firms, with their relatively low levels of disposable income and lack of access to long term debt institution, are often not capable of maintaining such large investments. In such cases firms may be unable to maintain their competitiveness and will eventually discontinue exporting. Table 4.3 further shows that firms that exit the export market are more likely to reduce their investment in capital below even those of non-exporting firms. This illustrates just one of the poor outcomes associated with an exit from exporting.

Table 4.3. Change in Capital Investment between 2005 and 2008 – Exporters and Non-exporters

VARIABLES	**Plant and Machinery - ln(Change)**	**Research and Development - ln(Change)**
Entering Exporters	5.767***	5.580***
	(0.500)	(0.331)
Continuous Exporter	3.669***	2.622***
	(0.525)	(0.348)
Exiting Exporters	-0.320	-0.788*
	(0.660)	(0.438)
Log(employment)	-1.698	0.306***
	(5.649)	(0.102)
Observations	*2,643*	*2,643*

Exit from the export market is particularly prevalent among small exporters (in terms of proportion contributed to total exports). Table 4.4 illustrates the transition matrix for firms in the sample. Exporters are ranked based on the amount their exporters contribute to total aggregate exports. The results show that around 31 percent of small exporters exited exporting between 2005 and 2008 compared to 12 percent of large exporters. While this may be because large exporters are more likely to be specialist exporters and thus hold firmer positions within international markets, these firms are still more likely to exit exporting than to drop in to a lower level of exporting. Despite the large number of exiting firms, the table also shows a relatively high degree of stability with regards to where a firm is in the distribution.

Table 4.4. Transitions over the Period between Export Groups

	2008 Status					
2005 Status	top 1%	2-5%	6-10%	Other exporter	Non-exporter	Not in sample
top 1%	70.59	5.88	0	5.88	11.76	5.88
2-5%	8.93	32.14	7.14	14.29	23.21	14.29
6-10%	0	18.92	22.97	27.03	21.62	9.46

Other exporter	0.38	1.13	2.89	50.44	31.12	14.05
Non-exporter	0.61	1.25	1.41	20.27	46.69	29.77
Not in sample	0.07	0.29	0.63	24.72	74.29	–

Overall the data shows that exporting in South African manufacturing conforms to the stylised facts discussed previously. Exporters are larger, more productive and capital- and skill-intensive, and pay higher wages than non-exporting firms. However, exit rates are high and export survival is likely to be low particularly among small exporters. In order to understand how to assist these small exporters it is first necessary to examine the characteristics of small businesses.

4.4.2. South Africa's Small Business Sector

In 2005 a firm qualified as a small business corporation (SBC) if it had a gross income not exceeding R6 million. In 2008 this qualifying threshold was increased to R14 million. Table 4.5 below illustrates a number of characteristics of South African SBCs and non-SBCs for 2005 and 2008 based on this data.

In 2005, almost 15 percent of the sample of firms qualified as an SBC. The increase in the qualifying threshold from R6 million to R14 million in 2008 increased the number of firms qualifying as SBCs to approximately 27 percent. As is to be expected, SBCs hire significantly fewer employees than non-SBCs regardless of the year or qualifying threshold. In terms of output per worker, firms that qualified as SBCs were less labour productive than non-SBCs both in 2005 and 2008. Table 4.5 further indicates that SBCs, as expected, are significantly less capital intensive than non-SBCs.

An interesting observation from the table shows that in 2005 SBCs were more likely to take on debt than non-SBCs. However, in 2008, SBCs were noticeably less likely to take on debt than non-SBCs. The table further shows that the proportion of firms with both long term and short term debt has decreased over the years. This reduction in debt is considerably apparent among smaller firms. A possible explanation for this may be the significant tightening of credit standards as a result of the global economic downturn of 2007/08. This was particularly worse for small and medium firms. According to Fuchs *et al* (2011), the tightened credit conditions negatively impacted on small and medium firms who not only saw a 23 percent decrease in loan applications, but also a decline in loan approval rates (from 61% to 45%).

Finally, table 4.5 indicates that SBCs are less likely to be exporters than non-SBCs. However, given that a firm exports, exporting SBCs tend to export a higher proportion of sales than exporting non-SBCs.

Overall, SBCs employ fewer workers, are less capital intensive and have less debt than non-SBCs. In addition, they are less likely to export, but when they do they export a higher proportion of sales than non-SBCs.

Table 4.5. Characteristics of South Africa's Small Business Sector (Balanced Panel)

	≤ R6 million		**> R6 million**		**≤ R14 million**		**> R14 million**	
	2005	**2008**	**2005**	**2008**	**2005**	**2008**	**2005**	**2008**
SBCs								
Total number	587	534	3348	3401	1326	1057	2609	2878
Proportion	14.92	13.57	85.08	86.43	33.7	26.86	66.3	73.14
Average Employment+	19	30	308	273	30	34	354	295
	(14)	(16)	(90)	(90)	(23)	(23)	(113)	(98)
log(output/labour) +	5.33	5.36	6.35	6.37	5.59	5.56	6.45	6.43
	(5.32)	(5.24)	(6.28)	(6.29)	(5.55)	(5.54)	(6.39)	(6.35)
log(capital/labour) +	3.11	2.80	4.04	4.09	3.26	3.07	4.15	4.16
	(3.32)	(2.89)	(3.99)	(4.08)	(3.39)	(3.17)	(4.11)	(4.14)
Debt (% of firms with) ++								
Any	92.50	22.10	86.00	61.89	92.98	27.53	83.92	67.13
Long term	87.73	17.23	80.50	55.78	88.60	23.75	78.00	60.39
Short term	79.73	11.42	56.88	39.11	75.62	13.81	52.5	43.26
Exporting								
% of firms exporting	17.14	4.49	31.81	40.12	19.47	7.95	33.67	45.32
Proportion exported if export+	20.89	22.15	24.81	14.77	18.97	14.98	25.52	14.88
	(5.98)	(5.62)	(7.69)	(5.57)	(6.52)	(5.46)	(7.73)	(5.59)

Notes: + Median given in parenthesis. ++ Long term debt includes long term loans; short term debt includes overdrafts and creditors

4.5. Methodology

To investigate the impact of the Section 12E tax incentive on employment, physical capital stock (defined as plant and equipment) and exports between firms that qualified for the tax incentive (SBCs) and those firms that did not qualify (non-SBCs), we exploit the panel nature of the dataset. A limitation of the current dataset is that of an inability to determine whether a firm did indeed take advantage of the accelerated depreciation allowance or not. Therefore, this chapter uses whether a firm qualifies as an SBC or not to proxy for the tax treatment. In essence, the following estimates approximate the intention to treat.

The effect of the tax incentive is assessed using difference- in- differences. In terms of the impact of employment, the following specification was run

$$\Delta emp_{it} = \alpha + \emptyset_1 Qualified_{it} + \emptyset_2 \ln(\frac{K}{L})_{it} + \emptyset_3 Industry_{it} + \varepsilon_{it} \tag{4.1}$$

where *i* is the firm subscript, *t* is the time subscript, *emp* is the natural logarithm of the change in employment, $Qualified$ is a dummy variable which takes on the value 1 if the firm qualified for the tax incentive (turnover ≤ R14 million) and 0 otherwise, $\ln\left(\frac{K}{L}\right)$ is the natural logarithm of the capital labour ratio, $Industry$ is a dummy variable which controls for industry effects at SIC 3 digit level and ε_{it} is the residual. The identifying strategy for this analysis relies on the assumption that the right hand side variables in the above, and following, equations are independent of the $Qualified$ variable, since the change in the qualifying threshold is exogenous.

Similarly, in order to estimate the impact of the tax incentive on physical capital stock, the following specification was run

$$\Delta plant_{it} = \alpha + \emptyset_1 Qualified_{it} + \emptyset_2 \ln(\frac{K}{L})_{it} + \emptyset_3 Industry_{it} + \emptyset_4 \ln(employ)_{it} + \varepsilon_{it} \tag{4.2}$$

where *i* is the firm subscript, *t* is the time subscript, $plant$ is the natural logarithm of the change in the level plant and machinery, $\ln(employ)$ is the natural logarithm of total employment and $Qualified$, $\ln\left(\frac{K}{L}\right)$, $Industry$ and ε_{it} are defined as before.

In addition to estimating the impact on capital in terms of plant and machinery, this study also estimates the impact on capital in terms of research and development (R&D). The specification is the same as equation (4.2) with the exception of $plant$ which is replaced by RD – the natural logarithm of the change in the level of research and development. The reason for this is to check the robustness of the results. Section 12E does not pertain to R&D but only physical stock of capital and therefore the estimations should not show similar changes in R&D seen for plant and machinery.

The estimator $\emptyset_1$ in equation (4.1) is the average effect of the tax incentive on the change in employment over time. In equation (4.2) it is the average effect of the tax incentive on the change in plant and machinery over time. The effect is identified by difference- in- difference variation i.e. it says how the tax incentive affected the level of employment (or capital) of SBCs relative to non-SBCs. In implementing this incentive the government expected a $\emptyset_2$>0 for both a change in employment as well as a change in capital.

In addition to examining the impact of the tax incentive on small firms, this chapter also examines the impact of the tax incentive on exporters. Therefore, a second set of equations is estimated.

Once again the outcomes of interest are the change in employment, plant and machinery and R&D and the effect of the tax incentive is assessed by difference- in- differences estimation. The following estimation was run

$$\begin{aligned}\Delta Y_{it} &= \alpha + \emptyset_1 Qualified_{it} * exporter_{it} \\ &\quad + \emptyset_2 \ln(\tfrac{K}{L})_{it} + \emptyset_3 Industry_{it} + \emptyset_4 \ln(employ)_{it} \\ &\quad + \varepsilon_{it}\end{aligned} \tag{4.3}$$

where i is the firm subscript, t is the time subscript, Y is the outcome of interest (i.e.: the natural logarithm of the change in the level of employment or in the level of capital), $exporter$ is a dummy variable which takes the value of 1 if the firm exported and 0 otherwise, $\ln(employ)$ is the natural logarithm of total employment (included only if Y is defined as a change in capital), $Qualified$, $\ln\left(\frac{K}{L}\right)$, $Industry$ and ε_{it} are defined as before.

In this case the estimator $\emptyset_1$ in equation (4.3) is the average effect of the tax incentive on the change in the outcome of interest (employment or capital) over time for exporting SBCs. The effect is once again identified by difference- in- difference variation and allows for the comparison of exporting SBCs to non-exporting SBCs, exporting on-SBCs and non-exporting non-SBCs. In implementing this incentive the government expected a $\emptyset_2$>0 for both a change in employment as well as a change in capital.

A third set of regressions was run in order to examine how the behaviour of exporters changed relative to non-exporters over the period among only those firms that qualified for the tax incentive. Therefore, restricting the sample to include only those firms that qualified, the following specification was run

$$\Delta Y_{it} = \alpha + \emptyset_1 exporter_{it} + \emptyset_2 \ln(\tfrac{K}{L})_{it} + \emptyset_3 Industry_{it} + \emptyset_4 \ln(employ)_{it} + \varepsilon_{it} \tag{4.4}$$

where all variables are defined as before with the difference being that the regression is run only if $Qualified = 1$.

The estimator $\emptyset_1$ in equation (4.4) is the average effect of the tax incentive on the change in the outcome of interest (employment or capital) over time for exporters relative to non-exporters only among SBCs.

Finally, given the potential impact on capital among exporting SBCs of the tax incentive and given the importance of capital to exporters, this study further estimates the effect of the tax incentive on the probability of exporting in the future. In order to do so, the following Linear Probability Model (LPM) was run

$$Exporter2008_{it} = \alpha + \emptyset_1 Qualified * exporter2005_{it}$$
$$+\emptyset_2 Industry_{it} + \emptyset_3 \ln(employ2005)_{it}$$
$$+\varepsilon_{it} \tag{4.5}$$

where $Exporter2008$ is a dummy variable which takes the value 1 if the firm exported in 2008 and 0 otherwise, $exporter2005$ is a dummy variable which takes the value 1 if the firm exported in 2005 and 0 otherwise and $employ2005$ is total number of employees in 2005.

Very large firms are different to very small firms. Thus in order to compare firms which are as similar as possible, only firms around the qualifying threshold of R14 million are considered for analysis. Therefore the sample is restricted to include only firms that generated a turnover within five specified bands. The first band includes only those firms with a turnover greater than R6 million, but no less than R14 million. The second band ranges from R10 million to R18 million; the third from R12 million to R16 million; the fourth from R12.5 million to R15.5 million; and the fifth from R13million to R15million.

A potential problem inherent in the dataset is due to the approach StatsSA uses to construct the sample. Large firms (size 1 in the Stats SA dataset) are fully enumerated in the sample. A smaller sample of medium firms (size 2) is then drawn, followed by an even smaller sample of size 3 firms etc. In addition these cut-offs differ across sectors. As a result, the panel has relatively fewer smaller firms and the probability of a certain sized small firm being in the sample differs across sectors. Therefore, in order to control for this selection bias, as well as the potential exit of small firms, and to check for the robustness of the results, Heckman's two stage sample selection model is run.

The first stage estimates the probability of selection into the sample using Stats SA's categorical dummy for size as well as industry at SIC level 4 as the exclusion variables. The second stage estimates the relevant regression (or outcome) equations as used in equations (4.1) to (4.5) above. The results are presented in Appendix D and discussed in the following section.

4.6. Results

This section presents a discussion of the results of the above econometric analysis on the effects of the tax incentive for small businesses on SBCs and exporting.

The first set of results investigates the impact of the Section 12 tax incentive on all SBCs and Non-SBCs within the specified turnover bands. Table 4.6 presents the estimation results of equations (4.1) and (4.2) i.e.: the change in employment and plant and machinery.

Table 4.6. Estimated Impact of Qualifying for Section 12E on the level of Employment and Capital (Plant and Machinery) – All SBCs and Non-SBCs

			Turnover		
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Change in (log) Employment					
Qualified	0.126***	0.0463	0.0788	0.139**	0.199***
	(0.0265)	(0.0388)	(0.0557)	(0.0641)	(0.0657)
Log(real capital/employee)2005	0.0702***	0.0595***	0.107***	0.101***	0.180***
	(0.0116)	(0.0173)	(0.0261)	(0.0303)	(0.0299)
Constant	0.140	0.233	0.142	-0.191	-0.636*
	(0.484)	(0.486)	(0.430)	(0.467)	(0.361)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*1,533*	*809*	*391*	*308*	*195*
Change in Plant and Machinery					
Qualified	-0.597***	-0.896***	-2.034***	-1.468***	-0.393
	(0.184)	(0.227)	(0.383)	(0.349)	(0.420)
Log(employment)	-0.586***	-0.726***	-1.211***	-0.482*	-0.473
	(0.138)	(0.203)	(0.315)	(0.276)	(0.318)
Log(real capital/employee)2005	-0.831***	-0.978***	-0.923***	-0.805***	-0.0777
	(0.0833)	(0.106)	(0.181)	(0.177)	(0.307)
Constant	6.967***	8.053***	11.22***	6.220***	3.852
	(1.745)	(1.854)	(2.232)	(2.090)	(2.990)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*505*	*328*	*169*	*142*	*94*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

One motivation for the implementation of Section 12E was to foster employment among small firms (or SBCs). The results suggest that firms which qualified as SBCs and thus qualified for the tax incentive did indeed increase their level of employment relative to larger firms which did not qualify for the tax incentive. The results are significant and hold even after controlling for industry fixed effects.

A second area through which the South African government aimed to achieve growth via the tax incentive is increased capital investment, specifically physical capital. As table 4.6 indicates, qualifying firms did not increase the level of capital investment as expected, but instead significantly reduced their capital investment relative to larger firms. Again, these results hold after controlling for industry. What may be going on here is a reduction in the book value due to the higher level of depreciation.

The impact of qualifying for the tax incentive on levels of research and development was also estimated. This serves to check that the previous results were as a consequence of the introduction of the tax incentive and not just a reduction in all capital investments. As can be seen in table D1 in Appendix D, there is no evidence of a reduction in capital investment of research and development among firms that qualified relative to firms that did not. The results presented do not control for industry. Due to the low number of firms that reported figures for research and development, controlling for industry reduces the number of observations too low for regression analysis.

As it stands, the results suggest that while the tax incentive may have played a role in encouraging employment among small firms, it did little to stimulate investment in physical capital. Indeed, it may have facilitated disinvestment. One potential explanation for this could lie within the income- and substitution-effect literature. The Section 12E tax incentive, through the accelerated depreciation of capital, ultimately reduced the user cost of capital. Theory suggests that this would lead to a substitution away from other factors of production (such as labour) towards the relatively cheaper capital. Thus, according to the substitution effect, one would expect to see an increase in capital investment and potentially a reduction in employment. However, the tax incentive also reduced the overall tax liability for small firms and therefore allowed firms to maintain a higher after-tax profit. Higher profit, or income, would allow for an increase in potentially all factors of production. This, in turn, would increase in output, thereby facilitating growth. The choice of which factors of production to increase depends on, among other things, the relative factor intensity of the firms. Small firms, as previously discussed, are more labour intensive than capital intensive. The income effect among small firms would therefore likely be an increase in employment potentially by more than capital. The question then would be which effect is greater. This question, while interesting, is beyond the scope of this study.

A second potential explanation for the above results is that firms are heterogeneous. Some firms, even within the same industry, may react differently to the tax incentive. One such example is that of exporting firms. As previously mentioned, exporters are different to non-exporters across a number of characteristics. One difference that is relevant for this discussion is that exporters are more capital intensive than non-exporters. A tax incentive that directly affects capital is therefore likely to have a different impact on exporting SBCs as it does on non-exporting SBCs. In addition, the reduction in the overall tax burden on SBCs may be more 'valuable' to exporters than non-exporters, since exporters operate in a highly competitive market. In such a market higher disposable income could be useful for the survival of exporters. The second set of regression results therefore includes an interaction between qualifying for the tax incentive and being an exporter. The results are presented in table 4.7.

Table 4.7. Estimated Impact of Qualifying for Section 12E on the level of Employment and Capital (Plant and Machinery) – Exporters versus Non-exporters

	Turnover				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
<u>Change in (log) Employment</u>					
Qualified	0.148***	0.0947**	0.0968	0.114	0.0638
	(0.0311)	(0.0436)	(0.0622)	(0.0721)	(0.0727)
Exporter	-0.0493	-0.164***	-0.208***	-0.286***	-0.461***
	(0.0460)	(0.0584)	(0.0785)	(0.0879)	(0.0813)
Qualified X exporter	-0.123**	-0.336***	-0.233*	-0.226	0.158
	(0.0620)	(0.0910)	(0.127)	(0.143)	(0.123)
Log(real capital/employee)2005	0.0735***	0.0723***	0.136***	0.131***	0.223***
	(0.0116)	(0.0168)	(0.0264)	(0.0297)	(0.0285)
Constant	0.114	0.169	0.0887	-0.359	-0.873***
	(0.482)	(0.469)	(0.418)	(0.446)	(0.329)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*1,533*	*809*	*391*	*308*	*195*
<u>Change in Plant and Machinery</u>					
Qualified	-0.834***	-1.154***	-2.627***	-2.051***	-1.205***
	(0.200)	(0.247)	(0.419)	(0.411)	(0.425)
Exporter	-0.0839	0.0332	-0.945	-1.033	-0.655
	(0.271)	(0.358)	(0.752)	(0.664)	(0.766)
Qualified X exporter	1.409***	1.425***	3.134***	2.495**	4.594***
	(0.420)	(0.509)	(1.114)	(1.004)	(1.199)
Log(employment)	-0.653***	-0.736***	-1.529***	-0.767***	-1.214***
	(0.139)	(0.202)	(0.323)	(0.292)	(0.318)
Log(real capital/employee)2005	-0.895***	-1.056***	-0.985***	-0.933***	-0.728**
	(0.0832)	(0.105)	(0.177)	(0.181)	(0.299)
Constant	7.573***	8.448***	13.27***	9.469***	11.11***
	(1.729)	(1.829)	(2.270)	(2.146)	(3.014)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*505*	*328*	*169*	*142*	*94*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

In terms of employment, the results suggest that exporters were likely to decrease employment in general, relative to non-exporters. In addition, there is evidence to suggest that firms that qualified for the tax incentive reduced employment levels if they exported, but increased employment if they did not export.

Table 4.7 also presents the results for the impact on capital, in terms of plant and machinery. After controlling for industry, the results suggest that exporters who qualified for the tax incentive significantly increased their investment in capital. This is relative to qualifying non-exporters who experienced a significant reduction in capital investment and relative to firms that did not qualify which show some evidence of reduced capital investment (see figure 4.2)[19].

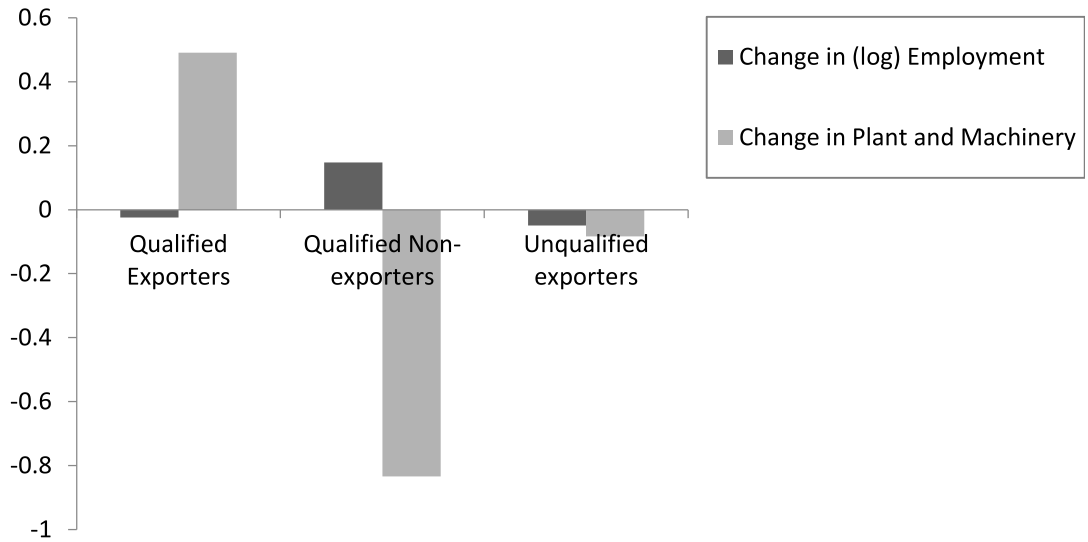


Figure 4.2. Illustrative Impact of Qualifying for Section 12E on the level of Employment and Capital – Exporters versus Non-exporters

In contrast, some evidence is found that exporters who qualified for the tax incentive reduced their spending on research and development relative to non-exporters and firms that did not qualify (see table D2 in Appendix D).

The previous set of results suggested that the tax incentive stimulated employment but discouraged investment in capital. However, once the estimations are broken down further between exporters and non-exporters, it is shown that these results are driven by non-exporting SBCs. In contrast, exporters who qualified as SBCs, and consequently qualified for the tax incentive, saw a reduction in employment and an increase in capital investment. This result is unsurprising considering that exporters are relatively more capital intensive than non-exporters. A reduction in the user cost of capital is beneficial to firms

[19] The capital:labour ratio is removed from these regressions to test for robustness. Overall conclusions remain unchanged. See an example in Appendix D.

that use capital relatively more intensely and it is therefore not difficult to expect exporters, who are capital intensive, to be encouraged to accumulate additional capital in response to the tax incentive.

Section 12E is targeted at small firms, or SBCs. Therefore a third set of regressions is estimated after limiting the sample to SBCs only. The results are presented in table D3 in Appendix D. As is to be expected, the results are consistent with those in table 4.7: qualifying exporters experienced a significant reduction in levels of employment and an increase in investment in plant and machinery relative to non-exporters.

So far the findings suggest that the tax incentive had the anticipated impact on employment for small non-exporters, but not small exporters. Additionally, it had the anticipated impact on levels of investment for small exporters, but not small non-exporters. Therefore, unlike many previous studies, this chapter finds some evidence to suggest a positive impact on levels of capital investment among small firms, in particular small exporters. This is an important finding, because not only does it suggest caution in treating firms as homogeneous when analysing the impact of certain policy interventions, but also because facilitating capital investment among small exporters may have further unintended, yet beneficial, effects for the affected firms beyond just an increase in employment and capital accumulation.

Table 4.8 presents an example of such an unintended consequence of qualifying for the tax incentive for small exporters. The results show that exporters who qualified for the tax incentive were significantly more likely to export in the future relative to exporters who did not qualify. Non-exporters who qualified for the tax incentive were the least likely to export in the future. This suggests that while a tax incentive which allows for an accelerated depreciation allowance on physical capital may not encourage non-exporters to enter the export market, it does seem to facilitate the survival of small firms who are already exporting. This is an important result given the negative effects for a firm of not surviving the export market.

Table 4.8. Estimated Impact of Qualifying for Section 12E on the Exporting in the Future

	Exporting in 2008				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Exporter 2005	0.151***	0.135*	0.0737	0.118	0.0875
	(0.0433)	(0.0698)	(0.125)	(0.147)	(0.208)
Qualified	-0.0244	-0.116**	-0.282***	-0.300**	-0.334*
	(0.0296)	(0.0504)	(0.0901)	(0.119)	(0.174)
Qualified X exporter 2005	0.150**	0.445***	0.564***	0.354*	0.165
	(0.0596)	(0.0960)	(0.167)	(0.198)	(0.308)
Log(employment) 2005	0.00781	0.00185	-0.0853*	-0.0936	-0.117
	(0.0166)	(0.0253)	(0.0485)	(0.0626)	(0.0915)

Constant	-0.298	-0.0103	0.474	0.520	0.648
	(0.362)	(0.360)	(0.450)	(0.498)	(0.652)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	957	399	160	123	76

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

The negative coefficient on the "Qualified" dummy is explained using the Meltiz (2003) framework which predicts that exporters tend to be larger than non-exporters. Furthermore the empirical literature finds that there is a strong positive correlation between exporting and size, and often this difference exists prior to entry into exporting. In addition, the qualified dummy refers to *small* business corporations (as opposed to larger firms). The negative coefficient on "Qualified" can be interpreted in the following way: non-exporters in 2005 who qualified for the tax incentive (small businesses/ SBCs) are less likely to export in 2008 than non-exporters who did not qualify (larger firms) potentially because of the size-exporting relationship which suggests that larger firms are more likely to export than smaller firms.

Up to this point the discussion has been conditional on survival of the firms over the period. However, survival may also be dependent on whether a firm was treated or, rather, whether a firm qualified for treatment. Table D5 in Appendix D illustrates the variables which affect the likelihood of attrition from the sample. As expected, firms with a higher profitability, higher labour productivity and more skilled employees are less likely to attrite. In addition there is evidence to suggest that a negative relationship exists between firms that qualified for the tax incentive and the likelihood of attrition. In other words, there is a positive effect of qualifying on seemingly survival. To check whether the previous results are robust not only to the sampling approach used by StatsSA, but also this potential treatment effect on survival a Heckman two stage sample selection model is run. The results are presented Appendix D.

Table D6 presents the results of the impact of the Section 12 tax incentive on all SBCs and Non-SBCs within the specified turnover bands. After controlling for selection from the two mentioned mechanisms, these results still indicate some positive impact on employment for firms which qualified for the tax incentive relative to those that did not, although this result loses significance as the turnover bands are tightened. There is also evidence to suggest that qualifying firms reduced their investment in capital. These results do not contradict the previous findings.

Once exporters are included in the outcome regression (table D7), it can be seen that, as the previous findings suggest, employment decreased among exporters that qualified for the tax incentive relative to non-exporters, however these results are not significant. In terms of the change in capital, the evidence suggests that qualifying exporters increased investment in capital relative to non-exporters.

Finally, looking at the impact on exporting in the future (table D8), the results indicate that exporters who qualified for the tax incentive were significantly more likely to export in the future than qualifying non-exporters and even, in some cases, exporters who did not qualify for the tax incentive.

Overall, after taking into account the potential selection bias, the preceding results are of the same sign, but only significant in some cases. This may be due to a reduction in sample size. In terms of employment, firms that qualified for the tax incentive decreased employment if they were exporting relative to firms that did not export. The results are less confirmatory for the impact on capital, although there is some evidence to suggest a positive impact among exporters. Further, there is a small amount of evidence to suggest that the ability to export in the future is positively influenced by the tax incentive.

4.7. Concluding Remarks

The international evidence on the success of tax incentives in stimulating investment and economic growth is mixed. Further, despite the common use of tax incentives in the developing world, relatively little empirical work has been done in these developing countries on the impact of such tax incentives.

This chapter adds to the current literature by analysing the impact of a tax incentive on small firms in the South African manufacturing sector. The motivation behind section 12E of the Income Tax Act No. 58 of 1962 was the stimulation of capital investment, employment and consequently growth within the small business sector. The incentive was available for all firms that qualified as small business corporations (SBCs) and included a lower and more progressive tax rate structure and accelerated depreciation on physical capital.

Making use of official firm-level panel data, this study estimated the effect of section 12E on levels of employment and plant and machinery among firms that qualified as SBCs. In addition, given the strong relationship between capital and exporting, the chapter examined the potential unintended effects on exporting firms that qualified as SBCs i.e.: small exporters.

The results indicate that in general, firms that qualified for the tax incentive experienced an increase in employment but a decrease in the levels of capital investment between 2005 and 2008. This reduction in capital accumulation goes against the intended aim of policy makers who expected to see an increase in levels of investment in capital among SBCs. In addition, it confirms the findings of other authors (for example Engelschalk, 2004 and Knittel, 2005) who find little evidence to suggest that tax incentive stimulate investment among small firms. However, as previously mentioned firms are not homogenous and are likely to react differently to incentives based on their firm-level characteristics. Case in point, once the sample is broken down further between exporters and non-exporters, it is found that these results are driven by non-exporting SBCs. Exporting SBCs, on the other hand, reduced employment and increased capital investment over the period.

Exporters are relatively sensitive to changes in levels of capital, since they are more capital intensive that non-exporters. It is therefore not unlikely that policies which foster capital accumulation and increase disposable income, particularly among small exporters, will consequently support the continued survival of small exporters. This chapter further shows that exporters that qualified for the tax incentive were significantly more likely to continue exporting over the period than exporters who did not qualify. This unintended impact on exporting firms is an interesting finding, particularly given the high levels of failure among small exporting firms, as well as the poor consequences associated with failure.

This chapter serves as a reminder that when evaluating the effects of policies at a firm level, firms should not be treated as homogenous. Where other studies have found no support, this study finds some. It would be further beneficial to measure the costs associated with section 12E and compare them to the estimated benefits in order to evaluate the cost effectiveness of this tax incentive, however this analysis is left for future research. The analysis in this chapter at least indicates that the effectiveness of tax incentives must not be so easily dismissed, particularly given their unintended consequences.

Chapter 5

Conclusions

5.1. Summary

A significant obstacle for enhancing the literature on international trade and firm performance in South Africa has been a lack of access to large comprehensive datasets that track firms over time. Since the publication of the seminal 1995 paper by Bernard and Jensen, which was the first to highlight the importance of substantial and official firm-level data for investigating the differences between exporters and non-exporters, the international literature on firm performance and trade activities has grown exponentially (Wagner, 2011). South African studies, in contrast, have had to be satisfied with the use of either aggregate country-level datasets, or data from small cross-sectional sample surveys, to study the dynamics of firms and international trade. A direct consequence of these limited studies is the inability to give sufficiently detailed advice for export policy. Without a deeper understanding of the dynamics of exporting firms, policies which aim to encourage export growth and consequently economic growth will remain vague and potentially ineffectual.

This thesis adds significantly to the discussion of international activities and firm performance in the South African context. It does so first and by constructing a comprehensive dataset of manufacturing firms obtained from official, tax administration records of the South African Revenue Service (SARS). The construction of the dataset involves combining detailed population firm level data from corporate income tax records, employee income tax certificates and customs transactions records to generate a final dataset of firms from which both performance and trade activities can be measured. Therefore, for the first time in South African studies, the relationship between firm performance and exporting dynamics can be examined for population of manufacturing firms. The construction of this dataset not only contributed to the analysis in this thesis, but further to the overall construction of the SARS-National Treasury Firm-Level Panel – a unique and extensive source of data for the study of South African firm dynamics (Pieterse, Kreuser & Gavin, 2016).

The thesis then makes use of this dataset, as well as the substantial firm-level data from StatsSA’s Large Sample Survey, to analyse a number of research questions on the behaviour and characteristics of South African firms and, in particular, exporters.

Chapter 2 uses these unique, substantial datasets to re-examine the stylised facts of exporting in South Africa, with a focus on the exporting-productivity relationship. Despite it being a stylised fact that exporting firms exhibit a significant productivity premium relative to non-exporting firms, a handful of studies find otherwise. This is the case for South African exporters who are found to be, in general, no more productive in terms of TFPR than non-exporters, despite being larger, more labour productive and paying higher wages. South African evidence, however, is based on small and limited survey datasets (Edwards *et al*, 2008). This chapter contributes to the expanding micro-trade literature by making use of the two substantial, official firm-level datasets to determine whether this finding of a missing productivity premium still holds for South African exporters. It does so by following the now standard Bernard and Jensen (1999) methodology of estimating the Cobb Douglas productivity equation using OLS, Fixed Effects and the more recent Levinsohn-Petrin estimation techniques.

A number of possible explanations for the missing productivity premium have been suggested in the literature, however given the previous data limitations, few of these explanations have been adequately tested in the South African context. One such explanation is simply that an underlying sample bias is inherent in the data used in existing studies. Chapter 2 finds no evidence of a missing productivity premium when using more substantial, representative data: South African exporting firms do seem to be more productive than non-exporters in general. Other possible explanation for why previous studies were unable to identify a productivity premium is linked to export destination-heterogeneity (Rankin 2001 & 2013). By making use of the firm and export transactions data this chapter highlights the degree of heterogeneity among exporters: firms that export outside of Africa exhibit higher premiums, as do multiple-destination exporters. Yet, there is little evidence to suggest that exporting within Africa only, and exporting to only a single destination, generates a productivity premium greater than that of non-exporters. The findings in this chapter highlight the importance of both firm and exporter heterogeneity particularly from a policy perspective. This will be discussed in more detail in the following section.

The availability of such large rich datasets calls for further exploration. Specifically, Chapter 3 uses cluster analysis, an unsupervised machine learning methodology, to develop a typology of South African firms and exporters, in order to assess whether the traditional groupings of observations as defined by *a priori* approaches in the existing literature are in fact present in the South African firm-level data. In Chapter 3 the data is left to determine the classification of firms without imposing any structural assumptions. The identification and evaluation of homogeneous, distinct groups of firms is helpful for identifying segments in the manufacturing environment who could benefit from targeted policy interventions. The study in this chapter is one of the first to analyse South African population data through unsupervised techniques such as cluster analysis. Further, given the nature of the datasets used (mixed qualitative and quantitative variables) this Chapter makes use of a relatively novel algorithm called ClustOfVar which allows for dimensionality reduction on mixed datasets. This provides Chapter 3 with the opportunity to identify which aspects are the most important for distinguishing firms into clusters. Again, this is one of the first studies to use such a technique on firm-level data and provides added nuance to the traditional classification of South African firms and exporters.

The results of the cluster analysis in Chapter 3 confirm the previous South African classifications. Firms are grouped into small and medium/large exporters and non-exporters and exporters are grouped into within-Africa exporters, and out-of-Africa exporters. Firm size, export status and productivity play an important role in distinguishing firms into groups. Without imposing any *a priori* assumptions, the data confirms that heterogeneity is present not only among firms in general, but also among exporters themselves but also confirms that the groupings used in Chapter 2 are broadly correct.

Finally, Chapter 4 uses the rich StatsSA dataset to investigate the intended and unintended impact of tax incentives for small businesses, with specific focus on exporters. The incentive evaluated refers to section 12 E of the Income Tax Act No. 58 of 1962, the aim of which is to provide relief to small businesses in the form of lower, progressive tax rates as well as accelerated depreciation, both of which are unavailable for non-qualifying companies. The paper uses the novel, official dataset for South African manufacturing firms that merges information from StatsSA's 2005 and 2008 Large Sample Surveys to form a panel. By making use of the panel element of the data, as well as the natural experiment provided for by a change in the qualifying threshold of the tax incentive in 2006, this chapter is able to assess the impact of these types of tax incentives on small businesses and in particular exporters. The analysis finds evidence to suggest that while the tax incentive did not have the desired effects on all small businesses in terms of the accumulation of capital, it was beneficial to exporters. Not only did it encourage capital investment among exporters, but it also enabled exporting firms to remain in the export market. This is an important finding given the high failure rates experienced among exporters, particularly small exporters, as well as the negative effects associated with failure including reductions in productivity, output and employment (Wagner, 1997; Rankin, 2013).

The results of the analysis in the three main chapters of this thesis have some implications for policy. These will be summarised next, followed by a brief discussion on potential future directions.

5.2. Policy

Job creation, increased investment and economic growth are high on the agenda of South African policy makers (National Planning Commission, 2013). Export growth, particularly in the manufacturing sector, has significant potential to stimulate investment, productivity, employment and income and it is for this reason that export stimulation is one of the key goals of South Africa's growth policy (see for example the NEDP, IPAP and MTFS 2014-2019). However, despite the emphasis on exports in government's growth strategies, limited details on how this will happen have been provided. Part of the reason for this is that very little is known about export dynamics at a micro-level in South Africa (Rankin, 2013). In order to advocate policies that support and foster exporting, policy makers require information about the underlying characteristics of exporting firms, particularly at the firm-level. This thesis contributes to this discussion on export policy by making use of comprehensive official data on

South African firms to develop a deeper understanding their dynamics, particularly in terms of exporting.

Chapters 2 and 3 highlight the presence of firm and exporter heterogeneity in the data. This has important consequences for policies and interventions that are too general in the sense that they treat all firms, or exporters, as homogeneous. The analysis in this thesis shows aspects that are important to any one group of firms may be quite different to what is important to a different group of firms. Consequently the true impact of a policy that does not distinguish between these groups is likely to remain unknown.

According to the cluster analysis of Chapter 3 size is a major factor for grouping firms followed by export status and productivity. Therefore, if a policy is to be geared towards targeting specific groups of firms then size, not sector, is the correct dimension to do this by. This finding is in direct contrast to South African industrial policy, such as the IPAP, where sector is an important consideration for targeting. Further, these results imply that if policy is to encourage smaller firms to become exporters, these firms would either need to grow or become more productive.

Productivity has further implications for policy, particularly in the presence of exporter heterogeneity. Chapter 2 shows that not only do exporting firms exhibit a significant productivity premium relative to non-exporting firms, but also that heterogeneity among exporters is evident in the data in terms of the destination of exports. This is related, in turn, to firm characteristics, such as size, wages, etc. as well as productivity. The policy implications of these results are as follows. Firstly, the finding of exporter-heterogeneity means that policies aimed at export growth need to acknowledge that there are different types of exporters, who have different behaviours and characteristics. Policy must therefore first identify the type of exporters it wants to support and then implement a more targeted strategy to encourage this type of behaviour.

Secondly, policies which aim to grow exports (whether through encouraging firms to become exporters or through stimulating exports among exporters), are not likely to perform well if the targeted firms do not already exhibit the underlying characteristics of successful exporters, particularly in terms of productivity. In general, the analysis in this thesis indicates that productivity is important for exporting, particularly when exporting to multiple destinations outside of Africa. Therefore, a policy which aims to facilitate an increase in productivity levels within firms will further enable more firms to enter exporting and expand their markets. Furthermore, policies that increase productivity have benefits beyond exporting: domestic firms with higher levels of productivity are better equipped to compete against foreign competition.

Acknowledging that firms and exporters differ among themselves is important not just for implementing policy, but also for assessing the impact of policy particularly at the firm-level. Chapter 4 of this thesis is an example of this. It illustrates how treating firms, and exporters, as homogeneous when analysing the effect of policy interventions can lead to inconclusive results. Many studies, for example, find no support for the use of tax incentives in stimulating investment and growth in small businesses. The analysis in Chapter 4 however does find evidence in support of the tax incentive, particularly when taking into account the firm-heterogeneity in the data. Furthermore, the analysis identifies a potential unintended consequence of the tax incentive for small exporters: exporters who qualified for the tax incentive were significantly more likely to continue exporting over the period than exporters who did not qualify. From a policy point of view, these results indicate that while such tax incentives may not stimulate growth among small businesses in general, they are a good starting point for policies aimed at supporting firm survival in the export market.

5.3. Future Directions

While the three main chapters of this thesis contribute significantly to the discussion of export dynamics in South Africa, each chapter could benefit from further research. Firstly, Chapter 2 concludes that destinations matter for exports. The existing literature on South African exporters further shows that exporters differ in terms of the amount and value of exports. It would therefore be of interest to analyse the levels of export-intensity among exporters in terms of the destinations served as well as the value of these exports. This would allow for a potentially interesting investigation into the dynamics of what type of exports are going to which destination. This in itself has important implications for policy. The firm and customs transactions data provide sufficient detail to answer this type of question.

Further, another topic often researched within the export-productivity literature discussed in Chapter 2 is that of causality. Bernard and Jensen (1999) propose two alternative, but not mutually exclusive, premises for why exporters tend to be more productive than non-exporters. The first is the self-selection hypothesis: firms that are productive choose to export. The second is the learning-by-exporting hypothesis: it is the entry of firms into the export market that raises their productivity. Again, this is important from a policy perspective: should export policies aim to improve productivity of firms if they wish to grow exporting, or is it the activity of exporting that ultimately leads to firms becoming more productive? This question has yet to be adequately researched in South Africa, given the data constraints. At the time of writing Chapter 2, there was an insufficient time dimension to allow for an investigation of these dynamics. However, as the panel dimension of the SARS dataset develops, it will enable the much needed future research into South African exporter dynamics at the firm-level.

Secondly, Chapter 3 not only identified groups of firms similar to those traditionally classified in the literature, but additionally identified a number of small outlier groups, for example: the high performing super-exporters, as well as a group of small, high technology non-exporters who pay substantially higher wages relative to a number of other types of firms. A next step could be to employ some type of discriminant analysis on these outlier groups to determine if they differ significantly in terms of their behaviours and dynamics. This analysis could lead to future acknowledgement of even higher degrees of heterogeneity among firms, which is important for determining where to target policy in order to appropriately stimulate economic growth and employment. Case studies, through qualitative interviews, of these groups should also help understand their behaviour and characteristics in more detail.

It is further noted that the analysis in Chapter 3 was plagued by missing data. It would be beneficial to re-examine the analysis on a more complete dataset. This could be done in two ways. The first way requires some imputation of the missing data, which could be done by a simple means-substitution, or by using a multivariate imputation by chained equations (from the `MICE` package in R, see van Buuren and Groothuis-Oudshoorn (2011)) where plausible data substitutions are drawn from a distribution specifically designed for each missing data point. Alternatively, given that the collection of the tax administration data is an ongoing process, and given the tendency of firms to hands in tax returns late, the analysis could be repeated on updated versions of the data as, and when, they become available. Validation of the results is indeed an important step in exploratory data analysis.

Finally, the results of Chapter 4 can be augmented by further research. In particular, the analysis found that while South Africa's tax incentive, section 12E, may have played a role in encouraging employment among small firms, it did little to stimulate investment in physical capital. Indeed, the chapter suggests that section 12E may have facilitated disinvestment. Potentially, this result could have something to do with the net income- and substitution-effect. According to the theory, the substitution effect of the tax incentive would be an increase in capital investment and potentially a reduction in employment in small firms. However, the tax incentive also reduced the overall tax liability for small firms and therefore allowed firms to maintain a higher after-tax profit. The income effect therefore would likely be an increase in employment potentially by more than capital. It would be of interest to determine which effect is greater. This would add another dimension to the debate on the effectiveness of tax incentives. Further, it would beneficial to estimate the costs associated with section 12E, and compare them to the estimated benefits in order to evaluate the cost effectiveness of this tax incentive.

This thesis has highlighted the importance of recognising heterogeneity among firms in general, and exporters in particular. It has done so through the use of comprehensive firm-level data. This data has enabled a detailed investigation into the dynamics of firms in South Africa, with a focus on exporting. However, this firm level data provides a great opportunity for future studies across many other topics on firm dynamics. This is particularly true of the tax administrative data which not only provides a much larger sample size that traditional survey data sources, but further is less like to suffer from the high non-response rates and under-reporting usually inherent in existing survey data sources due to the compulsory requirement of firms to submit his information (Pieterse, Kreuser & Gavin, 2016). In addition, the tax administration data is available for multiple years. This provides opportunities for researchers to track firms over time, which enables the more cutting-edge empirical research necessary for credible policy evaluation (Card et al, 2011).

5.4. Final Remarks

The development of the literature on South African firms has been significantly hampered by access to good, longitudinal firm-level data. This thesis overcomes this limitation by making use of two recent, unique sources of data from official government agencies to examine the dynamics of South African exporters at a level previously reserved for international studies. The findings of this thesis highlight the importance and usefulness of comprehensive firm-level data and serve to illustrate the useful conclusions that can be made about the behaviour of South African firms when data at this level has been provided for research. The final words of this thesis are therefore to implore government to increase the (secured) access to such data, so that researchers and policy-makers alike can obtained a more rigorous and detailed understanding of South African firms.

References

Abor, J., & Quartey, P. (2010). Issues in SME Development in Ghana and South Africa. *International Research Journal of Finance and Economics*(39), 218-228.

Ahmad, A., & Dey, L. (2007). A k-mean Clustering Algorithm for Mixed Numeric and Categorical Data. *Data & Knowledge Engineering, 63,* 503–527.

Alexandru, A., Strat, V., & Gogonea, R. (2012). Statistical Analysis of Romanian Insurance Market. A Gross Written Premiums Perspective. *Journal of Social and Economic Statistics, 2*(1).

Alvarez, R., & López, R. (2008). Entry and Exit in International Markets: Evidence from Chilean Data. *Review of International Economics, 16*(4), 692–708.

Amiti, M., & Khandelwal, A. (2013). Import Competition and Quality Upgrading. *Review of Economics and Statistics, 95*(2), 476-490.

Anderberg, M. (1973). *Cluster Analysis for Applications.* New York: Academic Press.

Andersson, M., Lööf, H., & Johansson, S. (2008). Productivity and International Trade: Firm level evidence from a small open economy. *Review of World Economics, 144*(4), 774-801.

Andrews, R., & Currim, I. (2003). Recovering and profiling the true segmentation structure in markets: an empirical investigation. *International Journal of Research in Marketing, 20*, 177-192.

Apili Ejupu, E., Kirsten, J., & Makhura, M. (1999). Food security in a Commercialising Rural Economy : Initial findings from a case study of two districts in Uganda . *Agrekon, 38*(4), 817-838.

Armeanu, D., & Lache, L. (2008). Application of the Model of Principal Components on Romanian Insurance Market. *Theoretical and Applied Economics*, 11-20.

Aw, B., & Hwang, A. (1995). Productivity and the Export Market: A Firm-Level Analysis. *Journal of Development Economics*, 313-332.

Aylward, D., & Glynn, J. (2006). SME Innovation within the Australian Wine Industry: A Cluster Analysis. *Small Enterprise Research: The Journal of the SMAANZ (Small Enterprise Association of Australia and New Zealand), 14*(1), 42-54.

Ayramo, S., & Karkkainen, T. (2006). Introduction to Partitioning Based Clustering Methods with a Robust Example. *Reports of the Department of Mathematical Information Technology, University of Jyvaskyla*.

Babu, G. J., & Feigelson, E. D. (1997). *Statistical Challenges in Modern Astronomy II.* Springer-Verlag, New York.

Bailey, K. (1992). Typologies. In *Encyclopedia of Sociology* (pp. 2188-2194). New York: Macmillan: E. F. Borgatta and M. L. Borgatta (eds.).

Bailey, K. (1994). Typologies and Taxonomies: An Introduction to Classification Techniques. *Sage University Paper series on Quantitative Applications in the Social Sciences*, 07-102.

Baldwin, R. E., & Robert-Nicoud, F. (2008). Trade and growth with heterogeneous firms. *Journal of International Economics*, 21-34.

Baldwin, R., & Harrigan, J. (2007). Zeros, Quality and Space: Trade Theory and Trade Evidence. NBER working paper 13214.Cambridge, MA: National Bureau of Economic Research. Retrieved March 7, 2015, from http://www.nber.org/papers/w10344

Bastos, P., & Silva, J. (2010). The Quality of a Firm's Exports: Where you Export to Matters. *Journal of International Economics, 82*(2), 99-111.

Baxter, M. (2009). Archaeological Data Analysis and Fuzzy Clustering. *Archaeometry, 51*, 1035–1054.

Beckstead, J. (2002). Using Hierarchical Cluster Analysis in Nursing Research. *Western Journal of Nursing Research, 24*(3), 307-319.

Bernard, A. B., Redding, S. J., & Schott, P. K. (2007). Comparative Advantage and Heterogeneous Firms. *The Review of Economic Studies, 74*(1), 31-66.

Bernard, A. B., Redding, S. J., & Schott, P. K. (2011). Multiproduct firms and trade liberalization. *Quarterly Journal of Economics, 126*(3), 1271-1318.

Bernard, A., & Jensen, B. (1995). Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987. *Brookings Papers on Economic Activity. Microeconomics*, 67-119.

Bernard, A., & Jensen, J. (1999). Exceptional Exporter Performance: Cause, effect, or both? *Journal of International Economics, 47*, 1–25.

Bernard, A., & Jensen, J. (2004, May). Why Some Firms Export. *The Review of Economics and Statistics, 86*(2), 561–569.

Bernard, A., & Wagner, J. (1997). Exports and Success in German Manufacturing. *Weltwirtschaftliches Archiv, 133*(1), 134-157.

Bernard, A., Eaton, J., & Kortum, S. (2003). Plants and productivity in international trade. *The American Economic Review, 93*(4), 268-1290.

Bernard, A., Jensen, J., Redding, S., & Schott, P. (2007). Firms in International Trade. *Journal of Economic Perspectives*, 105-130.

Bernstein, J., & Shah, A. (1993). Corporate Tax Structure and Production. *World Bank Policy Research Working Paper Series, WPS 1196*. Retrieved August 15, 2014, from http://econ.worldbank.org/external/default/main?pagePK=64165405&theSitePK=469372&piPK=64165420&menuPK=676714&searchTxt=&docType=1&toDate=&date=&fr_id=&docTY=620265%20563787%20581952&sortDesc=DOCDT&dAtts=REPNB,%09REPNB,DOCNA,DOCTY,BG_AUTHR,&startPoint=6501

Besedes, T., & Blyde, J. (2010). What drives export survival? An analysis of export duration in Latin America. *Draft, Inter-American Development Bank.*

Besedes, T., & Prusa, T. (2007). The Role of Extensive and Intensive Margins and Export Growth. *NBER Working Paper 13628*.

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J., . . . Zeufack, A. (2004). Do African Manufacturing Firms Learn from Exporting? *The Journal of Development Studies, 40*(3), 115-141.

Bird, R. (1996). Why Tax Corporations? *Technical Committee on Business Taxation, Department of Finance Working Paper 96-2*.

Bird, R. (2008). Tax Challenges Facing Developing Countries. *Institute for International Business Working Paper No. 9*. Retrieved March 13, 2014, from http://dx.doi.org/10.2139/ssrn.1114084

Blalock, G., & Gertler, P. J. (2004). Learning from Exporting Revisited in a Less Developed Setting. *Journal of Development Economics*, 397– 416.

Blažić, H. (2004). *Tax Compliance Costs of Small Business in Croatia.* Zagreb, Croatia: Institute of Public Finance. Occasional paper/ Institute of Public Finance, 22.

Boermans, M. (2013). Learning-by-Exporting and Destination Effects: Evidence from African SMEs. *Applied Econometrics and International Development, 13*(2), 155-173.

Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V., & Walther, G. (2007). Editorial:Advances in Mixture Models. *Computational Statistics & Data Analysis*, *51*, 5205 – 5210.

Booysen, N. (1997). The Impact of the South African Tax System on Small and Medium Sized Enterprises. *(Unpublished doctoral dissertation)*, Randse Afrikaanse Universiteit, Johannesburg. Retrieved August 26, 2014, from https://ujdigispace.uj.ac.za/handle/10210/5874

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees.* Monterey, Calif., U.S.A.: Wadsworth, Inc.

Brida, J., Fasone, V., Scuderi, R., & Zapata-Aguirre, S. (2014). ClustOfVar and the Segmentation of Cruise Passengers from Mixed data: Some managerial implications. *Knowledge-Based Systems*, *70*, 128–136.

Bunting, I., Sheppard, C., Cloete, N., & Belding, L. (2010). *Performance Indicators in South African Higher Education 2000–2008.* Wynberg: Center for Higher Education Transformation(CHET).

Cadot, O., Iacovone, L., Pierola, D., & Rauch, F. (2013). Success and Failure of African Exporters. *Journal of Development Economics, 101*, 284–296.

Calinski, T., & Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Communications, 3*(1), 1-27.

Card, D. R., Chetty, M., Feldstein, & Saez, E. (2010). Expanding Access to Administrative Data for Research in the United States. Paper submitted to the National Science Foundation as part of its SBE 2020 planning activity. Retrieved May 15, 2016, from www.nsf.gov/sbe/sbe_2020/

Castellani, D. (2002). Export behavior and productivity growth: Evidence from Italian manufacturing firms. *Weltwirtschaftliches Archiv, 138*(4), 605-628.

Castellani, D., Serti, F., & Tomasi, C. (2010). Firms in International Trade: Importers’ and Exporters’ Heterogeneity in Italian Manufacturing Industry. *The World Economy, 33*(3), 424-457.

Castellares, R. (2015). Competition and Quality Upgrading in Export Markets: The case of Peruvian Apparel Exports. *Banco Central de Reserva del Perú Working Paper Series 2015-010*. Retrieved October 20, 2016, from http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2015/documento-de-trabajo-10-2015.pdf

Cebeci, T. (2014). Impact of Export Destinations on Firm Performance. World Bank Policy Research Working Paper no. 6743. Retrieved August 10, 2015, from https://openknowledge.worldbank.org/bitstream/handle/10986/16819/WPS6743.pdf

Cebeci, T., Fernades, A., Freuend, C., & Pierola, M. (2012). Exporter Dynamics Database. *World Bank Policy Research Working Paper No 6229*.

Chakrapani, C. (2004). *Statistics in Market Research.* Arnold, London: Wiley.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software, 61*(6), NbClust Package. R package version 1.4 <http://CRAN.R-project.org/package=NbClust>.

Chavent, M., Kuentz, V., Liquet, B., & Saracco, J. (2012). ClustOfVar: an R package for the clustering of variables. *Journal of Statistical Software, 50*(13), 1-16.

Clerides, S. K., Lach, S., & Tybout, J. R. (1998, August). Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco. *The Quarterly Journal of Economics, 113*(3), 903-947.

Commission on Growth and Development. (2008). *The Growth Report: Strategies for Sustained Growth and Inclusive Development.* Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Corporal-Lodangco, I., & Leslie, L. (2016). Cluster Analysis of Philippine Tropical Cyclone Climatology: Applications to Forecasting. *J Climatol Weather Forecasting, 4*.

Cortese, J. D. (2000). The array of today: biomolecule arrays become the 21st century test tube. *The Scientist*, *14*.

Davis Tax Committee. (2014). *Small and Medium Enterprises: Taxation Considerations, Interim Report.* Retrieved September 15, 2014, from National Treasury: http://www.treasury.gov.za/TaxReviewCommittee/DTC%20SME%20Report%20for%20Public%20Comment%2014%20July%202014.pdf

Davis, D. (1995). Intra-industry trade: A Heckscher-Ohlin-Ricardo approach. *Journal of International Economics, 29*(3), 201-226.

De Loecker, J. (2007). Do Exports generate higher productivity? Evidence from Slovenia. *Journal of International Economics, 73*(1), 69-98.

Department for Economic Development. (2011). *The New Growth Path: The framework.* Retrieved May 6, 2013, from http://www.economic.gov.za/communications/publications/new-growth-path-series

Department of Planning, Monitoring and Evaluation. (2014). *Medium Term Strategic Framework: 2014-2019.* Pretoria: The Presidency.

Department of Trade and Industry. (2010). *Industrial Policy Action Plan: Economic Sectors and Employment Cluster—IPAP 2012/13–2014/15.* Pretoria, South Africa: dti.

Department of Trade and Industry. (2013). *National Exporter Development Plan.* Pretoria.

Deraniyagala, S., & Fine, B. (2001). New Trade Theory versus Old Trade Policy: a continuing enigma. *Cambridge Journal of Economics*, 809-825.

Diaz-Bonilla, E., Thomas, M., Robinson, S., & Cattaneo, A. (2000). Food security and trade negotiations in the World Trade Organization: A Cluster Analysis of Country Groups. *Trade and Macroeconomics Division International Food Policy Research Institute*, TMD Discussion Paper No. 59.

Dixit, A., & Pindyck, R. (1994). *Investment under Uncertainty.* Princeton, NJ: Princeton University Press.

Dornbusch, R., Fischer, S., & Samuelson, P. (1977). Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review, 67*(5), 823-839.

Driver, H., & Kroeber, A. (1932). Quantitative Expression of cultural relationships. *University of California Publications in American Archaeology and Ethnology 31*, 211-256.

Eaton, J., & Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica, 70*(5), 1741-1780.

Eaton, J., Eslava, M., Kugler, M., & Tybout, J. (2007). Export Dynamics in Colombia: Firm-Level Evidence. *National Bureau of Economic Research No.13531*.

Edwards, L., & Lawrence, R. (2012). *South African Trade Policy and the Future Global Trading Environment.* Johannesburg: SAIIA Occasional Paper No 128. Retrieved February 26, 2014, from http://www.saiia.org.za/occasional-papers/20-south-african-trade-policy-and-the-future-global-trading-environment/file

Edwards, L., Rankin, N., & Schöer, V. (2008). South African Exporting Firms: What do we know and what should we know? *Journal of Development Perspectives, 4*(1), 67-92.

Ee, C. Y. (2016). Export-Led Growth Hypothesis: Empirical Evidence from Selected Sub-Saharan African Countries. *Procedia Economics and Finance, 35*, 232 – 240.

Engelschalk, M. (2004). Creating a Favorable Tax Environment for Small Business. *Contributions to Economic Analysis, 268*, 275–311.

Eragbhe, E., & Modugu, K. (2014). Tax Compliance Costs of Small and Medium Scale Enterprises in Nigeria. *International Journal of Accounting and Taxation, 2*(1), 63-87.

Escaith, H., & Gaudin, H. (2014). Clustering Value-Added Trade: Structural and Policy Dimensions. *World Trade Organization Economic Research and Statistics Division Staff Working Paper No. ERSD-2014-08*.

Ester, M., Ge, R., Gao, B., Hu, Z., & Ben-Moshe, B. (2008). Joint Cluster Analysis of Attribute Data and Relationship Data the Connected k-Center Problem. *ACM Transactions on Knowledge Discovery from Data, 2*(2), 246-257.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). King's College London, UK: Wiley.

Falk, M., Hölzl, W., & Oberhofer, H. (2015). The Importance of Firm-Level Data for Empirical Research and Economic Policy Consulting. *WIFO Monatsberichte (monthly reports), 88*(11), pp. 845-857.

Farmer, A., McGuffin, P., & Spitznagel, E. L. (1983). Heterogeneity in Schizophrenia: A Cluster Analytic Approach. *Psychiatric Research, 8*, 1-12.

Faúndez-Abans, M., Ormeno, M. I., & de Oliveira-Abans, M. (1996). Classification of Planetary Nebulae by Cluster Analysis and Artificial Neural Networks. *Astronomy and Astrophysics Supplement, 116*, 395-402.

Fernandes, A., & Isgut, A. (2005). Learning-by-doing, learning-by-exporting, and productivity: evidence from Colombia. World Bank Policy Research Working Paper no.3544. Retrieved August 10, 2015, from http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2005/03/24/000012009_20050324092129/Rendered/PDF/wps3544.pdf

FIAS. (2006). *Sector Study of the Effective Tax Burden: South africa.* Retrieved August 26, 2014, from https://www.wbginvestmentclimate.org/uploads/SouthAfricaTaxBurdenStudy.pdf

FIAS. (2007). *South Africa: Tax Compliance Burden for Small Businesses: A survey of tax practitioners.* Retrieved August 29, 2013, from https://www.wbginvestmentclimate.org/uploads/FIAS_Tax_Practitioners_Report_-_FINAL_29+Aug.pdf

Freund, C., & Pierola, M. (2010). Export Entrepreneurs: Evidence from Peru. *World Bank Policy Research*, Working Paper 5407.

Freund, C., & Pierola, M. (2012). Export Superstars. *World Bank Policy Research Working Paper No. 6222*.

Friedman, H., & Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association, 62*(320), 1159-1178.

Friedman, J., Mulder, N., Faúndez, S., Caldentey, E., Yévenes, C., Velásquez, M., . . . Reinecke, G. (2011). Openness, Wage Gaps and Unions in Chile: A Micro Econometric Analysis. *OECD Trade Policy Working Papers*, No. 134.

Fuchs, M., Iacovone, L., Jaeggi, T., Napier, M., Pearson, R., Pellegrini, G., & Sanchez, C. (2011). *Financing Small and Medium Enterprises in the Republic of South Africa.* Washington DC: The World Bank (AFTFP) Department.

Garanti, Z., & Zvirbule - Berzina, A. (2013). Regional Cluster Initiatives as a Driving Force for Regional Development. *European Integration Studies, 7*.

García, F., Grigonytė, I., & Oliver, J. (2015). Trade Relationship Analysis Among EU Members by Means of Cluster Analysis. *1st International Conference on Business Management.* Universitat Politècnica de València: DOI: http://dx.doi.org/10.4995/ICBM.2015.1313.

Ghironi, F., & Melitz, M. (2005). International Trade and Macroeconomic Dynamics with Heterogeneous Firms. *Quarterly Journal of Economics, 120*, 865-915.

Goolsbee, A. (2000). The Importance Of Measurement Error In The Cost Of Capital. *National Tax Journal*, 215-228.

Gower, J. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics, 27*(4), 857-874.

Granér, M., & Isaksson, A. (2009). Firm efficiency and the destination of exports: evidence from Kenyan plant-level data. *The Developing Economies, 47*(3), 279–306.

Greenaway, D., & Kneller, R. (2007). Firm Heterogeneity, Exporting and Foreign Direct Investment. *The Economic Journal*, F134–F161.

Greenaway, D., Gullstrand, J., & Kneller, R. (2005). Exporting May Not Always Boost Firm Productivity. *Review of World Economics, 4*, 561–582.

Guenther, G. (2012). *Section 179 and Bonus Depreciation Expensing Allowances: Current Law, Legislative Proposals in the 113th Congress, and Economic Effects.* Congressional Research Service. Issue 7-5700.

Gumede, V. (2000). Growth and Exporting of Small and Medium Enterprises in South Africa Some Thoughts on Policy and Scope for Further Research. Paper presented at the Trade and Industrial Policy Secretariat Annual Forum 18 – 20 September. Retrieved April 9, 2014, from www.ajournal.co.uk/EFpdfs/EFvolume6(2)/EFVol.6%20(2)%20Article%203.pdf

Hall, R., & Jorgenson, D. (1967). Tax policy and Investment Behavior. *American Economic Review, 57*(3), 391-414.

Hallak, J., & Sivadasan, J. (2006). Productivity, Quality and Exporting Behavior Under Minimum Quality Requirements. *Unpublished NBER (University of Michigan and Universidad de San Andres)*. Retrieved August 10, 2015, from https://mpra.ub.uni-muenchen.de/24146/1/MPRA_paper_24146.pdf

Hasset, K. A., & Hubbard, R. G. (2002). Tax Policy and Business Investment. In A. Auerbach, & M. Feldstein, *Handbook of Public Economics 3* (pp. 1293-1343). Amsterdam: North- Holland.

Hausmann, R., Hwang, J., & Rodrik, D. (2007). What you export matters. *Journal of Economic Growth, 12*(1), 1-25.

Hausmann, R., & Klinger, B. (2008). South Africa's Export Predicament. Economics of Transition. *Economics of Transition, 16*(4), 609-637.

Hay, P. J., Fairburn, C. G., & Doll, H. A. (1996). The Classification of Bulimic Eating Disorders: A Community-Based Cluster Analysis Study. *Psychological Medicine, 26*, 801–812.

Henn, C., Papageorgiou, C., Spatafora, N. (2015). Export quality in advanced and developing economies: Evidence from a new dataset. *WTO Staff Working Paper ERSD-2015-02*. Retrieved October 20, 2016, from https://www.wto.org/english/res_e/reser_e/ersd201502_e.pdf.

Hollenstein, H. (2003). Innovation Modes in the Swiss Service Sector A Cluster Analysis Based on Firm-level Data. *Research Policy, 32*, 845-863.

Howard, E., Newman, C., Rand, J., & Tarp, F. (2014). Productivity-Enhancing Manufacturing Clusters: Evidence From Vietnam. *WIDER Working Paper 2014/071*, Helsinki: UNU-WIDER.

Iacovone, L., & Javorcik, B. (2012). Getting Ready: Preparation for Exporting. CEPR Discussion Paper 8926. Retrieved October 20, 2016, from http://users.ox.ac.uk/~econ0247/Tequila.pdf.

Income Tax Act No. 58 of 1962. (n.d.).

Isgut, A. (2001). What's Different about Exporters? Evidence from Colombian Manufacturing. *Journal of Development Studies, 37*(5), 57-82.

Jain, Murty, & Flynn. (1999). Data Clustering A Review. *ACM Computing Surveys, 31*(3), 264-323.

James, S. (2009). Incentives and Investments: Evidence and Policy Implications. *FIAS, The World Bank Group*.

Jang, W., & Hendry, M. (2007). Cluster Analysis of Massive Datasets in Astronomy. *Journal of Statistics and Computing, 17*(3), 253-262.

Johnson, R., & Wichern, D. (1998). *Applied Multivariate Statistical Analysis* (Fourth ed.). Prentice Hall.

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

Kaufman, L., & Rousseeuw, P. (2005). *Finding Groups in Data. An introduction to cluster analysis.* Hoboken, NY: Wiley.

Kharlamova, G., & Vertelieva, O. (2013). The International Competitiveness of Countries: Economic-Mathematical Approach. *Economics & Sociology, 6*, 39-52. DOI: 10.14254/2071-789X.2013/6-2/4.

Kiers, H. (1991). Simple structure in Component Analysis Techniques for mixtures of qualitative and quantitative variables. *Psychometrika, 56*, 197–212.

Klemm, A. (2010). Causes, Benefits and Risks of Tax Incentives. *International Tax and Public Finance, 17*(3), 315-336. doi:10.1007/s10797-010-9135-y

Klemm, A., & Van Parys, S. (2009). Empirical Evidence on the Effects of Tax Incentives. *International Monetary Fund (IMF), Working Paper WP/09/136*.

Knittel, M. (2005). Small Business Utilization of Accelerated Tax Depreciation: Section 179 expensing and bonus depreciation. *National Tax Association Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 273-286.

Kox, H., & Rojas-Romagosa, H. (2010). Exports and Productivity Selection Effects For Dutch Firms. *De Economist*, 295-322.

Krugman, P. R. (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics*, 469-479.

Krugman, P. R. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, 950–959.

Krugman, P. R., & Obstfeld, M. (2006). *International Economics: Theory and Policy* (Seventh ed.). New York: Pearson Addison Wesley.

Lagendijk, A., Charles, & D. (1999). Clustering as a New Growth Strategy for Regional Economies? A Discussion of New Forms of Regional Industrial Policy in the United Kingdom. In *Boosting Innovation: the Cluster Approach* (pp. 127-153). Paris: OECD Publishing.

Larson, J. S., Bradlow, E. T., & Fader, P. S. (2005). An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing, 22*(4), 395–414.

Lawrence, R., & Edwards, L. (2008). South African Trade Policy Matters: Trade Performance & Trade Policy, *Economics of Transition*, *16*(4): 585-608.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies, 70*(2), 317–41.

Lileeva, A., & Trefler, D. (2010). Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants. *The Quarterly Journal of Economics, 125*(3), 1051-1099.

Lombardo, R., & Falcone, M. (2011). Crime and Economic Performance. A Cluster Analysis of Panel Data on Italy's NUTS 3 Regions. *Università della Calabria, Dipartimento di Economia, Statistica e Finanza (Ex Dipartimento di Economia e Statistica)*, Working Paper No 201112.

Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate Analysis.* London: Academic Press Inc.

Martincus, C., & Carballo, J. (2010). Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and the Extensive Margins of Exports. *Inter-American Development Bank*, IDB working paper series 201.

Masso, J., & Vahter, P. (2011). Exporting and Productivity: The Effects of Multi-Market and Multi Product Export Entry. University of Tartu Faculty of Economics and Business Administration Working Paper No. 83 - 2011. Retrieved November 10, 2015, from http://www.mtk.ut.ee/sites/default/files/mtk/RePEc/mtk/febpdf/febawb83.pdf

Matthee, M., & Krugell, W. (2012). Barriers to Internationalisation: Firm-Level Evidence from South Africa. *Studia UBB Oeconomica, 57*(1), 3-20.

Matthee, M., Farole, T., Naughtin, T., & Rankin, N. (2016). South African Exporters and the Global Crisis: Intensive Margin Shock, Extensive Margin Hangover. *South African Journal of Economics, 84*, 183–198. doi: 10.1111/saje.12094.

Mayer, T., & Ottaviano, G. (2008). The Happy Few: The Internationalisation of European Firms . *Intereconomics*, 135-148.

Mayer, T., Melitz, M., & Ottaviano, G. (2011). Market size, competition, and the product mix of exporters. NBER Working Paper 16959. Cambridge, MA: National Bureau of Economic Research. Retrieved March 14, 2015, from http://www.nber.org/papers/w16959.pdf

McParland, D., Gormley, I., McCormick, T., Clark, S., Kabudula, C., & Collinson, M. (2014). Clustering South African households based on their asset status using latent variable models. *The Annals of Applied Statistics, 8*(2), 747–776.

Melitz, M. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 1695–1725.

Melitz, M. (2008). *International Trade and Hetrogeneous Firms.* New Palgrave Dictionary of Economics, 2nd Edition. Palgrave Macmillan.

Melitz, M. J., & Ottaviano, G. I. (2008). Market Size, Trade, and Productivity. *Review of Economic Studies, 75*, 295–316.

Melitz, M., & Redding, S. (2014). *Heterogeneous Firms and Trade. Handbook of International Economics* (Fourth ed., Vol. 4). Elsevier.

Milligan, G., & Cooper, M. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika, 50*(2), 159-179.

Minas, C., Jacobson, D., Antoniou, E., & McMullan, C. (2014). Welfare Regime, Welfare Pillar and Southern Europe. *Journal of European Social Policy, 24*(2), 135-149.

Montalbano, P., & Nenci, S. (2011). Trade Patterns and Trade Clusters: China, India, Brazil and South Africa in the Global Trading System. Working paper 125/2011, University of Roma Tre (Italy).

Montalbano, P., & Nenci, S. (2014). The Trade Competitiveness of Southern Emerging Economies: A Multidimensional Approach Through Cluster Analysis. *The World Economy, 37*(6), 783-810.

Mooi, E., & Sarstedt, M. (2011). Chapter 9: Cluster Analysis. In *A Concise Guide to Market Research* (pp. 237-284). Berlin Heidelberg: Springer Texts in Business and Economics.

Moroko, L., & Uncles, M. D. (2009). Employer branding and market segmentation. *Journal of Brand Management, 17*(3), 181–196.

Mucha, H., Bartel, H., & Dolata, J. (2005). Model-based Cluster Analysis of Roman Bricks and Tiles from Worms and Rheinzabern. In *C.Weihs and W.Gaul (eds.). Classification - the Ubiquitous Challenge* (pp. 317-324). Berlin: Springer.

Mukherjee, S. (2014). The Impact of Trade Liberalization on SMEs versus Large Firms. *Paper accepted in 10th Annual Conference on Economic Growth and Development.* Delhi Centre, India. Retrieved October 2, 2015, from http://www.isid.ac.in/~epu/acegd2014/papers/SubhadipMukherjee.pdf

Murray, M. (2016). Chapter 9: Factor analysis, cluster analysis, and nonparametric research methods for heterodox economic analysis. In *Handbook of Research Methods and Applications in Heterodox Economics* (pp. 190–209). Cheltenham UK: Edward Elgar Publishing.

Murtagh, F. (2014). History of cluster analysis. In *M. Greenacre and J. Blasius, eds., The Visualization and Verbalization of Data.* CRC/Chapman and Hall, in press.

Murtagh, F., & Kurtz, M. (2016). The Classification Society's Bibliography Over Four Decades: History and Content Analysis. *Journal of Classification, 33*(1), 6-29.

Muûls, M., & Pisu, M. (2009). Imports and Exports at the Level of the Firm: Evidence from Belgium. *The World Economy, 32*(5), 692–734.

National Planning Commission. (2013). *National Development Plan 2030 "Our Future - make it work".* Pretoria: The Presidency, Republic of South Africa.

National Treasury. (2001). *Budget Review.* Pretoria: National Treasury of the Republic of South Africa.

National Treasury. (2005). *Budget Review.* Pretoria: National Treasury of the Republic of South Africa.

National Treasury. (2014). *Budget Review.* Pretoria: National Treasury of the Republic of South Africa.

Naudé, W. (2000). The determinants of South African exports: An econometric analysis'. *South African Journal of Economics, 68*(2), 103–113.

Naughtin, T., & Rankin, N. A. (2014). South African super-exporters: Are they different and what does this mean for policy? Retrieved August 2015, 10, from http://www.tips.org.za/files/south_africas_super_exporters_-_naughtin_and_rankin.pdf

Neary, J. P. (2009). Putting the "New" into New Trade Theory: Paul Krugman's Nobel Memorial Prize in Economics. *Scandinavian Journal of Economics, 111*(2), 217-250.

Niakšu, O. (2013). Calculating distance measure for clustering in multi-relational settings. *16th international multiconference information society. Vol. A.* Ljubljana. Retrieved February 10, 2016, from http://ailab.ijs.si/dunja/SiKDD2013/Papers/Niaksu%20-%20Calculating%20distance%20measure%20for%20MRDM%20clustering.pdf

Niakšu, O. (2013). Calculating distance measure for clustering in multi-relational settings. *16th international multiconference information society. Vol. A.* Ljubljana. Retrieved February 10, 2016, from http://ailab.ijs.si/dunja/SiKDD2013/Papers/Niaksu%20-%20Calculating%20distance%20measure%20for%20MRDM%20clustering.pdf

OECD. (2009). *Clusters, Innovation and Entrepreneurship.* Paris: Local Economic and Employment Development (LEED), OECD Publishing.

Ojeka, S. (2011). Tax Policy and the Growth of SMEs: Implications for the Nigerian Economy. *Research Journal of Finance and Accounting, 2*(2), 16-26.

Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263–1298.

Onodi, A., & Pecze, K. (2014). Behind the Exporters' Success: Analysis of Successful Hungarian Exporter Companies From a Strategic Perspective. *Managing Global Transitions, 12*(4), 325–346.

Organisation for Economic Co-operation and Development. (1997). *Small Business, Job Creation and Growth: Facts, Obstacles and Best Practices.* Paris: OECD.

Organisation for Economic Co-operation and Development. (2009). *Clusters, Innovation and Entrepreneurship.* Paris: OECD.

Ottaviano, G. (2011). Firm Heterogeneity, Endogenous Entry, and the Business Cycle. *NBER International Seminar on Macroeconomics 2011. 8(1),* pp. 57 - 86. University of Chicago Press.

Papageorgeou, I., Baxter, M., & Cau, M. (2001). Model-based Cluster Analysis of Artefact Compositional Data. *Archaeometry, 43*(4), 571-588.

Park, A., Yang, D., Shi, X., & Jiang, Y. (2010). Exporting and firm performance: Chinese exporters and the Asian financial crises. *Review of Economics and Statistics, 92*(4), 822-842.

Pienaar, L., & Traub, L. (2015). Understanding the smallholder farmer in South Africa: Towards a sustainabile livelihoods classification. Paper No 212633. *International Association of Agricultural Economists.* Milan, Italy.

Pieterse, D., Kreuser, C. F., & Gavin, E. (2016). Introduction to the South African Revenue Service and National Treasury firm-level panel. WIDER Working Paper 2016/42. Helsinki: UNU-WIDER. Retrieved June 23, 2016, from https://www.wider.unu.edu/publication/introduction-south-african-revenue-service-and-national-treasury-firm-level-panel

Pisu, M. (2008). Export destination and learning-by-exporting: Evidence from Belgium. National Bank of Belgium Working Paper No 140. Retrieved August 10, 2015, from https://www.nbb.be/doc/ts/publications/wp/wp140en.pdf

Popa, I., & Stefan, S. (2015). 'Why do some nations succeed and others fail in international competition?' Factor analysis and cluster analysis at European level. *Annals of the University of Oradea : Economic Science*, 1149–1157.

Porter, M. E. (1990). The Competitive Advantage of Nations. *New York: Free Press* , (Republished with a new introduction, 1998).

Porter, M. E. (2007). Clusters and Economic Policy: Aligning Public Policy with the New Economics of Competition. *ISC White Paper November 2007*, Harvard Business School, Institute for Strategy and Competitiveness.

Rankin, N. (2001). The Export Behaviour of South African Manufacturing Firms. Munich Personal RePEc Archive paper No 16904. Retrieved August 10, 2015, from http://mpra.ub.uni-muenchen.de/16904/1/MPRA_paper_16904.pdf

Rankin, N. (2013). *Exporting and Export Dynamics Among South African Firms.* South African Institute of International Affairs (SAIIA): Occasional Paper No 149.

Rankin, N., & Schöer, V. (2013). Export Destination, Product Quality and Wages in a Middle-Income Country. The Case of South Africa. *Review of Development Economics, 17*(1), 64-73.

Rankin, N., Soderbom, S., & Teal, F. (2006). Exporting from Manufacturing Firms in Sub-Saharan Africa. *Journal of African Economies*, 671-687.

Redding, S. (2010). Theories of Heterogeneous Firms and Trade. NBER Working Paper 16562. Retrieved May 21, 2013, from Available from http://www.nber.org/papers/w16562

Roberts. (2000). Understanding the Effects of Trade Policy Reform: The Case of South Africa. *South African Journal of Economics, 68*(4), 270–281.

Saint-Arnaud, S., & Bernard, P. (2003). Convergence or Resilience? A Hierarchical Cluster Analysis of the Welfare Regimes in Advanced Countries. *Current Sociology, 51*(5), 499-527.

Salomon, R., & Shaver, J. (2005). Learning by Exporting: New Insights from Examining Firm Innovation. *Journal of Economics & Management Strategy, 14*(2), 431–460.

SBP. (2013). *SBP Alert: Developing a New Path for SMEs in South Africa.* Retrieved June 26, 2014, from http://www.sbp.org.za/uploads/media/SBP_Alert_-_Developing_a_new_path_for_SMEs_in_South_Africa.pdf

Scott, A., & Symons, M. (1971). Clustering Methods Based on Likelihood Ratio Criteria. *Biometrics*, 387-397.

Serti, F., & Tomasi, C. (2009). Self Selection Along Different Import and Export Markets. *Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies*, Working Paper 2009/18.

Shirotori, M., Tumurchudur, B., & Cadot, O. (2010). *Revealed factor intensity indices at the product level.* Available at: http://unctad.org/en/docs/itcdtab46_en.pdf [accessed May 2016].

Shome, P. (2004). *Tax Administration and the Small Taxpayer.* Washington, DC: IMF Policy Discussion Paper No. 04/2 .

Silva, A., Afonso, O., & Africano, A. (2010). International Trade Involvement and Performance of Portuguese Manufacturing Firms: causal links. *Instituto Politécnico do Porto, ESEIG, mimeo-July*.

Singh, T. (2010). Does International Trade Cause Economic Growth? A Survey. *The World Economy, 33*(11), 1517-1564.

Smulders, S., Stiglingh, M., Franzsen, R., & Fletcher, L. (2012). Tax compliance costs for the small business sector in South Africa - establishing a baseline. *Ejournal of Research, 10*(2).

Sneath, P., & Sokal, R. (1973). Numerical taxonomy: the principals and practice of numerical classification. *San Francisco: Freeman*.

Srholec, M., & Verspagen, B. (2008). *The Voyage of the Beagle in Innovation Systems Land. Explorations on Sectors, Innovation, Heterogeneity and Selection.* Maastricht : UNU-MERIT : Working paper series / United Nations University, UNU-MERIT ; 2008,008.

Stern, R., & Barbour, P. (2005). *Designing a Small Business Tax System that Enhances Growth: Lessons from Africa.* Retrieved July 14, 2014, from http://www.businessenvironment.org/dyn/be/docs/78/Session3.2SternDoc.pdf

Tanzi, V., & Shome, P. (1992). The Role of Taxation in the Development of East Asian Economies. In T. Ito, & A. Krueger (eds), *The Political Economy of Tax Reform, NBER-EASE* (pp. 31 - 65). Chicago: University of Chicago Press.

Thomas, R., & Narayanan, K. (2012). Productivity Heterogeneity and Firm Level Exports: Case of India Manufacturing Industry. *Presented at The 11th Annual GEP Postgraduate Conference 2012 Leverhulme Centre for Research on Globalisation and Economic Policy (GEP).* United Kingdom: University of Nottingham.

Tiri, M., Peeters, L., & Swinnen, G. (2006). Are Innovation Patterns in Manufacturing and Services Converging? A multivariate analysis of CIS-3 data. *DRUID Summer Conference on Knowledge, Innovation and Competitiveness: Dynamics of Firms, Networks, Regions and Institutions.* Hasselt University, Belgium: KIZOK, Research Centre for Entrepreneurship and Innovation.

Trofimenko, N. (2008). Learning by Exporting: Does It Matter Where One Learns? Evidence from Colombian Manufacturing Firms. *Economic Development and Cultural Change, 56*(4), 871-894.

Tryfos, P. (1998). Chapter 15 Cluster Analysis. In *Method for Business Analysis and Forecasting:Text and Cases.* Wiley.

Tryon, R. (1939). *Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality.* Ann Arbor, Mich., Edwards brother, inc., Lithoprinters and Publishers.

Tvaronavičienė, M., Razminienė, K., & Piccinetti, L. (2015). Aproaches towards cluster analysis. *Economics and Sociology, 8*(1), 19-27. DOI: 10.14254/2071- 789X.2015/8-1/2.

Tyler, Z., & Gopal, S. (2010). Sub-Saharan Africa at a Crossroads: A Quantitative Analysis of Regional Development. The Pardee Papers/ No. 10.

Unal, Y., Kindap, T., & Karaca, M. (2003). Redefining the Climate Zones of Turkey Using Cluster Analysis. *International Journal of Climatology, 23*(9), 1045–1055.

Vacek, P. (2010). Productivity Gains From Exporting: Do Export Destinations Matter? IES Working Paper: 18/2010. Retrieved August 10, 2015, from https://ideas.repec.org/p/fau/wpaper/wp2010_18.html

Valpola, H. (2000). Bayesian ensemble learning for nonlinear factor analysis. *Acta Polytechnica Scandinavica, Mathematics and Computing Series*, No. 108.

van Averbeke, W., & Mohamed, S. (2006). Smallholder farming styles and development policy in South Africa: The case of Dzindi Irrigation Scheme. *Agrekon, 45*(2).

Van Biesebroeck, J. (2005). Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics*, 373–391.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*.

Van Parys, S., & James, S. (2010). The Effectiveness of Tax Incentives in Attracting Investment: panel data evidence from the CFA Franc zone. *International Tax Public Finance, 17,* 400-429.

Vázquez, S., & Sumner, A. (2012). Beyond Low and Middle Income Countries: What if There Were Five Clusters of Developing Countries? *IDS Working Papers, 2012*(404), 1-40. DOI: 10.1111/j.2040-0209.2012.00404.x.

Vella, V., & Vichi, M. (1997). Identification of Standards of Living and Poverty in South Africa. *Unpublished*.

Venkatesan, R. (2007). Cluster Analysis for Segmentation. In *Technical Note M-0748 / Published March 22, 2007.* University of Virginia, Charlottesville, VA: Darden Business Publishing.

Verdugo, A. (2005). Tax Incentives and Business Investment New Evidence from Mexico. *MPRA Paper 2272*. Retrieved from http://mpra.ub.uni-muenchen.de/2272/

Verhoogen, E. (2008). Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector. *The Quarterly Journal of Economics, 123*(2), 489-530.

Wagner, J. (1995). Exports, Firm Size, and Firm Dynamics. *Small Business Economics, 7*, 29-39.

Wagner, J. (2007). Exports and Productivity: A Survey of the Evidence from firm-level Data. *The World Economy*, 60-82.

Wagner, J. (2011). International trade and firm performance: A survey of empirical studies since 2006. *Review of World Economics, 148*(2), 235-267.

Wagner, J. (2012). German multiple-product, multiple-destination exporters: Bernard-Redding-Schott under test. University of Lüneburg Working Paper Series in Economics No. 242. Retrieved August 10, 2015, from http://www.leuphana.de/fileadmin/user_upload/Forschungseinrichtungen/ifvwl/WorkingPapers/wp_242_Upload.pdf

Wagstaff, K., & Laidler, V. (2005). Making the Most of Missing Values: Object Clustering with Partial Data in Astronomy. In *Astronomical Data Analysis Software and Systems XIV* (pp. 172–176). Pasadena, California, USA.

Westhead, P., Wright, M., & Ucbasaran, D. (2001). The Internationalization of New and Small Firms: A resource-based view. *Journal of Business Venturing, 16*, 333–358.

Westphal, L. (2002). Technology Strategies for Economic Development in a Fast Changing Global Economy. *Economics of Innovation and New Technology*, 275–320.

World Bank. (2014). South Africa Economic Update: Focus on Export Competitiveness. *South Africa economic update, issue no. 5*, World Bank Group: Washington, DC.

Wziatek-Kubiak, A., Balcerowicz, E., & Peczkowski, M. (2009). Differentiation of Innovation Behavior of Manufacturing Firms in the New Member States: Cluster Analysis on Firm-Level Data. *CASE Network Studies and Analyses No. 394.*, http://dx.doi.org/10.2139/ssrn.1518349.

Yasar, M., Nelson, C., & Rejesus, R. (2006). Productivity and Exporting Status of Manufacturing Firms: Evidence from Quantile Regressions. *Review of World Economics, 142*(2), 675-694.

Yohannes, Y., & Webb, P. (1998). Classification and regression trees: A user manual for identifying indicators of vulnerability to famine and chronic food insecurity. *International Food Policy Research Institute*, Washington, D.C. Mimeo.

Yokoi, S., Takayabu, Y., Nishii, K., Nakamura, H., Endo, H., Ichikawa, H., . . . Watanabe, M. (2011). Application of Cluster Analysis to Climate Model Performance Metrics. *Journal of Applied Meteorology and Climatology, 50*, 1666 -1675.

Yu, T., & Zhang, M. (2007). Multivariate Analysis Of Foreign Multivariate Analysis Of Foreign. *Journal of Applied Business Research, 23*(2), 21-30.

Zee, H., Stotsky, J., & Ley, E. (2002). Tax Incentives for Business Investment: A Primer for Policy. *World Development, 30*(9), 1497-1516.

Zubin, J. (1938). A Technique For Measuring Likemindedness. *Journal of Abnormal and Social Psychology, 3*, 508-516.

Appendix A

SARS Tax Administrative Data Setup

The National Treasury (NT) of the Republic of South Africa and the World Institute for Development Economics Research of the United Nations University (UNU-WIDER), in collaboration with the South African Revenue Service (SARS) have embarked on a research project (over the period 2014 to 2018) that makes use of tax administrative data to conduct firm level analysis. This forms part of UNU-WIDER's project "Regional Growth and Development in Southern Africa", firm level data analysis. This project enables a select few research groups access to these currently restricted, confidential population data sets with the aim of understanding what drives high productivity firms in order to contribute policy-relevant ideas related to job-creation and growth.

The datasets made available via SARS include both employer and employee tax data with accompanying metadata. These datasets are large and, in some cases, complex and require more technical methods in order to be assessed. The data include:

- Corporate Income Tax (CIT) (IT14 and ITR14)
- Personal Income Tax (PIT)/ PAYE (IRP5)
- Assessed Tax
- Employee Tax Incentive (ETI)
- Monthly and Bi-annual Employer Declaration (EMP201 and EMP501 respectively)
- VAT Returns
- Customs Transactions data

A significant amount of time and effort was involved in preparing the data in general for the UNU-WIDER/SARS/NT project[20] and in particular for the purpose of this analysis. For the purpose of this chapter, three of these datasets namely: PIT, CIT, and Customs were utilized. A brief discussion of each dataset and the techniques employed to ready the datasets for our analysis is presented below.

[20] Acknowledgement of this author's contribution was made note of in: Pieterse, D., Kreuser, F. and Gavin, E. (2016), "*Introduction to the South African Revenue Service and National Treasury firm-level panel*" WIDER Working Paper 2016/42. Available from: https://www.wider.unu.edu/publication/introduction-south-african-revenue-service-and-national-treasury-firm-level-panel (accessed on 25 May 2016).

CIT14 and CITR14 Data

The information available in corporate income tax data (CIT) is captured from two forms, namely the IT14 and ITR14. These cross-sectional datasets include information on company characteristics, balance sheet items (capital, equity and liabilities), income statements (debit and credit items), as well as tax deductions and allowances for the tax years 2007 to 2014.

A challenge with assessing the CIT data stems from the fact that the IT14 form was officially replaced by the ITR14 form on 4 May 2013 as part of SARS' modernization of corporate income tax. Unlike the IT14 which was issued in the same format to all corporations regardless of its nature, the ITR14 return first identifies the status of the company (dormant, Share Block Company or body corporate, a micro business, a small business, or a medium to large business) and then customises the fields and questions accordingly. The ITR14 contains fewer questions for smaller firms, and many more questions for the larger firms, compared to the IT14. In addition, the ITR14 requires more detailed information on international tax aspects, transactions between connected persons as well as disclosure relating to investments made by corporates.

To further complicate matters, any overdue returns as of May 2013 would have to be completed using the new ITR14 format. As a result for most, if not all, tax years there is CIT data available from both forms and therefore, in order to obtain a complete panel of CIT data, the IT14 dataset must be appended to the ITR14 dataset for each year. It is noted here that an entity is not able to complete both an IT14 and an ITR14 form for each tax year, there are no company duplicates between the datasets.

Since micro, small and medium and large firms have different levels of detail for certain variables in the ITR14, these variables need first to be aggregated into single variables. For example, property plant and equipment is captured in one variable for micro and small firms named 'Property, plant and equipment". However, for medium and large firms, property plant and equipment is split into 'Fixed Property' and 'Fixed Assets (Plant and equipment)'. It is therefore necessary to first aggregate these medium and large variables into a single property, plant and equipment variable in order for it to be comparable to the smaller firms. For the purpose of this particular paper, this type of cleaning procedure needed to be done for the capital variable (property, plant and equipment), output variable (sales/turnover), labour cost (wages) and intermediate inputs (cost of sales) all which were disaggregated to varying degrees depending on the size of the firm.

Once this task of aggregating the ITR14 variables by firm size was complete, it was then necessary to aggregate these size variables into a single variable that encompassed information for all firms regardless of size. This was necessary in order for the ITR14 variables to be consistent with the IT14 variables (which did not distinguish between firms sizes).

Finally, the ITR14 and IT14 datasets were appended to create a consistent panel of key productivity variables (output, capital, labour cost, and intermediate inputs, industries, etc), including a company type dummy variable that would be used to distinguish the data by firm size. The firm size categories were generated based off the SARS definition of firm size, namely:

- A firm is classified as a micro firm if its gross income is no more than R1 million and total assets amount to no more than R5 million;
- A firm is classified as a small firm if its gross income is no more than R14 million and total assets amount to no more than R10 million;
- A firm is classified as a medium/large firm if its gross income is greater than R14 million and total assets amount to more than R10 million;

This analysis made use of data for the tax years 2010 to 2013. The reasoning for this is that the 2009 data provided was incomplete at the time – there seemed to be only a third of the observations of any other year. In addition, SARS made note that the IT14 forms only came into use some time during the 2009 tax year and information prior to that year was not reported or captured reliably. In order to have a consistent year-to-year panel it was decided to focus on the data after 2009. In addition, at the time of the analysis, the 2014 tax-year data was incomplete (since many firms fill in their returns late), as such it was decided to concentrate on the tax years from 2010 to 2013.

PIT/PAYE Data

A key productivity variable missing from the CIT data is the number of persons employed by the firm. This information was therefore extracted from the personal income tax (PIT) data. The PIT is obtained from the records of completed IRP5 certificates by employers on behalf of their employees. The IRP5 certificate data includes cross-sectional data for the tax years 2007 to 2014 and includes information on employee demographics, allowances (travel, subsistence, uniform, telephone, etc.), benefits (meals, bursaries, medical schemes, etc.), income items (taxable and non-taxable income, commission, overtime, annuities, etc.) and other items (UIF contributions, SDL contributions, provident fund contribution, SITE, PAYE, etc.).

Employment was calculated as the weighted number of employees over the years per firm. This was done by first calculating the length of employment for all workers within a year, summing up within the firm and dividing by twelve, essentially generating the stock of worker inputs into production for the whole year per firm.

We note here that IRP5 censoring may be inherent in the data, since IRP5 forms and numbers are only issued when an employee earns above a certain threshold (see table A1). As such, total employment per firm will be undercounted in all firms (micro, small, medium and large), since employees earning less than the threshold may not be included in the dataset. Further, firms which pay less on average will report lower employment than firms paying higher amounts.

In addition to the potential under-counting problem due to censorship, we are also under-counting employees due to the inability of identifying foreign employees since, at the time of the analysis, passport numbers were unavailable in the PIT dataset. As such, only employees with a South African identification number earning above the specific tax threshold were counted in total employment.

Table A1. Tax Thresholds

	Tax Year			
	2013	**2012**	**2011**	**2010**
Under 65	R63 556	R59 750	R57 000	R54 200
65 an older	R99 056	R93 150	R88 528	R84 200
75 and older	R110 889	R104 261	R88 528	R84 200

Source: SARS website

Once employment per firm was calculated, this data was then merged onto the CIT dataset through a conjunction table. This conjunction table was provided by SARS for the purpose of linking the various datasets through different identifiers. For the CIT-PIT merge, the conjunction table linked the unique firm identifier in the PIT data (the PAYE reference number) to the unique firm identifier in the CIT data (the tax reference number) which then enabled the merge of the employment data to the other key productivity variables in the CIT data.

Customs Transactions Data

The final dataset used for this analysis was the customs transactions data. This dataset contains information regarding trade transactions (by both importers and exporters), such as: the country of origin, export and destination, statistical quantity and value, as well as duties and tariffs (HS4, HS6 and HS8). At the time of analysis this data was available for the tax years 2009 to 2013 and the number of transactions totalled between 9 and 11 million for around 33 000 to 41 000 unique traders.

For the purpose of this paper, only exporter information was required from the dataset. Transactions were linked to unique exporters and then merged onto the CIT-PIT dataset through the conjunction table. Specifically the exporter-transactions subset was merged onto the conjunction table through the unique customs identifier. From that table a corresponding tax reference number was found which enabled the data to be merged to the CIT-PIT data. A number of exporters did not have a match in the conjunction table and could therefore not be linked to the CIT-PIT dataset. By the end of this process 23 000 and 27 000 exporters were merged into the CIT-PIT dataset.

Final CIT-PIT-Customs data

The final, usable dataset contained around 700 000 unique firms with employment information per year. This dataset included information on firm characteristics, key productivity variables, employment and exporting activities for the tax years 2010 to 2013. In terms of the merge rates: of the approximately 700 000 firms in the CIT data, around 22 percent had a match to the IRP5 employment data. Of those firms that did not have a match, 38 percent were dormant companies, while 42 percent were classified as micro businesses. In addition, 33 percent of the unmatched firms had zero turnover. Of the firms that did have a match to the employment data, 27 percent were micro businesses, 44 percent small businesses and 20 percent medium to large businesses (the remaining 9% was made up of dormant companies or body corporate/share block companies). Finally, of those firms with matching IRP5 data, approximately 15% had a match in the Customs data.

Potential Bias

In terms of the Stats SA's Large Sample Survey, estimates from this data source may suffer from both sampling and non-sampling errors. Since the data presented in the thesis is obtained from a random sample of manufacturing enterprises, it is subject to sampling variability. That is, estimations on this data may differ from estimations made from data on the full population of manufacturing enterprises. In the presence of sampling variability, inferring conclusions to the whole from the sample should be treated with caution. Potential sources of non-sampling errors in this dataset include out of date sample frame; incorrect definitions or classifications; inaccuracies from misreporting or collection of data as a result of misunderstanding, bias, negligence, or dishonesty on the part of either the interviewer or responder; incorrect demarcation of units; non-responses; and data capturing errors. Stats SA report that every effort has been made to minimise these non-sampling errors by careful design and implementation of their questionnaires.

Estimates from the SARS dataset may suffer from selection bias in the sense that only firms that submit a tax return will appear in the dataset, i.e.: only tax-registered firms are included. Therefore this dataset includes very little information from unregistered firms such as informal firms as well as very small or very new firms. An additional source of selection bias comes from the IRP5 data. As mentioned above, only employees who earn above a certain threshold are issued with an IRP5 form. In addition, foreign employees were not identifiable in the dataset at the time of assessment. As such, the data is likely under-counting the number of employees in all firms. Finally, potential bias or errors may arise from the collection and aggregation of the data. For instance, measurement error may arise from the intentional (tax minimization) or unintentional (misunderstanding) misreporting of variables. In addition, measurement error may be present as a result of the aggregation of ITR14 variables to IT14 variables.

Despite these sources of error and potential bias both the LSS data as well as the SARS data remain important sources for firm level analysis. In addition, the SARS data is closer to the population of firms in the (formal) South African economy than previously used survey datasets and therefore provides a better basis for firm-level analysis than before.

Appendix B

Supplementary Tables for Chapter 2

This Appendix contains additional tables as referenced in section 2.5. The first set of tables (table B1 to B4) contain the results of estimated characteristic differences, or export premia, as defined by equation 2.1, for certain subsets of the data. The next set of tables (tables B5 to B8) follow the estimation of productivity premiums as per the methodology set out in section 2.6 for the same subsets of data and are referred to in section 2.7. Finally, figure B1 illustrates the productivity premium by industry.

Table B1. Medium to Large Exporters -Export Premia

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
LSS data						
Exporter	0.127***	0.109***	0.0150	0.158***	0.215***	0.0457
	(0.0441)	(0.0401)	(0.0314)	(0.0242)	(0.0438)	(0.0328)
Exporter*2008	0.0807	0.0807	0.00190	-0.0220	0.196***	-0.0135
	(0.0551)	(0.0502)	(0.0393)	(0.0303)	(0.0549)	(0.0411)
2008	-0.307***	-0.227***	-0.0658***	0.157***	-0.103***	-0.145***
	(0.0356)	(0.0322)	(0.0255)	(0.0194)	(0.0354)	(0.0264)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*8,447*	*8,635*	*8,408*	*8,632*	*8,528*	*8,631*
SARS data						
Exporter	0.519***	0.248***	0.164***	0.544***	-0.0320	0.203***
	(0.0307)	(0.0360)	(0.0322)	(0.0369)	(0.0666)	(0.0368)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*7,561*	*5,159*	*5,124*	*5,038*	*4,861*	*5,008*
Controlling for firm-size						
Exporter			0.258***	0.198***	-0.0179	0.302***
			(0.0290)	(0.0320)	(0.0668)	(0.0335)
SARS - Controlling for firm size						

Industry controls	Yes	Yes	Yes	Yes
Firm size control	Yes	Yes	Yes	Yes
Observations	*5,124*	*5,038*	*4,861*	*5,008*

Source: Authors own calculation using the LSS data and SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Table B2. Medium to Large Exporters – Export Premia within Africa

	Output	**No of employees**	**Output per worker**	**Labour Cost**	**Capital per worker**	**Intermediate Inputs per worker**
Exporter	0.819***	0.474***	0.233***	0.870***	0.120	0.257***
	(0.0380)	(0.0420)	(0.0379)	(0.0455)	(0.0781)	(0.0434)
Africa Only Exporter Dummy	-0.579***	-0.435***	-0.132***	-0.631***	-0.294***	-0.105**
	(0.0440)	(0.0426)	(0.0385)	(0.0525)	(0.0794)	(0.0439)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*7,561*	*5,159*	*5,124*	*5,038*	*4,861*	*5,008*
Controlling for firm-size						
Exporter			0.423***	0.369***	0.154*	0.459***
			(0.0341)	(0.0377)	(0.0789)	(0.0396)
Africa Only Exporter Dummy			-0.309***	-0.321***	-0.327***	-0.293***
			(0.0345)	(0.0382)	(0.0801)	(0.0400)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations			*5,124*	*5,038*	*4,861*	*5,008*

Source: Authors own calculations using SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Table B3. Medium to Large Exporters – Export Premia for Multiple Destinations

	Output	No of employees	Output per worker	Labour Cost	Capital per worker	Intermediate Inputs per worker
Exporter	0.0111	-0.150***	0.0743	0.0273	-0.130	0.105*
	(0.0549)	(0.0565)	(0.0509)	(0.0659)	(0.106)	(0.0580)
Multiple Dest. Exp Dummy	0.627***	0.492***	0.111**	0.637***	0.121	0.121**
	(0.0564)	(0.0543)	(0.0488)	(0.0675)	(0.102)	(0.0556)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*7,561*	*5,159*	*5,124*	*5,038*	*4,861*	*5,008*
Controlling for firm-size						
Exporter			0.00972	-0.0127	-0.143	0.0340
			(0.0454)	(0.0503)	(0.106)	(0.0524)
Multiple Dest. Exp Dummy			0.310***	0.263***	0.156	0.336***
			(0.0439)	(0.0486)	(0.103)	(0.0506)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations			*5,124*	*5,038*	*4,861*	*5,008*

Source: Authors own calculations using SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Table B4. Medium to Large Exporters – Export Premia for Multiple Destinations within and outside Africa

	Output	No of employees	Output per worker	Labour Cost	Capital per worker	Intermediate Inputs per worker
Exporter	-0.0117	-0.0570	0.0505	0.0818	-0.101	-0.0344
	(0.127)	(0.126)	(0.114)	(0.155)	(0.241)	(0.129)
Africa Only Exp Dummy	0.0233	-0.115	0.0266	-0.0685	-0.0389	0.164
	(0.137)	(0.134)	(0.121)	(0.167)	(0.255)	(0.137)
Multiple Dest. Exp Dummy	0.887***	0.568***	0.195*	0.839***	0.234	0.311**
	(0.130)	(0.126)	(0.114)	(0.158)	(0.242)	(0.130)

Interaction:Afri*Multi-dest	-0.557***	-0.263*	-0.159	-0.513***	-0.282	-0.278*
	(0.145)	(0.141)	(0.128)	(0.176)	(0.268)	(0.145)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm size control	No	No	No	No	No	No
Observations	*7,561*	*5,159*	*5,124*	*5,038*	*4,861*	*5,008*
Controlling for firm-size						
Exporter			0.0200	0.0836	-0.116	-0.0713
			(0.101)	(0.113)	(0.240)	(0.116)
Africa Only Exp Dummy			-0.0177	-0.120	-0.0397	0.120
			(0.107)	(0.120)	(0.255)	(0.123)
Multiple Dest. Exp Dummy			0.433***	0.306***	0.288	0.569***
			(0.102)	(0.114)	(0.242)	(0.117)
Interaction:Afri*Multi-dest			-0.271**	-0.185	-0.312	-0.400***
			(0.113)	(0.126)	(0.268)	(0.130)
Industry controls			Yes	Yes	Yes	Yes
Firm size control			Yes	Yes	Yes	Yes
Observations			*5,124*	*5,038*	*4,861*	*5,008*

Source: Authors own calculation using SARS data
Notes: ***p<0.01**p<0.05*p<0.1
(Is significant at the 1% level, 5% level and 10% level respectively)
Values are given in natural logarithms.

Table B5. Exporter Productivity Premium by OLS, Fixed Effects (F.E.) and Levinsohn-Petrin (L.P.) estimation – Medium to Large Firms

	Standard			Africa			No.of Destinations			Multi-destinations		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**	**OLS**	**F.E.**	**L.P.**
2013	0.00767	0.0405***	0.0330*	0.00698	0.0406***	0.0324*	0.00451	0.0405***	0.0299*	0.00749	0.0413***	0.0329
	(0.0170)	(0.0155)	(0.0182)	(0.0170)	(0.0155)	(0.0172)	(0.0169)	(0.0155)	(0.0173)	(0.0170)	(0.0156)	(0.0202)
Exporter	0.0556***	0.0311	0.0713***	0.0757***	0.0319	0.0948***	0.00227	0.0260	0.0232	0.0136	0.0263	0.00290
	(0.0181)	(0.0302)	(0.0218)	(0.0201)	(0.0334)	(0.0231)	(0.0191)	(0.0315)	(0.0216)	(0.0290)	(0.0352)	(0.0380)
Exporter*2013	-0.0321	-0.0297	-0.0401*	-0.0208	-0.0287	-0.0416*	-0.0221	-0.0299	-0.0315	-0.0143	-0.0607	0.00630
	(0.0205)	(0.0189)	(0.0207)	(0.0229)	(0.0206)	(0.0229)	(0.0204)	(0.0189)	(0.0212)	(0.0333)	(0.0394)	(0.0390)
Africa_only				-0.0429**	-0.000865	-0.0522***						
				(0.0206)	(0.0268)	(0.0172)						
Africa_only*2013				-0.0158	-0.00278	0.0114						
				(0.0235)	(0.0230)	(0.0168)						
No. dest							0.00873***	0.00303	0.00801***			
							(0.00127)	(0.00527)	(0.00110)			
(No. dest)^2							-5.75e-05**	-4.34e-05	-6.10e-05***			
							(2.32e-05)	(9.50e-05)	(2.29e-05)			
Multi-dest.										0.0516*	0.0315	0.0835***
										(0.0275)	(0.0334)	(0.0286)
Multi-dest. *2013										-0.0209	0.0330	-0.0553**
										(0.0314)	(0.0374)	(0.0279)
ll	-0.0575***	-0.373***	-0.0385***	-0.0608***	-0.373***	-0.0414***	-0.0714***	-0.374***	-0.0508***	-0.0591***	-0.374***	-0.0405***
	(0.00380)	(0.0248)	(0.00507)	(0.00385)	(0.0249)	(0.00545)	(0.00406)	(0.0249)	(0.00489)	(0.00385)	(0.0248)	(0.00469)
lkl	0.0232***	-0.0115	0.0647**	0.0228***	-0.0114	0.0650**	0.0224***	-0.0113	0.0235***	0.0232***	-0.0108	0.0652**
	(0.00204)	(0.00900)	(0.0315)	(0.00204)	(0.00903)	(0.0275)	(0.00203)	(0.00901)	(0.00807)	(0.00204)	(0.00901)	(0.0304)
lil	0.803***	0.521***	0.477***	0.802***	0.521***	0.477***	0.794***	0.521***	0.841***	0.802***	0.522***	0.477***
	(0.00389)	(0.0163)	(0.0957)	(0.00390)	(0.0163)	(0.0876)	(0.00399)	(0.0163)	(0.0764)	(0.00391)	(0.0163)	(0.0873)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,514	6,514	6,514	6,514	6,514	6,514	6,514	6,514	6,514	6,514	6,514	6,514

R-squared	0.919	0.702	0.919	0.702	0.920	0.702	0.919	0.702
Number of id		5,097		5,097		5,097		5,097

Source: Authors own calculation using SARS data | Notes: ***p<0.01, **p<0.05, *p<0.1 is significant at the 1% level, 5% level and 10% level respectively.

Table B6. Exporter Productivity Premium by OLS, Fixed Effects (F.E.) and Levinsohn-Petrin (L.P.) estimation – Micro to Small Firms

	Standard			**Africa**			**No.of Destinations**			**Multi-destinations**		
	(1) **OLS**	(2) **F.E.**	(3) **L.P.**	(4) **OLS**	(5) **F.E.**	(6) **L.P.**	(7) **OLS**	(8) **F.E.**	(9) **L.P.**	(10) **OLS**	(11) **F.E.**	(12) **L.P.**
2013	0.0408***	0.00525	0.0337**	0.0408***	0.00524	0.0337**	0.0407***	0.00524	0.0335***	0.0408***	0.00477	0.0338**
	(0.0150)	(0.0138)	(0.0145)	(0.0150)	(0.0139)	(0.0140)	(0.0150)	(0.0138)	(0.0130)	(0.0150)	(0.0139)	(0.0161)
Exporter	0.0291	-0.00602	0.0131	0.0524	-0.00286	0.0252	-0.0227	-0.0221	-0.0335	-0.00396	-0.00831	-0.00338
	(0.0212)	(0.0277)	(0.0195)	(0.0323)	(0.0427)	(0.0309)	(0.0242)	(0.0355)	(0.0211)	(0.0294)	(0.0321)	(0.0254)
Exporter*2013	-0.0110	0.0224	0.00487	-0.0366	0.0213	-0.00997	-0.00880	0.0199	0.00680	-0.0198	0.0377	-0.0180
	(0.0242)	(0.0231)	(0.0213)	(0.0368)	(0.0352)	(0.0345)	(0.0242)	(0.0232)	(0.0211)	(0.0337)	(0.0360)	(0.0330)
Africa_only				-0.0341	-0.00403	-0.0178						
				(0.0357)	(0.0416)	(0.0301)						
Africa_only*2013				0.0377	0.00173	0.0219						
				(0.0408)	(0.0407)	(0.0341)						
No. dest							0.0205***	0.0123	0.0184***			
							(0.00527)	(0.0181)	(0.00702)			
(No. dest)^2							-0.000271	2.73e-05	-0.000232			
							(0.000236)	(0.00132)	(0.000541)			
Multi-dest.										0.0567*	-0.00897	0.0292
										(0.0338)	(0.0402)	(0.0295)
Multi-dest. *2013										0.0156	-0.0205	0.0387
										(0.0388)	(0.0397)	(0.0348)
ll	-0.177***	-0.315***	-0.133***	-0.177***	-0.315***	-0.133***	-0.180***	-0.315***	-0.136***	-0.179***	-0.315***	-0.135***
	(0.00560)	(0.0238)	(0.00923)	(0.00560)	(0.0239)	(0.00904)	(0.00561)	(0.0238)	(0.00981)	(0.00561)	(0.0238)	(0.00905)
lkl	0.0208***	0.00259	0	0.0207***	0.00259	0	0.0206***	0.00255	0	0.0208***	0.00270	0
	(0.00207)	(0.00562)	(0.0130)	(0.00207)	(0.00563)	(0.0129)	(0.00206)	(0.00562)	(0.0157)	(0.00206)	(0.00562)	(0.0129)
lil	0.616***	0.567***	0.724***	0.616***	0.567***	0.723***	0.616***	0.566***	0.895***	0.616***	0.567***	0.719***
	(0.00508)	(0.0168)	(0.202)	(0.00510)	(0.0168)	(0.215)	(0.00508)	(0.0168)	(0.318)	(0.00508)	(0.0168)	(0.176)
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,095	6,095	6,095	6,095	6,095	6,095	6,095	6,095	6,095	6,095	6,095	6,095

R-squared	0.842	0.783	0.842	0.783	0.843	0.783	0.843	0.783
Number of id		4,992		4,992		4,992		4,992

Source: Authors own calculation using SARS data | Notes: ***p<0.01, **p<0.05, *p<0.1 is significant at the 1% level, 5% level and 10% level respectively.

Table B7. Multiple-destination/Africa Interaction Productivity Premium – Medium to Large Firms

VARIABLES	(1) **OLS**	(2) **F.E.**	(3) **L.P.**
nonexp2013	0.00694	0.0408***	0.0325**
	(0.0170)	(0.0156)	(0.0153)
sing_dest_out_afr_2013	0.0844**	0.0780	0.0704
	(0.0388)	(0.0695)	(0.0511)
sing_dest_out_afr_2012	0.0380	0.0266	0.0491
	(0.0518)	(0.0565)	(0.0514)
sing_dest_in_afr_2013	-0.00923	-0.0157	0.0354*
	(0.0215)	(0.0431)	(0.0189)
sing_dest_in_afr_2012	0.00421	0.0237	-0.0135
	(0.0324)	(0.0409)	(0.0350)
multi_dest_in_afr_2013	0.00863	0.0756*	0.0494***
	(0.0180)	(0.0385)	(0.0165)
multi_dest_in_afr_2012	0.0449*	0.0543	0.0664***
	(0.0239)	(0.0389)	(0.0216)
multi_dest_out_afr_2013	0.0523***	-0.0152	0.0874***
	(0.0135)	(0.0287)	(0.0174)
multi_dest_out_afr_2012	0.0340	-0.00392	0.0994***
	(0.0233)	(0.0289)	(0.0231)
ll	-0.0612***	-0.373***	-0.0422***
	(0.00389)	(0.0249)	(0.00488)
lkl	0.0228***	-0.0106	0.0653**
	(0.00204)	(0.00904)	(0.0256)
lil	0.801***	0.523***	0.477***
	(0.00392)	(0.0163)	(0.0820)
Industry controls	Yes	Yes	Yes
Observations	6,514	6,514	6,514
R-squared	0.919	0.703	
Number of id		5,097	

Source: Authors own calculation using SARS data
Notes: ***p<0.01, **p<0.05, *p<0.1 significance at the 1%, 5% and 10% level respectively.

Table B8. Multiple-destination/Africa Interaction Productivity Premium – Micro to Small Firms

VARIABLES	(1) **OLS**	(2) **F.E.**	(3) **L.P.**
nonexp2013	0.0409***	0.00558	0.0338**
	(0.0150)	(0.0139)	(0.0142)
sing_dest_out_afr_2013	-0.130***	-0.0353	-0.125**
	(0.0400)	(0.0591)	(0.0574)
sing_dest_out_afr_2012	-0.00644	0.00896	-0.0348
	(0.0722)	(0.0510)	(0.0728)
sing_dest_in_afr_2013	0.0436**	-2.662	0.0371**
	(0.0207)	(130.6)	(0.0157)
sing_dest_in_afr_2012	-0.00335	-2.402	0.00188
	(0.0312)	(15.02)	(0.0243)
multi_dest_in_afr_2013	0.0805***	0.00134	0.0726***
	(0.0213)	(0.0649)	(0.0147)
multi_dest_in_afr_2012	0.0431	0.00120	0.0162

	(0.0312)	(0.00746)	(0.0226)
multi_dest_out_afr_2013	0.0212	-8.723	0.0912***
	(0.0247)	(36.48)	(0.0243)
multi_dest_out_afr_2012	0.0236	0.00433	0.0399
	(0.0429)	(0.0181)	(0.0351)
ll	-0.179***	-0.315***	-0.135***
	(0.00561)	(0.0239)	(0.00847)
lkl	0.0207***	0.00249	0
	(0.00206)	(0.00564)	(0.0145)
lil	0.615***	0.566***	0.718***
	(0.00510)	(0.0169)	(0.239)
Industry controls	Yes	Yes	Yes
Observations	6,095	6,095	6,095
R-squared	0.843	0.784	
Number of id		4,992	

Source: Authors own calculation using SARS data
Notes: ***p<0.01, **p<0.05, *p<0.1 significance at the 1%, 5% and 10% level respectively.

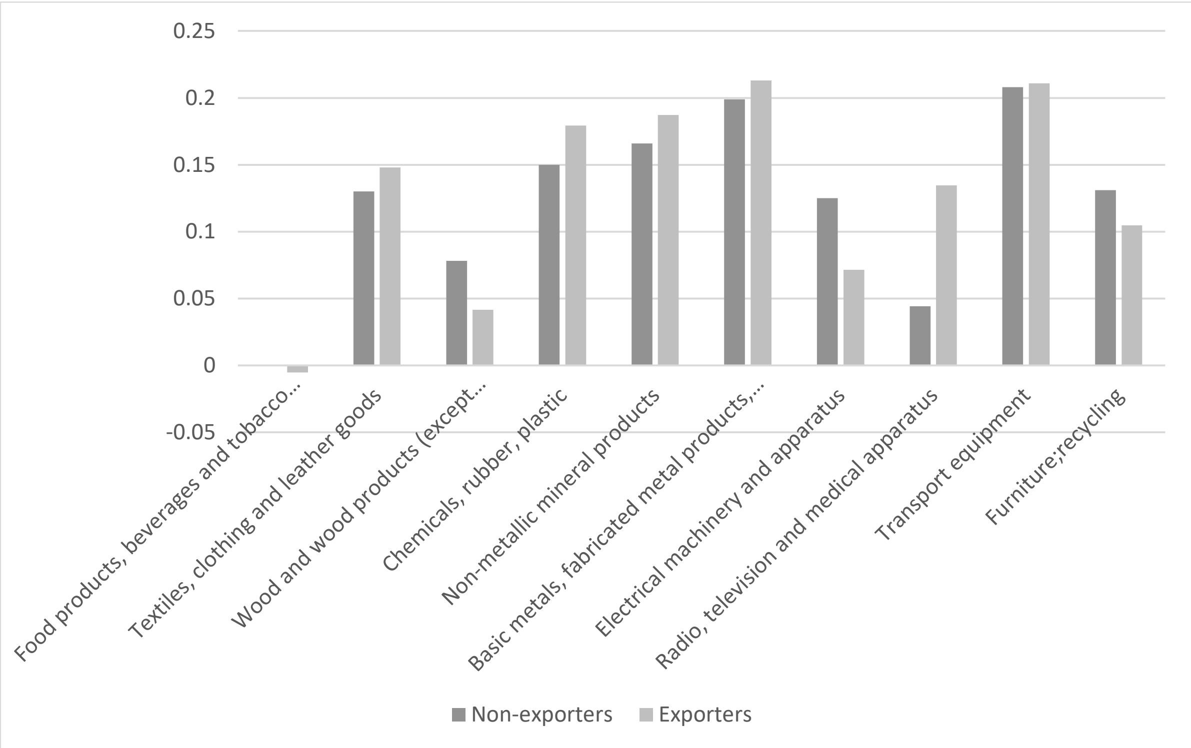


Figure B1. Productivity by industry for LSS 2005. Base category is non-exporters in the Food Products industry.

From figure B1 it can be seen that the LSS data shows that there are not significant differences in premia across industries. This is consistent with the results presented in the rest of the thesis, and theoretical models (such as the Melitz model), which indicate that it is firm size rather than industry which is a more important correlate with exporting than industry is.

Appendix C

Supplementary Tables for Chapter 3

This Appendix contains additional tables and figures as referred to in the cluster analysis of section 3.6.1.2. The first set of results (table C1) give the p-value output from the Student t-Test used to test for the significance of any characteristic differences in the variables between each cluster. The next set of tables (tables C2 and C3) present the median characteristics and p-values respectively of the SARS 6 cluster solution referred to in section 3.6.1.2.

Table C1. P-value output of Student's t-Test for LSS and SARS data

	LSS 2005						
	lryl	ll	lrkl	lril	ln_TFP_OLS	lrwl	exp_sales_%
Cluster1_Cluster2	0.004	0.000	0.694	0.018	0.145	0.797	n/a
Cluster1_Cluster3	0.000	0.000	0.000	0.000	0.001	0.000	n/a
Cluster1_Cluster4	0.000	0.000	0.000	0.000	0.001	0.000	n/a
Cluster1_Cluster5	0.000	0.250	0.001	0.000	0.014	0.545	n/a
Cluster1_Cluster6	0.001	0.000	0.001	0.002	0.034	0.005	n/a
Cluster2_Cluster3	0.000	0.000	0.000	0.000	0.602	0.000	n/a
Cluster2_Cluster4	0.000	0.000	0.000	0.000	0.598	0.000	0.613
Cluster2_Cluster5	0.000	0.407	0.001	0.000	0.035	0.570	0.000
Cluster2_Cluster6	0.002	0.000	0.001	0.003	0.055	0.005	n/a
Cluster3_Cluster4	0.105	0.001	0.012	0.020	0.977	0.774	n/a
Cluster3_Cluster5	0.000	0.207	0.005	0.000	0.044	0.774	n/a
Cluster3_Cluster6	0.005	0.000	0.002	0.009	0.068	0.002	n/a
Cluster4_Cluster5	0.000	0.147	0.007	0.000	0.044	0.037	0.000
Cluster4_Cluster6	0.005	0.000	0.003	0.008	0.069	0.002	n/a
Cluster5_Cluster6	0.790	0.003	0.868	0.883	0.770	0.004	n/a

	LSS 2008						
	lryl	ll	lrkl	lril	ln_TFP_OLS	lrwl	exp_sales_%
Cluster1_Cluster2	0.000	0.000	0.000	0.000	0.780	0.000	n/a
Cluster1_Cluster3	0.000	0.000	0.008	0.000	0.560	0.000	n/a
Cluster1_Cluster4	0.001	0.303	0.599	0.001	0.001	0.000	n/a
Cluster1_Cluster5	0.192	0.472	0.079	0.087	0.546	0.320	n/a
Cluster1_Cluster6	0.146	0.472	0.191	0.160	0.446	0.320	n/a
Cluster2_Cluster3	0.000	0.000	0.000	0.000	0.473	0.000	0.256
Cluster2_Cluster4	0.001	0.000	0.955	0.002	0.001	0.000	0.000
Cluster2_Cluster5	0.184	0.343	0.071	0.082	0.546	0.332	n/a
Cluster2_Cluster6	0.158	0.427	0.227	0.173	0.426	0.261	0.185
Cluster3_Cluster4	0.000	0.003	0.485	0.000	0.001	0.002	0.000
Cluster3_Cluster5	0.206	0.252	0.079	0.097	0.548	0.282	n/a
Cluster3_Cluster6	0.129	0.566	0.179	0.143	0.658	0.318	0.183

	lryl	ll	lrkl	lril	ln_TFP_OLS	lrwl	exp_sales_%
Cluster4_Cluster5	0.094	0.184	0.152	0.001	0.493	0.182	n/a
Cluster4_Cluster6	0.868	0.807	0.152	0.903	0.001	0.715	0.000
Cluster5_Cluster6	0.086	0.213	0.039	0.033	0.551	0.115	n/a

	SARS						
	lryl	ll	lrkl	lril	ln_TFP_OLS	lrwl	exp_sales_%
Cluster1_Cluster2	0.000	0.000	0.000	0.000	0.736	0.000	n/a
Cluster1_Cluster3	0.000	0.053	0.927	0.000	0.688	0.107	n/a
Cluster1_Cluster4	0.619	0.982	0.589	0.612	0.509	0.080	n/a
Cluster1_Cluster5	0.020	0.930	0.086	0.000	0.189	0.080	n/a
Cluster2_Cluster3	0.000	0.000	0.003	0.000	0.882	0.000	0.000
Cluster2_Cluster4	0.000	0.000	0.087	0.000	0.544	0.287	0.357
Cluster2_Cluster5	0.022	0.655	0.089	0.000	0.188	0.084	0.000
Cluster3_Cluster4	0.022	0.408	0.629	0.024	0.567	0.286	0.017
Cluster3_Cluster5	0.019	0.897	0.084	0.000	0.188	0.081	0.001
Cluster4_Cluster5	0.006	0.931	0.072	0.000	0.178	0.079	0.001

Table C2. Median Values of Variables on 6 cluster solution - SARS 2013 dataset

	Variables (Performance Indicators)														
	SARS														
	med_large	exporter	lryl	ll	lrkl	lril	ln_TFP	lrwl	exp_sales_p	industry	SDIA	SDOA	MDIA	MDOA	*Frequency*
Cluster 1	0	1	13.50	2.75	10.62	12.99	-0.01	11.51	1.55	25111	1	0	0	0	1020
Cluster 2	0	1	13.24	2.63	10.53	12.63	0.01	11.66	2.38	25994	0	1	0	0	185
Cluster 3	1	1	13.83	3.58	10.95	13.34	-0.01	11.80	8.53	25999	0	0	0	1	1849
Cluster 4	1	1	13.62	3.30	10.59	13.14	-0.01	11.63	4.09	25119	0	0	1	0	1633
Cluster 5	0	0	13.31	2.60	10.67	12.74	-0.04	11.45	0.00	25111	0	0	0	0	4190
Cluster 6	1	0.5	19.48	2.41	16.77	19.01	1.16	16.70	0.64	10605.5	0.5	0	0	0	2

Notes: SDIA refers to "sing_dest_in_afr_2013"; SDIO refers to "sing_dest_out_afr_2013"; MDIA refers to "multi_dest_in_afr_2013"; and MDOA refers to "multi_dest_out_afr_2013" variables.

Table C3. P-value output of Student's t-Test for SARS (6 cluster solution)

	SARS						
	lryl	ll	lrkl	lril	ln_TFP_OLS	lrwl	exp_sales_%
Cluster1_Cluster2	0.000	0.000	0.552	0.000	0.649	0.000	n/a
Cluster1_Cluster3	0.000	0.053	0.927	0.000	0.688	0.107	n/a
Cluster1_Cluster4	0.000	0.000	0.000	0.000	0.889	0.000	n/a
Cluster1_Cluster5	0.619	0.982	0.589	0.612	0.509	0.080	n/a
Cluster1_Cluster6	0.020	0.930	0.086	0.000	0.189	0.080	n/a
Cluster2_Cluster3	0.000	0.000	0.598	0.000	0.984	0.046	0.001
Cluster2_Cluster4	0.000	0.000	0.000	0.000	0.759	0.000	0.000
Cluster2_Cluster5	0.000	0.000	0.468	0.000	0.568	0.833	0.063
Cluster2_Cluster6	0.020	0.717	0.086	0.000	0.188	0.082	0.000
Cluster3_Cluster4	0.000	0.000	0.000	0.000	0.791	0.000	0.000
Cluster3_Cluster5	0.022	0.408	0.629	0.024	0.567	0.286	0.017
Cluster3_Cluster6	0.019	0.897	0.084	0.000	0.188	0.081	0.001
Cluster4_Cluster5	0.000	0.000	0.012	0.000	0.527	0.032	0.936
Cluster4_Cluster6	0.021	0.607	0.091	0.000	0.188	0.085	0.000
Cluster5_Cluster6	0.006	0.931	0.072	0.000	0.178	0.079	0.001

The following set of figures presents the cluster analysis with the inclusion of an import dummy. As mentioned in section 3.5.2, the decision to important may also influence the decision to export. To test if this is the case, we run a ClustOfVar analysis to see if the data identifies imports as a variable of importance. The number of clusters are again determined by the dendrogram and agglomeration schedule as four clusters of variables for both the 2005 and 2008 data.

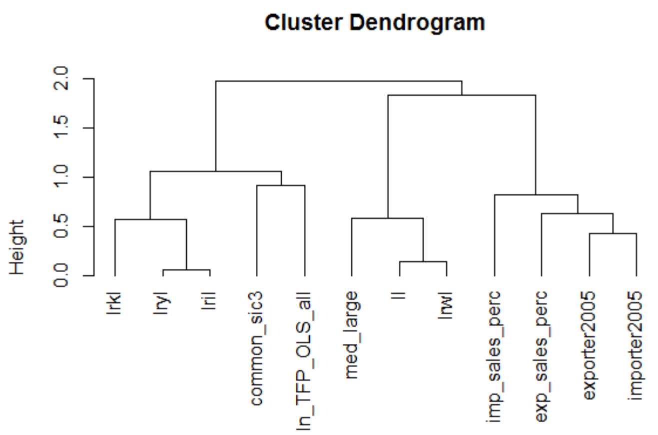


Figure C1. Cluster dendrogram – LSS 2005

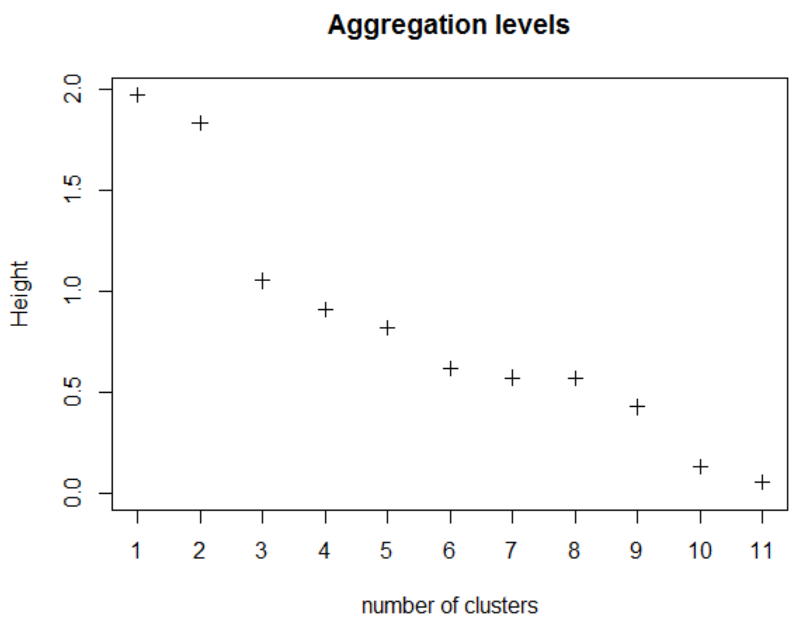


Figure C2. Agglomeration scree plot – LSS 2005

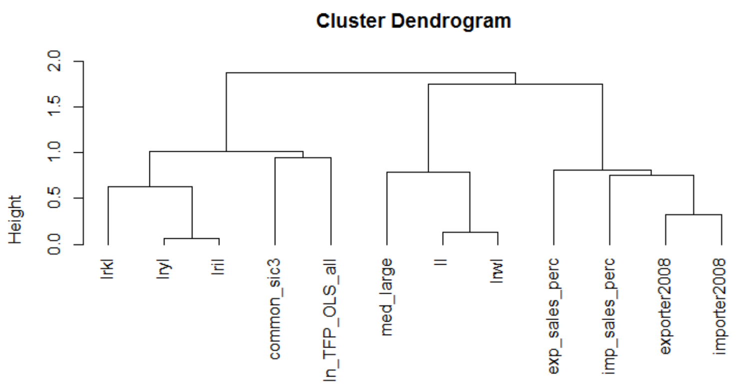


Figure C3. Cluster dendrogram – LSS 2008

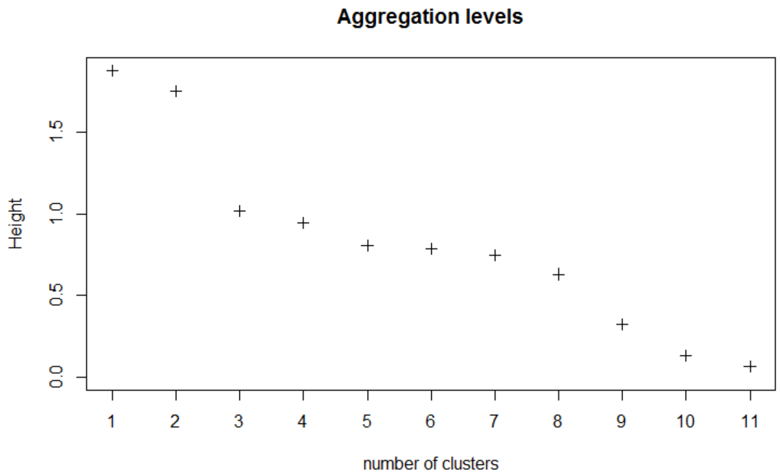


Figure C4. Agglomeration scree plot – LSS 2008

For the LSS 2005 data we can see from figure C5 below, the CART analysis still identifies firm size as the first most important discriminating variable, followed by export status. However, import status does appear as a discriminating variable too and implies that a small number of importers can be identified within the groups of medium to large exporters, small exporters as well as small non-exporters (if their intermediate input intensity is below a certain threshold).

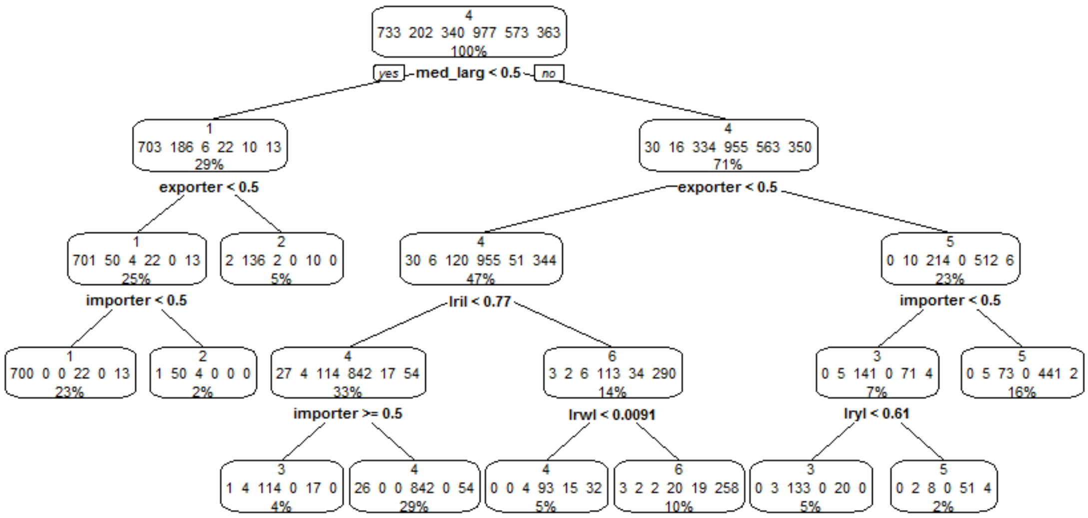


Figure C5. CART: Decision tree on the raw (original) LSS 2005 variables. Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 7 clusters (based on principal components).

The CART analysis on the LSS 2008 (figure C6) data identifies import status as the first most important variable. Firm size and export status also appear as important variables, but to a lesser degree. Indeed the 2008 data suggests that a large proportion of importing firms also export.

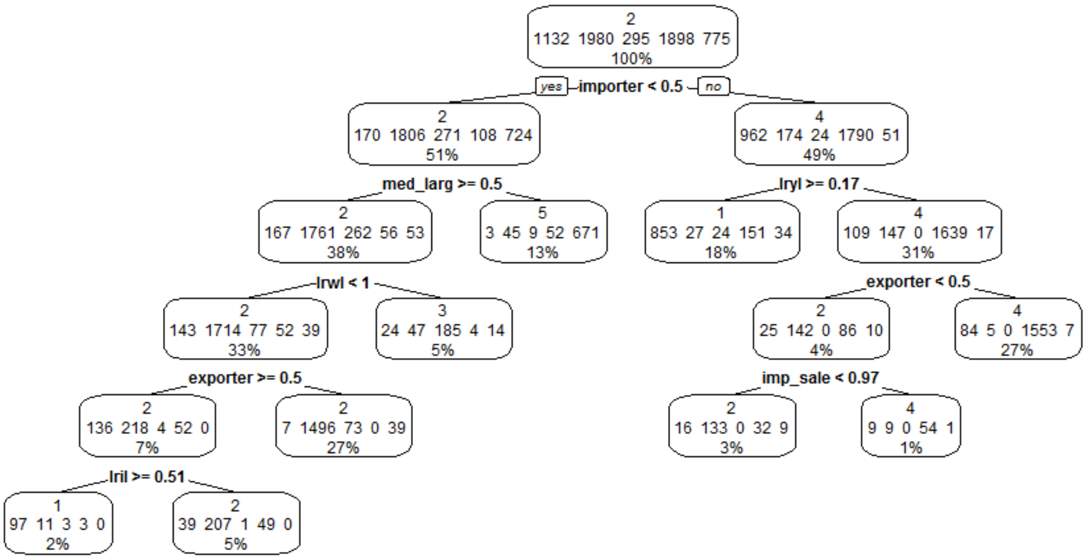


Figure C6. CART: Decision tree on the raw (original) LSS 2008 variables. Read from top to bottom, the topmost block (the root) contains the number of firms in each of the 7 clusters (based on principal components).

Appendix D

Supplementary Tables for Chapter 4

Supplementary tables as referred to in section 4.6 of Chapter 4 are presented in this Appendix. Tables D1 to D3 illustrate the impact of qualifying for the tax incentive on levels of research and development (represented by capital). Table D4 presents the robustness of removing the capital-labour ratio as a control. Table D5 presents the variables correlated with the likelihood of attrition from the sample. Finally tables D6 to D8 present results for the Heckman Two-Stage regressions as defined in section 4.5.

Table D1. Estimated Impact of Qualifying for Section 12E on the level of Capital (Research and Development) – All SBCs and Non-SBCs

	Change in Research and Development				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Qualified	0.0277	1.171	3.844**	3.844**	2.521
	(0.708)	(0.873)	(1.399)	(1.399)	(1.930)
Log(employment)	0.642*	0.420	2.887*	2.887*	2.037
	(0.370)	(0.417)	(1.348)	(1.348)	(1.680)
Log(real capital/employee)2005	-0.355	-0.436	0.327	0.327	0.0410
	(0.251)	(0.375)	(0.517)	(0.517)	(0.631)
Constant	-1.097	-0.162	-13.15*	-13.15*	-8.212
	(1.699)	(2.128)	(6.058)	(6.058)	(7.997)
Industry (SIC 3)	*No*	*No*	*No*	*No*	*No*
Observations	*46*	*33*	*16*	*16*	*12*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D2. Estimated Impact of Qualifying for Section 12E on the level of Capital (Research and Development) – Exporters versus Non-exporters

	Change in Research and Development				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Qualified	0.261	0.618	-3.104	-3.104	-17.27***
	(1.168)	(1.074)	(3.190)	(3.190)	(0.847)
Exporter	0.458	7.683**	26.01**	26.01**	213.3***
	(1.266)	(3.540)	(10.31)	(10.31)	(9.283)
Qualified X exporter	-0.795	-5.380	-18.15**	-18.15**	-175.2***
	(2.130)	(4.088)	(7.867)	(7.867)	(7.767)
Log(employment)	0.672*	0.120	-7.193	-7.193	-55.21***

	(0.389)	(0.392)	(4.146)	(4.146)	(2.430)
Log(real capital/employee)2005	-0.496	-3.079**	-9.655**	-9.655**	-76.52***
	(0.452)	(1.272)	(3.966)	(3.966)	(3.322)
Constant	-0.895	7.892*	52.11*	52.11*	401.3***
	(1.840)	(3.892)	(26.09)	(26.09)	(17.51)
Industry (SIC 3)	*No*	*No*	*No*	*No*	*No*
Observations	*46*	*33*	*16*	*16*	*12*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D3. Estimated Impact of Qualifying for Section 12E on the level of Employment and Capital (Plant and Machinery) – limited to SBCs only

			Turnover		
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Change in (log) Employment					
Exporter	-0.168***	-0.548***	-0.643***	-0.665***	-0.219**
	(0.0383)	(0.0717)	(0.0888)	(0.0881)	(0.0845)
Log(real capital/employee)2005	0.0427***	-0.0379	0.269***	0.294***	0.125***
	(0.0134)	(0.0233)	(0.0404)	(0.0461)	(0.0394)
Constant	0.300	0.399	0.0219	-1.977***	-1.793***
	(0.403)	(0.410)	(0.310)	(0.295)	(0.204)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*951*	*430*	*208*	*159*	*101*
Change in Plant and Machinery					
Exporter	1.500***	1.265***	2.212***	1.389**	7.003***
	(0.241)	(0.225)	(0.636)	(0.609)	(0.797)
Log(employment)	-1.024***	-1.919***	-2.509***	-1.485***	-3.151***
	(0.180)	(0.241)	(0.508)	(0.553)	(0.445)
Log(real capital/employee)2005	-1.824***	-1.567***	-1.294***	-0.992	-3.200***
	(0.122)	(0.114)	(0.364)	(0.771)	(0.614)
Constant	9.488***	13.05***	15.27***	7.403***	21.83***
	(1.361)	(1.337)	(2.404)	(1.453)	(1.941)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*255*	*156*	*82*	*71*	*43*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D4. Estimated Impact of Qualifying for Section 12E on the level of Employment – without capital:labour as a control.

	Turnover				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Change in (log) Employment					
Tax treated	0.153***	0.0580	0.0994	0.0892	-0.0950
	(0.0338)	(0.0439)	(0.0644)	(0.0746)	(0.0863)
Exporter 2005	-0.0263	-0.168***	-0.172**	-0.283***	-0.396***
	(0.0503)	(0.0587)	(0.0779)	(0.0874)	(0.0990)
Tax treated X exporter 2005	-0.116*	-0.294***	-0.196	-0.178	0.364**
	(0.0676)	(0.0920)	(0.133)	(0.149)	(0.149)
Constant	0.199	0.294	0.253	-1.249***	-0.0654
	(0.531)	(0.480)	(0.440)	(0.443)	(0.355)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*1,567*	*828*	*404*	*321*	*208*

Table D5. Variables Affecting Attrition from Sample

	Will Attrite in 2008				
VARIABLES	R6 - R22 million	R10 - R18 million	R12 - R16 million	R12.5 - R15.5 million	R13 - R15 million
Qualified	0.0434	0.0672*	-0.0914*	-0.210***	-0.285***
	(0.0375)	(0.0388)	(0.0511)	(0.0538)	(0.0676)
Output/labour	-0.0384***	-0.0318*	-0.0737***	-0.0351	0.0496
	(0.0125)	(0.0168)	(0.0262)	(0.0294)	(0.0436)
% of unskilled workers	-0.205***	-0.174***	-0.207***	0.228**	0.101
	(0.0298)	(0.0447)	(0.0755)	(0.0908)	(0.119)
Return on assets	0.000731**	-0.261***	-0.211***	-0.190***	-0.268***
	(0.000311)	(0.0438)	(0.0698)	(0.0691)	(0.0750)
StatsSA size 2	-0.0956***	-0.0685*	0.0757	-0.0133	
	(0.0361)	(0.0403)	(0.0779)	(0.0958)	
Constant	0.152	0.0856	0.459	0.457	-0.412
	(0.386)	(0.345)	(0.351)	(0.345)	(0.458)
Industry (SIC 4)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*2,439*	*1,097*	*548*	*403*	*262*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D6. Heckman Selection Results of the Impact on Employment and Capital – All SBCs and Non-SBCs

	Change in (log) Employment		**Change in Plant and Machinery**	
VARIABLES	R6 - R22 million	R10 - R18 million	R6 - R22 million	R10 - R18 million
Regression Equation				
Qualified	0.202***	-0.0709	-3.103**	0.506
	(0.0701)	(0.103)	(1.345)	(1.663)
Log(real capital/employee)2005	0.0950***	0.136***	-0.394	-1.136*
	(0.0245)	(0.0394)	(0.448)	(0.641)
Log (employment)			-0.00178	1.658
			(0.879)	(1.236)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*
Selection Equation				
Qualified	-0.659***	-0.704***	-0.435***	-0.796***
	(0.115)	(0.192)	(0.108)	(0.195)
Stats SA firm size control	*Yes*	*Yes*	*Yes*	*Yes*
Industry (SIC 4)	*Yes*	*Yes*	*Yes*	*Yes*
Lambda (Inverse Mills)	0.00426	0.0181	1.298	1.635
	(0.104)	(0.135)	(2.030)	(2.089)
Observations	*1,008*	*413*	*1,001*	*408*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D7. Heckman Selection Results of the Impact on Employment and Capital– Exporters versus Non-exporters

	Change in (log) Employment		**Change in Plant and Machinery**	
VARIABLES	R6 - R22 million	R10 - R18 million	R6 - R22 million	R10 - R18 million
Regression Equation				
Qualified	0.207***	-0.0842	-3.450**	-1.587
	(0.0769)	(0.121)	(1.501)	(1.996)
Exporter	0.0200	-0.250*	-0.884	-2.422
	(0.0990)	(0.135)	(1.956)	(2.247)
Qualified X exporter	-0.0107	-0.0348	1.323	7.508*
	(0.137)	(0.247)	(2.587)	(4.031)
Log(real capital/employee)2005	0.0950***	0.134***	-0.427	-1.062*
	(0.0249)	(0.0386)	(0.452)	(0.637)
Log (employment)			-0.0485	1.450
			(0.883)	(1.227)
Constant	-1.200***	-0.884*	10.63*	1.611

	(0.292)	(0.507)	(6.145)	(10.25)
Industry (SIC 3)	*Yes*	*Yes*	*Yes*	*Yes*
<u>Selection Equation</u>				
Qualified	-0.659***	-0.704***	-0.435***	-0.796***
	(0.115)	(0.192)	(0.108)	(0.195)
Stats SA firm size control	*Yes*	*Yes*	*Yes*	*Yes*
Industry (SIC 4)	*Yes*	*Yes*	*Yes*	*Yes*
Lambda (Inverse Mills)	0.00446	-0.00956	1.229	1.331
	(0.105)	(0.133)	(2.039)	(2.075)
Observations	*1,008*	*413*	*1,001*	*408*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1

Table D8. Heckman Selection Results of the Estimated Impact of Tax Incentive on the Exporting in the Future

	Exporting in 2008	
VARIABLES	R6 - R22 million	R10 - R18 million
<u>Regression Equation</u>		
Exporter 2005	0.353***	0.00924
	(0.0791)	(0.103)
Qualified	-0.254***	-0.287***
	(0.0587)	(0.0904)
Qualified X exporter 2005	0.0753	0.452**
	(0.105)	(0.185)
Log(employment) 2005	-0.0639*	0.0329
	(0.0338)	(0.0531)
Constant	0.304	-0.0422
	(0.209)	(0.418)
Industry (SIC 3)	*Yes*	*Yes*
<u>Selection Equation</u>		
Qualified	-0.543***	-1.068***
	(0.139)	(0.284)
Stats SA firm size control	*Yes*	*Yes*
Industry (SIC 4)	*Yes*	*Yes*
Lambda (Inverse Mills)	0.0415	-0.127
	(0.0729)	(0.113)
Observations	*690*	*275*

Notes: Standard errors are in parenthesis and *** p<0.01, ** p<0.05, * p<0.1