

**Land rights in sub-Saharan Africa: measuring
impact with satellite images, machine learning and
citizen science**

by

Tawanda Chingozha

Dissertation presented for the degree of

Doctor of Philosophy (Economics)

in the Faculty of Economic and Management Sciences

at the University of Stellenbosch



Supervisor: Prof Dieter von Fintel

March 2020

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.

This dissertation includes one original paper (Chapter 2) published in the peer reviewed *Economic History of Developing Regions* (EHDR), Volume 34(2), 132-155; and two unpublished manuscripts. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

With regard to Chapters 2, 3 and 4 the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
Cleaning and analysing data; estimation; write-up of literature review and analysis of results.	75%

The following co-authors have contributed to Chapters 2 and 3:

Name	E-mail address	Nature of contribution	Extent of contribution (%)
Dieter von Fintel	_____	Commented on all drafts of the Chapters and assisted in framing research questions and firming up analysis and identification.	25%

The following co-authors have contributed to Chapter 4:

Name	E-mail address	Nature of contribution	Extent of contribution (%)
Dieter von Fintel	_____	Commented on all drafts of the Chapter and assisted in framing research questions and firming up analysis and identification.	15%
Vanessa McBride	_____ _____	Assisted with setting up Zooniverse Project and data processing.	5%
Kevin Govender	_____	Assisted with setting up Zooniverse Project.	5%

Signature of candidate:

Date: March 2020.....

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 Chapters 2, 3 and 4.
2. No other authors contributed to Chapters 2, 3 and 4 besides those specified above, and
3. Potential conflicts of interest have been revealed to all interested parties and the necessary arrangements have been made to use the material in Chapters 2, 3 and 4 of this dissertation.

Signature of co-author:

Dieter von Fintel

Vanessa McBride

Kevin Govender

Date: March 2020

Copyright © 2020 University of Stellenbosch

All rights reserved

Abstract

The thesis employs satellite imagery to measure the impacts of land rights- in data-scarce sub-Saharan Africa (SSA). SSA governments are politically and financially constrained to provide objective and reliable research data at a reasonable spatial and temporal frequency. The thesis hence fills important data gaps. The research content highlights the importance of land tenure security enforcement and access to markets in rural SSA.

It is widely acknowledged that colonial institutions, particularly private property rights, continue to affect modern development. Across SSA, the majority of people rely on agriculture as a source of livelihood. Hence, agriculture has an important role to play in alleviating poverty and inequality. A consequence of extractive colonial institutions is that they selectively introduced property rights, with the majority of indigenous farming in SSA remaining under customary tenure system. This system limits the extent to which there can be effective market participation. Low investments in public goods (in particular roads and railways) and the relatively poor quality of the land in these areas compounds the problem. Chapter 2 of the thesis investigates access to markets as an important pre-condition for land titles to affect agricultural growth. Using the case of Southern Rhodesia, we investigate whether land titles incentivised African large-scale holders in the Native Purchase Areas (NPAs) to put proportionally more of their available land under cultivation than their counterparts in the overcrowded Tribal Trust Areas (TTAs). We create a novel dataset by applying a Support Vector Machine (SVM) learning algorithm on Landsat imagery for the period 1972 to 1984 - the period during which the debate on the nexus between land rights and agricultural production intensified. Our results indicate that land titles are only beneficial when farmers are located closer to main cities, main roads and rail stations or sidings.

In order to address past imbalances, SSA countries have attempted various reforms in the agricultural sector, including land redistribution and tenure reform. These reforms have not translated to agriculture growth; the main argument for stagnation post-reform is that governments do not follow through in enforcing property rights. Chapter 3 of the study focuses on Zimbabwe's 2000 Fast Track Land Reform Program (FTLRP) that reallocated more than 80% of land previously held by Europeans to the African majority. We rely on a novel, countrywide dataset of the amount of land under cultivation and crop quality [Normalised Difference Vegetation Index (NDVI)] as the endogenous variables. No study has measured the national impact of the programme on agriculture. The wide scale of the FTLRP offers a unique opportunity to interrogate how incomplete property rights (enforced land titles) affect crop cultivation in SSA post land reform, within a natural experiment design. Our Difference-in-Difference (DID) and Spatial Regression Discontinuity (RDD) estimates suggest lack of follow-through. Land redistribution reduces crop cultivation and crop quality significantly.

Measuring socio-economic change using remote-sensed data is also important within urban settings. The unplanned nature and unregistered status of commercial and residential informal establishments in urban SSA limit economic potential because of lack of land titles. Where settlements are unplanned and businesses are unregistered, trust lacks, land markets are imperfect and other opportunity costs arise. Owners or occupants cannot use informal establishments as collateral to access credit, for example. Informality also has direct costs if urban services and amenities buckle under pressure. High informality results from a migration rate that exceeds job creation in urban areas. Authorities can choose between destroying

informal establishments and efficient urban planning and enforcing tenure security to manage urban densification. The latter requires the development of cadastral databases and land use maps – an exercise which may be costly and resource intensive. Chapter 4 investigates the use of citizen science to classify the informal sector and different land use types from Very High Resolution (VHR) satellite images. It explores the conditions or factors affecting the precision of generating land cover maps/cadastral databases through citizen science. Cost minimization should not significantly sacrifice quality. The chapter presents a pilot study with a group of 41 Stellenbosch University students, who volunteered to classify different land use types. We use a sample of satellite images before and after Operation Restore Order (ORO) (a 2005 clean-up operation that destroyed informal structures in Zimbabwe's main urban areas). Estimates show that the higher the number of classifications, the better the precision axiom in accordance to Linus Law does not hold; possibly due to the irregular, small and sparsely distributed nature of informal structures. It is also shown that learning effect (experience and training), as well as the demographic variables of race, sex, nationality and student volunteer hands-on experience with the land use type in question are important factors affecting classification accuracy.

Opsomming

Die tesis gebruik afstandwaarnemingsbeelde om sosio-ekonomiese verandering in data-skaars Afrika suid van die Sahara (sub-Saharan Africa = SSA) te meet. Regerings van SSA word polities en finansiëel beperk om objektiewe en betroubare navorsingsdata teen 'n billike ruimtelike en tydfrekwensie te verskaf. Hierdie tesis vul gevolglik belangrike datagapings. Die navorsingsinhoud beklemtoon die belang van die handhawing van grondbesitsekereheid en toegang tot markte in landelike SSA.

Daar word wyd erken dat koloniale instellings, en veral privaat eiendomsreg, steeds moderne ontwikkeling beïnvloed. Oor die hele SSA heen vertrou die meerderheid mense op die landbou as 'n bron van hul lewensbestaan. Die landbousektor speel dus 'n belangrike rol by die verligting van armoede en ongelykheid. 'n Gevolg van die onttrekking van koloniale instellings is dat hulle eiendomsreg selektief ingestel het, terwyl die meerderheid van die inheemse boerdery in SSA onder die gewone besitstelsel bly funksioneer. Hierdie stelsel beperk die mate waarin daar effektiewe markdeelname kan wees. Lae beleggings in openbare goedere (veral paaie en spoorweë) en die betreklik swak gehalte van die grond in hierdie gebiede vererger net die probleem. Hoofstuk 2 van die tesis ondersoek toegang tot markte as 'n belangrike voorwaarde waarop grondtitels die groei in die landbousektor kan beïnvloed. Met behulp van die geval van Suid-Rhodesië ondersoek ons of grondtitels die aansporing insentief was wat Afrika se grootskaalse titelhouers in die inheemse aankopegebiede (Native Purchase Areas = NPA's) aangespoor het om proporsioneel meer van hul beskikbare grond te bewerk as hul eweknieë in die oorvol stamtrustgebiede (Tribal Trust Areas = TTA's). Ons skep 'n nuwe datastel deur 'n steunvektormasjien-leeralgoritme (Support Vector Machine = SVM) op Landsat-beeldmateriaal vir die tydperk 1972 tot 1984 toe te pas – die tydperk toe die debat oor die neksus tussen grondregte en landbouproduksie verskerp het. Ons resultate dui daarop dat grondtitels slegs voordeel inhou as boere nader aan die belangrikste stede, hoofweë en spoorwegstasies of sylyne geleë is.

Ten einde die wanbalanse van die verlede aan te pak, het die SSA-lande gepoog om verskillende hervormings in die landbousektor teweeg te bring, onder meer grondhervorming en besithervorming. Hierdie hervormings het egter nie tot groei in die landbou gelei nie; die belangrikste argument vir stagnasie ná hervorming is dat regerings nie voortgaan en eiendomsreg afdwing nie. Hoofstuk 3 van die studie fokus op Zimbabwe se 2000 Fast Track Land Reform-program (FTLRP) wat meer as 80% van die grond wat voorheen deur Europeërs besit is, aan die meerderheid Afrikane herverdeel het. Ons gebruik 'n nuwe, landwyse datastel wat die hoeveelheid bewerkte grond en gewasgehalte [Normalised Difference Vegetation Index = NDVI] as endogene veranderlikes bevat. Geen studie het nog die nasionale impak van die program op die landbou gemeet nie. Die wye skaal van die FTLRP bied 'n unieke geleentheid om uit te vind hoe onvolledige eiendomsreg (afgedwonge grondtitels) gewasverbouing in die SSA ná grondhervorming binne 'n natuurlike eksperimentontwerp beïnvloed. Ons beramings ten opsigte van verskil-in-verskil (Difference-in-Difference = DID) en ruimtelike regressie-diskontinuiteit (Spatial Regression Discontinuity = RDD) dui op 'n gebrek aan deurvoer. Grondhervorming verlaag die gewasverbouing en gewasgehalte aansienlik.

Die meting van sosio-ekonomiese verandering met behulp van afstandwaarnemingsdata is ook belangrik in stedelike omgewings. Die onbeplande aard en ongeregisteerde status van kommersiële en residensiële informele instellings in stedelike SSA beperk ekonomiese

potensiaal weens 'n gebrek aan grondtitels. Waar nedersettings onbeplan is en ondernemings nie geregistreer is nie, is daar 'n gebrek aan vertroue, is die grondmark problematies en ontstaan daar ander geleentheidskoste. Eienaars of bewoners kan byvoorbeeld nie informele instellings as aanvullende sekuriteit gebruik om toegang tot krediet te verkry nie. Informaliteit het ook direkte koste wanneer stedelike dienste en geriewe onder druk meegee. Hoë informaliteit is die resultaat van 'n migrasiekoers wat werkskepping in stedelike gebiede oorskry. Owerhede het die keuse tussen die vernietiging van informele instellings en doeltreffende stedelike beplanning, en die toepassing van verblyfsekerheid wat die bestuur van stedelike verdigting betref. Laasgenoemde vereis die ontwikkeling van kadastrale databasisse en grondgebruikskarte – 'n oefening wat duur en hulpbronintensief kan wees. Hoofstuk 4 ondersoek die gebruik van burgerlike wetenskap om die informele sektor en verskillende soorte grondgebruik aan die hand van satellietbeelde met 'n baie hoë resolusie (Very High Resolution = VHR) te klassifiseer. Dit ondersoek die toestande of faktore wat die akkuraatheid van die generering van grondbedekkingskarte of kadastrale databasisse deur middel van burgerlike wetenskap beïnvloed. Kosteminimalisering moenie veroorsaak dat gehalte beduidend ingeboet word nie. Die hoofstuk bied 'n loodsondersoek aan met 'n groep van 41 studente aan die Universiteit Stellenbosch, wat vrywillig aangebied het om verskillende soorte grondgebruik te klassifiseer. Ons gebruik 'n steekproef van satellietbeelde voor en ná Operation Restore Order (ORO) ('n 2005-opruimingsoperasie waarin informele strukture in Zimbabwe se belangrikste stedelike gebiede vernietig is). Ramings toon dat hoe hoër die aantal klassifiserings is, hoe meer hou die presisie-aksioom volgens die Linus-wet nie steek nie; waarskynlik as gevolg van die onreëlmatige, klein en yl verspreide aard van informele strukture. Daar word verder aangetoon dat die leereffek (ervaring en opleiding), sowel as die demografiese veranderlikes van ras, geslag, nasionaliteit en die studentevrywilligers se praktiese ervaring betreffende die betrokke grondgebruikstipe, belangrike faktore is wat die klassifikasie-akkuraatheid beïnvloed.

Acknowledgements

First, I would like to sincerely acknowledge my supervisor Prof. Dieter von Fintel to whom I owe gratitude for his guidance, vision, wisdom and expertise, encouragement and most importantly for believing that this work would be a success. I thank him for the challenges and opportunities without which success would not have been possible. I would also like to thank colleagues in the Economics Department at Stellenbosch University, particularly within the Research on Social Economic Policy (ReSEP) and Laboratory for the Economics of Africa's Past (LEAP) ambits. It is not possible to attach value to the feedback that I received in the ReSEP and LEAP brown bag lunches. I shall forever be grateful. Within these research groups, I am indebted to Prof. Servaas van der Berg for tapping into his wealth of experience to offer critical feedback on my work and for his financial support. I also thank Prof. Johan Fourie for his unparalleled contributions to my work during the LEAP sessions and the opportunity he provided me to share my research in a few of the Economic History lectures. I would like to thank Thomas Ferreira and Anderson Gondwe for their guidance as I began my studies. The Graduate School of Economic and Management School (GEM) bore the financial burden of my studies and Dr. Jaco Franken (our PhD Manager) is a man to whom I also owe my gratitude. I also acknowledge assistance from the Beit Trust in my last year of study 2019. Also, funding received through my supervisor (Prof. Dieter von Fintel) from the National Research Foundation (Grant number IFR170208222264) and the Elite Fund of the Faculty of Economic and Management Sciences is acknowledged.

Thanks also go to my colleagues within the GEM for their friendship and camaraderie and also their invaluable input into my work through the weekly GEM sessions. I am grateful to the International Astronomical Union (IAU) Office of Astronomy for Development (OAD) for extending a Development Economist Fellowship to me. Kevin Govender, Prof. Vanessa McBride and Ramasay Venugopal from the OAD helped to stretch the boundaries of what is possible, by encouraging and assisting me to venture into the generation of economic data through crowdsourcing using the Zooniverse platform. Papers out of this thesis have been presented at different international and local conferences. I would therefore like to thank Jörg Peters, who led the discussion on my paper in the Poverty and Inequality Parallel Session 6 at the 2017 Oxford CSAE conference and contributions from the session attendees. I also acknowledge Prof. Elizabeth Carlson of Penn State University, for facilitating my travel to the 2017 Penn State Conference on Data Driven Development. I thank Belinda Archibong and other conference attendees for the input I received. Further input from attendees at the ESSA Biennial Conference, the African Economic History Network (AEHN) meeting in Stellenbosch, the Regional Science CRUISE conference in Stellenbosch, the DataFirst

Conference on Data Quality in Cape Town, the IAU General Assembly Focus Meeting 15 (Astronomy for Development), the World Bank Conference on Land and Poverty (2019) and Stellenbosch University Economics PhD Conference was crucial in the development of this work. Last, but not least, without the support and encouragement from my wife Kina, as well as the regular lighter moments that my children Rutendo and Nevada created, this work would have been an insurmountable task. And finally, all glory be to God.

Table of Contents

Declaration	ii
Abstract	iv
Opsomming	vi
Acknowledgements	viii
Table of Contents	x
List of Tables.....	xiv
List of Figures	xvi
List of Acronyms and Abbreviations	xvii
CHAPTER 1 INTRODUCTION.....	1
1.1 Background and Context	1
1.2 Theoretical Background	2
1.2.1 Property Rights, Social Cost and Institutions.....	2
1.2.2 Economic History of Market Access.....	5
1.3 Importance of Land in Sub-Saharan Africa’s Socio-economic Development	6
1.4 Evolution of Land Institutions in SSA	7
1.5 Other Factors Affecting Socio-economic Development in SSA	8
1.6 Motivation and Research Aims	10
1.6.1 The complementarity between property rights and market access for crop cultivation in Rhodesia: Evidence from historical satellite data	11
1.6.2 Effects of poor land rights enforcement post-redistribution, on cultivation and crop quality: Satellite data evidence from Zimbabwe	12
1.6.3 A citizen science approach to classifying urban informality and other urban land use types using satellite imagery	14
1.7 An “African statistical tragedy”	15
1.8 Summary	16
CHAPTER 2 PROPERTY RIGHTS, MARKET ACCESS AND CROP CULTIVATION IN RHODESIA: EVIDENCE FROM HISTORICAL SATELLITE DATA	17
2.1 Introduction	17
2.2 Rhodesia Overview	19
2.2.1 The role of land in Rhodesia.....	19
2.2.2 Agriculture and the economy in Rhodesia	22
2.2.3 African agriculture in the 1970s and early 1980s.....	25
2.3 Data and Analysis Approach.....	26
2.3.1 Classification of images using machine learning	26

2.3.2 Approach	28
2.3.3 Model specification	28
2.3.4 Explanatory variables	29
2.4 Descriptive Data Discussion.....	31
2.5 Results and Discussion.....	32
2.6 Robustness Checks	38
2.7 Conclusion.....	39
CHAPTER 3 CULTIVATION AND CROP QUALITY AFTER THE FAST TRACK LAND REFORM PROGRAM: SATELLITE DATA EVIDENCE FROM ZIMBABWE.....	
3.1 Colonial Extractive Institutions Persistence, SSA Underdevelopment and Change	41
3.2 Importance of Land Rights in Poverty Alleviation	43
3.2.1 Property rights	44
3.2.2 Land titles and credit	45
3.2.3 Land titles and social contracts (trust).....	46
3.3 Land Reform in Developing Countries	46
3.3.1 Motivations for land reform	46
3.3.1.1 Political land reforms	47
3.3.1.2 Economic land reforms.....	47
3.3.1.3 Egalitarian (poverty reducing) land reforms	48
3.3.2 Approaches to land reform	49
3.3.3 Study contribution	50
3.3.4 The study setting.....	51
3.3.5 Zimbabwe Fast Track Land Reform Program (FTLRP)	52
3.3.6 Linking FTLRP, property rights and 99-year leases	55
3.3.7 Other factors affecting FTLRP	57
3.4 Data and Identification	57
3.4.1 Night lights data (NLD).....	58
3.4.2 Land cover data	60
3.4.3 Normalised Difference Vegetation Index (NDVI)	61
3.4.4 Classification of Landsat images	61
3.4.5 Support vector machines (SVM) classification algorithm.....	64
3.4.5.1 Image correction and supervised training points identification in QGIS	65
3.4.5.2 Classification using the SVM Algorithm in R.....	65
3.4.6 Difference-in-Difference identification (DID)	66
3.4.7 Regression Discontinuity Design identification (RDD)	68

3.5 Results and Discussion.....	69
3.5.1 Control variables	69
3.5.2 Descriptive data analysis	72
3.5.3 Discussion of results.....	75
3.6 Caveats, Assumptions and Robustness Checks.....	80
3.7 Conclusion.....	81
CHAPTER 4 A CITIZEN SCIENCE APPROACH TO CLASSIFYING URBAN INFORMALITY AND OTHER URBAN LAND USE TYPES USING SATELLITE IMAGERY	83
4.1 Introduction	83
4.2 Importance of Land Use Mapping.....	86
4.3 The Role of Citizen Science in Urban Land Use or Cadastral Mapping.....	87
4.3.1 Past examples of citizen science projects	88
4.3.2 Some drawbacks of the citizen science approach.....	89
4.4 Measurement of Urban Informality.....	90
4.4.1 Urban informality in Zimbabwe.....	93
4.4.2 The 2005 clean-up operation (ORO) in Zimbabwe.....	93
4.5 Experiment Setup, Implementation and Estimation Strategy.....	94
4.5.1 Processing classification results	99
4.5.2 Measuring classification accuracy.....	100
4.5.3 Correlates of classification accuracy	101
4.6 Results and Discussion.....	102
4.6.1 Classification accuracy estimates	106
4.6.2 Effects of demographic attributes estimates.....	111
4.7 Conclusion and Recommendations	114
CHAPTER 5 CONCLUSIONS AND RESEARCH IMPLICATIONS	116
5.1 Introduction	116
5.2 Research Contributions	116
5.3 Summary of Findings	118
5.3.1 The complementarity between property rights and market access for crop cultivation in Rhodesia: Evidence from historical satellite data	118
5.3.2 Effects of poor land rights enforcement post-redistribution on cultivation and crop quality: Satellite data evidence from Zimbabwe	119
5.3.3 A citizen science approach to classifying urban informality and other urban land use types using satellite imagery	120
5.4 Research Implications	121
5.5 Conclusions and Suggestions for Future Research.....	122

References	126
A. APPENDIX A: SUPPLEMENTARY TABLES AND INFORMATION.....	144
A.1 Chapter 2 Results Using Non-Linear Estimators.....	144
A.2 Chapter 3 Regression Discontinuity Design (RDD) estimates at 2 and 5km from border	146
A.3 Chapter 3 Allocation of different crop exports to Agro-ecological regions	148
B. APPENDIX B: DESCRIPTION OF CROP CULTIVATION DATA WORK.....	152
B.1 The Data Approach.....	152
B.2 Machine Learning.....	152
B.3 Classification	154
B.4 Processing Normalised Difference Vegetation Index.....	155
C. APPENDIX C: SOME ACCURACY VALIDATION	157
D. APPENDIX D: VALIDATION OF DID AND RDD ASSUMPTIONS.....	160
D.1 Difference in Difference (DID) – Parallel Trends Assumption.....	160
D.2 RDD Assumptions.....	161
i) Dis-continuity at Border.....	161
ii) Compound Treatments.....	161
iii) Dealing with asymptotic bias of standard errors.....	165
iv) Selection of Cutoffs	166
v) Demonstrating heterogeneity using Euclidian Distance	166
vi) Pre-treatment Covariate Balance.....	169
E. APPENDIX E: ZOONIVERSE DATA WORK	171

List of Tables

Table 2.1 Data Description.....	31
Table 3.1 Changes in the distribution structure of the land.....	52
Table 3.2 Timeline of land reform since majority rule.....	54
Table 3.3 Landsat 4-5 Thematic Mapper (TM) satellite image bands	64
Table 3.4 Correlation between endogenous variables and poverty estimates	72
Table 3.5 Descriptive Data Averages.....	74
Table 4.1 Experiment design.....	95
Table 4.2 Allocation of students to streams	96
Table 4.3 Fischer Tests for statistical balance.....	97
Table 4.4 Classification tasks	98
Table 4.5 Individual classification descriptive statistics	103
Table 4.6 Consensus data descriptive statistics.....	103
Table 4.7 MLogit estimation (marginal effects for the coefficient on the 2006 time dummy)	105
Table 4.8 Night lights data estimates	106
Table 4.9 Main effects estimates (Classification Accuracy 1)	108
Table 4.10 Main effects estimates (Classification Accuracy 2)	110
Table 4.11 Demographic attributes estimates.....	111
Table A.1 Class Effects of Colonial Land Policy (Fractional Probit Estimates).....	144
Table A.2 Class Effects of Colonial Land Policy (Tobit Estimates).....	145
Table A.3 RDD Regression Estimates (2km Buffer from Treatment Region Border).....	146
Table A.4 RDD Regression Estimates (5km Buffer from Treatment Region Border).....	147
Table A.5 Export crops allocation.....	148
Table A.6 Correlating trade in crops and exchange rates.....	148
Table A.7 European Commercial Agriculture.....	149
Table A.8 Comparison between EAs and NPAs	150
Table B.1 List of Landsat Imagery used for Chapter 2 by date.....	152
Table B.2 List of Landsat Imagery used for Chapter 3 by date.....	153
Table C.1 Time Series Correlations of European Commercial Agriculture with Aggregates from Classified Images.....	158
Table C.2 Correlations of Time Series of African Agriculture with Aggregates from Classified Images	158
Table D.1 Using <i>rdrobust</i>	166
Table D.2 Heterogeneity Robustness: EAs (Treated) TTAs (Untreated) - Midlands	169

Table D.3 Heterogeneity Robustness: EAs (Treated) TTAs (Untreated) – Mashonaland East	169
Table E.1 Area Classified by Type.....	172

List of Figures

Figure 2.1 The Land Apportionment Act Land Distribution.....	21
Figure 2.2 Ratio of cropland to total ward area (1972-1984).....	26
Figure 2.3 Distance variables	30
Figure 2.4 Proportion of land under crop cultivation (1972-1984)	33
Figure 2.5 Advantages of NPAs over TTAs in various periods and locations: results from rolling regressions (1972-1984)	35
Figure 3.1 The Land Apportionment Act of 1930 (LAA, 1930).....	53
Figure 3.2 Idealised spectral signatures for selected colours	62
Figure 3.3 Near-infrared and natural colour composites of agricultural land	63
Figure 3.4 Performance of the SVM algorithm in R	66
Figure 3.5 Natural regions of Zimbabwe.....	70
Figure 3.6 Comparing 1999 versus 2000 (NLD and NDVI).....	71
Figure 3.7 Raster Images for NLD and NDVI	73
Figure 4.1 Discerning small objects on satellite images using the Bezier tool	90
Figure 4.2 Areas sampled in the study	95
Figure 4.3 Classification using the Bezier tool.....	98
Figure 4.4 Creating multi-parts from single polygon to address crossing lines	99
Figure 4.5 Defining classification accuracy	101
Figure A.1 Southern Rhodesia 1962 Population	149
Figure A.2 1930 Land Apportionment Act	150
Figure A.3 Calculating Distance variables in QGIS.....	151
Figure B.1 Landsat Frames over Zimbabwe	154
Figure B.2 Cutting Wards at Landsat Boundary	155
Figure C.1 Comparison of machine-learning generated data to other sources.....	157
Figure C.2 Comparing Near Infrared composite and predicted raster.....	159
Figure D.1 DID Parallel Trends Assumptions.....	160
Figure D.2 RD plots with CI – Crop Cultivation	162
Figure D.3 RD plots with CI – NDVI	163
Figure D.4 Ward boundaries overlapping with LAA 1930	164
Figure D.5 Ideal conditions for GRD	164
Figure D.6 Controlling for District FEs to deal with compound Effects.....	165
Figure D.7 Mberengwa, Murehwa and Kwekwe Test Areas	168
Figure D.8 Pre-treatment covariate balance	170
Figure E.1 User classification polygons in QGIS map canvas.....	171

List of Acronyms and Abbreviations

BSAC	British South African Company
CS	Citizen science
DID	Difference-in-Difference
DN	Digital number
EA	European Areas
ESAP	Economic Structural Adjustment Program
FE	Fixed effects
FTLRP	Fast Track Land Reform Program
GDP	Gross Domestic Product
GE	Google Earth
GEM	Graduate Economic and Management
LAA	Land Apportionment Act
LEAP	Laboratory for the Economics of Africa's Past
LRAD	Land Redistribution for Agricultural Development
ML	Machine learning
MSS	Multi-Spectral Sensor
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
NLD	Night Lights Data
NOAA	National Oceanic and Atmospheric Administration
NPA	Native Purchase Areas
OLS	Ordinary Least Squares
ORO	Operation Restore Order
OSM	Open Street Map
QGIS	Quantum Geographic Information System
RDD	Regression Discontinuity Design
ReSEP	Research on Social Economic Policy
RF	Rhodesian Front
RNFU	Rhodesian National Farmers Union
RS	Remote sensing
SCP	Semi-Automatic Classification Plugin
SSA	sub-Saharan Africa
SU	Stellenbosch University
SVM	Support vector machine

TM	Thematic mapper
TOA	Top of the Atmosphere
TTA	Tribal Trust Areas
UK	United Kingdom
UN	United Nations
US	United States
USGS	United States Geological Survey
VGI	Volunteered geographic information
VHR	Very high resolution
ZIDERA	Zimbabwe Democracy and Economic Recovery Act
ZIMPREST	Zimbabwe Programme for Economic and Social Transformation

CHAPTER 1

INTRODUCTION

1.1 Background and Context

Poverty reduction takes centre stage within the United Nations (UN) Sustainable Development Goals¹ (SDG) framework. Goal 1 of the SDGs aims to “End poverty in all its forms everywhere”. Under Goal 1, Indicator 1.4.2 aims to achieve a higher “Proportion of total adult population with secure tenure rights to land, with legally recognised documentation and who perceive their rights to land as secure, by sex and by type of tenure” by 2030 (United Nations [UN], 2019). Indicator 1.4.2 is of crucial importance in sub-Saharan Africa (SSA), given the dominance of agriculture as a source of livelihood in the region (Davis, Di Giuseppe & Zezza; 2017). This thesis focuses on the role of land rights in Zimbabwe and the role they play in supporting poverty reduction. Within the broader theoretical discourse, the thesis deals with institutions, initial conditions and markets access as the three views that are widely accepted to influence the spatial variation of contemporary development (*see* Redding and Sturm, 2008). For institutions, the discussion mainly builds on the work of Acemoglu, Johnson and Robinson (2001), while the disease environment and climate [based on Bloom et al., (1998)] are dealt with in the empirical strategy sections of Chapters 2 and 3 as initial conditions. Lastly, within this broader frame of reference, the study dwells on market access - building mainly from the work of Krugman (1991) as well as prominent literature in the economic history of transport institutions such as the work of Donaldson and Hornbeck (2016). In forming an understanding of how land rights played a role in defining poverty and other socio-economic outcomes in Zimbabwe and other developing countries, it is important to discuss the theoretical background of property rights and market access. The next section thus presents some theoretical background on the institutions and historical context of property rights (including change in the institutions) and market access.

¹ Accessed <http://www.undp.org/content/undp/en/home/sustainable-development-goals.html>

1.2 Theoretical Background

1.2.1 Property Rights, Social Cost and Institutions

Poorly defined property rights negatively affect economic development. Hornbeck (2010) argues that introduction and near total adoption of barbed wire in the US between the period 1880 and 1900 allowed increased settlement in non-woody areas, and the ability of farmers to protect their land from encroachment. This resulted in agricultural growth. At the same time Ashton (1997) postulates that the private enclosure of common lands in Britain might have been the trigger for “The Industrial Revolution” through ramping up the output from agriculture as well as the supply of labour to other sectors. Hence property rights are important in the agriculture discourse.

Coase (1960) and Demsetz (1974) indicate that property rights give consent to be able to act in certain ways, with the holder of the rights expecting the community to prohibit others from interfering with his/her actions as long as the actions are allowed in the conditions governing the rights. In the words of Demsetz (1974), “property rights specify how persons may be benefited and harmed, and, therefore, who must pay whom to modify the actions taken by persons”. Demsetz (1974) sets forth the argument that a property rights regime is about the absorption or allocation of social costs (externalities) arising out of an activity or transaction. In the main, Coase (1960) emphasizes the reciprocal nature of the social costs, contending that “restraining A” (a smoke producing factory, for instance) so as to avoid harming B (a neighborhood resident affected by the factory smoke) also harms A: - which then exposes a flaw in the Pigouvian model of addressing social costs. According to Coase (1960) the problem of social cost has to be considered in “total and at the margin”, which implies that any two alternative choices of allocating/internalizing social cost should be weighed against each other in terms of importance with the eventual end goal of arriving at a compromise. Chapter 3 of the thesis investigates the effect of redistributive land reform on crop cultivation and crop quality. Within that context, it may be important to consider how the social cost arising from dispossession of land can be internalized or compensated as a way of properly guiding incentives (Demsetz, 1974).

A key takeaway from the work of Coase (1960) is that conflicted parties are able to negotiate a mutually acceptable level of economic activity whether the party creating the negative externality is liable for damages or not. With the enforceability of legal damage, the responsible party can agree to compensate the party affected, whereas without liability of damage, the

affected party could agree to make a certain payment to reduce the level of production to a level in which they are not negatively affected. In both cases, only the wealth distribution of both parties is affected. Coase (1960) argues that regulating the right to use a commodity (factor of production) might not be necessary because those economic agents who require the commodity must have an incentive to pay potential competitors in order to maintain that exclusive right to use it. Property rights can take various forms. Whereas the enforcement of property rights conventionally happens through the legal channel², but they can also be entrenched by certain technologies. A good example is barbed wire adoption in the US between 1890 and 1900 that protected farmlands from cattle encroachment (Hornbeck, 2010).

Building on Coase (1960), Farrell (1987) mentions that assumption of competitive markets is not necessary for efficiency to be achieved, and Calabresi (1968) adds that as long as there are no restrictions to negotiation, then economic agents will negotiate until efficient outcomes are reached. Still, perfect competition rarely exists in practice and transaction costs do affect the internalization of social cost. This therefore warrants some discussion on transaction costs. Coase (1960) explains that in executing a transaction on the market it is imperative to know who the deal counterpart would be, convey the intention to deal plus the respective terms to the deal counterpart, write up the contract as well as put in place inspection mechanisms to ensure that the terms are being adhered to. As Coase (1960) postulates, a rearrangement of legal rights can only take place if the marginal production (value) is greater than the transaction costs. There are three potential ways to deal with transaction costs namely the creation of a firm (as long as the administrative costs of running the firm are less than the transaction costs), government regulation or “doing nothing” (Coase, 1960).

Property rights evolve over time (Demsetz, 1967). The evolution of property rights is important in the thesis because institutional changes in property rights over the course of the 20th Century have had an important implication on contemporary development as investigated in Chapters 2 and 3. For instance, while in 1910 the Shona and Matabele of Rhodesia were active agriculturalists (*see* Phimister, 1974), further creation and growth of agricultural markets (to a large extent due to World War II induced demand) as the colony developed might have influenced further competition for land in the Tribal Trust Areas (TTAs), notwithstanding the overcrowding in these areas. The TTAs were the area designated for Rhodesia’s Africans, based on the Land Apportionment Act of 1930 (LAA 1930) as in discussed in greater detail in

² Rhodesia’s Land Apportionment Act of 1930 (LAA 1930) is an important piece of legislation that discussion and empirical strategy for Chapters 2 and 3 are premised on.

Chapter 2. As the thesis discusses in Chapter 3, events in the socio-economic and socio-political discourse again inspired institutional changes in land ownership in 2000 (twenty years after the end of colonial rule and creation of the independent state of Zimbabwe). This brings the discussion to induced institutional change. Ruttan and Hayami (1984), Grabowski (1988) and Lin (1995) posit that in explaining different patterns of growth in agricultural productivity – the theory of induced institutional change is important. The theory of institutional change involves analyzing institutional innovations that take place in developing nations as technical innovation and growth.

In explaining the neo-classical institutional economics perspective, North and Thomas (1970) posit that institutional innovations occur when individuals/groups in a given society can reap positive rewards out of their investment in making change happen. In the context of Rhodesia, the Land Apportionment Act of 1930 (LAA 1930) formalized land tenure security on a racial basis – benefitting Europeans at the expense of Africans. Yet a small group of Africans benefitted from the secure land rights in the Native Purchase Areas (NPAs) as is investigated in Chapter 2. The NPAs were small land pockets in Rhodesia where Africans could hold land titles. The forces driving that innovation within the realm of property rights in 1930 had various incentives, among them to keep African competition in agriculture markets at bay. Seventy years later (in the year 2000), the ‘now under self-rule’ African elites would reconfigure the land tenure system through the Fast Track Land Redistribution Program (FTLRP) in order to achieve different objectives (among them political ends) as discussed in Chapter 3. Yet, Grabowski (1988) cautions that the discounted future returns must outweigh the costs of any changes to the institutional status quo (in this case land rights) are to be made.

Casting the induced institutional change reasoning within the thesis context - Chapter 2 begs the question whether NPA farmers had a higher proportion of land under cultivation than their counterparts in the customary tenure Tribal Trust Areas (TTAs) using market to access as an important condition. Andersson & Green (2016) mention that sometimes the policies of colonial administrations resulted in unintended consequences. They portray African farmers in the NPA areas of Rhodesia as having greatly benefitted from the LAA 1930 – yet that was not the express intention of the colonial government. As Chapter 2 investigates, the right to title (as a result of LAA 1930) may have resulted in NPAs achieving better cultivation outcomes compared to TTAs. In that case, this would be an example of how profits brought on by LAA 1930 resulted in institutional innovation. Grabowski (1988) posits that as long as the discounted returns of enforcing property rights and other institutional innovation outweigh the costs – the

system will remain in equilibrium. Yet, the racial character of LAA 1930 potentially fueled the incentives of the Zimbabwe government to reverse LAA 1930 seventy years later. Also, Olson (2008) that postulate institutions must change if development is to take place, especially if the old institutional order safeguards the vested interests of some at the expense of development as the greater good. At the same time Boserup (1996) brings the perspective that land pressures (as was the case in overcrowded TTAs before and after Zimbabwe's majority rule in 1980, and until 2000) can also induce institutional change (FTLRP being a case in point). Given the lack of consensus on the effects of Zimbabwe's FTLRP, Chapter 3 of the thesis carries out nationwide analysis using innovative, objective datasets as discussed in much greater detail in subsequent sections.

1.2.2 Economic History of Market Access

Donaldson (2018) notes that in 2017, one fifth of World Bank funding targeted the development of transportation infrastructure, inspired by the need to reduce transport costs. This is inspired by history. History has shown that periods of infrastructure construction (such as railroads) are associated with economic boom, for example in Japan, the United States (US) and Western Europe (Banerjee, Duflo and Qian, 2012). To support this argument, there exists a substantial literature on the positive effect of infrastructure on economic outcomes (Donaldson, 2018; Donaldson and Hornbeck, 2016; Banerjee, Duflo and Qian, 2012; Mlambo, 1994). Banerjee et al., (2012) evaluated the effect of transport infrastructure on Gross Domestic Product (GDP) growth in China and found that growth in infrastructure positively affected GDP levels.

The reduction of susceptibility to local production shocks, reduction in the costs of trade as well as ensuring smoother and higher income flows for farmers are some of the channels through which railways (and other transport infrastructure) can stimulate development (Donaldson, 2018). Donaldson and Hornbeck (2016) investigated the impact of railroad expansion in the US in 1890, and found that it had an important positive effect on agricultural expansion. According to Donaldson and Hornbeck (2016), agricultural land values in counties increased on the back of improved market access brought about by the railroad expansion between 1870 and 1890. Therefore, access to markets is important. Redding and Sturm (2008) provide evidence that before unification, the decline of Eastern Germany cities (population-wise) that were closer to the border with Western Germany was mainly due to lack of market access in the former.

In the US the country's emergence as the leading global economy was accompanied by intensified railroad construction towards the end of the 19th Century and the immediate effect was that areas that were close to this rail infrastructure achieved more prosperity (Donaldson and Hornbeck, 2016). In the example of India (geographical context referring to India, Pakistan and Bangladesh), Donaldson (2018) reveal that the designing and building of a railroad network by the British government greatly transformed technology of conducting trade. In the context of Rhodesia, the British colonial government also greatly invested substantial amounts of money in rail (Lunn, 1992) and road infrastructure. Railroad construction in the British territories was a means of "expanding and entrenching" the British empire and by 1911 (the end of the primary construction phase) the railway line that traversed through Northern Rhodesia had extended to the Congo border and the copper mines of Katanga (Lunn, 1992). For Rhodesia, the value of a rail and road transportation system need not be overemphasized because without a port of its own, the country needed to transport its produce to the coast (Mlambo, 1994). Internally, an efficient transportation system was required to link different mining and agriculture areas - thereby creating an internal market as well.

1.3 Importance of Land in Sub-Saharan Africa's Socio-economic Development

Having laid out the foundations of scholarly work on the institutions of property rights and market access, it is useful to now consider the importance of land in SSA's socio-economic development since the land has been a central issue that has ignited (or is likely to ignite) induced institutional change. Despite the growth of major metropolises in Africa, the majority of countries on the continent are still predominantly rural (Boone, 2017), which makes land an important livelihood source (Lund, 2011; Meinzen-Dick, Quisumbing, Doss & Theis, 2017). Pichel (2017) cites Ellen Johnson Sirleaf (former president of Liberia), who indicated that unless states secure land rights for smallholders, the African continent would continue to suffer from high levels of poverty and hunger incidence. At the same time, Shipton and Goheen (1992) identify three contemporary issues of land governance in Africa, namely, political (power), economic (wealth) and cultural (meaning). Therefore, in thinking about issues of poverty and increasing incomes, issues of land are prominent.

However, Keswell and Carter (2014) and Udry (2011) observe that there is no overwhelming evidence to support the effect of property rights on poverty alleviation. Yet, the argument that

property rights are essential for poverty reduction persists. The central argument put forward in this thesis is that the absence of well-defined private property rights in both rural and urban spaces has constrained sub-Saharan Africa's (SSA) socio-economic development. The majority of land in rural Africa is under customary tenure. Due to the lack of property rights enforcement in these customary tenure areas (and overall poor institutions), SSA has remained poor and underdeveloped. Notably, the context of Zimbabwe is used to show when property rights are effective at achieving these goals and when not.

Beside the issue of land tenure security, poor institutions constrain development in various other aspects in SSA (e.g. governments' administrative capacity). Today, many SSA governments fail to provide timely, objective and reliable data to enable proper assessment of land policies and socio-economic progress. An important secondary theme of this thesis is to generate alternative data sources from satellite images. In this way, institutional statistical capacity is extended without the use of expensive and often elusive survey data. The next section presents an overview of the evolution of land institutions in SSA.

1.4 Evolution of Land Institutions in SSA

Historically, Africa has not in general had landlords and tenants as in other parts of the world (for example, southern Asia) (Shipton & Goheen, 1992). Instead, it has been characterised by the scenario in which individual claims to land holdings are always linked to broader social groupings along family, clan, ancestral and ethnic lines (Shipton & Goheen, 1992; Place & Hazell, 1993). This is known as customary tenure and it dominates land governance in the SSA region (Chimhowu, 2019; Deininger, 2003; Lund & Boone, 2013). Chimhowu (2019) defines customary tenure as a scenario in which land is collectively owned by a society under the overall custodianship and control of some traditional authority (chief or large family/clan). An important characteristic of such land tenure systems is that collective ownership is earned by birth right and ethnicity (Boone & Nyeme, 2015), although in certain circumstances, one can acquire ownership by paying a sum of money (Chimhowu, 2019; Chimhowu & Woodhouse, 2006).

Prehistorically, customary tenure land systems were a critical aspect of state formation in Africa (Chimhowu & Woodhouse, 2006), yet they were altered and modified by both colonial and post-colonial governments (Boone & Nyeme, 2015). Through controlling the land – to stamp authority and control labour (Berry, 2002; Lund & Boone, 2013) – European settlers influenced the present day customary tenure systems in southern Africa, unlike in West Africa,

where colonisation depended more on local institutions (Deininger, 2003). Hence communal land institutions in settler economies such as Zimbabwe (the country used for the case study in this thesis) may be referred to as neo-customary tenure (Boone & Nyeme, 2015).

Shipton and Goheen (1992) and Deininger (2003) posit that customs are not constant but evolve over time. Hence there have a number of shifts in African customary tenure. Yngstrom (2002) asserts that this evolution is influenced by scarcity, since the market mechanism (through individual tenure recognition) is the best way to allocate the land under scarcity. However, land exchanges necessitated by scarcity have been characterised by disputes after the transaction has taken place (Colin & Woodhouse, 2006), highlighting weak property rights in customary/neo-customary land tenure systems.

The current status quo of customary land tenure in SSA is that communal land is registered in the name of the state³ (Adams, Sibanda & Turner, 1999), with transactions characterised by informality (Colin & Woodhouse, 2010) thereby making it difficult to discern the nature and values of customary land market transactions. Other challenges include that customary tenure systems are complex and malleable in nature to the extent that any contracts under them might be difficult to enforce contractually – resulting in little or no productive investment (De Soto, 2000). Thus, there may be a need to reform customary tenure systems as this thesis investigates. Yet, Yngstrom (2002) and Adams et al. (1999) are of the view that the state often causes insecurity and suggest that customary tenure reform may need to be community driven. Sender and Johnston (2004) support this view and indicated that poorer families may actually be negatively affected by tenure reform policies (land redistribution). As indicated by Deininger (2003), the move away from customary tenure systems is risky as they have evolved over a long period and stood the test of time (particularly due to the fact that people trust them). The thesis investigates the conditions under which reforming customary tenure systems (through land titles) may result in increased agriculture activity. The next section briefly discusses other factors affecting SSA's socio-economic development.

1.5 Other Factors Affecting Socio-economic Development in SSA

The East Asian growth miracle (Auty, 1994; Bhattacharya, 1995) and Africa's recent performance have inspired hope that the region may be the world's next growth frontier. Rodrik

³ This is with particular reference to Southern Africa. Boone and Nyeme (2015) note the heterogeneous nature of customary tenure within and across national borders.

(2016) and McMillan, Rodrik and Verduzco-Gallo (2014) note the rapid growth that SSA has achieved over the past several years. Rodrik (2016), attributes this to entrepreneurship, a burgeoning middle class and capital injections from China. On the other hand, McMillan et al. (2014) find that economic growth in Africa has mostly resulted from labour shifting to more productive manufacturing and services sectors, as agriculture has declined. However, despite the spurt in Gross Domestic Product (GDP) growth over the past several years, SSA's developmental progress is still to converge with peers around the world and the region arguably presents the greatest challenge to reducing poverty and achieving inclusive growth than any other region in the world. An important remaining constraint on which this thesis focuses, is incomplete private property rights, which arguably sustain short-run growth trajectories over the long term.

For the case study under investigation, Zimbabwe, the important question is why a country that showed so much opportunity moved from being the 'breadbasket of Africa' to a 'basket case'. The research more generally adds to the debate about why countries with seemingly abundant resources and opportunities perform dismally. This is an important debate in literature. For example, Sokoloff and Engerman (2000) question why the territories of Canada and North America turned out to be among the world's best performers yet their prospects seemed worse than nearby territories at the time of European colonisation. Importantly, the central argument in this thesis is that land governance institutions (respect for property rights) in both urban and rural areas are an important constraint that has stymied economic growth and socio-economic development in SSA as discussed in the previous section.

Without land reforms (accompanied by property rights enforcement or assertion) in the agricultural sector, opportunities for the poor may remain constrained. Some argue that to realise the full potential of the agricultural sector in SSA, the most important challenge is the establishment and enforcement of property rights (Besley & Ghatak, 2010), given that customary tenure systems are unable to effectively allocate land as population growth induces more scarcity (Yngstrom, 2002). Security of tenure is also important even within the context of urban areas. First, Shipton and Goheen (1992) note that rural folk in SSA vacillate between different occupations and livelihood sources, including migration to urban areas, where they may work or live in informal jobs or areas, respectively, hence security of tenure is critical in urban areas as well. Second, Deininger (2003) stated that many peri-urban areas fall under the jurisdiction of traditional authorities and are therefore characterised by informal land transactions. While conflict and disputes characterise these transactions, peri-urban informal

settlements also put urban services under enormous pressure. In urban areas poverty leads to insecure tenure (as the homeless settle in informal areas) and tenure insecurity increases poverty (Durand-Lasserve, 2006). The absence of security of tenure for informal business owners and residents of informal settlements or informal backyard structures imposes a number of rigidities in credit access (Aryeetey & Udry, 2010), in creating social contracts (Bezabih, Kohlin & Mannberg, 2011) and in promoting land market efficiency (Deininger, Ali, Holden & Zevenbergen, 2008; Markussen, 2008).

1.6 Motivation and Research Aims

The thesis explores the role of well-defined property rights to achieve poverty reduction and narrow inequality. First, the necessary conditions required (specifically access to markets) for property rights in the agriculture sector to increase crop cultivation are investigated. As has already been highlighted, neo-customary tenure is a result of Europeans modifying the traditional tenure regimes in settler colonies in order to control the workforce and assert their authority (Berry, 2002; Lund & Boone, 2013; Deininger, 2003); hence colonial institutions created imbalanced opportunities. In Rhodesia (colonial Zimbabwe), the Land Apportionment Act of 1930 (Floyd, 1962; Pollak, 1975) created skewed opportunities where only a few African farmers in the larger Native Purchase Areas (NPAs) had land titles, while the majority of Africans cultivated the overcrowded land of the Tribal Trust Areas (TTAs) without security of tenure.

Two decades after the end of colonial rule, such an imbalance of opportunities persisted, resulting in the Zimbabwe government embarking on a programme of land restitution – the Fast Track Land Reform Program (FTLRP) – although it has been hinted that the real objective behind it was political survival, see Berry (2002) for example. Hence, second, the effect of incomplete land reform/land restitutions (in terms of property rights enforcement and integration into the market economy) is investigated. Finally, the thesis turns to methodological ways to enhance institutional capacity that can aid in developing urban cadastres as a first step in defining property rights in urban areas; and investigates the conditions under which citizen science approaches can be implemented to identify/classify informality and other land use types from satellite imagery. The thesis presents three research papers to investigate these research questions as follows:

1.6.1 The complementarity between property rights and market access for crop cultivation in Rhodesia: Evidence from historical satellite data

Bhattacharya (1995) and Valdés and Foster (2010) suggest that it may be better for states to promote non-agricultural sources of income, as opposed to expensive agriculture reform (by expanding agriculture to marginal lands, for example). In Thailand – a developing country – Rigg, Salamanca and Parnwell (2012) found that agricultural incomes are declining relative to non-farm sectors – a finding which concurs with Bhattacharya (1995) and Valdés and Foster (2010). In an analysis of multidimensional poverty in Zimbabwe, Stoeffler, Alwang, Mills and Taruvinga (2016), found that in the wake of land reform policy in 2000, poverty increased, with the trend only reversing from 2007 onwards. Despite the scepticism reflected in Stoeffler et al. (2016), agriculture is still the dominant livelihood source in SSA. As highlighted in Chapter 3 of the thesis, it is the lack of enforcement of property rights as a way of integrating newly resettled farmers into the market economy that might have worsened the poverty situation as established by Stoeffler et al. (2016). Chapter 2 of the thesis illustrates how ill-defined property rights of the neo-customary tenure regime (which had been created by the colonial government) in the TTAs negatively affected crop cultivation in Rhodesia.

While the encouragement of smallholder farming is pro-poor (Collier & Dercon, 2014), it is important to ensure that land reform programmes that seek to promote more equal access to land guarantee property rights enforcement in the post-phase. This particular point is investigated in the third chapter. However, Chapter 2 (the first substantive paper of this thesis) goes beyond that and emphasises the role that is played by access to markets (proximity to infrastructure such as roads and railways) for the success of smallholder farmers.

The study builds a new dataset for 1970-80s' Rhodesia (the territory which is Zimbabwe today) using historical satellite imagery and machine learning methods. At that time, the country had three spatially distinct property rights regimes, whose historical names are still commonly referred to: European Areas (EAs), where only settlers enjoyed full tenure and access to infrastructure; Tribal Trust Areas (TTAs), where customary property rights resembled those of precolonial times and where indigenous populations were geographically isolated from formal markets; and finally, Native Purchase Areas (NPAs), where indigenous populations were granted private property rights, but had poor access to markets and infrastructure. Comparisons between NPAs and TTAs illustrate the effects of a typical land tenure reform, as the main difference between these areas was having title or not. However, comparisons between NPAs

and EAs emphasise the role of physical and market infrastructure; private property rights are necessary to function in a market economy – geographic isolation from these markets leaves private property rights without effect. In their study on the impact of US railroads in 1890, Donaldson and Hornbeck (2016) reveal that without railroads farms would have incurred high and prohibitive costs (using wagons and waterways) to get their produce to the market. Hence, given the general problems experienced in relation to land reforms – see Bhattacharya (1995), Valdés and Foster (2010) and Stoeffler et al. (2016) – the thesis investigates whether modern states should consider promoting access to markets for farmers. Importantly, this chapter emphasises that colonial land policies that denied indigenous populations property rights and enabled extractive institutions had a negative effect on the extent of smallholder production. In addition, this chapter shows that property rights in isolation do not necessarily improve this situation; rather, market access and private property rights are complementary in promoting the success of smallholder farming.

1.6.2 Effects of poor land rights enforcement post-redistribution, on cultivation and crop quality: Satellite data evidence from Zimbabwe

Agriculture has overarching importance for households in SSA and the role of property rights (land titles) supports its role in poverty alleviation. Davis et al. (2017) analysed household surveys for 22 African countries and found that in areas where the agro-climatic profile suited farming, agriculture was the preferred occupation for households. Poverty reduction and achieving pro-poor growth could start with agriculture. However, land access in SSA is, in many instances, still limited to customary tenure. In impoverished, agriculture dependent countries, reviewing and promoting access to land for the poor can be an important poverty alleviation strategy (Bezemer & Headey, 2008; Eastwood & Lipton, 2000; López & Valdés, 2000). Using representative data at the national level from 1990-2000 survey data for Ethiopia, Kenya, Rwanda, Mozambique and Zambia, Jayne et al. (2003) found that around 25% of households that relied on agriculture did not have access to land. Otsuka (2007) argues that landless people do not have incentives to work hard or invest in the land, which negatively affects economic growth. Reyes (2002) found that land restitution positively affects poverty. Yet Sikor and Müller (2009) and Besley and Burgess (2000) argue that state-led land reforms are limited in success, because bureaucratic structures are top-down in nature.

Chapter 2 shows that indigenous Zimbabweans in the NPAs had a higher proportion of land under crop cultivation due to security of tenure (in the form of land titles) as long as they had

access to markets (Chingozha & von Fintel, 2019). However, historical land imbalances continued post-independence, with the majority of rural dwellers working on the land in customary TTA areas that are distant from main roads, railways and major cities. In a market economy such as Zimbabwe, where property rights are required to effectively participate in the market, white farmers were at an advantage relative to their black counterparts since the former had land titles and the latter did not. In 2000, Zimbabwe embarked on the land restitution programme FTLRP to address these imbalances and transferred around 80% of European held land to indigenous Zimbabweans (Scoones et al., 2011). However, FTLRP did not address two key constraints in that the newly settled farmers did not enjoy the same property rights as Europeans who had formerly cultivated the lands; and the TTAs remained isolated from markets and also did not have secure private property rights.

In investigating the effect of poor property rights enforcement on crop cultivation and quality post-land reform, we employed data generated by applying a Support Vector Machine (SVM) learning algorithm to Landsat satellite imagery covering the entire Zimbabwe during the period 1997–2003. Identification relies on overlaying the Zimbabwe Statistical Agency (Zimstat) 2012 Census shape file on Rhodesia's Land Apportionment Act (1930) map. Thus, we were able to compare the proportion of land under crop cultivation in areas affected by land redistribution; European Areas (EAs) – treated regions – to areas not affected, Tribal Trust Areas (TTAs) and Native Purchase Areas (NPAs) – the control regions. Night Lights Data (NLD) and Normalised Difference Vegetation Index (NDVI) were also used as proxies for welfare and crop quality respectively. This chapter makes an important contribution because it is the first study that quantifies the effect of poor property rights enforcement post-FTLRP implementation using objective satellite data as well as robust econometric identification techniques for the whole country. Prior studies were qualitative and geographically limited, only focusing on particular small areas within the countries and hence not considering the nationwide view. Their conclusions are mixed and do not provide definitive answers for the macroeconomic impact of the reforms. The Difference-in-Difference and Regression Discontinuity (using distance to the border of region land class boundaries as per the 1930 Land Apportionment Act) estimates show that FTLRP negatively affected both crop cultivation and quality. This study emphasises the importance of the enforcement of property rights post-land reform to ensure effective participation on the market for the land beneficiaries.

1.6.3 A citizen science approach to classifying urban informality and other urban land use types using satellite imagery

Drawing on Bhattacharya (1995), the thesis argues that poorly implemented land reform (in which property rights are not asserted in the wake of the reform) can result in industrial contraction. In Zimbabwe, this situation partly influenced the growth of the informal sector – economic contraction had already started with the Structural Adjustment Programs (SAPs) of the early 1990s (Potts, 2006; Ncube, 2000). By the year 2000, an estimated 1.7 million Zimbabweans were working in the informal sector (Kumbawa, 2002). This does not imply that the informal sector is unimportant for the economy, but that its emergence led to new dimensions of activity and the concomitant need for new institutions that could facilitate market interactions. In explaining why the dual sector models fail to adequately explain development, Bhattacharya (1995) indicates that it is an oversimplification to assume that ancient societies only produced food. Rather, they were also engaged in the production of other goods and services which can be loosely referred to as ‘informal economy’ in present day terminology. In the context of developing nations, people in both urban and rural areas find themselves making different artefacts and services for their own and commercial purposes due to inherent market failures. Therefore, agricultural contraction (as the case appears to be in Zimbabwe) and growth in informality are inseparable from each other in agro-based economies. Becker (2004) associates informality with poverty in modern settings.

While the informal sector has an important economic role to play in SSA, unplanned residential and business premises expansion may put urban services under pressure. Additionally, the unplanned and unregulated nature of informal business establishments and residential areas is fraught with high opportunity costs given the absence of a formalised tenure regime. There is therefore a need for authorities to be able to map these informal areas and develop cadastral databases as a way towards formalising urban services planning management as well as enforcing property rights. Chapter 4 investigates the conditions under which citizen science can be used to correctly detect or classify (or detect changes in) the informal sector and other land use types from satellite imagery in urban areas.

The study is placed in the context of Operation Restore Order (ORO), a 2005 clean-up operation that affected close to a million people in Zimbabwe’s urban centres. The study employed the assistance of 41 Stellenbosch University (SU) students to classify 180 images for areas affected and not affected by ORO before the operation (2004) and afterwards (2006)

in order to investigate whether citizen science works best where there have been changes in informality and other land use types or not. The paper also investigates whether or not more classifications per image, learning effects and demographic attributes affect classification accuracy. The results shown in Paper 3 may assist urban councils in SSA to cost effectively develop land use maps and cadastres that can be used in planning and urban tenure enforcement as a way of reducing poverty and inequality.

1.7 An “African statistical tragedy”

Each of the Chapters outlined above also make a data contribution. Some researchers have been cynical about Africa’s positive growth in recent decades, including Rodrik (2016), McMillan et al. (2014) and Jerven (2010). The latter used Botswana as a case study and found that there has been upward bias in national income reporting for some years, while Harttgen, Klasen and Vollmer (2013) found no evidence to that effect. To highlight the problem put forward by Jerven (2010), Devarajan (2013) notes that statistical reporting in Africa is adversely affected by old standards, absence of capacity and lack of political will to promote the release of objective statistics (referring to this phenomenon as ‘Africa’s statistical tragedy’). Although not a paper on its own, this is a key aspect that cuts across the entire thesis.

To investigate the effect of market access and land titles on crop cultivation (Paper 1) and to investigate how the absence of property rights restoration (enforcement) post-land reform in Zimbabwe affects the proportion of land under cultivation (Paper 2), there was a need for granular spatial (at least at ward level) agricultural data for the entire country and with the right temporal frequency. Besley and Burgess (2000), note that empirical investigations on the effects of land reform are confronted by the need to have pre- and post-intervention data points; this poses a challenge, especially in developing countries. Census, household and labour force survey microdata is not publicly available in Zimbabwe. Therefore, this study relied on the application of a Support Vector Machine (SVM) learning algorithm on remotely sensed satellite imagery to generate crop cultivation data that was used as an endogenous variable. Owing to the unique, intricate rural/urban linkages in SSA, the outcome of which is a large informal sector, there needs to be sufficient attention paid to measuring this sector. Hence, Paper 3 investigates the conditions under which high quality urban land cover or cadastral maps can be generated through citizen science. Addressing these data shortcomings and creating (or contributing towards creating as in Paper 3) novel datasets are important contributions of this research.

1.8 Summary

The thesis contributes to the discussion around issues of land access, land rights, land reform and market access in Africa. These issues are juxtaposed against the background of a data scarce context, in which machine learning (ML) and citizen science (CS) are brought to the fore. Grindle (2004) describes developing countries as a world in which “all good things cannot be pursued at once”. Hence, it is important to consider both positive and negative effects of interventions such as agrarian reform; in particular, the thesis uncovers some of the related preconditions for land and tenure reform, namely market access, property rights enforcement, full tenure and the institutional capacity for managing land rights. The thesis presents several articles from which important insights can assist policymakers to choose between different priorities as they make critical trade-offs.

CHAPTER 2

PROPERTY RIGHTS, MARKET ACCESS AND CROP CULTIVATION IN RHODESIA: EVIDENCE FROM HISTORICAL SATELLITE DATA⁴

2.1 Introduction

From the mid-1970s, there was a general belief that land titling would lead to agricultural growth in developing countries. However, researchers found very little evidence to support this view (Udry, 2011). There have been various modern-day attempts to transition from traditional communal tenure systems to land titling in sub-Saharan Africa (SSA). The goal has been to promote access to land, but also to increase bankability of cultivated land that is not individually owned. Despite a lack of consensus, a section of the literature continues to argue that individual property rights can improve household agricultural production in some circumstances (Deininger & Jin, 2006; Goldstein & Udry, 2008; Newman et al. 2015). Several proposed mechanisms include incentives to invest, less time spent in conflicts over land with multiple claims of ownership, better access to credit, reduction in land transfer transaction costs through proper registration, as well as incentives to innovate (Muchomba, 2017; Fenske 2011). Some empirical studies support these arguments, showing that titling enhances farm yields (Abdulai et al., 2011; Newman et al., 2015) and leads to more intensive cultivation of available land (Do and Iyer, 2008) in certain circumstances. In addition, secure property rights have reduced conflict over land, as well as transactions costs⁵ in the land market (Aryeetey & Udry, 2010; Deininger et al., 2008; Deininger, Jin, Xia & Huang, 2014).

However, the lack of consensus on the role of property rights in agricultural growth prompts further investigation. Is the transition only successful when other complementary conditions are also in place? For instance, the effect of property rights on agricultural production is stronger in areas with better infrastructure and market access (Markussen, 2008). A separate literature documents the role of infrastructure and market access in the growth of rural economies, without taking land tenure into account. For instance, historical investments in

⁴ This chapter is published in the peer reviewed Economic History of Developing Regions (EHDR), Volume 34(2) of 2019, 132-155.

⁵ With land titles, it is much easier for land ownership to change from one person to another.

railways positively influence economic activity (including agricultural production) and settlement patterns (Herranz-Loncán & Fourie, 2017; Jedwab & Moradi, 2016). In Ghana, colonial railway construction encouraged the cultivation of cocoa due to reductions in transportation costs (Jedwab & Moradi, 2016). Jedwab, Kerby and Moradi (2017) illustrate the path dependence and persistence of city location in Kenya, even after the decline of colonial railroads. Historical infrastructure investments therefore determine the spatial distribution of current economic activity directly (by providing means to transport goods to market) and indirectly (by creating markets which agglomerated and remain in transport nodes, even in the cases where infrastructure became outdated). Contemporary studies find that enhancing roads and transportation infrastructure is key in improving market access, especially for smallholder farmers (Jordaan et al., 2014; Masuku et al., 2001; Ahmed et al., 2016; Senyolo et al. 2009; Boughton et al., 2007; and Poulton et al., 2006). If better infrastructure reduces transactions costs, farmers tend to sell their produce in more profitable, distant markets than at the farm gate (Abu, Issahaku & Nkegbe, 2016). Hence, distance to market proxies for transport costs and ultimately the decision of small farmers to embark on profitable transactions in central markets.

This study draws these two strands of literature together in a historical context. Using data from Rhodesia, we argue that land titling is only effective if property owners are integrated into markets in which they can sell their goods. Reforming land rights unconditionally may therefore have little impact on small producers' livelihoods. Multiple property rights regimes existed simultaneously in Rhodesia: our analysis compares groups that were farming in the same period and broader political context, rather than studying a transition from traditional rights across time; we therefore study the outcome of colonial policies once their implementation had been in place for a number of decades. Rhodesia's Land Apportionment Act of 1930 (LAA 1930) formalised racial land segregation in favour of European farmers. Under LAA 1930, the majority of Africans were forced to live in overcrowded Tribal Trust Areas (TTAs) under a traditional communal tenure system. However, the LAA also allowed some Africans to farm in small parts of the country demarcated as Native Purchase Areas (NPAs), where formal individual land rights were granted to the African population; this exception was designed to ease potential tensions between Africans and Europeans. Inhabitants in designated European Areas (EAs) and NPAs were the only groups that were granted land titles. While farmers in TTAs had relatively weak property rights, those in NPAs also did not enjoy the advantages of market proximity and high quality land that was found in EAs. Because of geographic and racial distinctions in property rights, the LAA 1930 is a central colonial

policy that subsequently determined agricultural fortune in Rhodesia. Various other confounding factors (such as land quality and climatic conditions) are discussed in the data section and, where possible, controlled in the analysis.

The majority of papers that have investigated the effects of land titles have relied on household survey datasets and therefore consider modern transitions to more secure property rights (Abdulai et al., 2011; Fenske, 2011; Goldstein & Udry, 2008; Markussen, 2008; Ravallion & Van De Walle, 2006). These data are largely unavailable for many countries and in particular in early periods. We therefore create a novel dataset by classifying Landsat satellite imagery into areas that were cultivated and those that were not using a Support Vector Machine (SVM) algorithm for the period 1972-1984. This period corresponds to the time that substantial emphasis on a nexus between land titles and productivity emerged (Udry, 2011). It also represents a time long after colonial policies were enforced: we therefore do not rely on measuring short-run effects of changes in property rights, but analyse long-established differences resulting from discriminatory land allocation. This data is used to test how differential land rights affected farmers' cultivation decisions in the various regions. We compare the TTAs, NPAs and EAs to assess whether land rights in and of themselves provided incentive for production, or whether NPAs faced other obstacles that EAs did not. We explore whether NPA farmers' property rights enabled similar cultivation intensity as for Europeans, taking into account their relatively poor access to markets. Alternatively, we assess whether isolation from markets meant that NPA farmers faced similar production incentives as those in TTAs.

The rest of the paper is structured as follows: Section 2 presents an overview of Rhodesia, from colonialization and how that culminated in the segregated apportionment of land. A discussion of African agriculture production (in general and in the 1970s) is also presented in this section. Section 3 discusses the data, analysis procedure and results. Lastly, Section 4 presents the conclusions.

2.2 Rhodesia Overview

2.2.1 The role of land in Rhodesia

After the Berlin Conference of 1884-1885 (which serves as a historical marker for the so-called "Scramble for Africa"), the British Empire lacked the finances to seek more territory in Africa. However, Cecil Rhodes' (the prime minister of the Cape Colony) British South African

Company (BSAC) had the necessary private financial and military resources to further British interests. The company expanded into current-day Zimbabwe in 1890, prompted by prospects of large gold deposits (Andersson & Green, 2016; Green, 2016; Palmer, 1971; Rifkind, 1969). However, settlers soon realised that estimates of potential mineral resources were inflated. Over-optimistic views regarding the prospects for mining persisted for one and a half decades. In the meantime, the BSAC invested in other infrastructure (such as railways) and the land itself to recover the costs of moving into the region (Arrighi, 1967). The company encouraged the development of a white agricultural class, in order to increase the value of land, railways, mines and other assets that they had established in the region. The pioneers therefore shifted their attention away from mining to agricultural land (Andersson & Green, 2016; Duggan, 1980; Frankema, Green & Hillbom, 2016; Moyana, 1975; Pollak, 1975; Weinrich, 1979; Arrighi, 1967).

The BSAC was granted the Royal Charter to administer the colony until 1915, although this would later be extended by another ten years (Rifkind, 1969). All land became the property of the Crown and the BSAC acquired large tracts of land to sell to European immigrants (Floyd, 1962; Pollak, 1975). Pollak (1975) claims that the company acquired land by “treaty, occupation and conquest”. After the so-called “war of dispossession”, the British Government gave an Order-in-Council that appointed a Land Commission to deal with disputed territory in Matabeleland (Floyd, 1962; Moyana, 1975; Stocking, 1978). This Commission recommended the creation of Gwai and Shangani for resettlement of the defeated Ndebele (Floyd, 1962; Moyana, 1975; Palmer, 1971). These two reserves were only a quarter of the size of the previous Ndebele kingdom and the soil was barren and dry as observed by the British Deputy Commissioner in 1897 (Moyana, 1975).

Despite the land imbalances that arose after the BSAC conquest, Africans were able to purchase land in the colony. An Order-in-Council of 1898 ruled that an African could purchase, hold and sell land under the same conditions as those of the non-native (Jennings & Huggins, 1935; Pollak, 1975). However, few Africans could afford to buy land at the time (Floyd, 1962; Jennings & Huggins, 1935). By 1921, European settlers had acquired 31 million acres of land versus the 40 000–47 000 bought by African farmers (Arrighi, 1967; Moyana, 1975; Pollak, 1975).

Motivated by the need to deter competition, the colonial administration did little to promote commercial African agriculture (Machingaidze, 1991). Continued displacement from their land

aggrieved the native population (Duggan, 1980; Herbst, 1991; Moyana, 1975; Pollak, 1975). In 1925, the Morris Carter Commission was formed to bring the issue of the land to finality and to consolidate the advantage that Europeans held over the African population.

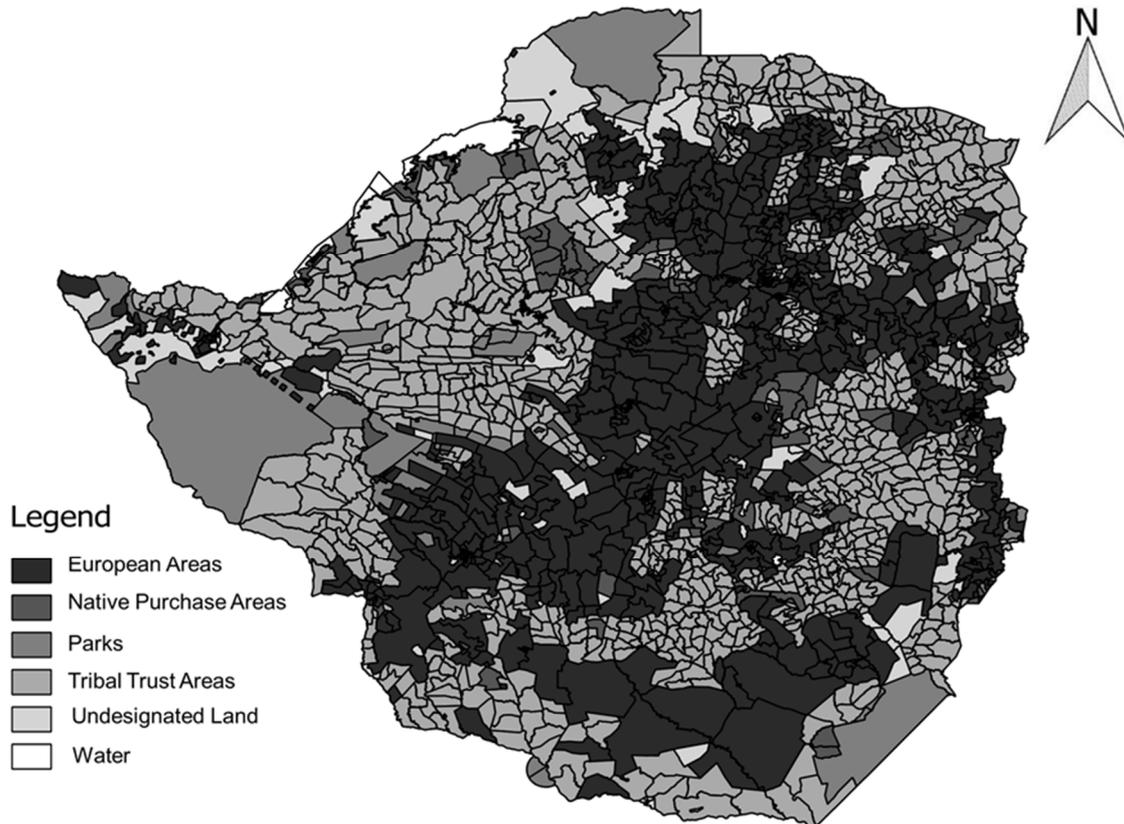


Figure 2.1 The Land Apportionment Act Land Distribution

Source: Ward level image showing the different land use classes. It is based on Federal Department of Trigonometrical and Topographical Surveys (1963). Available online at: <https://digitalcollections.lib.uct.ac.za/collection/islandora-25210> and shown in Figure A.2 in the Appendix A

The commission recommended that racial tension in the colony was a product of the contact between the European and African races. Therefore, the proposed solution was complete separation (Floyd, 1962; Jennings & Huggins, 1935; Pollak, 1975). The recommendations of the Morris Carter Commission were promulgated as the Land Apportionment Act (LAA) in 1930 (Floyd, 1962; Pollak, 1975). The LAA is one of several pieces of legislation that the colonial government had enacted to protect their interests and to formalise racial segregation (Ncube, 2000; Ranga, 2004; Jennings & Huggins, 1935). By 1963 the native African

population (of more than 2.5 million) was concentrated within 50 000 square miles of land, while about a fifth of a million settlers owned the other 75 000 square miles.

At the same time, Rhodesia pursued limited equality to align politically with the United Kingdom (UK). A compromise was achieved in setting up the so-called Native Purchase Areas (NPAs), where Africans could continue to buy land (Andersson & Green, 2016). The colonial administration believed that very few Africans would have the financial resources to acquire land and therefore limited the geographic extent of the NPAs. Africans' rights to the land were thus effectively suspended, except in the 81 NPAs (Floyd, 1962; Jennings & Huggins, 1935). Other Africans who could not afford to buy land in the NPAs were forcibly moved to Tribal Trust Areas (TTA) where communal land tenure existed. At inception of the LAA, around half of the land area was designated for whites, while only 30% had been set aside for Africans (Herbst, 1991). Figure 1 presents the land distribution that was formalised by the Land Apportionment Act.

2.2.2 Agriculture and the economy in Rhodesia

Rhodesia became an agricultural and economic leader in Africa. In the 1950s, the country was the main producer of tobacco in Africa (Haviland, 1953) and it was the most prosperous state in British Africa (Andersson & Green, 2016). Good (1976) argues that Rhodesia had enormous advantages because of its relatively rich endowment of natural resources such as iron, steel, asbestos and various crops. By 1979 – 1980 the territory became a net agricultural exporter; this was a rare achievement for an African country (Munslow, 1985).

State intervention by the colonial administrations of South Africa and Rhodesia was focused on improving the conditions of white farmers. They were primarily aimed at achieving maize price stabilisation in relation to international prices (Jayne & Jones, 1997). On the other hand, Machingaidze's (1991) view is that "it was not state policy to encourage Africans to produce for market"⁶. Transport infrastructure was also planned to achieve these ends. Roads and rail lines were constructed to connect European farming areas to markets and urban centres (Austin, 1975). The main railway route traversed the European agricultural highveld, although this was supposedly "sheer coincidence" (Punt, 1979). TTAs and NPAs did not receive the same

⁶ Jayne and Jones (1997) state that state intervention sought to "prevent African farmers from eroding the viability of the less efficient European producers". It is acknowledged that Africans had traditionally been successful agriculturalists (Phimister, 1974). For example, Phimister (1988) [in Andersson and Green (2016)] observe that the native commissioner of Chilimanzi (now Chirumanzu) wrote that by 1904 the Africans produced 90% of the country's crop production available for market.

investments, which exacerbated the effects of the weaker property rights that applied in those regions. In the US, Atack et al., (2014) indicates that railways had an important role in development, and that they established “commercial centers” and made pristine areas along the rail infrastructure more probable candidates for settlement. In the case of the US, Fishlow (1965) raises the issue of the exogeneity of railroads and questions whether the advent of railroads put in place the conditions for development or railroads were actually spurred by profitable circumstances. An important question that arises then in this chapter is whether railways really followed the EAs by coincidence. In Atack et al., (2014), they indicate that there was a high probability that counties that already had a bank would see the railroad coming through them in the next decade, while areas with railroads usually got a bank with two years of the railroad construction. This means that in the case of Rhodesia there might have been such bi-directional causality between EAs attracting railways and other non-EAs areas achieving development as a result of railway establishment. Altering tribal land tenure rules had long-run negative consequences for African farmers (Hughes, 1971). In the TTAs, the Act enforced permanent land for grazing and farming. Farmers had to switch to continuous agriculture, resulting in soil erosion and reducing the fertility of their land (Arrighi, 1967; Duggan, 1980; Machingaidze, 1991). Moyana (1975) asserts that the distribution of land limited the extent to which Africans could actively participate in the economic development of the country. Many Africans and their families therefore settled on European farms, primarily in exchange for labour supply (Arrighi, 1967; Moyana, 1975; Youé, 2002).

Almost three-quarters of the land allocated to Africans was drought prone, dry and more suitable for only extensive livestock and crop production. In contrast, European Areas (EAs) enjoyed good rainfall that allowed intensive crop and animal production (Floyd, 1962; Herbst, 1991; Palmer, 1990; Pollak, 1975). Moyana (1975) observes that African areas were located in the dry lowveld whose soils did not have much potential; while Floyd (1962) emphasises broken terrain, dryness, lack of water and tsetse fly infestation in these areas. This is despite the fact that the majority of Africans⁷ relied directly on agriculture, in stark contrast to only 30% of Europeans. Settlers held almost all areas that were suitable for dairy farming (Machingaidze, 1991) and more than three-quarters of land that was suitable for intensive farming (Clarke, 1975). Phimister (1974) notes that African agricultural returns were too small

⁷ Clarke (1975) puts 60-70% as the figure for Africans living and depending on rural land.

to re-invest; they could not undersell the European farmers because in many cases the latter sold produce on behalf of the former.

Farmers in NPAs faced similar disadvantages to those in TTAs; they too were isolated from transport and communication infrastructure, lacked water and the soils were largely infertile (Pollak, 1975). Only 25% of African farmers were located within 25 miles of a rail line, while this was the case for 75% of their European counterparts (Ndelela & Robinson, 2007). African market access was deliberately constrained (Abu et al., 2016; Ahmed et al., 2016). Usually [for example between 1860 and 1912 in Europe as discussed by Bogart (2009)] the development of rail systems followed high demand in densely populated areas. Infrastructure development in Rhodesia therefore limited market access of African farmers, despite the large market potential in the areas where they were located.

However, Andersson & Green (2016) posit that development paths pursued by Africa's former colonial governments had unintended consequences. Although the policies that were implemented by settler governments were meant to promote European farming, some sections of African society experienced positive spill over effects (Andersson & Green, 2016). For instance, NPA farmers had more secure property rights and cultivated more land compared to those in the TTAs. Farmers in these areas therefore constituted an African "middle class". This study adds empirical support to these claims.

Nevertheless, the hypothesised success of agriculture in the NPAs relative to the TTAs may remain disputed based on a number of obstacles that property rights would not have been able to resolve. The NPAs did not only lack good soils, market access and infrastructure, they also did not have the effective organisation of the Rhodesian National Farmers Union (RNFU). This organisation was established in 1949 by European farmers (Herbst, 1991). The RNFU had substantial influence on the colonial government, since the latter relied on the former's increased crop production to meet the requirements resulting from WWII (Herbst, 1991). Haviland (1953) offers that after WWII, African evictions from European designated areas cleared land for allocation to ex-service men. The RNFU was able to pressure the administration to pass the Farmers Licensing Act, which compelled farmers to purchase licences from the union, effectively shutting out African farmers from state sponsored research and other benefits of union lobbying (Herbst, 1991). Furthermore, farmers in the NPAs were excluded from access to finance. For example, the Land and Agriculture Bank that had been formed in 1924 served "persons of white descent only" (Machingaidze, 1991).

2.2.3 African agriculture in the 1970s and early 1980s

The proportion of land under cultivation fluctuated significantly over time in the NPAs, TTAs and EAs (see Figure 2). In Rhodesia, land policies were a major reason for the decline in yields of African agriculture (Arrighi, 1970; Binswanger, Deininger & Feder, 1995), but the fluctuations in the intervening periods for both African and European areas may be explained by production disturbances owing to the nationalist war of independence. The nationalists were a group of independence activists and later political leaders who organised and led the bush war against white minority rule in Rhodesia. Weinrich (1979) points out that after 1976, a large number of white men exited the economy to join the war. On the back of that, government expenditure focused mostly on military and defence spending, while emigration rates increased to 1 000 per month (Weinrich, 1979). The colonial administration forced thousands of Africans into 'protected villages' in order to cut off their contact with the nationalist fighters (Duggan, 1980). By 1977, more than 1 million Africans had been resettled, disturbing existing production (Weinrich, 1979).

These factors potentially explain agricultural production fluctuations. Apart from that, Sithole (1972) and Whillow (1980) argue that dense settlement in the TTAs impeded agricultural growth. African landholders were customarily obliged to accommodate many of their kinsmen (such as a deceased brother's wife or divorced sister and their children) on their plots (Chavunduka, 1975). This custom intensified existing overcrowding. Riddell (1978) argues that the TTAs exceeded their carrying capacity by one-fifth. One method of correcting overuse of the land and other resources is husbanding (Demsetz, 1974), hence the Native Land Husbandry Act that was introduced in Rhodesia 1951. However, conditions in NPA areas appeared to have been less severe. By 1982, African farmers who held title contributed 4% to the country's agricultural output (Munslow, 1985). For the 1975-1979 period, Zinyama (1986) puts this contribution at between 2 and 3 per cent.

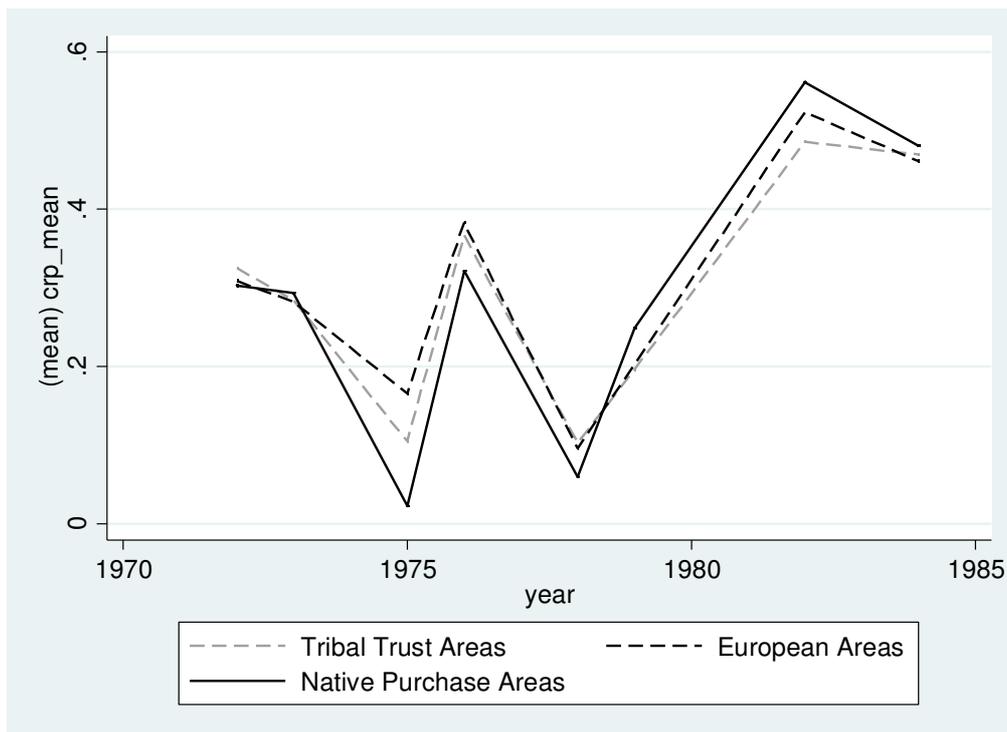


Figure 2.2 Ratio of cropland to total ward area (1972-1984)

Source: Own Illustration-using data generated from Landsat imagery through machine learning

2.3 Data and Analysis Approach

2.3.1 Classification of images using machine learning

Very little quantitative data is available to represent the various geographic regions defined by the 1930 LAA. Extricating the agricultural contribution of NPAs from that of the TTAs and the other regions has therefore traditionally been difficult. We therefore leverage historical satellite data to monitor land cover in colonial Rhodesia. The closest classification of land in the early colonial period is the work by Robbins (1934) in neighbouring Northern Rhodesia. Robbins (1934) used aerial photographs taken at the altitude of 10 000 feet by the Aircraft Operating Co. of South Africa. The main objective of the exercise was to assist in quicker area mapping; this study takes forward that work and classifies cropland in the NPAs, TTAs and EAs using machine learning and images captured in the colonial era by the Landsat Multispectral Scanner (MSS) sensor (see Appendix B for details on the data source). We first analyse data from 1972 to 1979, which excludes the confounding influence of the transition

from minority rule in 1980. However, our analysis is then extended up to 1984 to assess whether the relationships we estimate change beyond that time.

When electromagnetic energy from the sun hits an object on the surface of the earth, some of that energy is reflected back to satellites. Red (R), Blue (B) and Green (G) bands are the most common and are visible to the human eye. Of particular importance is the invisible near-infrared band: crops are most likely to reflect this band back to satellites than any other objects on the earth's surface. Crop classification takes advantage of this property, together with the fact that objects on the earth's surface have different spectral signatures of electromagnetic energy reflection (Chamunorwa, 2010; Eastman, 2003).

We employ the Support Vector Machine (SVM) algorithm [developed by Cortes & Vapnik (1995)] to classify images from the Landsat 1–5 MSS (see Appendix B). In a supervised classification, we train the samples used to distinguish between cropland and other forms of land cover (mainly consisting of natural forest) in Quantum Geographic Information System's (QGIS) Semi-Automatic Classification Plugin (SCP). Final image classification in R produces a raster image with pixels 1 and 0 denoting cropland and all other types of land cover respectively. Our unit of analysis is current-day Zimbabwean wards (classified into NPAs, TTAs and EAs). The predicted binary images are aggregated to these geographic demarcations; our dependent variable is therefore the proportion of the ward area covered by crops. While this indicator does not measure the output obtained from land, it does give an indication of the extent of land under production. In Appendix C, the data created through machine learning are compared with aggregated productivity time series data from Jayne & Jones (1997) and Thirtle et al. (1993). The agricultural data created in this study correlates highly⁸ over time with data from the other sources. We are therefore confident in the quality and reliability of the data. In addition, Appendix C discusses statistical metrics that establish that the classification was reliably conducted. For the first time, we have disaggregated indicators of crop production for small geographic areas in Rhodesia's history. Furthermore, while greater areas of crop production do not necessarily entail superior yields, the strong correlations between these quantities over time do suggest that studying only the former does provide valuable insights into variations in agricultural success of various regions. While we therefore cannot conclude

⁸ The crop coverage for TTAs, NPAs and EAs are the preferred variables used in the analysis. The highest correlation is between TTA crop hectareage and communal area output at 0.9994.

whether property rights raise incomes from yields, we can rely on the extent of land covered by crops as an indicator of welfare.

2.3.2 Approach

In investigating how holding land titles in Rhodesia affected agricultural production, we compare the area under cultivation of farmers in the TTAs to those in the NPAs. The main channel that we consider is differential access to markets [assuming that farmers in NPA areas unintentionally benefitted from better access to markets, as suggested by Andersson & Green (2016)], although we include many other control variables in our models to account for potential selection effects⁹. Most variables such as soil quality, lack of access to credit and research are similar (at the region level) across these region types. More pertinently, we introduce an unconventional dataset to test how property rights and differential market access (as measured by distance to rail, road and urban centres) affected agricultural production.

2.3.3 Model specification

Conventional economic wisdom from Krugman (1995)'s general equilibrium (GE) trade theory tells us that trade intensity between two neighbouring regions is inversely proportional to the Euclidian distance between them (Chaney, 2018). Following Donaldson and Hornbeck (2016), we rely on a reduced form model of the GE trade theory and estimate how enhanced market access increased the proportion of land under cultivation in a ward. The endogenous variable is the proportion of land under cultivation in a given region. Our main specifications are estimated using Ordinary Least Squares (OLS) that account for spatial dependence in the error terms (Conley, 1999). However, following Papke & Wooldridge (1996), we also adopt a fractional probit model because the dependent variable ranges between 0 and 1 and the linearity assumption of Ordinary Least Squares may not be suitable (Schwiebert, 2018; Gallani & Krishnan, 2017)¹⁰. Our model is expressed as:

$$Crop\ Share_{wft} = \alpha_0 + \alpha_1 Land\ Type_{wt} + \alpha_2 X_{wt} + C_f + \lambda_t + \mu_{wft} \quad \dots (2.1)$$

where w and t index geographic wards and time respectively; $LandType$ represents a set of dummy variables to distinguish NPAs from TTAs and European Areas; X_{wt} is a vector of

⁹ Differences in crop cultivation between the different land regimes may also be due area specific differences such as rainfall and soil quality. We introduced these covariates to deal with some potential bias.

¹⁰ For the sake of robustness, we also estimate the models with a Tobit specification, acknowledging that the dependent variable is censored from the bottom at 0 and from the top at 1.

explanatory variables (namely regional population, precipitation, temperature, distance to main road, distance to secondary road, distance to rail station or siding, distance to any road and a caloric suitability index). C_f captures agro-ecological and frame¹¹ fixed effects, while λ_t incorporates year¹² and season¹³ fixed effects. Five agro-ecological zone fixed effects, premised on FAO (2016), account for the fact that various parts of the country are suitable for specific crops and farming systems, which may naturally lead to higher crop cover. Frame fixed effects account for the fact that data is downloaded in 22 path/row frames from the US Geological Survey (USGS), each with its own measurement anomalies (see Appendix B, Figure B.1). Landsat 1-5 represent some of the first generation satellites deployed by the US National Aeronautics and Space Administration (NASA); later versions introduced improvements that enabled more reliable data analysis. Earlier satellites only recorded a handful of images per year over the chosen area, many of which are unusable due to cloud cover. Thus, year on year it is impossible to download imagery for the same month within the same season. Month and year fixed effects attempt to filter out any confounding effects of seasonality¹⁴.

2.3.4 Explanatory variables

We digitise the Land Apportionment map in Figure 1 to allocate current-day Zimbabwean wards to the various land classifications¹⁵. We furthermore use historical shape files to calculate distances to main roads, secondary roads, any roads, rail stations and sidings and main cities, as shown in Figure 2.3 and Figure A.3. Distance is calculated in Quantum GIS (QGIS)

¹¹ As shown in Appendix B (Figure B.1), Landsat images are available as frames or tiles. To obtain an image mosaic for the whole country, 22 tiles (that represent areas with different geographies and potentially differential data quality) are pieced together. Frame Fixed Effects (FEs) are introduced into the specification to absorb any systematic characteristics related to these images.

¹² The period 1972 – 1979 is rather eventful and the war of independence pitting the African nationalist fighters and the Rhodesian army is the major highlight. We introduce year FEs in an attempt to account for these fluctuations.

¹³ Southern Rhodesia (Zimbabwe)'s rain fed agricultural season starts around late October and ends around April. Ideally, the machine-learning algorithm should be applied on images for this period. However, the MSS 1-5 represents first generation sensors and the image time frequency is low. For some frames and years, images are selected even if they fall outside the agricultural season in order to ensure full geographic coverage of the country. Even where images are available for the same month (within the farming season), some are unusable due to cloud cover and the only option is to select another image from another month, preferably within the farming season. The season, month and frame FEs attempt to correct these differences in image acquisition dates.

¹⁴ It is also for this reason that our inference does not account for serial autocorrelation in addition to spatial dependence.

¹⁵ As at 2012, there are 1980 wards in Zimbabwe. These demarcations were inherited from the colonial administration and they have not changed. The analysis however uses fewer than 1980 wards due to absence of cloud free images for some areas.

from the centre of each ward to the nearest road, railway or main city. These variables proxy for market access in each ward.

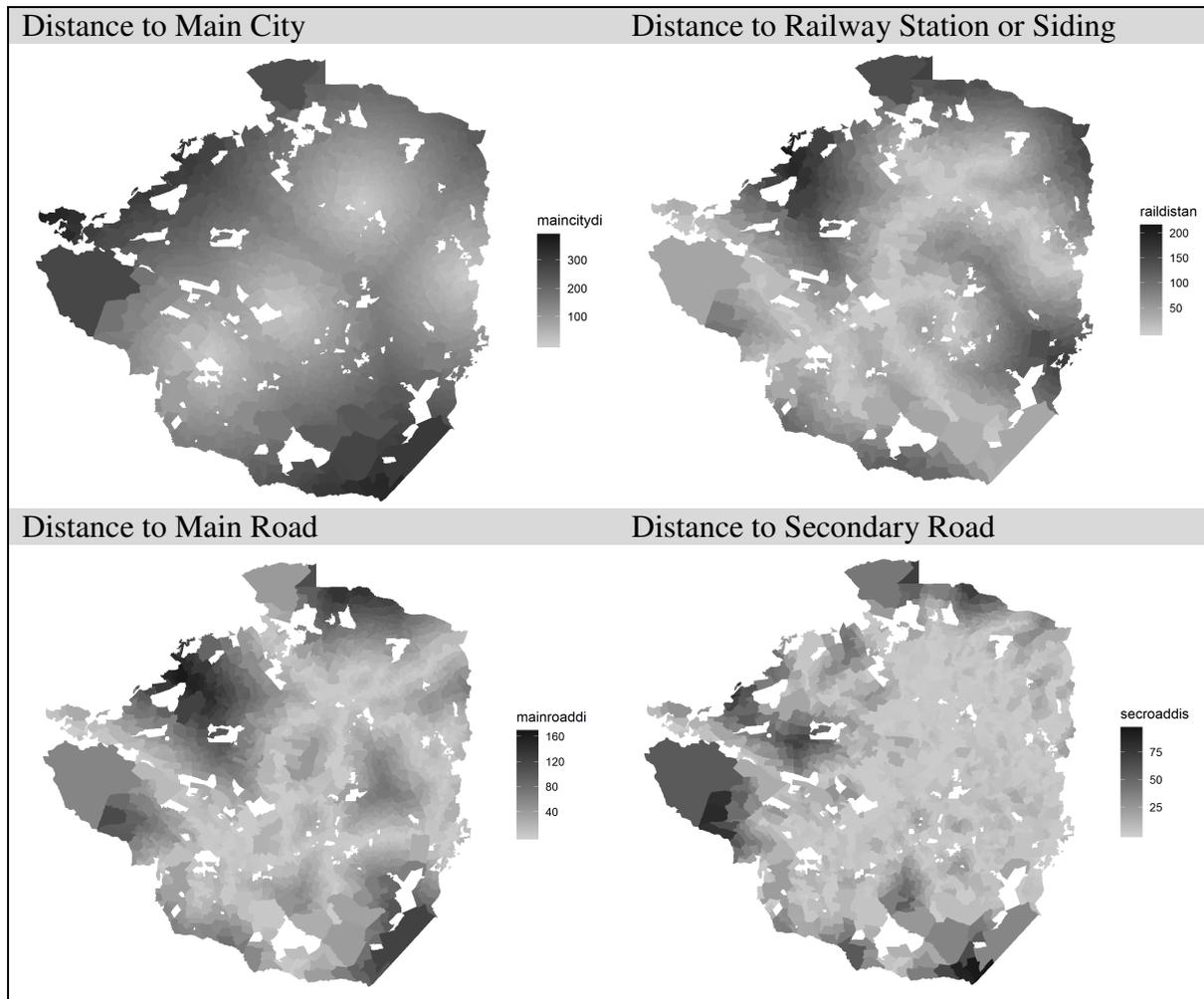


Figure 2.3 Distance variables

Source: Own illustrations

Note*: *maincitydi*, *raildistan*, *mainroaddi* and *secroaddi* are distance to main city, distance to rail station, distance to main road and distance to secondary road respectively. White values mean no data.

Control variables are introduced to account for selection of populations into various regions with varying conditions suitable to agriculture. Local population numbers are based on the Federal Department of Trigonometrical and Topographical Surveys, Rhodesia and Nyasaland (1962) map shown in Appendix A (Figure A.1). Each blue dot represents 1 000 African heads of household. Hence, we are able to compute a rough estimate of the working population in each ward in 1962. While the levels would have changed since that time and our period of analysis, we must assume (for lack of other data) that spatial variation has remained constant – in other words, we assume that migration resulting from conflict did not significantly change

the relative geographic concentrations of populations between 1962 and our period of analysis. This assumption may not be realistic. However, we take this approach because we do not have disaggregated population figures closer to our period of analysis. For the sake of robustness, we employ data aggregated by land class for 1970 and 1982 by Zinyama & Whitlow (1986). For 1970 and 1982, the data allocation to wards is based on the 1962 figures – the intermediary years are computed by relying on a geometric progression (see Table A.6 in Appendix A). The central findings of our models remain robust to using both imperfect approaches.

Even after controlling for population, a number of favourable characteristics may have driven selection into regions. We therefore construct further variables to account – as far as possible – for any bias in our models. Historical rainfall and temperature data are obtained from Willmott & Matsuura (2001), while a soil caloric suitability index is sourced from Galor, Özak & Sarid (2016), Galor & Özak (2014) and Galor et al. (2016). Precipitation, temperature and caloric suitability raster grids were processed in QGIS to match current-day Zimbabwean wards.

2.4 Descriptive Data Discussion

Table 2.1 Data Description

Class	Crop	Pop	Rain	Temp	CSI	Rail	City	Road	Area
TTAs	0.34	1885.42	725.80	20.65	1869.01	62.34	138.62	40.13	169.11
EAs	0.36	2322.18	759.30	19.61	1887.30	38.36	100.29	22.38	328.17
NPAs	0.35	1596.77	755.08	20.35	1951.63	62.26	142.25	40.86	193.58
F-Test	0.69	31.04	10.73	125.1	3.52	137.68	130.81	111.58	153.17
<i>p</i>	0.50	0.0***	0.0***	0.0***	0.03**	0.0***	0.0***	0.0***	0.0***

Notes*: Class = Land class, Crop = proportion of land under cultivation, Pop = a thousand African household heads, Rain = rainfall, Temp = temperature, CSI = Caloric Suitability Index, Rail = Distance to Rail, City = Distance to main city and Road=Distance to Main Road, Area = ward area in km²

Table 2.1 shows mean values of the ratio of land under cultivation as well as the main covariates by land class. The table shows that EAs have the highest proportion of land under cultivation (0.36), followed by NPAs (0.35) and lastly TTAs (0.34) although the differences are not statistically significant. TTAs were the least able to put land under cultivation due to competing land uses and less available land relative to EAs and NPAs. In terms of population, the NPAs were the least populated with about 1600 African heads of households per ward, while EAs

were the most populated with an average 2322 African heads of households. It is worth noting that TTA wards were generally much smaller in land area as compared to EAs. As shown in Table 2.1, TTAs has the highest rainfall and lowest temperatures. According to the representation in Table 2.1 NPAs had the highest mean caloric suitability index (CSI), while TTAs had the worst. For the transport infrastructure, EAs we located closest to railways, main cities and main roads with mean distances in kilometres of 38.36, 100.29 and 22.38 respectively. On average, NPAs were second to EAs in terms of proximity to rail infrastructure only, recording mean kilometre distances of 62.26, 142.25 and 40.86 to the nearest railway line, main city and main road respectively. TTAs were located nearer to main cities and main roads than NPAs, recording mean distances of 138.62 and 40.13 respectively. TTAs also had the smallest land area per ward, while the EAs has the most land area per ward. The average ward areas were 169.11, 193.58 and 328.17 for TTAs, NPAs and EAs respectively. Descriptive statistics therefore show that European Areas were ideally situated in relation to infrastructure, but did not have any natural agricultural advantages.

2.5 Results and Discussion

Before we present regression estimates, we descriptively assess whether farmers with titles in the NPAs cultivated more of their land than those without individual property rights in TTAs. We indirectly test the assertions of Anderson & Green (2016), who describe NPA farmers as a middle class relative to those in the TTAs; however, we cannot directly test whether more intensive production also improved yields and incomes, even if these are strongly correlated in aggregate over time. Figure 2.4 shows a choropleth map depicting the proportion of each ward area that was cultivated, for the pooled period 1972 – 1984. The central highlands of Zimbabwe, which were allocated to Europeans (EAs), had higher agricultural activity per land area than the TTAs and NPAs. It is more difficult to see differences in crop production across the latter two region types on the map. Our regression analysis will distinguish whether these effects are the result of land classifications only, or whether other contributing factors (such as suitable agricultural conditions) result in the distinction.

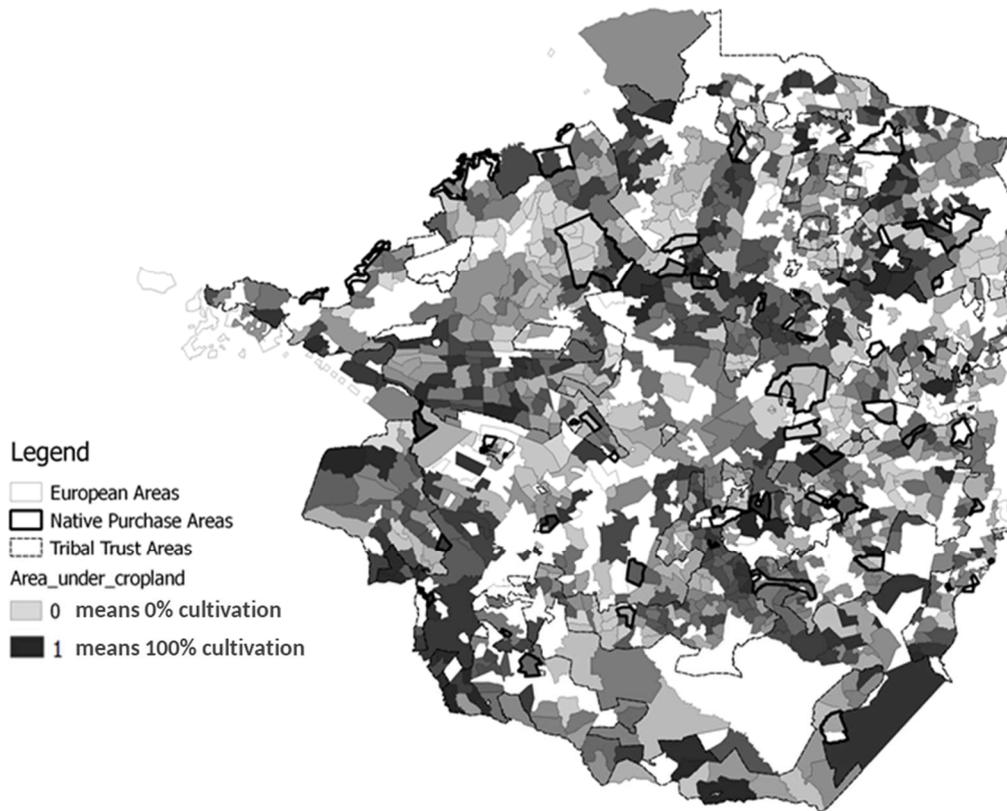


Figure 2.4 Proportion of land under crop cultivation (1972-1984)

Source: Own compilation

We estimate regressions to establish whether NPA farmers in Rhodesia had a higher proportion of land under cultivation as compared to their counterparts in the TTAs. The regressions also explore whether advantages in crop production were contingent on access to markets and infrastructure in the form of major cities, main roads and railways lines. Core results are presented in Table 2.2 using Ordinary Least Squares (OLS) models that adjust for spatial dependence following the procedure by Conley (1999) and implemented by Nunn and Wantchekon (2011). Tables A4.1 and A4.2 repeat these results using fractional probit and tobit specifications. Estimated marginal effects from the fractional probits (which are not shown) are very close to those from OLS estimates. Tobit models also yield similar results. We therefore limit our discussion to Table 2.2, which allows us to interpret coefficient magnitudes with ease and also allows us to assess the role of spatial dependence in inference.

Table 2.2 Class effects of colonial land policy (OLS estimates with SEs adjusted for spatial dependence)

	(1)	(2)	City (3)	Road (4)	Rail (5)	(6)	(7)	City (8)	Road (9)	Rail (10)
	<i>Period: 1972-1979</i>					<i>Period: 1980-1984</i>				
European Areas	0.026 (0.013)** [0.004]***	0.028 (0.013)** [0.004]***	-0.047 (0.092) [0.030]	0.026 (0.031) [0.007]***	0.068 (0.046) [0.025]***	0.052 (0.023)** [0.007]***	0.056 (0.023)** [0.006]***	0.065 (0.169) [0.052]	-0.044 (0.060) [0.017]***	0.032 (0.087) [0.041]
NPA	-0.019 (0.024) [0.005]***	-0.020 (0.023) [0.005]***	0.603 (0.291)** [0.109]***	0.389 (0.110)*** [0.020]***	0.568 (0.125)*** [0.059]***	0.040 (0.047) [0.013]***	0.038 (0.047) [0.012]***	0.766 (0.487) [0.413]***	0.280 (0.209) [0.055]***	0.651 (0.285)** [0.078]***
ln(dist city)			0.000 (0.015) [0.006]	0.005 (0.013) [0.004]	0.005 (0.013) [0.005]			0.019 (0.029) [0.020]***	0.011 (0.025) [0.007]*	0.016 (0.025) [0.007]**
ln(dist main road)			0.007 (0.006) [0.002]***	0.008 (0.007) [0.002]***	0.007 (0.006) [0.002]***			-0.018 (0.010)* [0.002]***	-0.028 (0.012)** [0.003]***	-0.017 (0.010) [0.002]***
ln(dist rail)			-0.002 (0.008) [0.004]	-0.001 (0.008) [0.004]	0.007 (0.010) [0.007]			0.025 (0.015)* [0.003]***	0.025 (0.015)* [0.003]***	0.025 (0.017) [0.007]***
EAs x Distance			0.018 (0.020) [0.007]**	0.003 (0.010) [0.003]	-0.009 (0.013) [0.007]			-0.001 (0.036) [0.011]	0.036 (0.019)* [0.005]***	0.009 (0.025) [0.011]
NPAs x Distance			-0.130 (0.060)** [0.022]***	-0.122 (0.032)*** [0.006]***	-0.154 (0.032)*** [0.016]***			-0.152 (0.102) [0.028]***	-0.068 (0.060) [0.014]***	-0.162 (0.075)** [0.019]***
Constant	0.346 (0.076)*** [0.021]***	0.400 (0.104)*** [0.025]***	0.395 (0.122)*** [0.030]***	0.360 (0.115)*** [0.027]***	0.338 (0.115)*** [0.031]***	0.164 (0.116) [0.031]***	0.420 (0.193)** [0.039]***	0.231 (0.232) [0.071]***	0.295 (0.219) [0.047]***	0.240 (0.221) [0.062]***
FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ward Controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
N	2108	2108	2108	2108	2108	696	696	696	696	696

NOTES*: Individual ward controls are population, rainfall and temperature and caloric suitability index. FEs denote region, frame, month and year fixed effects. White robust standard errors are reported in parentheses while Conley (1999) spatial dependence adjusted standard errors are shown in square brackets. Conley SEs are computed using distance cut-off of 100 km. Results remain robust at 16km (*see* Jeanty, 2012) and 50 km. Base category = Tribal Trust Areas (TTAs). Columns 3 and 8; 4 and 9; 5 and 10 presents results from the regression that interacts land class with distance to main city, distance to main road and distance to rail station or siding respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



Figure 2.5 Advantages of NPAs over TTAs in various periods and locations: results from rolling regressions (1972-1984)

Source: Own compilation. Results represent the NPA advantage in crop cover over TTAs at various distances from the relevant infrastructure. The 1972 regression only includes that year. Estimates for later years include all prior years. Dashed lines are 90% confidence intervals

Conley standard errors never reduce the statistical significance of our coefficient estimates, regardless of the distance threshold chosen. Our results are therefore robust to our mode of statistical inference. Specifications capture three different channels that affected agricultural production in Rhodesia, namely distance to main cities, distance to main roads and distance to rail station or siding.

We start with estimates that only include the period before 1980. Column (1) provides a base specification, which includes only relevant fixed effects to account for seasonal and frame-specific measurement errors¹⁶. Crop coverage was 2.6 percentage points higher in European Areas compared to TTAs. We measure no significant differences in production between NPAs and TTAs using White robust standard errors, but with Conley adjustments the effect is significant. On average (before taking into account access to markets) NPAs had significantly less proportion of land under cultivation in comparison to TTAs. Controlling for population, climatic variables and caloric suitability in Column (2) only marginally influences these results. Area under cultivation was therefore much greater in the less densely populated European regions. Our finding is therefore consistent with literature that suggests that African areas were overcrowded and experienced high competition for resources. Many Africans had few options but to find employment either in cities or European farming areas. However, differentials in property rights, farming conditions and settlement patterns do not fully explain this gap: we estimate a disadvantage for NPA farmers whose property rights were less well-defined than those in TTAs.

If climatic and soil advantages do not explain the greater extent of cultivation in European areas, then one important omitted factor is proximity to markets and infrastructure. We control for distance to a main road, to railways and to a main city. However, we posit that these market access variables may have heterogeneous effects across the various land classifications. Columns (3) to (5) explore this possibility with interaction terms. The main effects reveal that NPA farmers that (hypothetically) lived next to¹⁷ main cities, main roads and railways cultivated 60.3, 38.9 and 56.8 percentage points more of their total available land respectively, compared to similar farmers in TTAs. Significant negative coefficients on interactions reveal

¹⁶ Without fixed effects, we detect no differences in crop coverage between the various land classes. This, however, results because usable images were non-randomly distributed across various seasons in the three areas.

¹⁷ Main effects in Table 2.2 represent areas that live 1km from the various infrastructures, since $\ln(\text{distance}) = \ln(1) = 0$ and this eliminates the interaction term.

that the advantages of individual property rights decay the further farmers were located from these key infrastructures.

European Areas coefficients are positive and significant only in specifications (1), (2), (6) and (7), which shows EAs had a higher proportion of land under cultivation after only controlling for FEs and ward controls (but without distance) for the periods before and after 1980. In Columns (4) and (5), the European Areas coefficient is positive and statistically significant, implying that EAs close to roads and railways had more land under cultivation than farmers in the TTAs. In Column (9) the European Areas coefficient is statistically significant but negative, whereas the EAs x Distance effect is also significant but positive. Column (9) represents the period after 1980 (a change period) but this result may be driven more by lack of statistical power, since EAs close to infrastructure would be expected to do better. The EAs x Distance interaction effect is also significant and positive in Column (3). Specifications (6) and (7) show that in the period after 1980, NPAs had a higher proportion of land under cultivation before controlling for distance to markets relative to TTAs [a reversal of the results in (1) and (2) for the period before 1980], and this was potentially due to NPAs farmers now able to access research and extension services with the advent of “majority rule”. This is supported by the F-test comparisons of EAs against NPAs in Table A.8 in Appendix A. The F-tests show statistically significant differences between EAs and NPAs before 1980, whereas there appear to be a convergence afterwards. The distance to main city [$\ln(dist\ city)$] is positive but only significant in columns (8) to (10) – a reflection of the fact that agriculture is concentrated in TTAs (the base category) that are far from towns in general. A similar argument can also be made about the coefficient of distance to main road [$\ln(dist\ main\ road)$] which is positive and significant in columns (3) to (5) and columns (8) to (10). The coefficient for distance to rail [$\ln(dist\ rail)$] is also positive and statistically significant in columns (8) to (10).

These results suggest that access to transport infrastructure and main cities (markets) was crucial for cultivation decisions of farmers in NPAs. Those who were located close to roads and rails were more likely to cultivate greater expanses of land, while those further away more closely resembled their counterparts in TTAs. Lack of property rights stymied cultivation in TTAs; however, even land rights in NPAs did not guarantee production that is more expansive. Crucially, areas where Africans enjoyed individual land rights only achieved more expansive cultivation if they were better connected to external markets. Hence, the notion that NPA farmers performed better than TTAs was conditionally true: land rights assisted them if they

could also easily transport their goods for sale beyond local markets. Our results suggest that land tenure reform can only be successful if other market failures (which also arose from exclusionary, isolationist policies) are simultaneously addressed.

2.6 Robustness Checks

A number of potential confounders may drive our results. Firstly, local political power of European farming unions is difficult to control for. However, as discussed above, estimated European advantage becomes statistically insignificant if we consider models that are heterogeneous in access to infrastructure. As a result, we do not consider union power to be a remaining confounder; in fact, the ability to lobby for the infrastructure that we control for represents this political influence. National political power of the Rhodesian Front (RF) may also be influential in determining our results. To test this assertion, we re-run all our regressions for a sample that includes only the period after 1980 [see Columns (6) to (10)]. Our core coefficients are stable; we conclude that regime change did not deteriorate the higher cultivation areas in NPA regions that were adequately closely located to relevant transport infrastructure. Implementing our empirical methodology for the period after the attainment of 'self-rule' in 1980 and observing that the estimates remain robust importantly show that the results are not affected by other events.

To test this further, we conduct rolling regressions across time. In each year we run a regression that includes data from the prior and subsequent period, so that we have a three year rolling window in each sub-sample¹⁸. Model predictions at selected distances from railways, main roads and cities are represented over time in Figure 2.5. Results focus on the premium which NPA farmers exhibit vis-à-vis TTA farmers. In all cases, farmers in NPAs have a premium in crop production over TTAs if they are located within 1km of the respective infrastructures. The relationship is robust over time, except towards the very end of the period. The magnitude of the NPA advantage diminishes as farmers are located further from cities, railways and roads. Again, this result is stable over time, except towards the end of the period. Our results also include 90% confidence intervals: in the case of distance to railways and cities, we do not have sufficient statistical power to distinguish year-specific estimates from zero. Close proximity to roads is, however, advantageous (from an economic and statistical perspective) for the extent

¹⁸ The first and last year naturally only comprise a 2-year window. We pool regressions to smooth over idiosyncratic fluctuations in specific years, and to gain statistical power, as the year-specific analyses contain few NPA observations relative to TTAs.

of NPA cultivation in all periods; the reduction of the effect with remoteness is also statistically significant and robust to the period of analysis. Overall, the magnitudes of our estimates suggest that the results found in pooled analysis are also stable over time; where we have statistical power, we can also show that they are significant. The most convincing estimates emphasise proximity to roads as a precondition for property rights to change farmer decisions. We conclude that in all periods property rights only incentivised land production if located close to infrastructure; in contrast, NPA property rights made no difference (*vis-à-vis* TTAs) to the extent of production if access to markets and infrastructure was absent. Our estimates are not spuriously driven by specific events, such as transition to majority rule in 1980.

2.7 Conclusion

The majority of people in SSA live in rural areas and rely on agriculture for their sustenance. Land titling can serve as a possible mechanism to promote access to land and improve agricultural growth. However, previous research has not provided unequivocal evidence for these links (Udry, 2011). This paper contributes to this debate. We turn to a historical setting, where we are able to compare three groups of interest: African farmers in TTAs operated in a context of communal ownership, while Africans in NPAs were granted individual title; EAs were also governed by individual property rights. Moreover, the two regions allocated to Africans were located in inferior agricultural areas and were isolated from markets and key infrastructures (Machingaidze, 1991). In contrast, European farmers were advantaged by the 1930 LAA by being granted land that was agriculturally suitable and also proximate to markets. From historical accounts, we anticipated a production hierarchy that emphasised the advantages of both location and property rights.

Our study investigated the importance of access to markets over and above putting in place individual property rights, using newly constructed crop production data obtained from historical satellite imagery. The paper finds evidence that the farmers in the NPAs had more land under cultivation, as long as they were located close to main cities, main roads and railway stations or sidings. Access to land rights was not sufficient to create the incentives for more expansive cultivation. Instead, access to infrastructure that linked to other markets was important for farming decisions. Our results highlight that proximity to transport infrastructure was important for NPA farmers. The results show that as distance increases from key infrastructure, NPA farmers tended to perform worse than those in TTAs. TTAs were overcrowded with high population density, while the opposite was true in NPAs. In explaining

the process of development Boserup (1996) argues population intensity (in the TTAs in this case) leads to the take up of uncultivated land for farming and settlement – hence it is no surprise that the TTAs had a higher ratio of land under cultivation. Access to profitable, external markets was therefore more important for NPA farmers than those in TTAs and combined with private property rights to incentivise production. In principle, land titles may increase the area of land under crops. Yet, access to markets complements individual property rights in contributing to greater areas of land under cultivation.

While our empirical analysis cannot conclude on agricultural yields and incomes of the various farmer groups; and while we are unable to account for levels of mechanization and other potentially confounding un-observables that may reduce the magnitude of our coefficients, time series evidence shows that proportion of area under cultivation is strongly correlated with the former over time (Appendix C, Table C.1 and C.2). We therefore tentatively infer that this analysis extends to other (unmeasurable) indicators of agricultural success. Our results, however, have other contributions to make. They are the first to use reconstructed regional micro data for analysis in Rhodesia; the benefit is that we could control for a host of observables that may have driven the selection of superior land for advantaged groups. Time series aggregates – which are more readily available – are not sufficient to account for such biases. The historical context within which this study is located provides an ideal setting to weigh up the roles of market access and land titles in agricultural production; this is because these characteristics were purposefully allocated to the various groupings through legislation and could be clearly identified. Our study, however, also adds quantitative estimates to understand an existing historiography on Rhodesia. Our models are able to uncover heterogeneity in the impact: in particular, we emphasise the necessary role that market access plays in the process of reforming land rights in sub-Saharan Africa.

CHAPTER 3

CULTIVATION AND CROP QUALITY AFTER THE FAST TRACK LAND REFORM PROGRAM: SATELLITE DATA EVIDENCE FROM ZIMBABWE

3.1 Colonial Extractive Institutions Persistence, SSA Underdevelopment and Change

Relative to other developing regions, sub-Saharan Africa (SSA) has been characterised by weak development (Nunn, 2008). Acemoglu, Johnson & Robinson (2001) single out poor institutions, not geography, as the reason. The persistence of historical institutions to influence present day contemporary underdevelopment has been well documented¹⁹ [see (Grier, 1999; Lange, 2004; Yoo & Steckel, 2010)]. Acemoglu et al. (2001), explain that in high mortality colonies, European settlers set up extractive institutions that did not provide for property rights and government accountability and the effects of these have persisted over time. In settler colonies (Amanor-Wilks, 2009), Europeans mimicked European Institutions that guaranteed property rights and established checks and balances, much of this literature citing Australia, New Zealand and North America as the prominent examples (Acemoglu et al., 2001). An important extension to this argument may be that institutions in settler economies such as Zimbabwe [the “Neo-Europes” in Crosby (2004)] were also to some extent extractive: they were engineered to serve the interests of a few settlers and excluded indigenous people from these rights; Rhodesia’s Land Apportionment Act (1930) (LAA 1930) is a case in point. In contrast, the Belgian colonisation of the Congo, for example, perfectly embodies the term extractive institutions. Hence, in settler economies (Neo-Europes), it is also expected that extractive colonial institutions can influence modern underdevelopment.

A few studies have investigated the channels through which this happens. Bambio and Agha (2018) note the role played by land institutions in promoting agricultural investments. Chingozha and von Fintel (2019), as well as Dell (2010) showed that absence of land tenure property rights and poor access to public goods (transport networks) have negatively affected crop cultivation and contemporary development respectively. Therefore, land reform

¹⁹ More discussion on the importance of institutions: Grier (1999) shows that former British colonies outperformed their French and Spanish counterparts, which shows the importance of institutions. Acemoglu, Johnson and Robinson (2001) also highlight the stark differences between South and North Korea, as well as between East and West Germany in explaining the importance of effective property rights in a private market.

programmes that seek to create or restore balanced property rights are an attempt to ensure effective market participation on the part of farmers. Similarly, as this study examines, redistributive land reform programmes should enforce property rights in the post-phase.

Land rights are important because they allow for the internalization of external cost (Coase, 1960; Demsetz, 1974). Land rights can ensure that farmers realise the full cost (*see* Demsetz, 1974) of working the land, thereby avoiding wasteful practices which culminates in inefficiency. Alston et al., (1996) concurs, mentioning that ill-defined property can influence rampant deforestation among other “wasteful” land uses. As put forward by Demsetz (1974) land ownership through land titles put the title holder in the position of a broker who has to consider and trade off current use/consumption against discounted future benefit. In this chapter, an understanding of the absorption of an activity’s transaction costs may be useful in thinking about the costs and benefits of Zimbabwe’s land reform program. Zimbabwe’s land reform program did not result in an internalization of the negative cost of losing farms on the part of the former owners in a Demsetz (1974)’s “buy-him-out” fashion since the land was acquired without compensation. The land was acquired from white farmers without negotiation (*see* Demsetz, 1974), without the full cost of the transaction bearing on the right person as the so called “willing-buyer-willing-seller” could have ensured.

Institutions evolve (Demsetz, 1967). As explained by North and Thomas (1970) institutional innovations occur when particular individuals or groups in a given society can reap positive rewards out of their investment in making change occur. Hence the attempt at land reform in SSA may be viewed as a form of induced institutional change, and it is the effects of such change on cultivation and crop quality that this chapter investigates. This study investigates the effects of attempts to correct the persistence of selective land rights enforcement by colonial institutions in Zimbabwe (formerly Rhodesia): - specifically how the lack of follow-through (among other factors) in subsequent corrective land reform efforts has affected crop cultivation and crop quality. We used the Fast Track Land Reform Program (FTLRP) as a case study. The Fast Track Land Reform Program was a redistributive land reform programme that distributed 80% of land in EAs (resulting from the LAA, 1930) to the indigenous African majority with the aim of correcting the imbalance caused by historical institutions (Scoones et al., 2011a). Given the lack of enforcement of property rights enforcement post the FTLRP exercise and its wide-scale nature, this paper partly contributes to the literature explaining stagnation in post-independent SSA states, particularly those that have attempted various forms of agrarian

reform. Probably owing to “Africa’s statistical tragedy”, a scenario in which governments in Africa lack the resources or political will to provide objective datasets with high temporal and spatial frequency for research (Devarajan, 2013), the majority of studies that have investigated the effects of FTLRP are qualitative and covered specific (small) geographic areas [*see* Mutangi (2010), Marimira (2010), Mandizadza (2010)]. Therefore, another critical contribution of this study is that it employs novel nationwide cultivation data generated from applying an SVM learning algorithm on Landsat imagery as well as Normalised Vegetation Index (NDVI) (as a proxy for crop quality).

The rest of the chapter is structured as follows: Section 3.2 provides an overview of the importance of property rights (particularly land rights) in poverty alleviation as well as explaining the different channels through which land titles promote agricultural growth. Section 3.3 defines and describes various issues regarding land reforms and the study contribution. The section also describes the study setting (Zimbabwe) and the relevant land tenure security issues in the country as well as providing an explanation of FTLRP. Section 3.4 provides a description of the remotely sensed data (and its advantages) that were used in the study and the machine learning that is adopted. The section also describes the identification approaches and model specification. Section 3.5 discusses the analysis and results obtained from the Difference-in-Difference (DID) and Regression Discontinuity (RDD) estimates, while Section 3.6 presents the conclusions.

3.2 Importance of Land Rights in Poverty Alleviation

Sokoloff and Engerman (2000) argue that North America and Canada achieved more industrialisation than countries in Latin America because in the latter, political and socio-economic elites restricted wealth and kept opportunities among themselves, suppressing the potential of the rest of the population; and this highlights the importance of colonial institutions in contemporary underdevelopment and incidence of poverty. Hazell, Poulton, Wiggins and Dorward (2010) state that rural dwellers number three billion in developing countries, with nearly 67% of them domiciled on almost 500 million farms of less than two hectares (Lowder, Skoet & Raney, 2016; Vitoria, Mudimu & Moyo, 2012). At the same time, Rigg (2006) observed that there is an intricate tie between agriculture and rural Global South livelihoods. Although we cannot necessarily associate land abundance with prosperity in the rural south (Rigg, 2006), poverty reduction and property rights alignment through agrarian reform are related (Besley & Burgess, 2000).

In sub-Saharan Africa, agriculture is central to economic activity (Peters, 2009; Vitoria et al., 2012). From this perspective, the most critical aspect, also from a poverty alleviation viewpoint is the land (Holden & Ghebru, 2016). Land redistribution seems to be a sensible way to ensure equitable access to land, yet this occurs without properly configuring the legal rights to it in the wake of agrarian redistribution. Newman, Tarp and Van Den Broeck (2015), note the importance of upholding land property rights as an important ingredient in poverty alleviation. We estimate the effect of misconfigured land rights post 2000's Fast Track Land Reform Program (FTLRP) on welfare, cultivation and crop quality using remotely sensed data in the hope that other countries can learn from this experience.

3.2.1 Property rights

Put simply, property rights encourage investment because landowners have the confidence that they will enjoy the benefits or profits from the investment (Markussen, 2008). Empirical evidence shows that land titles as a way of enforcing property rights enhance the chances of financial and time investment on the land – thereby raising agricultural output (Deininger et al., 2008; Muchomba, 2017; Newman et al., 2015; Place, 2009). To explain why more time was spent in agriculture, Deininger et al. (2014) used longitudinal data on a sample of 1 200 Chinese households and found that land titling and formal certification discouraged farmers from leaving agriculture. Second, property rights ensure tradability in land and production gains since the market will help determine that it is held by the most efficient farmers (Colin & Woodhouse, 2010; Deininger et al., 2008; Markussen, 2008; Yngstrom, 2002). As Deininger et al. (2008), Aryeetey and Udry (2010) and Deininger et al. (2014) argue, property rights reduce transaction costs in the transfer of land for agricultural production purposes. Deininger and Jin (2006), established that a guarantee that land was safe from redistribution enhanced farm income by approximately 1,5 per cent.

Markussen (2008) investigated the effect of property rights in Cambodia and found that households that held proper documentation to their land holdings had higher agricultural production – even in environments where the government did not have a lot of capacity. At the same time, Do and Iyer (2008), Abdulai, Owusu and Goetz (2011), Newman et al. (2015) and Abdulai et al. (2011), showed that land titling improved agricultural yields among Vietnamese and Ghanaian farmers, respectively.

Land tenure security is viewed as an important challenge to poverty alleviation (Place, 2009). In the case of China, Deininger et al. (2014) observed that while agricultural expansion had aided poverty reduction efforts, institutional factors impeded smooth land transferability to more efficient producers. On the other end of the spectrum, creating an efficient market could result in landlessness. However, Ravallion and Van de Walle (2008) analysed Vietnamese data consisting of four household surveys for the period 1993-2004 and found that landlessness occurs in upper income strata, which implies that land titling and land commodification by no means exacerbate rural poverty. In Zimbabwe's case, institutional factors have also impeded the bankability and transferability of land and this chapter investigates the effects. Guardado (2018) asserts that property rights are a crucial determinant of income in rural areas and reforming customary land tenure systems is an important strategy in correcting colonial institutions. However, for Zimbabwe, land reform meant reclaiming abundant land that was held by farmers in European Areas (EAs) (see Figure 3.1).

3.2.2 Land titles and credit

An important means by which property rights and land titles can enhance agriculture production is through the credit channel. Deininger (1999) suggested that the long-term success of any agrarian reform also relies heavily on the participation of the private sector in implementation. For instance, banks have an important role in the provision of working capital to newly resettled farmers, yet the state-led nature of Zimbabwe's FTLRP did not promote such a scenario. Without land titles, ability of farmers to access credit is constrained. Yet, credit constraints prevent farmers from achieving the full opportunity out of their land (Markelova et al., 2009).

Despite the importance of land titles as collateral to access credit (Aryeetey & Udry, 2010) various studies do not find supporting empirical evidence (Deininger, Savastano & Xia, 2017; Do & Iyer, 2008; Markussen, 2008; Udry, 2011). Meinzen-Dick et al. (2017) indicate that many of these studies rely on household level data, a level at which it is difficult to establish the nature of property rights. In Vietnam, Do and Iyer (2008) did not find any significant effects in borrowing following land titling. Markussen (2008) used data from the Cambodia Household Socioeconomic Survey for 2003/2004 and found that formal land titles do not increase propensity to access credit, but rather allow holders to pay lower interest rates. Deininger et al. (2017), explain that the customary setup of African agriculture – characterised by high land

endowment, low-capital intensity production and employment of family labour does not create an environment that necessitates credit access.

On the other hand, this evidence may not necessarily reflect unwillingness to obtain credit but rather lack of access due to financial constraints such as incomplete property rights bestowment (Abdulai et al., 2011). Goldstein and Udry (2008) and Markussen and Tarp (2014) found that politically connected farmers in Ghana and Vietnam, respectively, were able to build investments on their land most likely because their connections buttressed their “de facto property rights and access to credit and transfers”. In Brazil Alston et al., (1996) found that the land title is the acknowledged “institutional device” that facilitates land transfers. Therefore, land titles are important in facilitating the land market.

3.2.3 Land titles and social contracts (trust)

Bezabih et al. (2011) suggested that social contracts are an important aspect in the functioning of any society, adding that land titling helps in cementing trust – an important ingredient in the functioning of social contracts. Bezabih et al. (2011) used a DID identification strategy to analyse the effects of Ethiopia’s land titling programme and found that there was increased interpersonal trust in land transactions (between sellers and buyers) as well as towards the authorities due to land certification. By extension, the institutional environment surrounding Zimbabwe’s FTLRP and the subsequent provision of 99-year leases may have been marred by high levels of mistrust between the government, the private sector (including agricultural financiers) and the farmers themselves.

3.3 Land Reform in Developing Countries

3.3.1 Motivations for land reform

State-led agrarian reform has been thought to create the momentum, or “big push” in the words of Rosenstein-Rodan (1961), needed to achieve development (Sikor & Müller, 2009). Adams (1995) and Zarin and Bujang (1994) defined land reform as the reallocation of property or rights for the benefits of previously disadvantaged groups such as tenants, peasants and farm labourers. There are diverse views behind the reasons for land reform. Zarin and Bujang (1994), offer three motivations for agrarian reform. The political motive is when a government uses land reform either to gain or retain power; the social motive targets a more egalitarian society

while the economic one centres around efficiency (Zarin & Bujang, 1994). Adams (1995) and Zarin and Bujang (1994) concur that politics are inseparable from land.

3.3.1.1 Political land reforms

A number of studies acknowledge land reform as a political survival instrument. For example, using a panel of Mexican states for the period 1917-1992, Albertus, Diaz-Cayeros, Magaloni and Weingast (2012) found empirical evidence that the objective of the country's agrarian reform was political survival, leaving peasants much more dependent on the state. Albertus et al. (2012), observed that land redistribution efforts were high during election years in Mexico. Berry (2002) indicates that the same could be said for Zimbabwe since the government embarked on FTLRP to survive pressure from the opposition. Land reform may also be implemented just to break landlord power. In Iraq, for example, Warriner (1969) asserts that land reform did not take the country forward in any significant way, but only achieved to break the power of the sheikhs. Apart from the political motive, land reform has a very strong political character. The Bolivian and Cuban agrarian reforms in the 1950s attracted economic sanctions (Barraclough, 1999; Seligson, 1984). According to Barraclough (1999), land reform in Cuba attracted a trade embargo from the US. Similarly, a number of restrictions were placed on the Zimbabwean economy in the aftermath of FTLRP; notably ZIDERA²⁰ that prohibited US firms from engaging in any business with Zimbabwe. The political consequences of land reform may thus be an important cause of hesitation by post-independent states to redistribute land.

3.3.1.2 Economic land reforms

Agrarian reform can increase agriculture returns to scale through eliminating the diseconomies of scale associated with larger farms and allowing the agility and innovation of smaller farm holdings (Adams, 1995; Barraclough, 1999; Cotula, Toulmin & Quan, 2006; Zarin & Bujang, 1994). While larger farms can enjoy economies of scale, Cotula et al. (2006) posit that mechanisation returns to scale are evident in crops such as sugar cane, some cereals and soya, while crops such as rubber, fruit and vegetables produce better yields under manual labour intensive conditions.

Better efficiency under smaller farms may also be due to the outputs gains from the self-employment incentive of family farms (De Janvry, Sadoulet & Wolford, 2001). Hazell et al.

²⁰ ZIDERA – Zimbabwe Democracy and Economic Recovery Act. It was passed by the US in 2000. ZIDERA instructed the director of any US financial institution to block any grants to Zimbabwe or any reduction in debt.

(2010), support the efficiency and equity argument and added that the role that small farms play in ending poverty remains compelling. In an input use and financing survey for Malawi, Nigeria and Uganda, Adjognon, Liverpool-Tasie and Reardon (2017) found that smaller farms applied more inputs per hectare than their medium/large counterparts to compensate for the lack of land. On the contrary, De Schutter (2011) is of the view that large-scale farms can result in increased production and crop exports, thereby positively affecting welfare. Similarly, Collier and Dercon (2014) argue that larger-scale farmers push the limits of technology, which could improve yields, for example. Warriner (1969), adds that if the costs of land reform are too high, output tends to decline and land reform cannot contribute to economic growth under such circumstances. Rather, rapid economic growth is a necessary precondition for a fruitful land reform programme. Importantly, Collier and Dercon (2014) highlight the contradicting priorities of the need to reduce poverty and ensure pro-poor growth by focusing on smallholder agriculture, yet for Africa to realise economic development the modern sector should extract vast populations out of (especially smallholder) agriculture.

3.3.1.3 Egalitarian (poverty reducing) land reforms

Land and poverty are related, thus addressing land imbalances is a strategy towards a more egalitarian society. A number of authors agree that there is a strong positive correlation between rural poverty and landlessness (Adams, 1995; Christodoulou, 1990; Cotula et al., 2006; Reyes, 2002; Seligson, 1984; Tarisayi, 2013; Zarin & Bujang, 1994). Tarisayi (2013), argues that land reform can influence upward mobility of the previously underprivileged because land ownership increases their asset base. Therefore the question of land redistribution is important in any country where a significant proportion of society is poor as access to land is one of many things needed to alleviate poverty although it is only a starting point (Tarisayi, 2013). Cotula et al. (2006), mention that the World Bank's Poverty Reduction Strategy Papers (PRSPs) have (to various degrees) linked poverty to landlessness; for example, in Burkina Faso, Mongolia, Honduras, Cambodia, Lao and southern Africa. Similarly, Cotula et al., (2006) suggest it will be difficult to lift communities out of poverty unless land reform is implemented on a foundation of commitment and strong political will by the government, as well as effective reorganisation, orderliness and the provision of support and incentives.

Mellor and Johnston (1984), argued that dualistic based policies that prioritise the funding of industry and large-scale farming leaves out the broader population base. Focusing on smaller farmers improves employment and poverty through increased productivity and innovation

related to their labour intensive production (Birner & Resnick, 2010; Mellor & Johnston, 1984; Mellor & Malik, 2017). The increased demand for goods and services from the rural non-farm sector by smallholder farmers reduces poverty. In rural Mexico Finan et al. (2005) found that the welfare of smallholders increased by a mean of 1.3 times their earnings per additional hectare of land. Reyes (2002) found that being a beneficiary of land reform reduced the probability of being poor in the Philippines. Ciamarra (2004) argued that providing more access to land improves the welfare of the poor but cautions that this is only possible when both state-led expropriation and market-led policies are simultaneously pursued.

3.3.2 Approaches to land reform

“Land tenure reform, external inducements, external controls and confirmation of title” are the major variants of agrarian reform (Adams, 1995). Land tenure reform involves realignment of reciprocal property rights between owners and gives informal tenants formal property rights as a way of balancing the tenant-landlord relationship (Adams, 1995; Cotula et al., 2006). In contrast to land tenure reform, external inducements are market-based incentives (for example, credit to allow land transfers) that are put in place by the state to lead the agrarian property rights structure in a certain way, such as Zimbabwe’s willing buyer willing seller model prior to 2000 (Adams, 1995; De Villiers, 2003). External controls describe agrarian reform as a set of legislative controls or prohibitions on property rights including, but not limited to, nationalisation, restitution or land expropriation with or without compensation on the grounds of underutilisation, excessive size, landlord absenteeism or just to correct historical imbalances (Adams, 1995). Lastly, land reform may involve the verification and confirming of titles of those already in possession of landholdings (Adams, 1995).

In order to secure welfare, the four principal forms of agrarian reform may need to be considered holistically and regarded as a cycle through which the external control version of land reform should pass. In Latin America, López and Valdés (2000) found provision of credit to small farmers and land titling were the most effective means for poverty reduction. In an evaluation of the success of land markets from the perspectives of efficiency and equity, Otsuka (2007) suggested that tax and subsidies may be preferable instruments of land reform which are not detrimental to production efficiency as is likely under outright expropriation.

3.3.3 Study contribution

There is not much debate on the definition for land reform. Tarisayi (2013) suggests that while there appears to be consensus with respect to the definition of land reform, it is the approach and justification that has remained the subject of debate. Therefore, this study attempts to evaluate Zimbabwe's approach to land reform and lack of follow-through in property rights enforcement in the post-phase effects of crop cultivation and crop quality. Due to the persistence of colonial institutions in contemporary underdevelopment, the expectation that post-independence states must address colonial land imbalances has always been alive and it is the aim of this study to provide these countries with useful insights using the case of Zimbabwe. These countries include Namibia, South Africa and many others in the developing world.

Enhancing understanding on the best approach to land reform and highlighting potential pitfalls of not enforcing property rights is important because land occupies a central place in the economic development of any nation, aside from its relevance in poverty reduction and inclusive growth. Moyo, Jha and Yeros (2013) assert that because no country can absolutely guarantee the food security of its people and that every investment question, one way or another, falls back to the land, it is important to understand land tenure security and agrarian reform. Branca, McCarthy, Lipper and Jolejole (2011) stress that the sustainable management of agricultural areas is indispensable in developing countries, given the critical role it plays in the economy.

Against the backdrop of Devarajan's "African statistical tragedy" (2013), this chapter also makes a data contribution. It adds to the debate on the importance of securing private property rights for effective market participation in the post-land reform phase by estimating at the ward level the effects of this lack of follow-through in property rights enforcement in the wake of agrarian reform using remotely sensed data. Although the Zimbabwe government provide some form of "tenure security" after its land reform program, Fenske (2011) and Zikhali (2008) suggest that these were unmarketable and unbankable (as discussed later in the chapter) – hence the lack of follow-through on the part of government. We used Night Lights Data (NLD) [following Pinkovskiy and Sala-i-Martin (2014)] from The National Oceanic and Atmospheric Administration (NOAA); Normalised Difference Vegetation Index (NDVI)²¹ and also applied

²¹ The study uses an historical NDVI from the NOAA vegetation monitoring series that spans the period 1981 to 2016.

an SVM machine learning algorithm on Landsat imagery [following Fernandes (2015)] from the US Geological Survey (USGS) to generate crop cultivation data. The lack of publicly available nationally representative survey data means that it is not possible to track standard welfare measures for the entire economy. To fill this gap, this study used unconventional, innovative datasets to estimate changes in welfare for small areas (wards). Identification relies on Difference-in-Difference (DID) and Regression Discontinuity (RDD) econometric approaches; that is, we measure differences in crop cultivation, NLD and NDVI before and after the implementation of FTLRP, but also remove the effects of time changes in regions that were not primarily targeted by the policy.

3.3.4 The study setting

Moyo et al. (2013) stated that

[...] the land movement in Zimbabwe may have been the most successful in reclaiming land, but the depth of the political work that has been underway on all continents has set the stage for consideration of 're-peasantisation' as a modern, sovereign project in the twenty-first century.

This chapter considers Zimbabwe not only because it is the most recent twenty-first century case study, but also it is the most large-scale (see Table 3.1) from the point of view of acquiring land from European farmers for distribution among the landless indigenous population. This is in contrast to its southern neighbour, South Africa, whose land reform programme has been a “disappointment” according to Binswanger-Mkhize (2014), if the number of beneficiaries is considered. There has not been consensus on the effects of incomplete land titles on welfare and agriculture production in Zimbabwe and this research conducts a nationwide, “bird’s eye view” investigation to contribute to this debate. Additionally, the majority of the studies (Abdulai et al., 2011; Fenske, 2011; Goldstein & Udry, 2008; Markussen, 2008; Ravallion & Van De Walle, 2006) that have investigated the effects of land titling programmes have relied on household survey datasets. In the main, we used a novel dataset created from classifying Landsat imagery using an SVM algorithm for the period 1997-2003.

Table 3.1 Changes in the distribution structure of the land

Land Category	1980	2000	2010
	Area (million ha)	Area (million ha)	Area (million ha)
Communal areas	16.4	16.4	16.4
Old settlement	0.0	3.5	3.5
New resettlement: A1	0.0	0.0	4.1
New resettlement: A2	0.0	0.0	3.5
Small-scale commercial farms	1.4	1.4	1.4
Large-scale commercial farms	15.5	11.7	3.4
State farms	0.5	0.7	0.7
Urban land	0.2	0.3	0.3
National parks and forest land	5.1	5.1	5.1
Unallocated land	0.0	0.0	0.7

Source: Scoones et al. (2011b)

Table 3.1 shows that 78% of the land that was European large commercial farms in 1980 had been reallocated to Africans by 2010²². This confirms that Zimbabwe land reform programme was indeed the most wide-scale in as far as transferring the land from a privileged class to their less fortunate counterparts.

3.3.5 Zimbabwe Fast Track Land Reform Program (FTLRP)

Acemoglu et al. (2001) point out that colonial institutions have persisted even after independence in previously settler economies worldwide. Table 3.1 shows evidence of such persistence since the status quo of colonial distribution of land in Zimbabwe had hardly changed until the year 2000. Until 2000, land distribution had been shaped by LAA 1930. The enactment of LAA 1930 (see Figure 3.1) was motivated by settlers' need to reduce agricultural competition from indigenous farmers and to control labour (Machingaidze, 1991). These factors parallel those of Peru, where indigenous farming was restricted to communal farming in the *mita*²³ districts; this was done to ensure that there was surplus labour to work in the mercury mines in these regions (Dell, 2010; Guardado, 2018).

²² By 2010, 15.5 million ha of commercial farms were reduced to 3.41 million equating to 78% of the total

²³ *Mita* were forced labour mining districts which existed during the colonial era in Peru [see Dell (2010)].

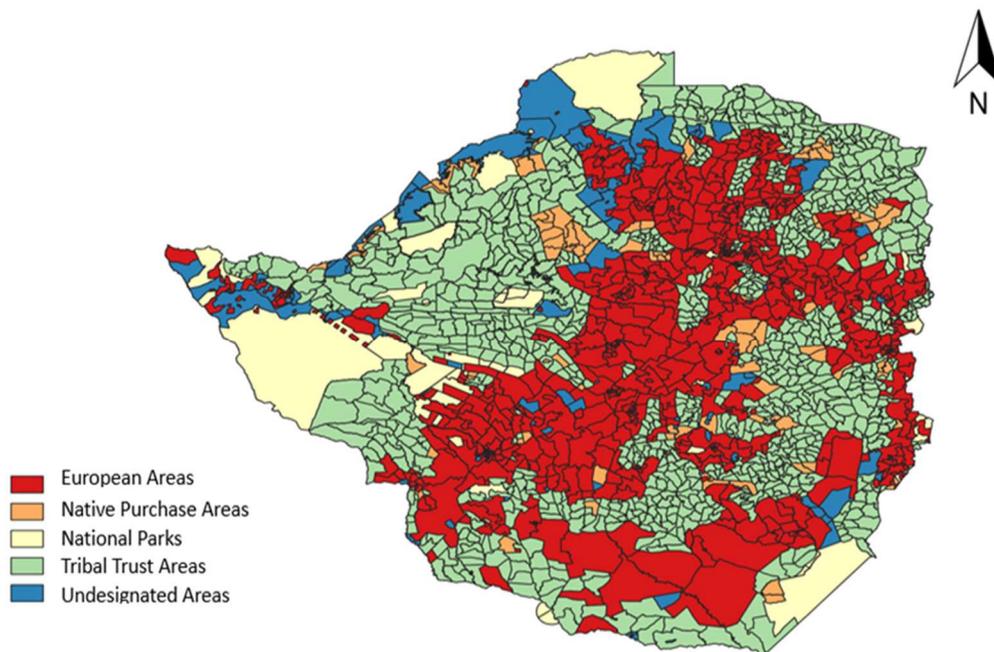


Figure 3.1 The Land Apportionment Act of 1930 (LAA, 1930)

NOTES*:

Figure 3.1 shows the distribution of land in Zimbabwe (formerly Rhodesia), based on the Land Apportionment Act of 1930 – superimposed on the Zimbabwe Level 3 ward level shape file. The areas in red were designated for Europeans (EAs) while those in green were the Tribal Trust Areas (TTAs) (for Africans). Those in brown are the Native Purchase Areas (NPAs) (African large-scale farms).

Upon independence in 1980, Zimbabwe inherited the legacy of an unequal land distribution (Tarisayi, 2013). As at 1980, European farmers owned most of the fertile, rain fed highveld in the middle of the country (about half the size of the country), with Africans occupying mostly the sandy, dry soils of the lowveld (see Figure 3.1). However, tenets of the Lancaster House Conference (the treaty that ushered in majority rule) demanded the implementation of land reform on a willing buyer willing seller basis, with Britain and the US meeting half the cost. Partly because these pledges were never really fulfilled (De Villiers, 2003) – drastically slowing down the process of agrarian reform between 1980 and 2000 – and due to increased pressure from liberation war veterans and the general masses, the Zimbabwe government implemented the Fast Track Land Reform (FTLRP) in 2000. The FTLRP allocated smaller A1 and larger A2 farms, the latter mostly being the preserve of ruling elites (Scoones et al., 2011a). Table 3.2 shows the timeline of land reform in Zimbabwe.

Table 3.2 Timeline of land reform since majority rule

Phase	Hectares acquired	Families resettled	Notes
LRRP ²⁴ I 1980-1998	3 498 444	71 000	Based on willing buyer willing seller (WB-WS) principle. The programme was characterised by meticulous planning and beneficiary selection and input and infrastructure support. The programme slowed at the end of the 80s due to movement towards structural adjustment.
LRRP II 1998-2000	168 264	4 697	Also based on WB-WS, but slowed down by the Economic Structural Adjustment Program (ESAP) and the follow-up Zimbabwe Programme for Economic and Social Transformation (ZIMPREST). The programme was guided more effectively by the Land Acquisition Act of 1992. The Act provided guidance in designation, gazetting, valuation and compensation. Programme implementation was also adversely affected by withholding of donor funds over differences with government.
FTLRP	8 300 000	-	Programme was characterised by violence, forceful expropriation and the tacit approval of state security apparatus. Farms were seized without compensation.

Source: Based on information in Waeterloos and Rutherford (2004) and Scoones et al. (2011b)

In 2000, Zimbabweans voted in a constitutional referendum in which the government had proposed that the country compulsorily acquire land from the European farmers without compensation (Waeterloos & Rutherford, 2004). The government lost this referendum, but would still have its way in FTLRP. This programme commenced soon after the constitutional referendum and in the period leading to elections in 2002 (Tarisayi, 2013); the initial momentum of the programme was built by disgruntled war veterans who “spontaneously” attacked and occupied white-owned commercial farms in 2000 (Berry, 2002; Marimira, 2010). War veterans led almost the entire programme until the government made new constitutional provisions that would seal the expropriation of the land from white commercial farmers.

²⁴ Land Reform and Resettlement Program

Land reform mainly focused on commercial farms that were involved in crop production and as a result a number of farms in the Eastern Highlands were left untouched because they consisted mainly of plantations (Mutangi, 2010). Under the FTLRP almost 4 000 white-owned commercial farms were expropriated and reallocated to the indigenous African population (Richardson, 2007). Chigumira (2010), reveals that by February 2010 the Zimbabwe government had resettled 156 000 households on almost seven million hectares of land.

There are divergent views regarding the efficacy of FTLRP and its effects. Chigumira (2010), Mutangi (2010) and Zikhali (2010) argue that FTLRP did not improve welfare, while Mandizadza (2010), Mbereko (2010), and Moyo, (2010) found that the programme resulted in positive gains in productivity and welfare. Each of these studies was localised and did not assess the large-scale effects of the policy. Consequently, this study aims to examine the effects of lack of property rights enforcement post-FTLRP on welfare (with NLD as proxy), crop cultivation and crop quality (using NDVI as a proxy) for *all* of Zimbabwe. This chapter makes an important contribution because it offers the first nationwide view of the effects of FTLRP on welfare, crop cultivation and crop quality using innovative datasets.

3.3.6 Linking FTLRP, property rights and 99-year leases

After implementation of FTLRP, the Zimbabwe government realised that the violent nature of the land repossession left the newly resettled farmers in a quandary due to market failure in the property market. Parallels can be drawn between agricultural stagnation in Zimbabwe following FTLRP and Mexico as De Janvry, Gonzalez-Navarro and Sadoulet (2014) highlight with the rural poverty and agricultural impasse that followed reform in the Latin American country which led to a formal land titling and certification programme in 1992. After land reform, the Zimbabwe government implemented a certification programme where it offered 99-year leases to farmers so that they could make the land bankable. Bezabih et al. (2011) describes land certification as a process that is designed to enhance land tenure security for farmers. However, as at 2018 the 99-year lease agreements had not been accepted by the banking sector as collateral. Among the various reasons adversely affecting the bankability of the permits, Matondi and Dekker (2011) argue that the land permits post-FTLRP were not based on a formal land survey and neither were they registered with the government deeds office; and that under they did not allow the new land claimants to establish permanent structures or to the dispose the land occupancy rights. This is an example of African governments enacting land laws to help the poor out of poverty, but in the absence of effective

implementation and follow-through they fail to achieve the intended objective (Deininger et al., 2008).

At the same time Deininger and Castagnini (2006), Bezabih et al. (2011), Fenske (2011) and Muchomba (2017), argue that while land titling has been found to increase agriculture production in Asia and Latin America, the same cannot be said about the sub-Saharan region. Perhaps titling programmes in SSA do not create enough trust between the various stakeholders such as banks for the “farmer support” social contract to be smoothly executed, following Bezabih et al. (2011).

De Janvry et al. (2014), argued that the conundrum of lack of follow-through to cement land rights in the wake of reform is a result of political uncertainty. Politicians may speculate that creating a land market through land titling and certification post-land redistribution may lead voters to identify with leftist opposition politics (which is analogous to Zimbabwe’s ZANU PF government [in power since 1980] that implemented FTLRP), yet Mexico’s “*Procede*” land certification exercise benefitted the right-wing political establishment. De Janvry et al. (2014), contend that ruling elites fear that completely granting land rights to farmers will wean them off government and reduce electoral influence. Long before the FTLRP, Zimbabwe’s colonial history also offers important political lessons. Christopher (1971) and Nelson (1982), posit that the 1962 elections (before the end of white minority rule in 1980) were contested around the issue of land and the United Federation Party’s (UFP) ambition to provide more land access to the African majority through a reform programme contributed to its election loss. Thus, complete land reform may be politically risky.

Regardless, there is speculation that in Zimbabwe, politically connected people do manage to secure lines of credit despite the weakness in bankability in the 99-year lease agreements. As argued by Scoones et al. (2011a; 2011b) the FTLRP ruling elites benefitted from the larger A2 farms while the less politically networked benefitted under the redistribution of smaller A1 farms. Fenske (2011) contends that the A1 “offer permits” system allowed heirs to inherit the land but did not provide for marketability; whereas the A2 farmers could use their landholdings as collateral on the basis that there would be a future land market. Still, Zikhali (2008) notes that the lack of clarity over the terms of the A2 permits also rendered them unbankable. As a result, costs of accessing agricultural finance in Zimbabwe are astronomical at present and the majority of small-scale farmers do not have access to credit (Vitoria et al., 2012). Against the backdrop of improperly configured property rights in the form of the 99-year lease agreements,

we investigate the effects of FTLRP on welfare, the amount of land under cultivation and crop quality.

3.3.7 Other factors affecting FTLRP

The preceding discussions revealed not only how important land titles and property rights in general are. Whereas the lack property rights enforcement in the wake of FTLRP is indeed significant and is our main argument, there are factors that could also have adversely affected the effectiveness of the program. These include “the learning curve effect”, the violent nature of the dispossession, insufficient mechanization, lack of extension services, training and other support. In a study of factors affecting land utilization for FTLRP beneficiaries in Zimbabwe’s Mashonaland Central Province, Musemwa and Mushunje (2012) found a positive correlation between herd size (for use as draught power) and land utilization – which presents mechanization as one of the important factors to ensure agricultural success especially in the post reform phase.

In the immediate short term, success of FTLRP (as a change process) might also have been affected by “teething problems” and other natural disturbances towards a new equilibrium. This might be so especially in light of the “violent nature” of the program as put forward by Cliffe et al., (2011). Additionally, the lack of education and training might have been an additional significant factor affecting the success of FTLRP (measured by way of crop cultivation and crop quality in this chapter) given that Musemwa and Mushunje (2012) identified training needs in their case study of the country Mashonaland Central Province after FTLRP. Obi (2011) and Murisa (2011) also found “limited” farm mechanization as an important input in the farm production function in Bindura and Goromonzi Districts of Zimbabwe respectively.

3.4 Data and Identification

Census agricultural data is seldom available for fine geographical units such as farms (Lowder et al., 2016). For Zimbabwe, agricultural time series data is not publicly available for our study period at the small areas level (ward in our case). The study therefore relies on remotely sensed data. Fernandes (2015), defines remote sensing as the “collection of information [pertaining to] an object without making [any] physical contact with it”. Remotely sensed data has proven to be an important resource in the spatial examination of economic phenomenon, because it is

available over time even in parts of the world that would otherwise be lacking in available census and other types of data (Elvidge et al., 2009; Henderson, Storeygard & Weil, 2012; Li, Ge & Chen, 2013). Another distinct benefit of these data is that they provide the ability to track welfare for very fine spatial units on an annual basis. Remote sensing data that is obtained by satellites has immense benefits as it provides for the temporal and spatial examination of environmental variables. The data has also been widely used in crop classification and the estimation of yields (Dhumal, YogeshRajendra & Mehrotra, 2013). Ustuner, Sanli, Abdikan, Esetlili and Kurucu (2014) mention that due to the expanding need for quick, cost effective data on land cover, many countries have launched several satellites, for example, RapidEye, GeoEye-1, WorldView-2, Landsat8, SPOT-7, TerraSAR-X (2007), Sentinel-1A and ALOS-2 (2014).

Estimation of the FTLRP's effect on welfare, crop quality and crop cultivation was carried out using NLD, NDVI and Landsat image cultivation data, respectively. In this section, NLD, NDVI as well as some literature on the classification of Landsat images using machine learning and the techniques used are discussed. The DID and Regression Discontinuity Design (RDD) identification strategies employed by the study are also discussed.

3.4.1 Night lights data (NLD)

Satellites measure the luminosity of night lights per pixel at regular time intervals. Pinkovskiy and Sala-i-Martin (2014) define a pixel as one square kilometre and each pixel is assigned a Digital Number (DN) that represents its brightness. The DNs are measured on a scale of 1 to 63 (Pinkovskiy & Sala-i-Martin, 2014). Chen and Nordhaus (2011) indicated that there are three versions of the NLD, namely raw, stable and calibrated light. Most researchers use stable lights data because it removes veld fires and other noise. For the same reason, this research uses stable lights data.

Night Lights Data has been used over the past several years because of its high correlation with economic welfare measures. According to Henderson et al. (2012) light is required for the consumption of any good or service at night so increases in light intensity may imply increases in consumption; which may in turn increase GDP, economic activity or welfare. Elvidge et al. (2009) and Henderson et al. (2012) concur that satellite sensors address the problem of inconsistency by availing NLD datasets on a yearly basis and thus economic phenomenon can be traced over time.

Chen and Nordhaus (2011) assert that the other advantage of NLD is that there is relative objectivity (versus survey data) as well as the ability to consider geographical variations that would inadvertently affect national income. Literature suggests that several ways have been used by economists to find proxies that can estimate GDP and economic welfare at very fine geographic levels (Henderson et al., 2012). Night lights data has been found to be a better alternative among a number of proxies for economic activity. Against a backdrop of high rates of informality, NLD can also be a useful solution to the problem of data unavailability from this low income sector. To support this argument, Henderson et al. (2012) indicate that in developing countries a significant proportion of economic activity takes place in the informal sector where there is very poor collation of statistics. Thus, NLD is a useful tool to measure spatial economic activity over fine geographical areas in unmeasured economies such as Zimbabwe.

Night lights data can make for very effective spatial analysis of economic phenomenon. Henderson et al. (2012) indicate that the data can show light intensity over very fine geographical areas, which makes it a very useful tool for spatial examination of the economic phenomenon, adding that the data is available at higher and more consistent frequency and thus is a good tool to measure the effects of shocks and other events on economic phenomenon. Hentschel (1998) argued that estimates that are available using poverty maps – or in other words, for the smallest administrative unit of a country – are an indispensable avenue to target and refine policy interventions within smaller geographical areas of a country that have different needs. That is why NLD is used for spatial analysis of almost 2 000 wards in Zimbabwe.

The other important application of NLD is that it is a good indicator of stagnation. Henderson et al. (2012) present an analogy between North and South Korea and light intensity for the south clearly resonates with the more than 100%²⁵ economic growth that the country has gained between 1992 and 2008, while for the North there is absolute absence of any change in light intensity, which may indicate economic stagnation. Thus NLD could be effective in assessing possible stagnation of the economy of Zimbabwe after FTLRP as undertaken by Li et al. (2013), who used night lights imagery and found that mining and agricultural towns were worst affected by Zimbabwe's economic decline.

²⁵ Henderson, Storeyhead and Weil (2012)

Although Elvidge et al. (2009) and Henderson et al. (2012) commend NLD for its consistency over time, Jian and Weifeng (2013) argue that such consistency may not be achieved if the NLD data is used in their raw state since the satellites lack “inflight calibration”. Using the time invariant region approach Jian and Weifeng (2013) calibrated NOAA night lights imagery datasets taken by different satellites in the period 1992-2010 and calculated a and b data adjustment coefficients to address (i) satellite sensor differences; (ii) disparity in data acquisition time that could result in spontaneous oscillation in the data taken by satellites in different orbits; and (iii) the saturation of pixels in urban areas. The study therefore uses the calibration model that is shown below, following Jian and Weifeng (2013).

$$DN_c = a \times (DN_m + 1)^b - 1 \quad \dots (3.1)$$

Where,

a and b are the adjustment coefficients, DN_c is the NLD after calibration and DN_m is the raw NLD.

3.4.2 Land cover data

The two land cover products that are used in this study are NDVI and Landsat images. An assumption was made that we would be able to measure the welfare of small areas by investigating changes in the quality of crops as measured by NDVI, as well as changes in the acreage of land under crops. Changes in the crop acreage and in the quality of the crops should tell us something about the ability of FTLRP’s newly resettled indigenous black farmers to match the intensive cropping systems of the dispossessed former white commercial farmers; and the extent to which they have been able to apply expensive fertilisers and chemicals. By using machine learning techniques to undertake a binary classification of Landsat images into cropland and natural forest, this study is thus able to measure the ratio of land under crops to total land hectarage per ward. This ratio is also used to filter out the natural forest that is captured in the “off the shelf” NDVI raster dataset that is employed in the study as a proxy for the quality of crops.

Stojanova, Panov, Gjorgjioski, Kobler and Džeroski (2010) points out that conventional approaches for the ground measurement and monitoring of vegetation takes time and requires financial resources. Nagendra, Munroe and Southworth (2004) posited that land cover data mostly comprises of snapshots or images of different parts of the earth. However, for the most part the measured pixel value does not have an obvious correspondence to real economic

phenomenon or variables. Therefore, the issue of classification for land cover data [as done in this chapter (and Chapter 2)] assumes a central role if it is to be used meaningfully.

Several methodologies have been proposed that can be used to classify land cover data to its usable form, (see Ahmad, Kalra & Stephen, 2010; Fernandes, 2015; Gislason, Benediktsson and Sveinsson (2006); McIver & Friedl, 2001; Rogan et al., 2008; Shao & Lunetta, 2012; Stojanova et al., 2010). This classification can be supervised or unsupervised. This study employed supervised classification (see Section 3.5.4), following Fernandes (2015).

3.4.3 Normalised Difference Vegetation Index (NDVI)

This study used NDVI and following Ahmad et al. (2010), postulates that higher agricultural productivity per hectare should reflect in higher vegetation quality and higher NDVI and vice versa. As explained by Ahmad et al. (2010), NDVI has been widely employed as a resource to assess ground vegetation cover. Ahmad et al. (2010) explain that the intuition that informs NDVI is the disparity in the reflectance of red and near red infrared frequencies back to the satellite increase as vegetation becomes denser and the index is defined as:

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad \dots (3.2)$$

Where,

NIR and *RED* are the near red and red frequencies respectively, and normalisation of the above expression results in negative NDVI values which represent bare to sparse vegetation, and positive NDVI values representing dense to very dense vegetation (Ahmad et al., 2010).

3.4.4 Classification of Landsat images

Classification is defined as the procedure in which an input image that has multilayers is transformed into a single layer thematic map (Dhumal et al., 2013). Dhumal et al. (2013) point out that satellite multispectral images contain several bands of colour that reveal useful information for classification; adding that the smallest bandwidth contains the finest information about crops. Ustuner et al. (2014) note that classification of remote sensing imagery is the cornerstone of crop monitoring since it gives precise, up to date and less expensive data about crop types and different spatial and temporal resolution.

As mentioned earlier, classification can broadly be categorised into supervised and unsupervised methods. Dhumal et al. (2013), explain that unsupervised classification is when

a researcher groups particular pixels according to similarity and then labels the related land features appropriately. In an unsupervised case, the researcher needs to have prior ground knowledge of the area (Dhumal et al., 2013). This usually involves training a sample dataset based on the researcher's knowledge and then predicting the land features for the rest of the images. The training data is usually obtained from ground truth (physically surveying farm fields and recording the geo-location of different land use classes). The training set for this study was obtained through image expert analysis given the historical nature (1997-2003) of the period of analysis.

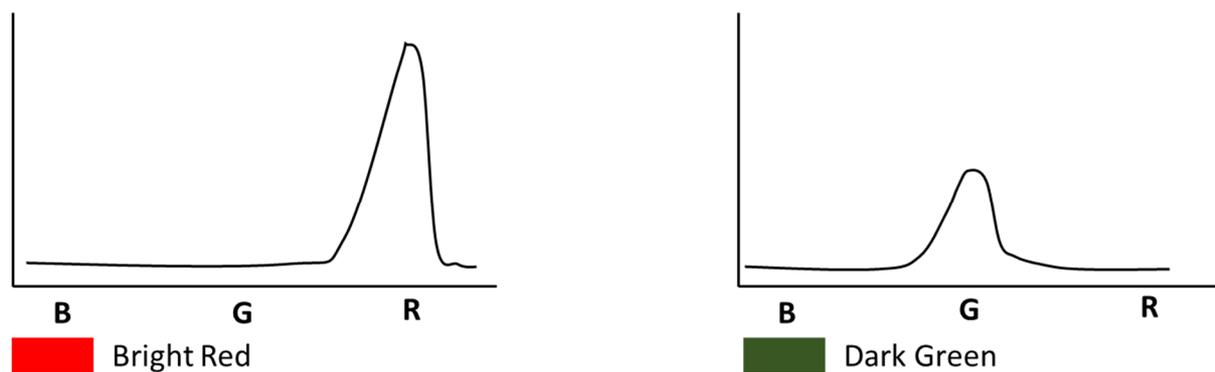


Figure 3.2 Idealised spectral signatures for selected colours

Source: Own illustrations, adapted from Eastman (2003)

Classification of remotely sensed images is made possible by the fact that different objects on the earth's surface have different spectral signatures. Reflection, absorption or transmission are the processes that result when electromagnetic energy reaches a material; and it is the reflection of the sun that is captured by the satellite sensor in remote sensing (Fernandes, 2015). Fernandes (2015) added that the pattern of spectral response pattern (signature) is a description of the extent to which energy is reflected in various areas of the electromagnetic spectrum. This is normally shown graphically as in Figure 3.2, following Eastman (2003). Figure 3.2 shows the signatures for the visual part of the electromagnetic spectrum (Fernandes, 2015). The graph on the left would be a bright red object absorbing the blue (B) and green (G) electromagnetic wavelengths and then reflecting the red (R) (Fernandes, 2015). The object represented by the graph to the right would be a dark green (as suggested by the low graph value) object as it is absorbing blue (B) and red (R) bands and reflecting green (G) back to the satellite sensor (Fernandes, 2015).

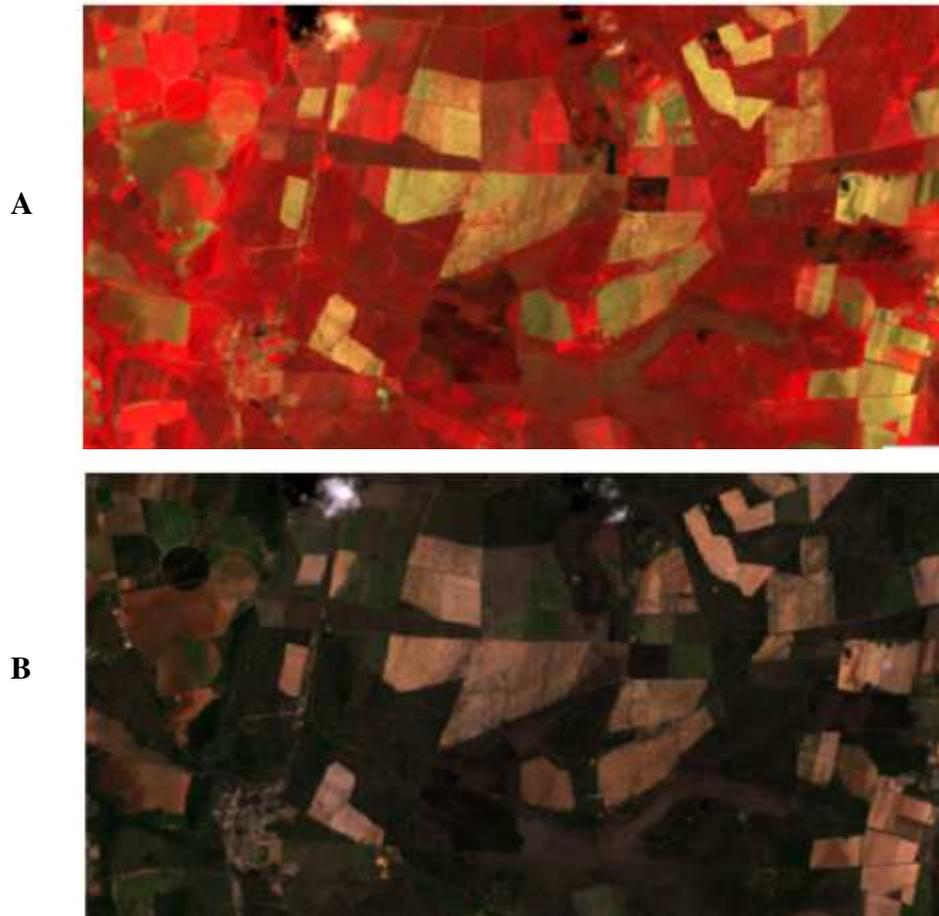


Figure 3.3 Near-infrared and natural colour composites of agricultural land

Source: Own composites created from Landsat footprints for Zimbabwe

It has to be noted that the bands that we can visualise (B, G, R) might not be enough to classify land features on their own – additional bands such as infrared and near-infrared might be necessary (Fernandes, 2015). Figure 3.3 shows near-infrared (Panel A) and natural (Panel B) colour composites created out of Landsat image bands. The images show the corresponding views of fields of farmland in the near-infrared or natural (false) colour composites.

Dhumal et al. (2013) explain that crops have different internal structures according to type and as such, they have different spectral signatures. It follows therefore that different crops with a somewhat similar structure would be much more difficult to distinguish, requiring hyper-spectral imagery that can enable such minute distinction. Although it might potentially be a useful control variable, the focus of this research is not to distinguish the different kinds of crops, as the interest (and capability through image expert analysis in developing training sets) is simply to delineate cropland from natural forest. The study builds on the observation by Fernandes (2015) that crops emit much of the near ‘not-visible to the human eye’ near-infrared

band. This is the centre piece of the classification, although Fernandes (2015) suggests that the best bands for vegetation classification are Blue, Green, Red, near-infrared, shortwave infrared 1 and shortwave infrared 2. In both Landsat 4-5 and Landsat 7 satellites, these bands are 1-5 and 7 as shown in Table 3.3.

Table 3.3 Landsat 4-5 Thematic Mapper (TM) satellite image bands

Landsat 4-5 Thematic Mapper (TM)	Bands	Wavelength (micrometres)	Resolution (metres)
	Band 1 – Blue	0.45-0.52	30
	Band 2 – Green	0.52-0.60	30
	Band 3 – Red	0.63-0.69	30
	Band 4 – Near-Infrared (NIR)	0.76-0.90	30
	Band 5 – Shortwave Infrared (SWIR) 1	1.55-1.75	30
	Band 6 – Thermal	10.40-12.50	120*(30)
	Band 7 – Shortwave Infrared (SWIR) 2	2.08-2.35	30

Source: USGS (2017)

The study considered a narrow period around the FTLRP (1997-2003) and classified the whole of Zimbabwe into cropland and natural forest using 24 image frames/footprints²⁶ for each year. The images were downloaded from the USGS website. The study mostly used data from the Landsat 4-5 satellites and Landsat 7 for the periods 1997-2000 and 2000-2003, respectively. Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+) sensor that is used to acquire images, the only difference between Landsat 4-5 Thematic Mapper (TM) and Landsat Enhanced Thematic Mapper Plus (ETM+) being that the latter has an additional panchromatic Band 8. The study did not use Band 8 to classify the images.

3.4.5 Support vector machines (SVM) classification algorithm

Ustuner et al. (2014) indicated that the SVM algorithm has proven to be superior to other classification algorithms. A Support Vector Machine (SVM) is a statistical learning methodology that is premised on fitting the optimal hyperplane separating the two classes (following Cortes & Vapnik, 1995; Huang, Davis & Townshend, 2002; Ustuner et al., 2014). The SVM is superior as an approach because it uses kernel functions to create an optimal hyperplane that cannot be accomplished linearly (Huang et al., 2002).

²⁶ See Figure B.1 in Appendix B

3.4.5.1 Image correction and supervised training points identification in QGIS

The study obtained cloud-free Landsat images from Jan, Feb, Mar, Nov and Dec for the years 1997-2003 as these months fall within Zimbabwe's rain fed agricultural season, to enhance the accuracy of classification. Some form of correction and calibration is always required for Landsat images (Fernandes, 2015). Images were adjusted for Top of the Atmosphere (TOA) correction in Quantum GIS's (QGIS) Semi-Automatic Classification Plugin (SCP) so that image comparison over time and from different satellites is possible following Congedo (2014) and Fernandes (2015). Congedo (2014) and Fernandes (2015) posit that there is a need to convert the Digital Numbers (DNs) on the image to Top of Atmosphere (TOA) reflectance values, because the electromagnetic energy measured by the satellite sensor is influenced by the scattering and absorption atmospheric effects; reflectance being the ratio between reflected and incident energy on a surface (Congedo, 2014). More than 20 000 training polygons (translating to hundreds of thousands of training pixels) for all the footprints were identified in QGIS. The study identified at least 150 training polygons per Landsat footprint (see Appendix B, Figure B.1).

For NLD and NDVI, the images were imported to QGIS, where zonal statistics were computed for every ward, based on the Zimbabwe Level 3 ward level shape file. The shape file was exported to STATA for analysis. The crop cultivation data derived from Landsat has a 30 m resolution. To filter out natural forest in the NDVI, the Landsat data was downscaled to the resolution of the NDVI. Natural forest and land under cultivation in the Landsat data were assigned values of 0 and 1 respectively so that the product of NDVI and the downscaled crop cultivation data could give natural forest filtered NDVI raster images ready for data extraction and analysis.

3.4.5.2 Classification using the SVM Algorithm in R

Machine learning classification proceeds in R, following Fernandes (2015). The SVM algorithm was used for classification (Cortes & Vapnik, 1995; Huang et al., 2002; Ustuner et al., 2014). The two main factors affecting the quality of classification *ceteris paribus* was the selection of the classification algorithm and the training set (Machová, Barcak & Bednár, 2006). Therefore, it was necessary to ensure that the correct algorithm was selected as well as the right training dataset.

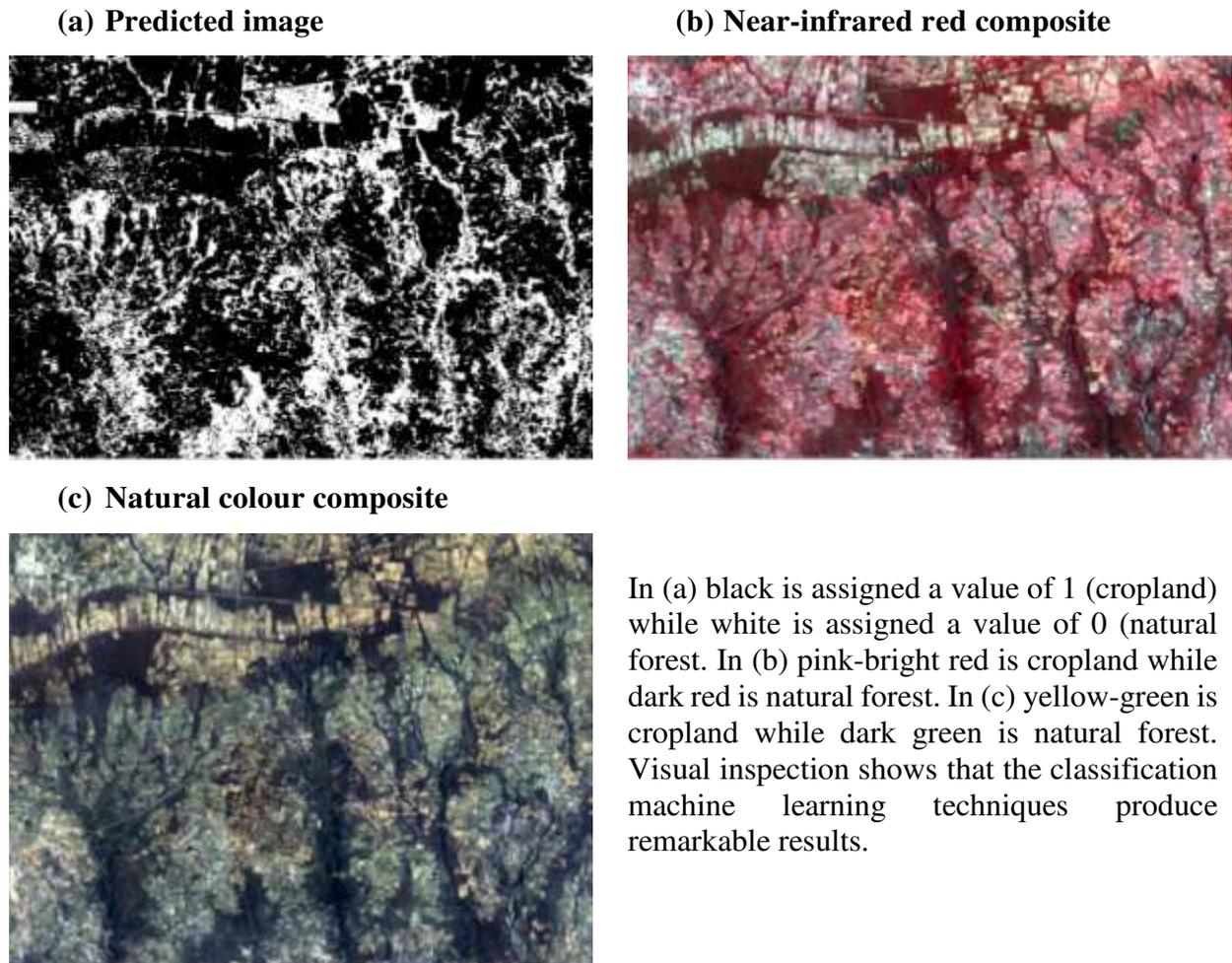


Figure 3.4 Performance of the SVM algorithm in R

Source: Based on USGS Landsat images

The high number of training pixels used in the study guaranteed high classification accuracy (see Appendix B – “Classification” sub-section for Kappa classification accuracy metrics), but they sacrificed computing speed. Classification produces predicted raster images and these are shown side to side with near-infrared and natural colour composites for the same area in Figure 3.4. The zonal (ward) mean was computed in R using the predicted raster images (where cropland pixels=1 and natural forest pixels=0). The derived mean is the ratio of area of crops under cultivation that was used as the endogenous variable in the analysis side to side with NLD and NDVI.

3.4.6 Difference-in-Difference identification (DID)

The study proceeded by following a DID approach (see Bertrand, Duflo and Mullainathan (2002), Bezabih et al. (2011), Muchomba (2017), Lechner (2010) and Wooldridge (2007)), taking the pre-2000 and post-2000 NLD and NDVI as the ‘before’ and ‘after’ land reform

periods, respectively. Land reform was launched in 2000. Various treatment groups were constructed. First, all wards that are located in EAs (see Figure 3.1) were regarded as the Treatment Group or area where land reform took place, while the Tribal Trust Lands (TTLs) (the areas reserved for indigenous Zimbabweans) were considered as the control group where reallocation of land did not take place. Second, the TTLs were replaced by the Native Purchase Areas (NPAs) as a control group. NPAs were also black farming areas. Third, EAs in the rest of the country, excluding the Eastern Highlands, were regarded as the Treatment Group while those located in the region were taken as the control group; following the assertion by Mutangi (2010) that land reform excluded this region.

Land reform targeted only the EAs while the TTLs and the NPAs that were reserved for blacks were not targeted by the programme. The Eastern Highlands were targeted for land reform, but was never pursued in this region because of the unsuitability for conventional crop farming. Therefore, in the spatial analysis the Tribal Trust Lands (TTLs), Native Purchase Areas (NPAs) and the Eastern Highlands (EHs) were used as the control groups. The model was specified following Ravi, Kapoor and Ahluwalia (2012) as follows:

$$Y_{it} = \alpha + \beta_1 Treat + \beta_2 Post + \delta Treat * Post + \beta_{3,\dots,n} X_{it\gamma} + D_i + \epsilon_{it} \quad \dots (3.3)$$

Where,

Y_{it} is the outcome of interest in ward i at time t which is represented by the ratio of crop cultivation, NLD and NDVI. $Treat$ is the Treatment Group dummy variable which equals 1 for regions targeted by land reform and 0 otherwise. $Post$ is the year dummy, which equals 0 before 2000 and 1 afterwards. X_{it} is a vector of time varying ward characteristics specified after (Garrison, 1982) as:

$$X_{it\gamma} = Population, Crops Trade, Rainfall, Temperature, \frac{Calories}{hectare}, Distance to Border \quad \dots (3.4)$$

And D_i is a vector of time-invariant FEs, specified in 3.5 as:

$$D_i = Agro - ecological Region, Frame \quad \dots (3.5)$$

Using the DID method, the coefficient of interest was δ and its OLS estimates measured the

causal effects of FTLRP on the outcomes of interest, which are changes to crop cultivation, NLD and NDVI.

$$\delta = (Y_{2002}^T - Y_{2002}^C) - (Y_{1999}^T - Y_{1999}^C) \quad \dots (3.6)$$

3.4.7 Regression Discontinuity Design identification (RDD)

The RDD is a pseudo-experimental design in which the chance of being treated changes discontinuously/dramatically as a function of at least one underlying attribute/variable (Hahn, Todd & Van der Klaauw, 2001). By taking advantage of discontinuity in the process of assigning treatment (Black, Galdo & Smith, 2005), the RDD mimics randomised experiments (Cook & Wong, 2008) and can thus be used to estimate causal effects of Zimbabwe's FTLRP: with the observable variable being distance to border of the treatment region from each ward centre. Following Dell (2010), this chapter takes advantage of the Land Apportionment Act's (1930) spatial variation in the assignment of land rights. As in Dell (2010), FTLRP in Zimbabwe was implemented only in EAs (see Figure 3.1), hence areas following outside the boundaries of EAs were not affected, which suggests a Regression Discontinuity Design (RDD) using distance to the boundary of the EAs region as the running variable. Following Dell (2010), we modelled the RDD specification as follows:

$$\begin{aligned} Y_{wt} = & \alpha_1 + \phi Post + \gamma Treat_w + \delta Treat_w \times Post + X'_{wt} \alpha_{2,\dots,n} \\ & + f(Euclidian Distance_w) \\ & + \mu_{wt} \end{aligned} \quad \dots (3.7)$$

Where,

Y_{wt} is the outcome variable of interest denoting either ratio of land under crops, NLD or NDVI in Ward w at time t ,

$Treat_w$ is a land reform treatment indicator taking the values of 1 and 0 for area affected and those not affected respectively,

X'_{wt} is a vector of control variables similar to $X_{it\gamma}$ in the DID specification,

$f(Euclidian Distance_w)$ is the distance from the centroid of each observed ward to the border of the treatment region,

α_1 is a regression intercept, γ is the treatment coefficient and $\alpha_{2,\dots,n}$ is the set of coefficients for the covariates, and

μ_{wt} is a stochastic regression error term.

Essentially this specification extends the difference-in-difference equation by controlling for the running variable. γ measures the regression discontinuity before the FTLRP, and should preferably be zero to motivate the approach. δ measures the change in the discontinuity relative to the pre-reform period, and is the coefficient of interest. As is standard in this literature, regression estimates are conducted in regions as close to the border as possible. We show estimates at 1km, 2km and 5km.

3.5 Results and Discussion

3.5.1 Control variables

The study controlled for several variables, namely Caloric Suitability Index, gridded rainfall and temperature, ward population, regional imports and exports, disease environment (following Bloom et al., 1998²⁷) and Euclidian distance to the border of the treatment region. This section briefly discusses the motivation for including these variables in the regressions, the sources from which they are obtained and how they were incorporated in the analysis. As mentioned previously, NLD is a good proxy for economic activity and development patterns, which includes population growth. Shipton and Goheen (1992), Jayne, Chamberlin and Headey (2014), and Chingozha and von Fintel (2019) have observed that high population density in customary areas translates to competition over land use – highlighting the need to control for population. Therefore, ward level population figures obtained from the 1992, 2002 and 2012 Zimbabwe population censuses were incorporated in the regression to remove the effects of population density. Following Besley and Burgess (2000) we estimated the intermediate years' ward population on the assumption that the yearly growth rate was uniform or constant.

Apart from that, the study used the Caloric Suitability Index, following Galor and Özak (2016). They argue that historic agro-climatic conditions have influenced the pace of economic development, thus we incorporated the Post-1500CE Caloric Suitability Index in order to estimate the extent to which these climatic conditions have affected crop cultivation, crop quality and welfare in Zimbabwe.

²⁷ Initial climatic conditions and the disease environment may be one of the factors that have influenced the spatial variation of contemporary development (Bloom et al., 1998).

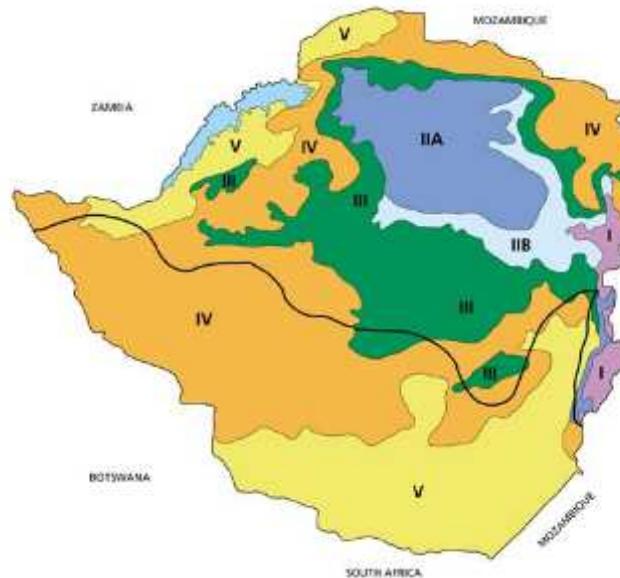


Figure 3.5 Natural regions of Zimbabwe

Source: FAO (2016)

The caloric index captures the spatial difference in potential agricultural yield in terms of calories per hectare (Galor & Özak, 2016). It is incorporated in our analysis to show that there was no selection effect in the implementation of land reform in Zimbabwe. There is a potentially valid argument that FTLRP took place in the areas that had good soil quality and soil potential, thereby affecting the randomization of assignment in our identification. To isolate this effect, we control for the caloric soil suitability index, and other covariates. To capture the effect of the disease environment, we also included a tsetse fly suitability index developed by Rogers and Williams (1994). Acemoglu, Jonson and Robinson (2001) noted the importance of the disease environment in explaining contemporary development. Another important control variable is distance from each ward to the border of the treatment region. We controlled for distance in both the DID (to capture potential spill overs) and as the running variable in our RDD estimates.

The timing of FTLRP coincided with the 2001/2002 drought. In order to separate the effect of drought from that of FTLRP, the study obtained gridded temperature and rainfall data created by Willmott and Matsuura (2015). It also included logged regional imports and exports data in order to separate the confounding effect of ‘Black Friday’.²⁸ We used the trade in crops as a

²⁸ In 1997, the Zimbabwe government was under immense pressure to compensate war veterans for liberating the country. As a result, Zimbabwe paid ZW 50 000,00 to each and every individual who could prove that they had

proxy for the exchange rate, whose downward spiralling was triggered by ‘Black Friday’. We obtained a longitudinal regional dataset of import and export values and quantities from FAOSTAT (2015).

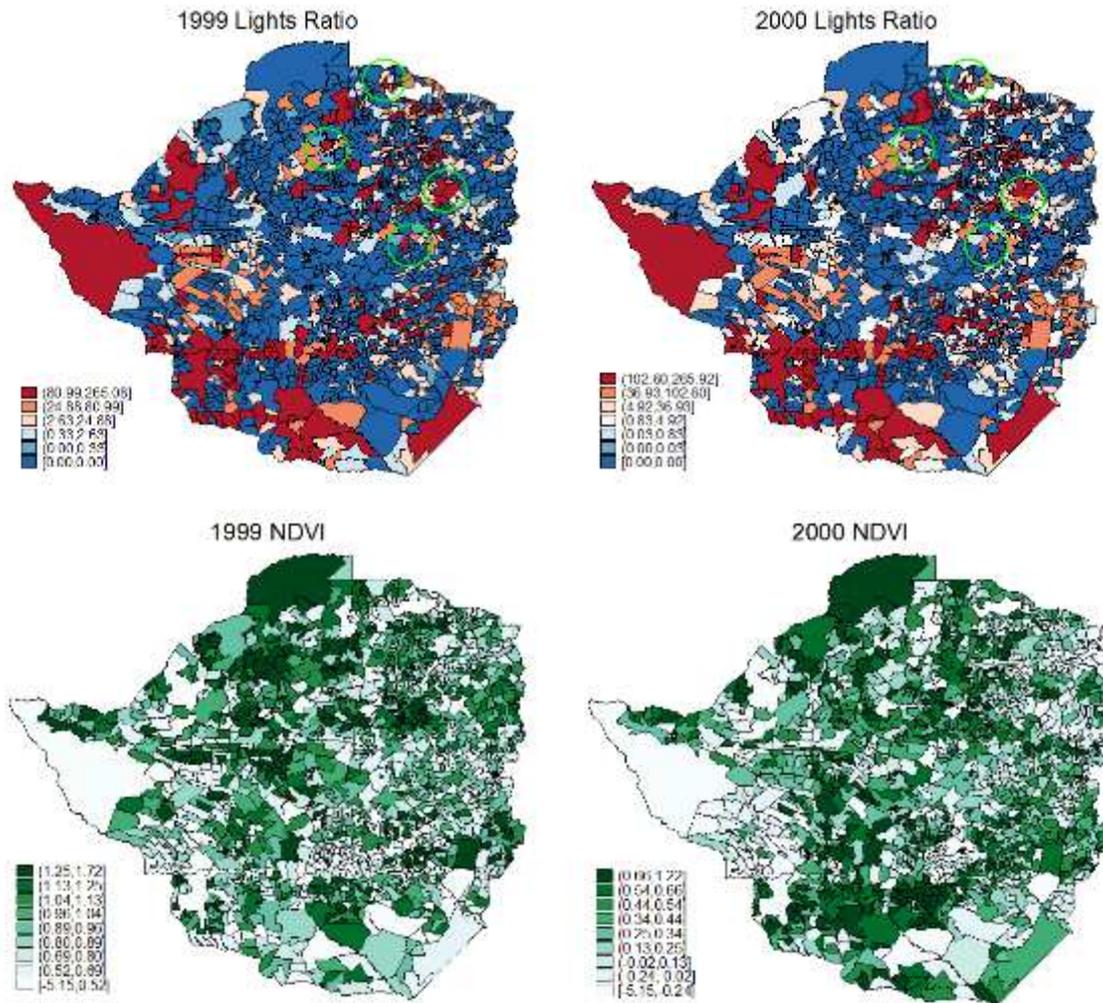


Figure 3.6 Comparing 1999 versus 2000 (NLD and NDVI)

Source: Own illustration

To convey regional variation to the data, we mapped different crop exports and imports to different regions using Figure 3.6. FAO (2016) indicates the different types of crops produced in each region. Based on that, we assigned different crops to different regions (without

played an active role during the war at different levels ranging from informer, collaborator to actual combat. On 14 November 1997, the announcement of these unbudgeted payments resulted in the Zimbabwe dollar losing 72% of its value and the Zimbabwe Stock Exchange (ZSE) fell by 46% as foreign investors scurried out (Bond, 1999; Maravanyika, 2007).

overlaps) as shown in Table A.5 in Appendix A and then added the different individual crop import or export values per natural region to arrive at the regional figures.

The data was then overlaid with the Zimbabwe Level 3 shape file so that the import and export data vary by ward. After ‘Black Friday’, both imports and exports correlate with the exchange rate (see Table A.6 in Appendix A). The reduced exports signify the negative performance (or welfare) of a particular natural region while increased imports also indirectly imply produce which would otherwise have been produced by a particular region prior to ‘Black Friday’. We regressed our three endogenous variables, namely NLD, NDVI and crop intensity (crop/total ward area) in turn. These correlate well with 2011 Census poverty estimates as shown in Table 3.4.

Table 3.4 Correlation between endogenous variables and poverty estimates

Variable	r (overall)	r (rural)
Night Lights (NLD)	-0.6663	-0.3025
NDVI	0.1483	0.0985
Crop Intensity	-0.0490	0.1468

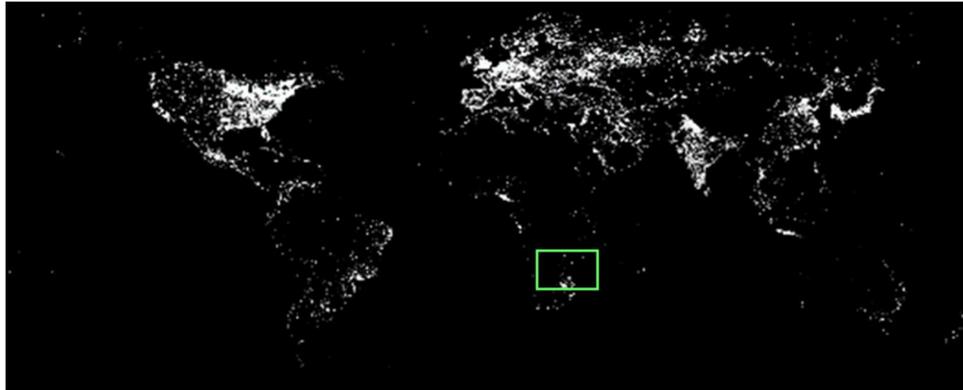
Table 3.4 shows a higher overall negative correlation between NLD and poverty. However, this relationship is weaker in rural areas due to lower access to electricity. There is a positive correlation between NDVI and welfare because farming areas usually have a higher incidence of poverty. By looking at the changes in NDVI and crop intensity over time, we were able to track changes in welfare at the small area level.

3.5.2 Descriptive data analysis

Raster images for Night Lights Data (NLD) and Normalised Difference Vegetation Index (NDVI) are shown in Figure 3.5. NLD has far less variation than NDVI, which makes the latter richer. The green and red boxes in Figure 3.5.1 show Zimbabwe’s approximate location. NDVI is approximately symmetrically distributed and therefore not transformed. NLD, on the other hand, is skew to the left, making a log transformation appropriate. However, a large number of observations is zero, which are discarded in estimation. Results are robust to other approaches

that are not shown (such as relying on Tobit regressions that incorporate bottom-censored zero observations).

The World at Night (Night Lights Data)



NDVI image of the World

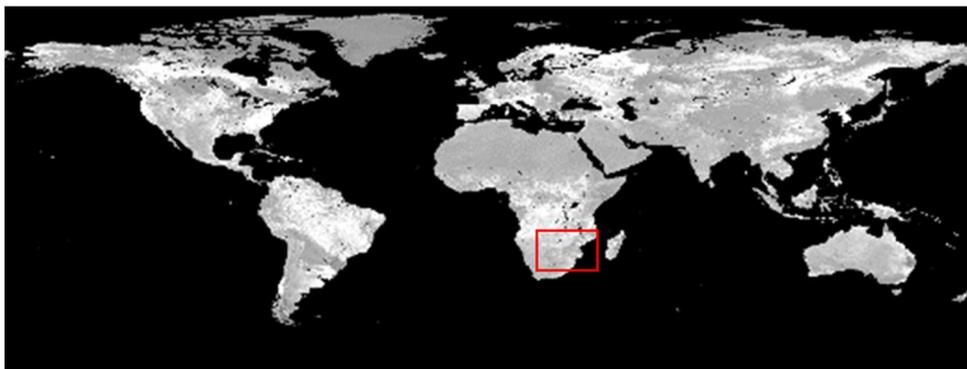


Figure 3.7 Raster Images for NLD and NDVI

Source: Using NOAA NLD and NDVI NCR

Table 3.5 presents additional information from descriptive exploration of the data for the pre and post phases of FTLLRP. In both pre and post periods Table 3.5 shows that TTAs generally had a higher proportion of land under cultivation than both EAs and NPAs. EAs, TTAs and NPAs had crop cultivation ratios of 0.41, 0.56 and 0.49 respectively while these were 0.25, 0.52 and 0.37 in the post phase (also respectively). These differences (by land class) are statistically significant as shown by the F-tests. It is also evident in Table 3.5 that all the three land classes experienced some decline in the ratio of land under cultivation in the post phase – although TTAs experienced the least decline. The intuition behind such a result the fact TTAs were overcrowded hence there was higher competition for land use – with agriculture making up the largest share of land use for consumptive purposes. In Table 3.5, EAs had a higher population than TTAs, with a mean ward population of 6730 and 7110 in the pre and post

phases respectively. The average ward population figures for TTAs were 3750 and 4000 for the pre and post phases respectively, while those for NPAs were 3140 and 3320 (also respectively). It is worth mentioning that TTA wards were significantly much smaller relative to their EA counterparts.

Table 3.5 Descriptive Data Averages

Class	Cultivation		NDVI		NLD		Population (000)		CSI (000)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
EAs	0.41	0.25	629.58	502.4	0.78	0.66	6.73	7.11	3.55	3.55
TTAs	0.56	0.52	841.36	874.8	0.13	0.11	3.75	4	3.50	3.50
NPAs	0.49	0.37	538.19	541.1	0.06	0.05	3.14	3.32	3.63	3.63
F-Test	105	629.4	74.1	137.5	233	226.8	329.2	312.2	11.8	11.78
<i>p</i>	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***

Notes*: Class = land class, Cultivation = proportion of land under cultivation, NDVI = Normalised Difference Vegetation Index, NLD = Night Lights Data, Population = ward population and CSI = Caloric Suitability Index

TTAs had the highest mean ward NDVI in both the pre (841.36) and post periods (874.8), followed by EAs (629.58 – pre and 502.4 -post) and NPAs (538.19 – pre and 541.1 - post). While NDVI was used in the study as a proxy for crop quality, differences in crop type between TTAs and EAs, and between TTAs and NPAs may also influence the ward mean NDVI. On average EAs has high night lights (NLD) than both TTAs and NPAs, and noticeable declines are evident in the post phase for all the three land classes.

Table 3.5 Descriptive Data Averages (continued...)

Class	Rain		Temperature		TSI		Exports (US\$000)		Imports (US\$000)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
EAs	847.94	699.55	27.85	30.45	0.07	0.07	228.13	217.8	224.49	52.74
TTAs	824.41	679.17	28.14	30.54	0.08	0.08	101.70	96.26	10.32	24.88
NPAs	842.36	693.36	28.02	30.31	0.12	0.12	116.26	109.9	12.10	27.78
F-Test	6.79	4.91	2.74	5.27	37.04	37.04	160	142.8	168.6	80.64
<i>p</i>	0.0***	0.01**	0.060	0.01**	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***

Notes*: Class = land class, Rain = Rainfall, TSI = Tsetse-fly Suitability Index, Exports = value of crops exported and Imports = value of crops imported

The ward average for CSI, total rainfall and mean temperatures were more or less the same across the three land classes (EAs, TTAs and NPAs). CSI and TSI remained the same across

the pre and post periods because they do not vary over time. However, there was an increase and decline in temperatures and rainfall across the three land class – indicative of the drought that affected the Southern African region around the year 2000. Unlike the other variables in Table 3.5, exports and imports did not vary by ward. As discussed earlier, exports and imports varied by agro-ecological region. Nevertheless, the information shows that EAs accounted for about half of the country's exports. In general exports declined while imports increased (though imports decline for the EAs).

3.5.3 Discussion of results

In Table 3.6 (DID estimates) and Table 3.7 (RDD estimates), Columns (1)-(3), (4)-(6) and (7)-(9) present estimation results for NLD (lights), Crops (ratio of land under crops) and NDVI, respectively. Columns (1), (4) and (7) show results for which TTAs are regarded as the control group while EAs are the Treatment Group. In (2), (5) and (8), NPAs and EAs are the control and treatment groups respectively. In (3), (6) and (9), European Areas elsewhere in the country (EAs-Else) are the control group while the European Highlands are the treated group. The multiple treatment and control groups ensure robustness of findings.

In all NLD specifications in all tables, columns (1)-(3) show that the coefficient of *Treat x Post*, the causal effect, is not statistically significant.²⁹ This is true with and without controls and also accounting for bottom-censoring with a Tobit regression (this is not shown). Therefore, an important secondary conclusion emanating from these results is that NLD may not be a viable data source in analysis measuring socio-economic development in *rural* SSA since many of these areas lack access to electricity.³⁰ The focus of the discussion therefore relies on results for crop cover and quality.

Columns (4)-(6) of Table 3.6, show DID estimates for crop cover. The causal effect of the FTLRP in Column (4) [for Tribal Trust Areas (TTAs) as control group] is a 9 percentage point

²⁹ RDD estimates deliver similar conclusions (1 km buffer results shown in Table 3.7 as well as 2 and 5 km buffers as shown in Tables A.3 and A.4)

³⁰ Existing research shows that night lights proxies are top-coded in industrial countries – hence understanding the underlying measured (that night lights proxies) since the maximum value of night lights does not exceed 63 no matter how high the light intensity is (Kulkarni, Haynes, Stough & Riggle; 2011). On the other end of the spectrum, very low levels of light are recorded as zero (the minimum value for night lights data). This “bottom coding” may exist due to little/no population (Lowe, 2014). However, in rural SSA this is due to lack of electricity access and very little has been done to analyse the limitations of bottom-coding in the rural parts of developing countries.

reduction in land under cultivation, significant at one per cent. In Columns (5) and (6), the causal effects are a 5 and 30 percentage point drop relative to NPAs and European Highlands, respectively. The coefficients for the former and latter are statistically significant at 5% and 1%, respectively. Hence, the difference in results between the sets of control and treatment groups demonstrates the negative effects of lack of property rights enforcement on the ratio of land under cultivation. Results remain robust to adding controls to all DID specifications in the bottom of the panel.

Columns (7)-(9) of Table 3.6 present estimates for vegetation quality/NDVI. Baseline DID specifications and those with full controls are robust. Coefficients of *Treat* x *Post* for TTAs and European Highlands are -261.6 and -455.8³¹ respectively and statistically significant at 1%. The effect is not significant using NPAs as control group. NPAs were also indirectly affected by FTLRP. Their markets were integrated with those of affected EAs and farmers relied on their marketing and supply chains. The with-controls specifications provide coefficients of similar results. NDVI (crop quality) reduced in the EAs that were affected by FTLRP. Newly resettled indigenous farmers did not receive title, and could therefore not access credit that provided access to expensive inputs such as chemicals and fertilisers (unlike the well capitalised white farmers who had previously farmed in these areas).

Turning to RDD estimates, impacts on the ratio of land under cultivation remain robust [using buffers of 1 km (Table 3.7), 2 km (Table A.3) and 5 km (Table A.4) around the border]. This is true for both the baseline and with-controls specifications shown in Column (4) of Table 3.6. The coefficient on *Treat* is insignificant using the narrowest buffer, highlighting that there was no discontinuity before the FTLRP and supporting this approach. However, the border discontinuity emerged after the reforms, as shown by the coefficient on *Treat* x *Post*. Relative to TTAs, cropland coverage reduced by 15 percentage points at a 10% level of significance. At all bandwidths, results for the post-reform discontinuity are not statistically significant when NPAs [Column (5)] and European Highlands [Column (6)] are each considered as the control group.

³¹ NOAA CDR NDVI takes minimum and maximum values of -9998 and 9998 respectively (<https://tinyurl.com/y37lyja8>).

Table 3.6 DID regression estimates

	<i>Lights</i>			<i>Crops</i>			<i>NDVI</i>		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control Group	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands
Treat	0.36*** (0.085)	0.71** (0.315)	-0.66* (0.347)	-0.18*** (0.009)	-0.047*** (0.017)	0.10** (0.041)	-228.2*** (17.536)	-45.4 (46.088)	128.4 (79.051)
Post	-0.32*** (0.095)	-0.79** (0.393)	-0.23 (0.343)	-0.055*** (0.007)	-0.10*** (0.021)	0.12*** (0.047)	339.4*** (14.819)	184.8*** (57.168)	516.8*** (89.360)
Treat x Post	0.0041 (0.121)	0.46 (0.401)	-0.10 (0.352)	-0.090*** (0.013)	-0.047** (0.023)	-0.28*** (0.048)	-261.6*** (25.307)	-102.8 (63.088)	-455.8*** (91.252)
Controls	N	N	N	N	N	N	N	N	N
<i>N</i>	3156	2032	1955	7316	3020	2455	7164	2908	2367
<i>R</i> ²	0.172	0.177	0.166	0.359	0.254	0.256	0.317	0.142	0.179
Treat	0.74*** (0.081)	1.05*** (0.287)	-0.031 (0.361)	-0.16*** (0.009)	-0.037** (0.017)	0.17*** (0.046)	-195.8*** (17.602)	-20.6 (45.179)	196.5** (86.964)
Post	-0.18* (0.097)	-0.48 (0.365)	-0.20 (0.310)	-0.092*** (0.009)	-0.11*** (0.022)	0.12*** (0.046)	243.0*** (17.565)	148.1** (58.457)	491.6*** (87.495)
Treat x Post	-0.0100 (0.113)	0.32 (0.367)	0.017 (0.317)	-0.10*** (0.013)	-0.054** (0.023)	-0.31*** (0.047)	-288.2*** (25.133)	-153.0** (61.646)	-505.5*** (89.288)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
<i>N</i>	3044	1976	1901	7015	2937	2383	6873	2835	2305
<i>R</i> ²	0.309	0.323	0.315	0.394	0.270	0.277	0.363	0.154	0.207

Notes*: In (1), (4) and (7) EAs are the Treatment Area while TTLs are the Control Area. In (2), (5) and (8) EAs are the Treatment Area while NPAs are the Control. In (3), (6) and (9), EAs Elsewhere are the Treatment Areas while European Highlands are the Control Area.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7 RDD regression estimates (1 km buffer from treatment region border)

	<i>Lights</i>			<i>Crops</i>			<i>NDVI</i>		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control Group	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands
Treat	1.10 (3.053)	-7.53*** (1.583)		-0.11 (0.081)	-0.00047 (0.204)	-1.05 (0.671)	-158.4 (259.723)	575.6 (400.415)	
Post	-0.13 (0.489)	0.53 (0.660)	-0.34** (0.167)	0.038 (0.073)	-0.11 (0.122)	-0.25 (0.328)	259.3** (120.976)	333.4* (200.595)	-56.3 (71.934)
Treat x Post	-0.22 (0.518)	-0.86 (0.679)		-0.15* (0.085)	0.0050 (0.128)	0.13 (0.331)	-378.7** (144.918)	-417.1* (215.727)	
Controls	N	N	N	N	N	N	N	N	N
<i>N</i>	132	126	130	255	204	193	155	127	120
<i>R</i> ²	0.774	0.800	0.772	0.364	0.379	0.358	0.550	0.417	0.488
Treat	14.3* (8.125)	-13.9*** (3.522)		-0.13 (0.096)	-0.11 (0.223)	-1.10 (0.745)	-100.9 (330.459)	241.9 (491.049)	
Post	0.13 (0.512)	0.33 (0.657)	0.011 (0.216)	0.029 (0.074)	-0.10 (0.118)	-0.15 (0.315)	214.5 (131.807)	292.3 (210.496)	-44.1 (93.483)
Treat x Post	-0.088 (0.525)	-0.29 (0.674)		-0.15* (0.080)	-0.019 (0.121)	0.022 (0.314)	-349.0** (145.927)	-409.1* (218.005)	
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
<i>N</i>	122	116	120	242	192	181	151	124	117
<i>R</i> ²	0.798	0.842	0.810	0.471	0.463	0.430	0.567	0.443	0.521

Notes*: In (1), (4) and (7) EAs are the Treatment Area while TTLs are the Control Area. In (2), (5) and (8) EAs are the Treatment Area while NPAs are the Control. In (3), (6) and (9), EAs Elsewhere are the Treatment Areas while European Highlands are the Control Area.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The RDD estimates also confirm a reduction in crop quality in TTAs and NPAs after FTLRP.³² Again, no significant discontinuities arose for the period before the reforms, regardless of the control group. The post-reform TTA and NPA border discontinuity is statistically significant in Columns (7) and (8) using a buffer of 1km (Table 3.7). With buffers of 2km (Table A.3) and 5km (Table A.4) only results for TTAs are statistically significant. Estimates relative to TTAs are therefore most robust. For NPAs, the mixed effects show that these regions were affected in similar ways to EAs. NPA farmers furthermore enjoyed title since the 1930 LAA. Even though these regions were not directly targeted by the reforms, they have historically been integrated into the same markets as EAs and have therefore experienced negative spill over effects from affected areas.

Across the different treatment and control groups, our results may³³ confirm that lack of follow-through in tenure security enforcement post-FTLRP had a negative effect on crop cultivation and crop quality. The RDD estimates provide better³⁴ estimates compared to the DID, on account of a better way of assigning treatment (through distance to border). Our results appear to be consistent with Besley and Burgess (2000), as well as Stoeffler et al. (2016). Our analysis was unable to establish changes with respect to welfare, since NLD proved to be of little usefulness in measuring the impacts of land rights in rural SSA due to limited observations (poor access to electricity). More generally, our main results (the RDD estimates) do not show improvements post-FTLRP, both in proportion of land under cultivation and crop quality. There are several differences between our approaches. Stoeffler et al. (2016), did not employ pre-intervention data. Keswell and Carter (2014) have likely inflated their estimations by excluding farmers whose characteristics do not show success potential. In addition, they analysed South Africa's Land Redistribution for Agricultural Development (LRAD) programme, which was implemented based on a 'willing buyer willing seller' basis, unlike FTLRP. Further, the South African government had more resources to support its newly resettled farmers, and LRAD farmers were given land titles. Perhaps most important, our endogenous variable is not the same. We used area level proxies (NDVI and crop activity),

³² Results for the Highlands control group do not estimate with narrow bandwidths, as insufficient treatment centroids are located very close to these borders.

³³ We maintain that tenure security post-FTLRP is the main driver of our results, but the potential effect of other confounding factors such as access to extension services and inputs cannot be discounted.

³⁴ Across the regressions, RDD provides higher explanation in endogenous variable variation. Also for the most part, the RDD estimates show absence of discontinuity before FTLRP.

³⁵ Results for the specification with District FEs is shown in the Appendices.

while in these studies and others, there is reliance on actual household data. We therefore miss significant farmer-level heterogeneity, but present a nationwide view.

3.6 Caveats, Assumptions and Robustness Checks

Before outlining the conclusions of the chapter, it is important to discuss some of the caveats and assumptions underpinning the results. In the same vein, it is also critical to present some analysis and arguments regarding the robustness checks/tests that were conducted for this chapter. Regarding model specification, we introduced several FEs (namely Agro-ecological region, Frame as well as District). All the FEs capture time-invariant location specific differences that may be potentially confounding on the results. Without including the FEs, our results would have some substantial omitted variable bias. In terms of specifics, the Agro-ecological FEs capture the different climatic conditions across five regions, whereas the Frame FEs account for sensor specific differences among the 24 Landsat image Frames that we stitch together to obtain nationwide data. The District FEs are useful in allowing us to account for compound treatment effects in the RDD estimation (*see* Appendix D) for more details.

The quasi experimental models (DID and RDD) adopted in this chapter rely on the validity of certain assumptions underpinning them. Hence in Appendix D we present results and discussion of various robustness checks/tests regarding the validity of the DID and RDD assumptions. Following Abadie (2005) we implemented a semi-parametric procedure to match sub-samples in our data where the parallel trends DID identifying assumption hold. After matching, the parallel trends assumption was only valid when we used NDVI as the endogenous variable, with the treated and untreated regions being EAs Elsewhere and EAs Eastern Highlands respectively. Figure D.1 in Appendix D shows the validity of the parallel trends assumption for the DID estimator for this region. The figure shows that the hypothesis that the “difference in difference” (the causal effect) is different from zero is not rejected before the 2000, whereas the hypothesis is negative and statistically after 2000 (after FTLRP).

The analysis evolved to the RDD as a direct result of the fact that the parallel trends assumptions only held for NDVI for EAs Elsewhere (treatment) and EAs Eastern Highlands (control) as shown in Figure D.1. However, as discussed in much greater detail in Appendix D, the RDD assumptions of continuity at the border, absence of compound treatment and pre-treatment covariance largely hold for the RDD estimates. In Appendix D, we also implemented the RDD estimator using the *rdrobust* and *mdrd* packages and the results were robust to the

RDD main estimates shown in the chapter in Table 4.7. The *rdrobust* and *mdrd* packages also dealt effectively with standard error asymptotic bias of the non-parametric polynomial estimator, and also used a data driven approach to select the optimum distance bandwidths. Standard error asymptotic bias can also be addressed by adopting narrow distance bands from the border (*see* Keele and Titiunik, 2014) and our 1km, 2km and 5km distance bandwidths were appropriate in that regard. After these and other robustness checks/tests presented in Appendix D, the general negative impact of FTLRP on crop cultivation and crop quality remain robust.

3.7 Conclusion

The study contributes to the debate around the efficacy of land reform in developing countries, given its importance and strong link to poverty reduction efforts. Even more importantly, it may highlight the significance of ensuring that private property rights are enforced in the wake of land redistribution/reform to ensure that there is effective market participation for the beneficiaries. The study found that Zimbabwe's FTLRP had a negative effect on crop quality (NDVI) and the amount of land under crops. Given the huge amount of racial and diplomatic polarisation FTLRP created, it can be viewed as a land conflict; losses in productivity can therefore be interpreted as the result of conflict, as illustrated by Deininger and Castagnini (2006). The discussion around conflict is important, especially given the violent and involuntary nature that characterised the dispossession. As a result, this may have influenced withdrawal of various forms of support [including the banning of Zimbabwean produce in the European Union (EU)] export market. In addition, the absence of proper training and mechanization on the part of the newly resettled farmers might also have influenced the decline in crop cultivation and crop quality. At the same time the effect of absence of land titles post-land reform cannot be discounted and is overarching. Also, the absence of compensation in the land redistribution exercise meant that the social cost was not adequately internalised. Without ward level data on other important factors such as access to extension services, levels of mechanization and agricultural training, it is difficult to attribute the decline in cultivation and crop quality solely to the poor property rights enforcement – although this is an important factor nevertheless.

Previous studies focusing on the effects of land reform in Zimbabwe have not examined the issue within a robust quasi-experimental design. This study used the DID and RDD approaches

to appropriately measure the effect of the lack of tenure security enforcement post-land reform on cultivation and crop quality by constructing various treatment and control groups in order to ensure that there is effective selection. The availability of data to fully comprehend the effects of land reform in Zimbabwe, as in any other developing country, is an important constraint that would inadvertently have affected the extent to which more empirical work could be done to understand the efficacy of FTLRP. The study exploits the advantages of remote sensing data as proxies for agriculture production, welfare and crop quality, and models these respectively using data derived from the classification of Landsat images using machine learning techniques, NLD and NDVI as endogenous variables. From our analysis it is clear that NLD suffers from potential bottom-censoring problems because it has low explanatory power when it comes to the analysis of economic phenomenon in a rural developing country. The study therefore also makes an important data contribution.

CHAPTER 4

A CITIZEN SCIENCE APPROACH TO CLASSIFYING URBAN INFORMALITY AND OTHER URBAN LAND USE TYPES USING SATELLITE IMAGERY

4.1 Introduction

De Soto (2001)'s seminal work argues that poor people in developing countries lack the ability to create capital not because of a lack of assets but because of the defective nature of these assets – the main defect being poorly defined property rights in informal settlements. Dwellers of informal settlements lack the land rights that would otherwise enable them to improve housing quality but also use their assets as collateral to access financing (De Soto, 2001) - hence their inability to create capital. Marx, Stoker and Suri (2013a) show that poorly defined/enforced property raise the incidence and occurrence of rent extraction. Using the Kibera slum of the Kenyan capital Nairobi as a case study, they found that tenants paid higher rents and invested less in the quality of their dwellings if they belonged in the same tribe with the local chief and vice-versa. Marx et al., (2013a) add that tenants were better off in areas of high unemployment since in these areas young gangs enforced tenancy rights. This sets in motion the argument why property rights are critically important within an urban discourse – despite their obvious importance in assisting service provision planning and delivery on the part of the local authorities as already discussed.

New “modernisation” theory (see Marx, Stoker and Suri, 2013b; Glaeser, 2013) portray slums/informal settlements as a necessary and transitory state of the development progress in fast growing economies and they eventually give way to formal housing as the profits of the development “trickle down” in later phases. It is only in these later stages of formalised housing where capital creation is possible because the legal representation and description of the physical and economic qualities of a piece of land in a title is what constitutes “birth of capital” (De Soto, 2001). Hence, this chapter investigates the conditions under which citizen science may be relied upon to create urban cadastres (that may assist in securing property rights in informal settlements) for future capital creation. According to Marx et al., (2013a) whereas

slum development usually takes place on “vacant” land, occupants in these settlements do pay rent in a system in which landlords strongly enforce and obligate this – thereby raising the issue of informal institutions within the urban discourse.

Aside from slums/informal housing, informal business also form a core part of the economies of the developing world. Weak land tenure systems and low rates of registration of informal business mean that data on both types of informality are rare; authorities do not typically collect administrative records on informal activity. Data gaps therefore limit knowledge on large sections of developing economies.

This chapter investigated the conditions under which it is possible to use citizen science in developing a precise urban business and residential informality dataset (and other land use types) from satellite imagery. Specifically, the study experiments with the conditions in which this may be possible by piloting a study with the assistance of a group of SU volunteers. A secondary objective was to quantify the magnitude and extent of the 2005 clean-up operation in Zimbabwe named Operation Restore Order (ORO) or Operation “*Murambatsvina*³⁶”, the outcome of which has the potential to be linked with other socio-economic data. Most importantly though, ORO allows us to test whether citizen science is able to detect changes in informal land use by including pre (2004) and post (2006) exercise images for areas that were affected and those that were not.

However, these approaches are not without their problems. Citizen science relies on the inputs of human subjects, who indicate where and when certain land classes are located. Fairbairn and Al-Bakri (2013) maintain that reliance on inaccurate and inconsistent Volunteered Geographic Information (VGI) generated data, is a much better scenario than to have “no mapping at all”. Yet for economists the combination of VGI and satellite imagery present an enormous opportunity to investigate important topics such as the informal sector in developing countries. Understanding the applicability of VGI on satellite imagery and its advantages and shortcomings is important so that economists and other researchers working in the data scarce settings of developing countries can know the extent to which they can rely on findings that are based on this approach. Therefore, it is necessary to develop an understanding of the factors and conditions that affect classification accuracy. Foody et al. (2013); Fonte, Bastin, See, Foody and Lupia (2015) and Albuquerque, Herfort and Eckle (2016) indicate that one way of improving the quality of data generated via citizen science is to rely on consensus among *many*

³⁶ Murambatsvina means “away with filth” in vernacular Shona.

volunteers for the same piece of work or area. Albuquerque et al. (2016) found that tasks with substantial agreement were 41 times more likely to be classified correctly in a crowdsourced project relative to those with low levels of agreement. This is understood as ‘Linus’s Law’. Linus’s Law is the supposition that the number of contributors and quality are positively related, thereby providing an intrinsic measure of quality assurance (Haklay, Basiouka, Antoniou & Ather, 2010).

However, the validity of this law has not been tested in classifying informality. Informal residence and business structures are irregularly shaped and small relative to other land use classes. Human classifiers may therefore be prone to make larger errors of inclusion and exclusion. If those errors are not mutually exclusive across subjects, then it is possible that the consensus will provide worse rather than better classifications. This chapter tested this proposition.

Another important variable among factors affecting classification accuracy is the user/volunteer demographic attributes. Classification (usually the first analytical task undertaken by participants in a crowdsourcing project) is associated with the deployment of interpretive skills and past/background (knowledge and experience) information to classify the piece of geographic information at hand (Albuquerque et al., 2016). Hence, user background information and the predefinition of values (for different class features on the image) is an important aspect in any citizen science project. Lastly, participant training (Fritz et al., 2017) is another key aspect that was also considered in the study.

We recruited 41 student volunteers at Stellenbosch University’s Stellenbosch campus for an experiment which tested whether Linus’s Law holds in the classification of urban informality. During the recruitment phase, students were asked to complete a questionnaire in order to collect some basic demographic information. This was not done to select individuals into different groups by perceived skills [for example, following Meier (2013)] but in order to check whether the different groups under the experiment were statistically balanced as well as to ensure that the effects of user skills, background [following Foody et al. (2013) and Albuquerque et al. (2016)] experience were tested. Students were randomly allocated into four experimental streams and a student could only classify images in their allocated stream. Streams 1 and 2 contained images for areas that were not affected by the 2005 clean-up operation (ORO), while Streams 3 and 4 contained images for areas affected by the operation. The experiment was set up such that an individual image was classified five times in Streams

1 and 3 and ten times in Streams 2 and 4 as discussed later in detail later in the chapter. Comparisons of the various experimental streams enabled the testing of Linus's Law in settings where the landscape has been both stable and unstable.

The chapter is structured as follows. Section 4.2 discusses the importance of land use and cadastral maps, for urban planning, as well as the usefulness of citizen science in the development of these maps/databases. Citizen science has been applied in the identification and classification of different land use types, however, given the potential of increased misclassification owing to the small and irregular nature of informal structures, this study mainly focused on the informal sector. Hence, in Section 4.3, a discussion of the measurement of both business and residential informality (and other land use types), and a description of the problem in Zimbabwe and the 2005 ORO clean-up operation is presented. In Section 4.4, the experiment setup is outlined in more detail, while Section 4.5 presents the results and discussion. The conclusion and recommendations are presented in Section 4.6.

4.2 Importance of Land Use Mapping

Regular and large population movements and changes in settlement patterns in developing countries highlight the importance of land cover mapping for urban and sprawl management (Hegazy & Kaloop, 2015; Malarvizhi, Kumar & Porchelvan, 2016; See et al., 2015; Vaz & Jokar Arsanjani, 2015). Although these geographical areas are often characterised by unorganised and unplanned urban expansion, Vaz and Jokar Arsanjani (2015) acknowledge the important role played by urban regions and major metropolises in future economic stability. However, cost effective, accurate mapping and development of cadastral data is required to achieve these objectives, especially in developing countries. Urban cadastral data obtained from land cover maps can assist in service provision and land title provision in informal settlements (Chitekwe-Biti, Mudimu, Nyama & Jera, 2012), for example. Although it may result in unintended long run cost distortions, utility provision in slums can yield immediate benefits (Harari and Wong, 2017) – hence the importance of urban land use mapping. In this context and against the background of an “African statistical tragedy” [see Devarajan (2013)], this study investigates the citizen science approach for the classification of the informal sector and other urban land use types. As previously mentioned, the study pilots land use type classification through citizen science with the assistance of a sample of SU volunteer participants.

4.3 The Role of Citizen Science in Urban Land Use or Cadastral Mapping

Advances in web technology have allowed volunteers to be at the forefront of producing different kinds of geographic data (Arsanjani & Vaz, 2015; Chapman, Bell & Bell, 2017). This is known as citizen science, crowdsourcing, human computation or Volunteered Geographic Information (VGI) (Albuquerque et al., 2016; Arsanjani & Vaz, 2015; Fonte et al., 2015). Wiggins and Crowston (2011) define citizen science as a kind of collaborative research in which scientific projects rely on public participation in order to address real-world problems. User contributions in VGI projects tend to involve tracing out specific land use types on georeferenced images, while in other cases this may involve the collection of GPS points by the users (Arsanjani & Vaz, 2015; Fonte et al., 2015; Neis & Zielstra, 2014). This study relied on the former approach (classification) and involved users tracing out different land use types on georeferenced Google Earth (GE) imagery uploaded on the Zooniverse³⁷ platform.

Citizen science and VGI has gained popularity in the mapping of urban land use types. Albuquerque et al. (2016), asserts that crowdsourcing may be superior to other land use classification approaches (such as machine learning), especially where features are heterogeneous and the classification typology is inconsistent. These tasks require human recognition where algorithms fail. The informal business sector in Zimbabwe is one such area of study, because its structures take on different (irregular) shapes and forms and are small – potentially rendering the crowdsourced approach more useful than automatic detection. Informal business activities may involve vending activities, different types of craftsmanship and artisan productive activities that are housed in different kinds of structures.

In the validation of land cover maps, Fonte et al. (2015) explain that physical ground-truthing (verification of land use features through field surveys, as opposed to remote observation using satellites) is laborious and may be impossible if researchers are located far away from the areas of study. Further, the method may suffer the drawback of the ground-based experts not agreeing on the classifications. VGI datasets and classifications can be used as reference or verification

³⁷ The Zooniverse platform (www.zooniverse.org) is a citizen science platform developed by astrophysicists at Oxford [initially as the Galaxy Zoo project (www.galaxyzoo.org) with the objective of identifying new galaxies from imagery taken by Sloan Digital Sky Survey at the Apache Point Observatory in New Mexico, USA]. The Zooniverse platform has evolved to include classification and related tasks for other non-space sciences such as the humanities and arts as well as nature and climate. It is well suited for citizen science applications in land use classification.

data to test accuracy of machine learning models,³⁸ resulting in efficiency and cost savings. An important cost-saving channel is the availability of free georeferenced JPEG format images available from the Google Earth (GE) platform, given that multispectral³⁹ band information is usually unnecessary in mapping driven by citizen science. Malarvizhi et al. (2016), note that GE provides free Very High Resolution (VHR) imagery that is suitable for image-based classification that crowdsourcing provides and it provides good temporal resolution.

Additionally the GE platform hosts VHR imagery with a good spatial resolution (such as 50 x 50 cm), which allows even non-remote sensing experts to classify different land use classes in a reliable and cost effective manner (Fritz et al., 2009). The GE platform therefore offers critical functionality for the validation of global land cover maps (Fritz et al., 2009) and was therefore an important source of imagery for this study.

4.3.1 Past examples of citizen science projects

There have been a number of papers that have applied crowdsourcing techniques in the measurement of land use, cultural and socio-economic changes [see Petteorelli, Gliozzo and Haklay (2016)]. In terms of web applications, Open Street Map (OSM) is among the most popular VGI projects among the academic and research community (Fonte et al., 2017). Neis and Zielstra (2014) explain that the objective of the OSM project is to establish an open source geographic information database for use in mapping, navigation and other case uses. Another prominent example is the Geo-Wiki project. It was established to promote global networking by volunteers with the end goal of enhancing global land cover maps through crowdsourcing (Foody et al., 2013; Fritz et al., 2017). The Zooniverse platform (on which this study relied) was developed by astronomers and physicists to aid the identification of new galaxies from thousands of night-time images of the sky taken by telescopes. Since then, the platform has hosted a plethora of different image classification projects from different fields.

³⁸ Machine learning models are algorithms that increase performance or knowledge as they acquire more experience (Flach, 2012). These can be used to identify informal structures and other land use types based on what they learn from a given training set.

³⁹ Satellites are fitted with multiple sensors that 'sense' information at different thresholds of the electromagnetic wavelength. The most popular bands are Red (R), Green (G) and Blue (B) which are used to create natural colour composites of images as they are shown on GE. However, machine learning applications require other bands not visible to the human eye such as near-infrared and aerosol. A description of some of the widely used bands is provided by the US Geological Survey (USGS) for Landsat imagery here: <https://tinyurl.com/y26ygjuf>

4.3.2 Some drawbacks of the citizen science approach

Despite the numerous advantages of citizen science, there are a few drawbacks. Explaining the importance of VGI data in land use mapping, Dorn, Törnros and Zipf (2015) and Foody et al. (2013) acknowledge the fact that the volunteers generating the data are untrained. Following Fritz et al.'s (2017) idea on the need to develop a training manual, this study created video⁴⁰ and in-application/on-demand tutorials in order to help volunteers on the projects to visualise the predefined typology and provide them with background information.

Apart from the need for training, participant inequality and time-consuming annotations are additional problematic issues in VGI/crowdsourcing and these may result in class noise. Class noise is basically error of classification while participant inequality is a scenario where some participants are disproportionately more closely involved in the classification work than others. Explaining some of the inherent issues with Open Street Map (OSM) data, Johnson and Iizuka (2016) and Johnson, Iizuka, Bragais, Endo and Magcale-Macandog (2017) highlight the problem of class noise in the data; this is caused by classification errors on the part of the users. The identification of small objects (potentially the identification of informal structures such as shacks) can exacerbate class noise because the identity of smaller objects is more difficult to discern relative to larger objects. As indicated by Albuquerque et al. (2016), there is little likelihood for volunteers to classify small objects correctly since they may be difficult to discern, as previously mentioned. Such issues are important to consider in the classification of the informal sector; structures are small and irregular and difficult to distinguish from other types of land use as shown in Figure 4.1.

In Figure 4.1, informal business structures (tuck shops, barber shops, hair salons and food stalls) are shown with red marking, while formal housing is shown with blue marking. Apart from the uniform and regular shape and related road network, formal housing structures are larger in comparison to informal structures. Without prior knowledge and sufficient training, it can be challenging to identify and discern what small structures represent. The extent of class noise in various circumstances was tested in this study through the calculation of classification accuracy rates.

⁴⁰ The study created video tutorials that are available at the following links: <https://www.youtube.com/watch?v=ycqxGdhd2qg>, <https://www.youtube.com/watch?v=OYdz0Zhx5sE> and <https://www.youtube.com/watch?v=807Lx2cOeKo>



Figure 4.1 Discerning small objects on satellite images using the Bezier tool

Source: GE imagery over Harare (Glen-View 8 Area); 17°53'36.51"S, 30°57'20.42"E; captured by satellite on 25 August 2004

Participant inequality is a critical issue that characterises VGI platforms (Neis & Zielstra, 2014; Nielsen, 2006; Stewart, Lubensky & Huerta, 2010). Known as the 90-9-1 rule, this is a scenario in which 90% of users on a VGI project never contribute anything; 9% have an irregular presence on the platform and 1% account for almost all the user contributed data on the project (Nielsen, 2006). Should the small group of individuals be prone to misrepresent information, they could exert substantial leverage on outputs from such classifications. VGI projects also take time and participants may choose either not to contribute, or make errors. Relative to machine learning, the classification of imagery using citizen science is essentially based on photo interpretation and it is time consuming (Srivastava, Lobry, Tuia & Vargas-Muñoz, 2018; Xing, Meng, Hou, Song & Xu, 2017). However, an important advantage of citizen science is its reliance on a bigger ‘crowd’ of volunteers that can process large amounts of information quickly; additionally, if substantial consensus among human classifiers exists, then the validity of classifications becomes more credible.

4.4 Measurement of Urban Informality

Housing and business informality are a result of unplanned expansion and labour market rigidities respectively, and these two form a significant part of the economy in developing

countries. Falco, Kerr, Rankin, Sandefur and Teal (2011) and Rothenberg et al. (2016) define the informal sector as all economic entities that neither pay taxes nor are registered with the government. ILO (1972), Rei and Bhattacharya (2008) and Batini, Levine, Kim and Lotti (2010) describe informal services as activities that are “unrecorded, unrecognised, unprotected and unregulated by the authorities”. According to Benjamin, Mbaye and Diop (2012) the informal sector is a group of small and organised producers who operate on the boundaries of the formal economy, while Becker (2004) views the informal sector as “the unregulated and non-formal element of the market economy that produces goods and services for other forms of remuneration”. A backyard structure offered for rental to a tenant satisfies a need (may be viewed as a housing service). It is highly likely that the transaction of letting a backyard structure to a tenant will be unrecorded and not result in tax remittance to the authorities. Hence, backyard structures fit into the general definitions provided by Falco et al. (2011), Rothenberg et al. (2016) ILO (1972), Rei and Bhattacharya (2008), Batini et al. (2010) and Becker (2004).

Bhattacharya (1995) stresses that dual economy models of development overemphasise the role of agriculture, while ancient societies were actually engaged in significant crafts-making. Informality has therefore always been a component of the pre-industrial economy and has in many cases been a launch for proto-industrialisation. The business informal sector is integral to the livelihood and sustenance of particularly the poor and it normally acts as the first point of employment for people migrating from rural to urban areas; it also sometimes becomes a hub for innovation (Bhattacharya, 1995). Despite all these positives, the informal sector is not well understood. By definition, the sector is unregulated (Falco et al., 2011; Rothenberg et al., 2016) therefore obvious sources of data from banking, tax returns and other activities are generally absent.

Often governments do not know the size of the informal sector and its monetary contribution; and this may partly explain negative state action in the form of clean-up operations against the informal economy. Clean-up campaigns are conducted by many developing countries – they are urban eviction programmes that remove individuals from their informal homes or business premises: for example, the 2005 “Operation Restore Order” in Zimbabwe (Potts, 2006; Tibaijuka, 2005), the 2006 “Operation Dongosolo” in Malawi (Riley, 2014) and the demolition of the Muoroto and Mwariro illegal settlements in Kenya in the 1990s (Arimah & Branch, 2011). The deficient recognition of the informal sector, potentially due to lack of awareness of their exact economic contribution or high quality institutions as in Shapiro (2015), may be an

influencing factor behind clean-up campaigns and other negative state action. Also, both national and by-laws that are used by African governments are often considered to be outdated and to represent fragments of a colonial inheritance that the political establishment retains to exploit citizens (Arimah & Branch, 2011; De Soto, 2000; Potts, 2006; Riley, 2014; Tibaijuka, 2005).

Tibaijuka (2005), Riley (2014), Arimah and Branch (2011) and Potts (2006) assert that clean-up operations still occur due to the proliferation and application of old, colonial legislation to achieve the political interests of present day political elites. Hence, apart from the absence of granular data on the size and growth patterns of the informal sector, the persistence of colonial legislation has resulted in clean-up operations as a response to unregistered residential and business establishments. Regardless of the motivation behind ORO, it provides an opportunity to compare user classifications in areas affected and those that were not in the pre and post periods.

The informal sector accounts for 70-78% of employment in sub-Saharan Africa excluding South Africa, indicating potentially high levels of competition with parts of the formal sector (Verick, 2006; Becker, 2004). Figures are slightly lower for Latin America and Asia at 60% and 59%, respectively (Becker, 2004). Benjamin et al. (2012), estimate that between 80-90% of employment in Africa as a whole is in the informal sector and that up to 60% of national income in West Africa is derived from informal activities. Clearly, the informal sector is significant in developing countries and that is why there is renewed interest in the subject beyond economics (Naik, 2009).

This study gathered data on both types of informal and other land use types. However, much of the study's attention is on the informal sector for two reasons. The first is that the study investigated the conditions affecting the accuracy of citizen science where features are small and irregular, and the second is that ORO, the policy event that we take advantage of in comparing classification accuracies across affected areas and those that were not (and across pre and post periods), was a campaign against informality. The discussion highlights the importance of measurement of the informal sector. Hence, the study examines the ideal conditions for using satellite imagery and citizen science in the classification of urban informality and other land use types – thereby offering a cost effective way to measure unrecorded sectors and to formulate policies that rely on this information. The study tests whether more classifications per image (Linus's Law) and user demographic attributes, as well

as learning (through training and experience) affect classification accuracy by analysing the classification work of a group of SU students.

4.4.1 Urban informality in Zimbabwe

Both business and housing informality are widespread in Zimbabwe. Historically, its economy has not been able to provide enough employment for the population (Ncube, 2000). This problem was already significant in the 1970s before independence; British sanctions and instability (due to the liberation war) harmed the Rhodesian⁴¹ economy. Growth after independence in 1980 was erratic, although the public sector played a critical role in absorbing large numbers of people (Ncube, 2000). However, The Economic Structural Adjustment Program (ESAP) called for a rationalisation of the civil service which resulted in lay-offs. As a result, the absorptive capacity of the public service was negatively affected and people were left with no choice but to join the informal sector.

According to Tibaijuka (2005), at independence the productive informal sector accounted for less than 10% of employment, but grew to 40% by 2005 due to the gradual economic decay experienced in Zimbabwe. Thus, Operation Restore Order (ORO) of 2005 (as described below) represented a direct negative shock that affected 40% of the employed labour force in 2005. The ORO clean-up operation targeted and destroyed informal/unregistered business establishments in Zimbabwe major urban areas in 2005. Informal housing (mainly backyard structures⁴²) was also affected by the operation, making it possible to investigate whether or not study participants (the group of SU student volunteers) could identify changes in both business and housing informality (and other land use types) by using pre and post ORO images. Zingoni, Love, Magadza, Moyce and Musiwa (2005), posit that population expansion against a background of fewer low cost housing options influenced informal settlements in Zimbabwe (although backyard structures within formal housing are the dominant class).

4.4.2 The 2005 clean-up operation (ORO) in Zimbabwe

Operation Restore Order was unprecedented and, according to Potts (2006), there are no other recorded urban eviction programmes that swept across entire metropolitan areas on the African continent, including apartheid South Africa. “To rid urban areas of uncontrolled expansion and

⁴¹ Zimbabwe is a former British colony. It was known as Rhodesia before it attained self-rule in 1980, and adopted the name “Zimbabwe”.

⁴² Backyard structures are informal/unplanned housing units/shacks that are erected next to the formal house either by the landlord or tenant (Turok and Borel-Saladin, 2016).

filth”, ORO was launched on 25 May 2005 in Harare, the capital of Zimbabwe, quickly spreading into other cities and towns. The operation was led by council employees, who were supported by the security forces in order to ensure compliance by the citizenry. By the end of the operation an estimated 700,000 people had been rendered homeless or without a source of livelihood. Government sought to quell all forms of illegality in the form of unregistered business and residential structures (Tibaijuka, 2005). According to Potts (2006), government put its own estimate of the total number of people affected at 570,000, with 92,460 dwellings being destroyed, 133,534 households losing their places of abode and 98,000 losing their sources of income. Poverty, destitution, deprivation and vulnerability increased as a result of the clean-up campaign (Potts, 2006).

Apart from investigating the conditions under which urban land use mapping from satellite imagery via citizen science is possible (i.e. it is more accurate), a secondary aim of the study was to buttress the survey findings of Tibaijuka (2005) and others, with objective satellite data from parts of Harare under study. Importantly though, ORO is a real event that allowed us to test under which conditions users in crowdsourcing projects can detect changes in urban informality and other land use types through the inclusion of the pre and post clean-up exercise images in the experiment.

4.5 Experiment Setup, Implementation and Estimation Strategy

As explained by Hegazy and Kaloop (2015), monitoring urban growth using Remote Sensing (RS) data involves detecting changes between two periods that are uncharacteristic of normal variation. ORO allowed us to do exactly that. Following Malarvizhi et al. (2016), 180 images (over Harare, Zimbabwe) were downloaded and georeferenced using the Elishayal Smart GIS tool. The same images, covering the same areas were downloaded for similar dates in 2004 and 2006, to provide a longitudinal spatial dataset (see Figure 4.2). The images used in the study cover a total area of 46 km².

The images were randomly allocated into four groups/streams as shown in Table 4.1. Areas that were not affected⁴³ by ORO were allocated to Streams 1 and 2, hence the structures on the images should look similar in both years. Areas affected by ORO were allocated to Streams 3

⁴³ To select areas not affected by the ORO, the study mainly considered areas dominated by land use types not targeted by the clean-up campaign, namely formal housing and formal industry. The study process included verification that the areas were indeed not affected by viewing imagery for the same area for 2004 and 2006 on the Google Earth platform. The classifications were made by experts who are familiar with Harare.

and 4; hence, the major differences between the 2004 and 2006 images are that informal housing and informal business land use types were destroyed during the 2005 clean-up exercise. The experiment was set up to test under which conditions citizen science can detect changes in urban informality and other land types and whether more user classifications per image attains better accuracy.

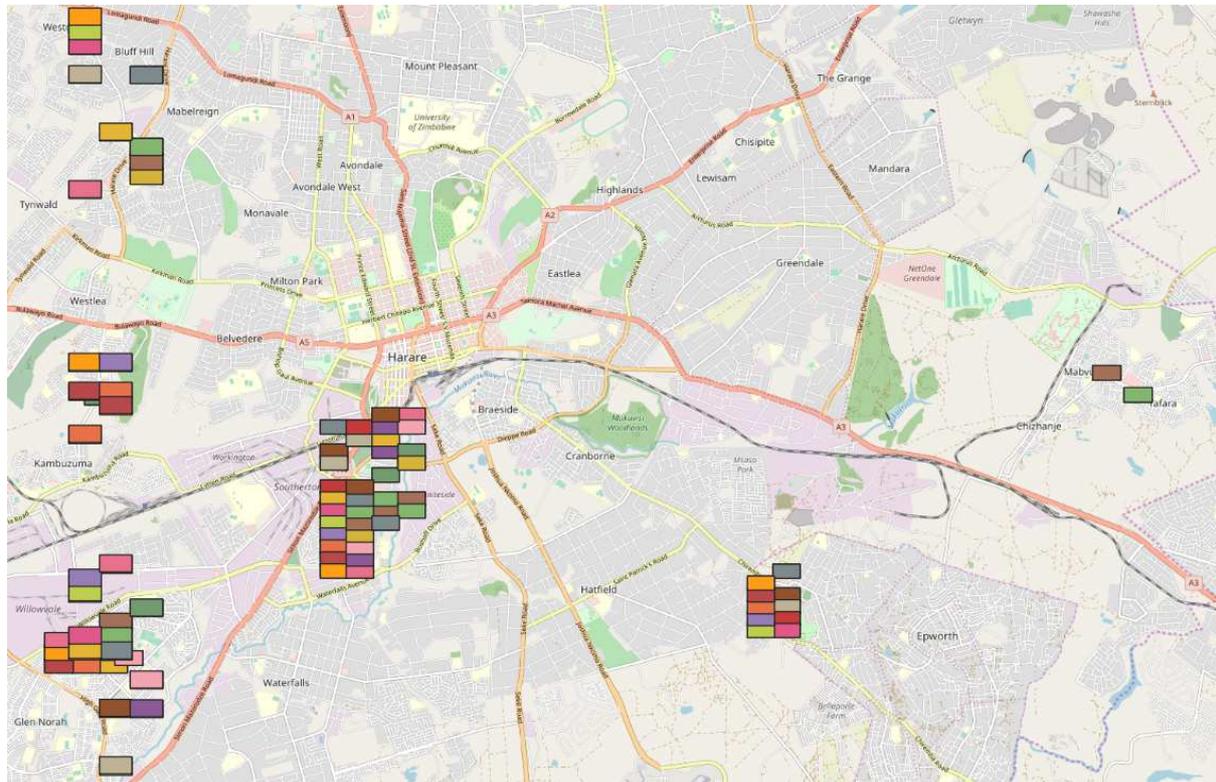


Figure 4.2 Areas sampled in the study

Source: Own polygons representing the positions of individual images that were used in the study. Base map obtained from Open Street Map (OSM).

Table 4.1 Experiment design

Year	No. of images			Intended Classifications per image	2004	2006	Totals
	2004	2006	Totals				
Stream 1 - No change	25	25	50	5	125	125	250
Stream 2 - No change	20	20	40	10	200	200	400
Stream 3 - Change	25	25	50	5	125	125	250
Stream 4 - Change	20	20	40	10	200	200	400
Totals	90	90	180				1300

Source: Own depiction of experiment setup

As mentioned earlier in the chapter, the quality of citizen science generated data can be increased by relying on consensus classifications (Albuquerque et al., 2016; Fonte et al., 2015; Foody et al., 2013). The design of the pilot study for Stream 2 and Stream 4 (such that an image is classified by ten users, instead of five) centred on Linus’s Law which is the supposition that there is a positive correlation between the number of contributors and quality, thereby providing an intrinsic measure of quality assurance (Haklay et al., 2010). This was a key factor under investigation in the experiment design. However, we also stratified treatment to regions where there were no known changes resulting from ORO and those that were affected. In essence the former scenario should have similar images in 2004 and 2006, so that consensus could also be assessed *across* years.

We ran a student recruitment campaign at Stellenbosch University’s Stellenbosch campus and 41 volunteers signed up for the project. For their classification efforts, students were paid ZAR5,00 per image. Albuquerque et al. (2016) recognise the need for citizen science projects to collect data on participant skills prior to the image classification work in order to inform the allocation of users to undertake specific tasks. As Foody et al. (2013) stated, volunteers may vary from eager but naïve and untrained individuals to those who are highly experienced and skilled. As mentioned, the volunteer students were asked to provide some demographic information by way of a questionnaire. Demographic attributes were not used as a basis for allotting individuals into the different groups/streams but rather to check whether the different groups under the experiment were statistically balanced. Students were randomly allocated into streams and a student could only classify images in the allocated stream.

Table 4.2 Allocation of students to streams

Stream	Number of students	Images/student (Intended)
1	8	31.25
2	12	33.33
3	9	27.78
4	12	33.33

Source: Own depiction of experiment setup

Table 4.2 shows the number of students allocated out of the 41 to each stream and the intended number of images each student needed to classify so as to ensure that the experiment worked and to prevent participant inequality. Table 4.3 shows the results of the balance tests. Fischer’s

Exact Test checks whether different treatments result in different outcomes in an experiment. In this case, the different treatments arms are constituted by the individuals in the four streams. Under Fischer's Exact Test, the null hypothesis (treatment affects outcome) is not rejected if the p-value is small. Table 4.3 shows that the allocation of participants into streams was balanced according to observable criteria.

Table 4.3 Fischer Tests for statistical balance

Variable	Fischer Test p-value
Frequently busy from informal business?	0.31
Informal business customer?	1.00
Owens informal business?	0.75
Frequently visits informal settlement?	0.38
Visited informal settlement?	0.12
Lived in informal settlement?	0.66
Household income	0.65
Education	0.31
Parent education	0.10
Sex	0.77
Age	0.26
Province	0.96
Nationality	0.37
Race	0.97

Source: Own depiction of experiment setup

Albuquerque et al. (2016), highlight the importance of background information on the skills and experience of the user, hence the study tested for balance across the streams. The demographic variables gathered in the study provided important information on each participant's background and experience. For example, questions on whether or not the participant had ever visited an informal settlement, whether or not they had ever bought goods/services from an informal entity, etc. are important experiences that may affect classification skill.

Table 4.4 Classification tasks

Task Number	Task	Instruction on platform
a	Informal businesses	Draw around the area(s) that appear to be informal businesses or business hubs.
b	Backyard structures within formal housing	Draw around the area(s) that appear to be backyard structures within formal housing.
c	Formal housing	Draw around the area(s) that appear to be formal housing.
d	Formal businesses	Draw around the area(s) that appear to be formal businesses or industry.

Source: Own depiction of experiment setup

There is no clear knowledge about the specific tasks most suitable for classification using citizen science (Albuquerque et al., 2016). In this study, participants were presented with five tasks,⁴⁴ as shown in Table 4. The tasks involved the identification of four different land use types namely a) informal businesses, b) formal housing with informal backyard structures, c) formal housing and d) formal businesses.

**Figure 4.3 Classification using the Bezier tool**

Source: Depiction of GE image inside Zooniverse platform

The participants completed the tasks by tracing polygon(s) around specific land use types in accordance with the task instructions on the platform. The platform carried clear instructions for participants to “do nothing” if they could not identify particular land use classes on an

⁴⁴ <https://www.zooniverse.org/projects/tchingozha/tracing-informality-through-citizen-science/classify>

image. Participants traced out the different land use types using a Bezier tool (see Figure 4.3). Tutorials to distinguish the different land use types and how to navigate the interface were inbuilt within the platform while videos were made available on YouTube as noted earlier.

4.5.1 Processing classification results

The classification output from the project was exported from the Zooniverse platform as Excel files for each stream. The files contained several columns of information such as participant usernames, image names and task numbers but, most importantly, the X and Y map coordinates for redrawing the user classification and reference⁴⁵ polygons.

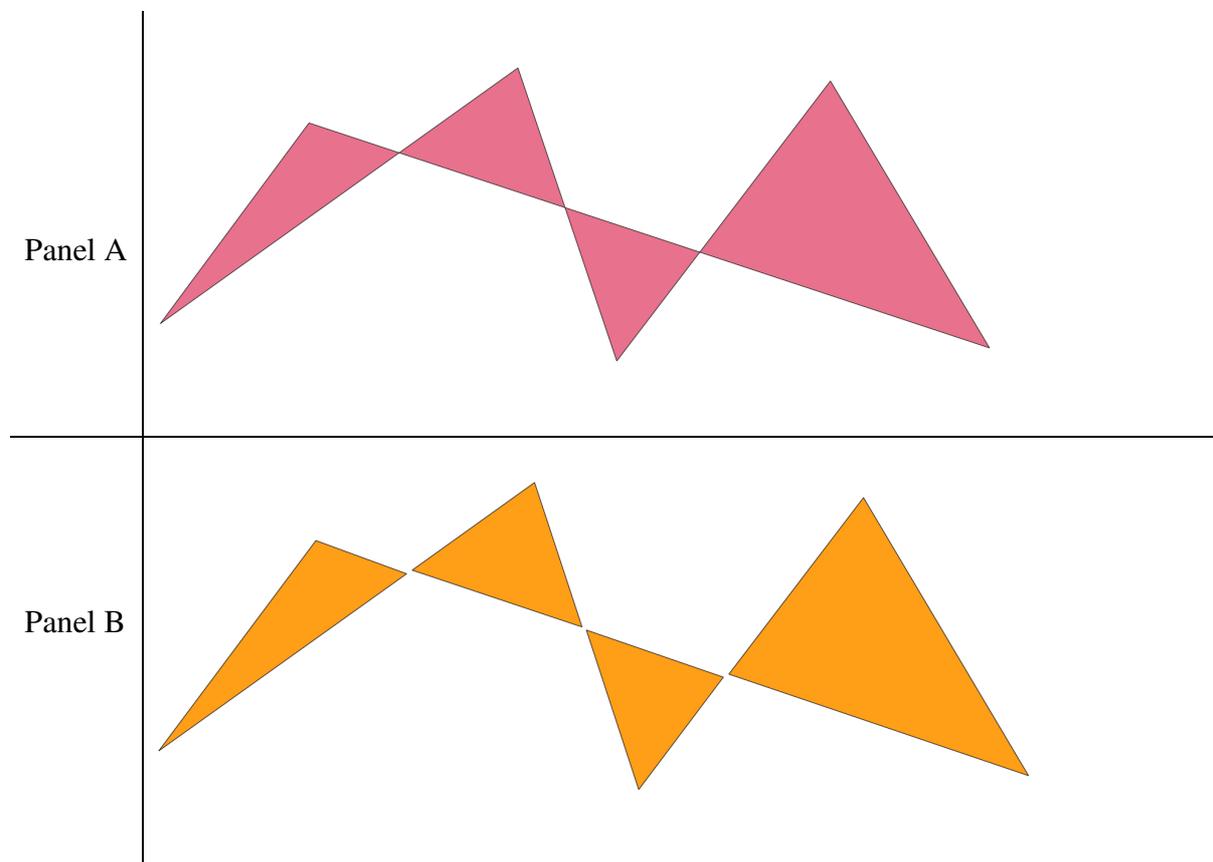


Figure 4.4 Creating multi-parts from single polygon to address crossing lines

Source: Illustration of QGIS geo-processing

⁴⁵ The authors performed an expert (reference) classification of all the images in the project, based on their familiarity and knowledge of the areas sampled out of Harare for the study – together with their knowledge of the classification typology.

The Excel files were prepared in STATA for eventual plotting and further cleaning in Quantum GIS (QGIS) software. Some of the polygons drawn by users violated some geometry properties of ESRI shape files (in some instances polygon lines crossed as shown in Figure 4.4). Such instances were corrected by manually retracing the polygons and ensuring that lines do not cross. Panel A in Figure 4.4 shows a polygon with crossing lines while panel B shows correction by splitting the polygon into multi-parts. Aside from the manual correction, subsequent geo-processing was performed as follows: i) the XY coordinates of each classification were plotted, ii) points were converted to polygons, iii) user and reference classifications were intersected, and iv) the areas of overlap were calculated programmatically using PYQGIS. The data was exported back to STATA for the final analysis.

4.5.2 Measuring classification accuracy

Quality and accuracy is a key issue in VGI (Fritz et al., 2009). Fonte et al. (2015), note that the accuracy of a land cover map is usually derived from the extent of its agreement to some gold standard of ground truth. We undertook an expert classification of the area under study given our in-depth knowledge and familiarity with it as mentioned previously. Classification accuracy was calculated for each individual classification of each image using Equations (4.1) and (4.2):

$$\text{Classification Accuracy 1} = \frac{\text{Overlap Area}}{\text{Reference Area}} \quad \dots (4.1)$$

Classification accuracy 1 (*accuracy_in*) is defined as the percentage of the reference area that has been correctly classified (see Figure 4.5, panel A). However, the formula in (4.2) overstates classification accuracy if the polygon drawn by the user goes beyond the boundary of the reference polygon (see Figure 4.5, panel B). To account for this, a second accuracy statistic [Classification Accuracy 2 (*accuracy_ex*)] was calculated in (3).

$$\text{Classification Accuracy 2} = \frac{\text{Overlap Area}}{\text{User Area}} \quad \dots (4.2)$$

Classification Accuracy 2 is defined as the percentage of the user/volunteer's classification that is correct. It penalises users who commit errors of inclusion, even if they classify the reference area correctly. These measures are used at the individual level to assess participant inequality in classification.

Measures of classification accuracy were extended to also measure consensus across users. This was to directly test the experimental setup and to understand whether relying on the inputs of more users improved accuracy. For the first measure, the user classification was limited to the intersected area, in which *all* users correctly indicated the land use class. On the other hand, for the second measure, the user classification was extended to the consensus/intersection of all users' polygons. Hence, if many users had non-overlapping errors of inclusion, the accuracy rate for a particular image was reduced as reflected in Table 4.5 (b).

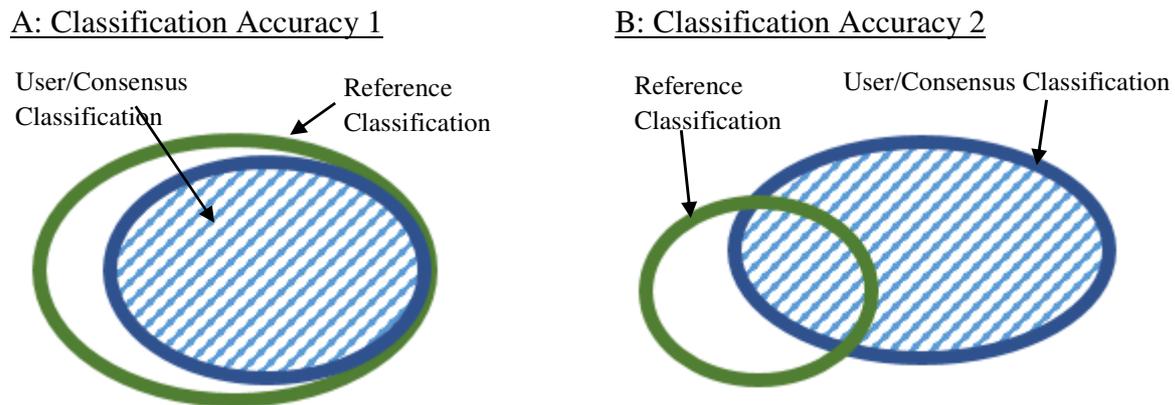


Figure 4.5 Defining classification accuracy

Source: Own Illustrations

NOTES*: In Figure 4.5 (B), for the consensus estimates, User Classification is the consensus classification of all the users classifying an image in a stream

4.5.3 Correlates of classification accuracy

Our main analysis implemented various OLS and fractional probit regressions to investigate the effects of several variables on classification accuracy. A first model, measuring accuracy for each classification task is specified in Equation (4.3). These correlations highlight demographic features that represent participant inequality in accuracy rates.

$$Acc_{imt} = \beta_0 + \beta_v y_t + \beta_1 X_i + \beta_2 X_m \quad \dots (4.3)$$

Where i indexes individuals, m images and t time. Acc_j is one of two endogenous variables representing classification accuracy (see below), y_t represents year dummies for 2004 and 2006, respectively.

Equation (4.4) measures the effect of consensus. The unit of analysis is the image. Accuracy is now determined by the intersection of all the classifications of one image by different study

participants. The intersection is the smallest part of the correct area which all users identified as being part of a particular land use class. This equation assesses directly the effects of the experiment and whether more classifications raise accuracy according to Linus's Law.

$$Acc_m = \beta_0 + \beta_v y_t \times (Many_m; Change_m; Number_{mt}) \quad \dots (4.4)$$

Where,

Many is a dummy variable taking the value of 0 and 1 if an image belongs to an experimental treatment stream with 5 or 10 assigned classifications per image, respectively,

Change is a dummy variable that takes the values of 0 and 1 if an image belongs to a “no change” and “change” stream, respectively,

Number is a continuous variable denoting the number of times an image is actually classified by different volunteers.

4.6 Results and Discussion

Before turning to the main analysis, a number of descriptive results are presented. First, descriptive statistics check the extent to which the intended setup of the experiment was achieved. Second, multinomial logit estimates are presented to descriptively show the changes in area coverage for the different land use types in the study using the reference data – this provides evidence on the extent to which ORO affected various land use types in the selected areas we sampled; this does not measure the full extent of the programme, as not the whole of Harare was classified. This is followed by a regression on NLD to establish the extent to which night lights might also reflect the changes in urban land use. Lastly, classification accuracy regressions estimates – the core of the analysis – are presented.

Table 4.5 and Table 4.6 show the descriptive statistics for the individual classification data and the consensus (intersected) data, respectively. In Table 4.5, the mean “*Number*” of classifications was greater for Stream 2 than Stream 1. At the same time, this was also greater for Stream 4 than it was for Stream 3. The expectation was that each image would be classified five times in Streams 1 and 3 (the “less classifications streams”) and ten times in Streams 2 and 4 (the “many streams”). The Zooniverse platform does not have the means to correct the number of classifications per image. However, instructions to users emphasised the upper bound number of images that each user/volunteer was supposed to classify in order to achieve

the 5 and 10 classification requirements. Some users stopped their tasks early while others exceeded the stream-specific quotas. The desired experimental design was therefore not perfectly achieved. Students who did not reach a minimum number of classifications did not receive payment, while payment was capped up to the maximum number of images. These conditions were explicit before the experiment was conducted.

Table 4.5 Individual classification descriptive statistics

Variable	Statistic	Stream 1		Stream 2		Stream 3		Stream 4	
		2004	2006	2004	2006	2004	2006	2004	2006
<i>Number</i>	Min	4	1	8	8	6	1	10	1
	max	21	22	26	32	26	38	42	42
	Mean	10.76	10.88	15.30	17.40	17.60	18.92	29.40	29.35
	Median	11.00	10.50	15.00	15.00	18.00	22.00	30.00	31.00
<i>Individual Classification Accuracy 1</i>	Min	0.18	0.06	0.22	0.24	0.40	0.21	0.38	0.19
	max	0.89	0.88	0.80	0.96	0.96	0.99	0.91	0.80
	Mean	0.61	0.59	0.51	0.63	0.70	0.65	0.64	0.50
	Median	0.67	0.67	0.50	0.64	0.67	0.68	0.62	0.51
<i>Individual Classification Accuracy 2</i>	Min	0.00	0.00	0.00	0.14	0.00	0.00	0.28	0.19
	max	0.99	0.99	0.99	0.83	0.77	0.89	0.72	1.00
	Mean	0.48	0.55	0.36	0.37	0.45	0.42	0.44	0.33
	Median	0.48	0.57	0.31	0.37	0.49	0.41	0.43	0.27

However, the fact that Stream 2's mean "Number of classifications" was greater than that for Stream 1 means the experiment design was loosely maintained. The same can be said about Stream 4 relative to Stream 3.

Table 4.6 Consensus data descriptive statistics

<i>F/M</i>	<i>A1(%)</i>	<i>A2(%)</i>	<i>Year</i>	<i>A1(%)</i>	<i>A2(%)</i>	<i>Change/No Change</i>	<i>A1(%)</i>	<i>A2(%)</i>
Few	0,001%	0,007%	2004	0,002%	0,015%	No Change	0,002%	0,007%
Many	0,002%	0,014%	2006	0,001%	0,006%	Change	0,001%	0,014%
F-stats	2.06	0.76	F-stats	4.67	1.03	F-stats	0.07	0.62
<i>p</i>	0.1522	0.3824	<i>p</i>	0.03**	0.3112	<i>p</i>	0.7929	0.4306

Notes*: F/M denotes the Few/No Change dummy. A1(%) is Classification Accuracy 1 and A2(%) is Classification Accuracy 2 – both expressed in percentage terms

The consensus classifications were obtained by intersecting all user classification polygons for a particular image and task. This produced small intersection polygons (whose areas were taken as numerator in calculating the accuracy of classification) due to outlier classifications. Hence an important take away from Table 4.6 is the very low accuracy rates obtained from the

consensus classifications. Furthermore, accuracy rates tended to be higher in the “Many” streams, providing preliminary evidence that Linus’s Law holds. Glancing over Table 4.6 also shows that the Year 2004 had better consensus estimates probably due to the fact that this was a “no-change” year as also shown in the “Change/No Change” section of the table. However, these differences are not statistically significant at baseline as shown by the p-values.

The study investigated the usefulness of measuring changes in urban land use using VGI and VHR satellite imagery. Although the sample area under investigation was self-selected, it is worth checking whether and to what extent the area occupied by different land use types changed in the aftermath of the ORO clean-up operation. To do this, the study adopted a multinomial logit specification. The results are shown in Table 4.7 and represent the marginal effect for movements in the probability of falling into various land classes from 2004 to 2006, as measured by the coefficients on time dummies.

The *Unassigned Class* was the base category (base land use type). The *Unassigned Class* (in most cases bare soil or natural vegetation) was created by subtracting the combined area of all land use types present (identifiable) on the image from the total image area. The coefficient of the *Unassigned Class* is not statistically significant, so there were no changes in this land type. *Informal Businesses* and *Backyard Structures* have coefficients of -0.0063 and -0.1672 respectively, both of which are statistically significant at one per cent. The probability of land being covered by backyard structures therefore reduced by nearly 20%. This shows, therefore, that the ORO clean-up operation had the effect of reducing⁴⁶ the size of the informal sector (in physical area terms). The coefficients of *Formal Housing* and *Formal Businesses* are 0.1229 and 0.0164 respectively, significant at the 1% and 5% levels. This signifies growth of formal housing and formal businesses, although the effect is rather small for the latter. It shows evidence of a substitution effect where evicted backyard housing dwellers (owing to ORO) ended up being settled in newly established formal housing units, as well as displaced informal business owners finally finding places to operate their businesses in newly constructed formal centres/hubs.

⁴⁶ Note: it is important to bear in mind that the areas sampled in the study were self-selected, hence this is, for all intents and purposes, a descriptive result that does not fully represent the impact of the ORO.

Table 4.7 MLogit estimation (marginal effects for the coefficient on the 2006 time dummy)

Predicted Dep. Var. Area	Marginal Effect/(SE)
Informal Businesses	-0.0063 (0.0020)***
Backyard Structures	-0.1672 (0.0382)***
Formal Housing	0.1229 (0.0102)***
Formal Businesses	0.0164 (0.0068)**
<i>N</i>	1643
pseudo R^2	0.0827

Notes*: Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, regression estimates of changes in night-time luminosity data between 2004 (pre-clean-up operation) and 2006 (post-clean-up operation) are presented in Table 4.8. The goal was to assess whether other types of economic activity are reflected by changes in urban land use types. Night Lights Data (NLD) has been used as a proxy for economic activity in many settings [see Henderson et al., (2012)], hence it may be expected to also show the displacement of informal activities.

We computed changes in average night lights luminosity for each image between 2004 and 2006. Out of the 180 images used in the study, 176 have non-missing lights data. Hence, the Lights estimates shown in Table 4.7 essentially track 88 images over the two years, 2004 and 2006. Due to a limited sample, we have low statistical power, but the results show that NLD might be relied upon to track changes in urban land use. Baseline land use characteristics were included as regressors. Only the coefficient of *Formal Housing* (-5.1008) is statistically significant (at the 1% level). Night Lights luminosity declined in areas that were occupied by formal housing, yet the marginal effects in Table 4.8 show that these areas grew in size.

Formal housing is located in areas with greater brightness at night; the decline provides an indication of the baseline reduction in economic activity between 2004 and 2006, which built up to the climax of Zimbabwe's economic crisis in 2008. Although not statistically significant, informal businesses have a coefficient of -1.0591, which correctly reflects the fact that these enterprises were displaced as shown in Table 4.7. However, the reductions do not reflect the large scale of ORO, as these areas were not sufficiently lit up to start with. Also not statistically

significant, backyard structures show an increase in night-time luminosity, while formal businesses and the no change streams suffered statistically insignificant declines, respectively.

Table 4.8 Night lights data estimates

Dep Var. Night Lights Difference	β /(S.E)
Informal Businesses	-1.0591 (2.7609)
Backyard Structures	1.6258 (2.6330)
Formal Housing	-5.1008 (1.8069)***
Formal Businesses	-2.619474 (1.6602)
No Change Streams	-0.3355 (0.7400)
Constant	-2.8037 (1.235)**
R^2	0.0751
N	76

Notes*: Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While these results may only be driven by a lack of statistical power, they highlight the importance of collecting granular land use data. Table E.1 in the appendices also show a similar result. Table E.1 shows the total area of features classified by classification type (based on expert classification). Night lights data popularly used by economists cannot track large changes in informality because, as noted earlier, informal areas tend to be poorly lit up to start with.

4.6.1 Classification accuracy estimates

A central objective of the study was to investigate how classification accuracy may be affected by different conditions and scenarios. Thus, the effect of stream and year are the main aspects that the study sought to investigate, although it was also important to capture demographic characteristics as well as image specific variation. As already discussed, the study focuses on two classification accuracy measures. These are Classification Accuracy 1 (*Accuracy_In*) and Classification Accuracy 2 (*Accuracy_Ex*). To recap, Classification Accuracy 1 captures the percentage of the reference area that is correctly classified (i.e. what percentage of the area that *should* have been classified has actually been classified?) while Classification Accuracy 2

captures the percentage of user/volunteer's classification that is correct (i.e. what percentage of what has *actually* been classified should have been classified?). Table 4.9 presents fractional probit estimates for the main effects (stream and year) for consensus Classification Accuracy 1 while the estimates for consensus Classification Accuracy 2 are presented in Table 4.10. Table 4.11 presents the effects of demographic attributes on classification accuracy (for both Classification Accuracy 1 and Classification Accuracy 2) at the individual classification level.

In Tables 4.9 and 4.10, Column (1) is the baseline specification that estimates classification accuracy on *Change Stream*, *Many Stream* and *Year* dummies. To account for a possible learning effect where the system serves 2004 images before those from 2006 or vice versa, a classification sequence (*Sequence*) is introduced from Column (2). While the *Many Stream* tests whether Linus's Law holds, the continuous variable *Number* captures the actual number of times an image was classified. *Number* is introduced to the specification from Column (3). The first three specifications are robust. Linus's Law apparently does not hold, even controlling for the actual number of times an image was classified. However, accuracy declined over time. This was not the result of the sequence in which images were presented to users.

Column (4) introduces *Year* and *Change Stream*; and *Year* and *Many Stream* interactions to the specification. Now *Many Stream* has a coefficient of 0.175, significant at 10% in the base year. The result confirms that the more there is consensus, the better the classification accuracy. The *Year 2006#Many Stream* interaction has a coefficient of -0.255, significant at a 5% level. This shows that classification accuracy was worse for the 2006 images relative to 2004 for the streams that had more users/volunteers classifying an image. This explains why Linus's Law did not hold in previous specifications.

Table 4.9 Main effects estimates (Classification Accuracy 1)

Dep. Var. Accuracy_In	(1)	(2)	(3)	(4)	(5)	(6)
Change Stream <i>*base = no change stream</i>	-0.034 (0.077)	-0.034 (0.077)	-0.050 (0.071)	0.009 (0.096)	-0.046 (0.078)	0.005 (0.087)
Many Stream <i>*base = less stream</i>	0.109 (0.070)	0.101 (0.072)	0.092 (0.074)	0.175* (0.095)	0.125 (0.117)	0.141 (0.122)
Year 2006 <i>*base = year 2004</i>	-0.19*** (0.065)	-0.19*** (0.066)	-0.19*** (0.065)	0.037 (0.091)	-0.061 (0.087)	-0.054 (0.086)
Sequence		-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.002)	-0.002 (0.002)	-0.003 (0.002)
Number			0.014 (0.017)	0.011 (0.018)	0.011 (0.017)	0.019 (0.017)
Year 2006#Change Stream				-0.186 (0.134)	0.025 (0.128)	-0.038 (0.133)
Year 2006#Many Stream				-0.26** (0.127)	-0.084 (0.152)	-0.096 (0.152)
Change Stream#Many Stream					0.082 (0.160)	0.030 (0.161)
2006#Change#Many Stream					-0.54** (0.273)	-0.507* (0.272)
Backyard Structures <i>*base = Informal Business</i>						-0.220 (0.138)
Formal Housing						-0.002 (0.118)
Formal Businesses						-0.108 (0.112)
Constant	-4.15*** (0.051)	-4.06*** (0.064)	-4.09*** (0.085)	-4.15*** (0.101)	-4.12*** (0.090)	-4.09*** (0.146)
<i>N</i>	379	353	353	353	353	353
pseudo <i>R</i> ²	0.009	0.009	0.009	0.013	0.015	0.020

Notes*: Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One possible interpretation is that large changes, which reduce commonly observed objects to small and scattered patches, reduce classification accuracy. Informal structures identifiable on 2004 imagery were absent on the 2006 imagery, opening up room for confusion on the part of the volunteers since they were expecting to see these structures in the 2006 imagery. The volunteers could not have had the intuition that these structures had been destroyed in the course of ORO, since they were not aware of the experimental design. This interpretation is reinforced in Columns (5) and (6) wherein the tri-interaction of *Year*, *Change Stream* and *Many Stream* is -0.541 and -0.507 respectively and all other coefficients are insignificant. We did not observe differences in classification accuracy by baseline land use type. Areas that experienced changes due to ORO had less accuracy in 2006; the inaccuracy was amplified if the image was classified often. Consensus views therefore reduce precision if the objects to be classified are small, irregular in shape and not easy to view. Linus's Law does not hold in these circumstances.

Table 4.10 presents fractional probit estimates of the accuracy estimates of Classification Accuracy 2 (the percentage of what has been classified that should have been classified). The results obtained in Table 4.10 are similar to those in Table 4.8. As in Table 4.9, the coefficient of Year 2006 is negative throughout, but is only significant (at 10%) in Column (3). In Columns (5) and (6), the triple interactions of *Year*, *Change Stream* and *Many Stream* reaffirm that areas whose land use types changed (in other words, informal structures were destroyed) recorded less classification accuracy relative to 2004. Very little else is significant. In Column (6), *Formal Housing* has a coefficient of 0.236 (significant at 10%), which shows that users struggled less in classifying formal housing relative to informal businesses. An important implication is that VGI projects are better able to correctly identify land use classifications if the target objects remain large in representation and are regularly shaped. Many classifications do not improve categorisation in these settings, but rather reduce them. Also important to note is that in comparison to the demographic attributes estimates in Table 4.11, the consensus estimates are much more limited sample wise since the individual volunteer classifications for the latter were collapsed/intersected together at image level to arrive at a single group classification as indicated earlier; unlike in the former where analysis is at an individual level.

Table 4.10 Main effects estimates (Classification Accuracy 2)

Dep. Var. Accuracy_Ex	(1)	(2)	(3)	(4)	(5)	(6)
Change Stream <i>*base = no change stream</i>	0.171 (0.160)	0.171 (0.164)	0.126 (0.205)	0.209 (0.256)	-0.175 (0.144)	-0.217 (0.163)
Many Stream <i>*base = less stream</i>	0.199 (0.174)	0.185 (0.168)	0.161 (0.191)	0.374* (0.224)	0.026 (0.144)	0.021 (0.153)
Year 2006 <i>*base = year 2004</i>	-0.220 (0.143)	-0.205 (0.131)	-0.213* (0.124)	0.280 (0.212)	-0.054 (0.136)	-0.048 (0.141)
Sequence		0.003 (0.004)	0.004 (0.003)	0.003 (0.003)	0.003 (0.003)	0.003 (0.003)
Number			0.033 (0.034)	0.027 (0.034)	0.024 (0.034)	0.014 (0.045)
Year 2006#Change Stream				-0.230 (0.239)	0.299 (0.190)	0.362 (0.233)
Year 2006#Many Stream				-0.66*** (0.236)	-0.093 (0.202)	-0.095 (0.209)
Change Stream#Many Stream					0.514* (0.302)	0.539 (0.348)
2006#Change#Many Stream					-1.19*** (0.364)	-1.21*** (0.378)
Backyard Structures <i>*base = Informal Business</i>						0.427 (0.310)
Formal Housing						0.236* (0.132)
Formal Businesses						0.187 (0.164)
Constant	-3.85*** (0.098)	-3.95*** (0.222)	-4.04*** (0.158)	-4.20*** (0.253)	-3.94*** (0.121)	-4.14*** (0.205)
<i>N</i>	379	353	353	353	353	353
pseudo <i>R</i> ²	0.021	0.022	0.024	0.039	0.047	0.060

Notes*: Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6.2 Effects of demographic attributes estimates

Table 4.11 Demographic attributes estimates

	Dep. Var. Accuracy_In				Dep. Var. Accuracy_Ex			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sequence	-0.000 (0.001)	0.003** (0.001)	0.003*** (0.001)	0.004*** (0.001)	-0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.002)
Coloured Race <i>*base = Black Race</i>			0.102 (0.154)	0.116 (0.152)			0.287* (0.167)	0.242 (0.160)
White Race			0.788*** (0.168)	0.802*** (0.160)			0.770*** (0.190)	0.767*** (0.185)
South Africa <i>*base = Malawi</i>			-0.326 (0.304)	-0.323 (0.299)			-1.13*** (0.347)	-0.78** (0.368)
Zimbabwe			-1.14*** (0.380)	-1.15*** (0.368)			-0.84** (0.425)	-0.775* (0.438)
Male <i>*base = Female</i>			-0.147* (0.077)	-0.17** (0.075)			-0.4*** (0.098)	-0.25** (0.104)
Household Income			-0.023 (0.018)	-0.024 (0.017)			-0.022 (0.020)	-0.038* (0.021)
Lived Informal Area			-0.010 (0.139)	0.006 (0.135)			0.294* (0.158)	0.187 (0.161)
Visit Informal Area			0.537*** (0.131)	0.524*** (0.126)			-0.014 (0.138)	-0.014 (0.138)
Backyard Structure <i>*base = Informal Business</i>				-0.33** (0.137)				0.628*** (0.094)
Formal Housing				-0.74*** (0.131)				2.189*** (0.109)
Formal Businesses				-0.40*** (0.128)				0.622*** (0.093)
Constant	0.200 (0.320)	-0.3** (0.139)	0.221 (0.436)	0.990** (0.447)	-0.258 (0.287)	-0.6*** (0.138)	0.437 (0.397)	-1.56*** (0.339)
FEs: Image	Y	N	Y	Y	Y	N	Y	Y
User/Volunteer	N	Y	N	N	N	Y	N	N
Class Type	N	N	N	Y	N	N	N	Y
Province	N	N	Y	Y	N	N	Y	Y
<i>N</i>	808	808	808	808	1612	1612	1612	1612
pseudo <i>R</i> ²	0.101	0.048	0.139	0.148	0.103	0.033	0.118	0.305

Notes*: Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The analysis now turns to participant inequality and analyses the accuracy of each *individual* classification task and not the image-level intersections. In Table 4.11, Columns (1) to (4) show demographic relationships with Classification Accuracy 1, while those for Classification Accuracy 2 are presented in Columns (5) to (8). The sample for the first set of estimates is about half that of the second, because the former is calculated based on the reference classification [with reference classification as the denominator (see Figure 4.5)]. In calculating Classification Accuracy 2 (*Accuracy_Ex*), the area of the user classification is taken as the denominator. Now, users erroneously classified more tasks than were present on an image (in as far as the reference classifications are concerned). The interpretation of Classification Accuracy 2 (*Accuracy_Ex*) is the percentage of the area classified that should have been classified. Hence there are more user classifications' polygon areas (denominator in Classification Accuracy 2) than there are reference classifications' polygon areas (denominator in Classification Accuracy 1). This explains why *N* is 808 for Columns (1) to (4) and 1 612 for Columns (5) to (8) in Table 4.11.

So far, it has been shown that classifications were worse in 2006 due to uncertainty among the volunteers since many informal structures existing in 2004 had vanished over the two-year period due to ORO; yet volunteers expected to see these informal structures since they were unaware of the experimental setup. Another possibility might be the existence of a “learning effect” wherein the Zooniverse platform could have served 2006 images first, which might have meant that the volunteers would have been better at classification by the time the 2004 images were served. Hence in Columns (1) and (5), Classification Accuracy 1 and Classification Accuracy 2, respectively, are regressed on *Sequence*, with image fixed effects (FE) included. In Columns (2) and (6), image FEs are replaced with user FEs. For Column (2), a coefficient of 0.003 is obtained for *Sequence* and is significant at 1% level. Hence a learning effect can be detected at the individual user level. Hence, training of participants is crucial in VGI projects and a testing phase should be conducted in advance of the experiment. In Columns (3) and (7), other demographic variables, including race, nationality, sex at birth, whether one ever lived in an informal area and whether one ever visited an informal area, come through as important in determining classification accuracy. The coefficients remain largely robust in Columns (4) and (8), with the only addition to the specification being the inclusion of class type to the specification.

In the experiment, *white* volunteers attained higher classification accuracy. In comparison to blacks, *White Race* has coefficients of 0.788, 0.802, 0.770 and 0.767 (all significant at 1%) for

Columns (3), (4), (7) and (8), respectively. In terms of nationality, Zimbabweans attained a worse classification accuracy in comparison to Malawians. A feasible explanation might be that the phenomenon of informal structures is even more widespread in Malawi relative to Zimbabwe, hence Malawians would have had more exposure to it. At the same time, it should be noted that the sub-samples for both Malawians and Zimbabweans are very small as compared to South Africans who dominate the sample. Hence the results should be interpreted with caution. For Columns (3), (4), (7) and (8), *Zimbabwe* has coefficients of -1.14, -1.15, -0.841 and -0.775 (all significant at 1% level except for Column (8) which is significant at 10%). South African volunteers performed worse than their Malawian counterparts only for *Accuracy_Ex* (Classification Accuracy 2). *South Africa* has coefficients of -1.13 and -0.78 respectively, significant at 1% and 5% for Columns (7) and (8), also respectively. Across the board, *Male* students performed worse than their female counterparts.

Column (8) shows a coefficient of -0.038 (significant at 10%) for *Household Income*. The higher the household income of the volunteer, the worse their classification accuracy. Higher income households might have less exposure to informal settlements/informal business settings. Whether one has ever lived in an informal area (*Lived Informal Area*) has a coefficient of 0.294 (significant at 10%) for Column (7), which confirms the assertion that exposure to informal settings is beneficial for accuracy. The coefficients of whether one has ever visited an informal area (*Visit Informal Area*) for Columns (3) and (4) are 0.537 and 0.524 respectively, which again reinforces better classification accuracy for volunteers who have had exposure to informal structures. Lastly, the negative coefficients for *Backyard Structures* (-0.326), *Formal Housing* (-0.74) and *Formal Business* (-0.40) in Column (8) (significant at 1% except for the one for backyard structures) show that classification accuracy was worse for all other land use types relative to informal businesses. Conversely, the results in Column (8) indicate that all other land use types were better classified relative to informal businesses. If the results in Columns (4) and (8) are viewed side to side, then it may imply that volunteers who could identify informal business structures were able to classify this particular land use type more accurately [in Column (4)], whereas those who struggled to identify informal structures classified other land use types better than informal businesses.

4.7 Conclusion and Recommendations

A number of findings from the study are useful for appreciating the conditions under which VGI/crowdsourcing projects could achieve the best classification accuracy. Assigning more volunteers/classifications per image resulted in worse classification accuracy when the target objects were removed after the ORO; they were more sparsely distributed, small and irregularly shaped, making them more difficult to identify. Adding more classifications in this setting reduced rather than improved consensus, so that the conventional wisdom put forth by Linus's Law does not hold. Classification of objects that are difficult to identify should therefore be approached with caution.

The study also found that volunteers attained lower classification accuracy in identifying changes in informality and other land use types, as compared to identifying land use features in no change areas. Importantly, the study highlights the importance of the learning effect, where volunteers get better at classification over time. This amplifies the need for rigorous training to improve classification accuracy. Findings from the study also suggest that race, nationality, sex at birth and other demographic characteristics do affect classification accuracy. Importantly, the study established that volunteers who once visited informal areas attained better classification accuracy compared to those who had not had that exposure. Hence, participant background information is a key determinant of classification accuracy. VGI studies of informality therefore cannot rely on the opinions of the public at large, but require the informed participation of individuals who have lived or worked in informal settlements and businesses. It is possible that using VGI with a select group of experienced participants is successful in building datasets that can be trained by machines to produce better classification.

The study suggested some substitution between the informal and formal sectors of the economy, wherein people/business owners who were displaced during the 2005 clean-up operation were later settled in formal residential or business areas. The logit estimates confirm the scale of the destruction brought about by ORO on Zimbabwe's urban landscape, which reinforces findings by Tibaijuka (2005) as well what the Afrobarometer 2005 survey showed. Also, there was significant variation in accuracy depending on category [as shown in Table 4.9 and 4.10, Column (7)]. This finding concurs with Foody et al. (2013) who also found class-specific differences in the quality of information obtained via VGI.

Considering the relatively low *consensus* classification accuracy rates [especially mean *Accuracy_Ex* in Table 4.5 (a) and Table 4.5 (b)] generated in this study, it is clear that

identification of informality and other land use types through citizen science has a few shortcomings. Some of these may include the difficulty faced by volunteers in identifying small objects on images with poor quality; furthermore, uncertainty may arise when some land classes become abruptly sparse after a large natural or human-induced shock (as was the case after ORO).

There are a few strategies that could potentially be implemented to better the quality of results in the future. The low accuracy rates shown in Tables 4.5 and 4.6 potentially signal the need for better participant training methods. Fritz et al. (2017), note that the Geo-Wiki platform could be used to train participants in crowdsourced projects. However, a potential setback may be that the platform might not have sufficient coverage of developing countries where informality is a massive phenomenon. Given the extent to which polygons drawn by volunteers extended beyond the boundaries of the reference polygon, it may be useful to get users to drop pins on various land use types as a way of identification rather than use the Bezier tool. Despite the low (consensus) to moderate (individual) rates of accuracy obtained in the study, it nevertheless provides an important starting point for future VGI projects. This is important for a number of reasons. For instance, in a typical citizen science/VGI project, researchers do not always have the privilege of having volunteers who are expert at the classification of the subject matter at hand. Training volunteers in this chapter consisted of online video tutorials as well as further tutorials/hints built into the Zooniverse classification platform itself. These are all standard approaches to upscaling the knowledge of usually untrained volunteers. The training that the volunteers who took part in the classification activities of this chapter received is to a large extent as much as any other VGI project could offer.

More broadly, this chapter showed that citizen science may not be appropriate where objects are small and not easily distinguishable. Other approaches, such as machine learning are required, but rely on the availability of very high resolution images. Even machine learning approaches are not perfect, thereby necessitating the need for composite type of approaches. Given that citizen science projects in most cases attract untrained “novices”, it is problematic because the consensus is affected by outliers and high misclassification error. This study is important because reliance on out-of-the-box and unconventional datasets may need to be considered with caution (as the results of this chapter showed) and it is imperative for economists to be cautious about the limitations of the data at their disposal and being more involved in the data generating process.

CHAPTER 5

CONCLUSIONS AND RESEARCH IMPLICATIONS

5.1 Introduction

The majority of people (about 70%) in sub-Saharan Africa (SSA) live in rural areas and rely on agriculture as a form of livelihood – which makes increased farm productivity key to reducing poverty and inequality. Private property rights (in the form of land titles) have been touted as a key part in unlocking the potential of the agricultural sector on the continent. Sub-Saharan African governments have engaged in various forms of agrarian reform to address historical land access imbalances, improve trust, open up agricultural credit and create land markets. This thesis identifies and investigates two missing links. First, access to markets is a precondition for land titles to result in increased crop cultivation; second, property rights that remain unenforced in the wake of land reforms are ineffectual. The thesis highlighted the pivotal role that security of tenure also plays in urban areas. It pointed out the need for urban councils to find cost effective ways for creating cadastres and land use maps and investigated the conditions under which citizen science/crowdsourcing may be considered a feasible option.

5.2 Research Contributions

The research makes an important contribution to the socio-economic discourse in SSA by explaining some of the unintended consequences of state intervention in the use of land for cultivation. First, in Chapter 2 the study notes the scenario in which land titles did not influence agricultural growth in SSA in expected ways. Second (in Chapter 3), it unravels why state-led land expropriations have not yielded the expected benefits as envisaged by their proponents. Land reforms without access to markets and with imperfect enforcement of property rights have zero to negative aggregate effects, as was shown in Chapters 2 and 3, respectively.

Land titles have not translated to increased agricultural growth in SSA (Udry, 2011), as initially expected. This outcome potentially contributed to the continued slow pace of tenure security reforms in the region. However, land reforms might only take effect when the conditions are right. Using the case of Rhodesia, Chapter 2 of the thesis investigates whether there exists a complementarity between market access and land titles. The study finds that access to markets offers one key missing link. High costs of transportation could outweigh the benefits of having access to land if the farmers in question are located far from major metropolises and without

access to good transport networks (main roads and railways). Hence, market access is an important channel through which land titles contribute to improving household welfare. Land titling is futile in motivating agricultural productivity if the infrastructure and markets are underdeveloped. However, these conditions are more difficult for policymakers to change (as they entail building new, extensive transport networks for a start), compared to faster legislative processes that give land titles and expropriate land. This could explain why land reforms have been ineffective in some settings, but have nevertheless been implemented as a political prerogative.

Chapter 3 presents the first quasi-experimental and nationwide estimation of the effects of Zimbabwe's 2000 FTLRP on crop cultivation. Previous studies have focused only on small geographic areas within the borders of the country and often relied on non-representative qualitative surveys. As a result, there has not been a general macro-level assessment of how the land reform programme affected crop cultivation and crop quality; while it has been known that the effects were negative, this study presents the first causal estimates. The results of Chapter 3 do not dismiss the need for land reform; however, they provide caution on the approach taken in pursuing land reform, particularly as it relates to guaranteeing that property rights are enforced in the post reform phase. Issuing bankable land titles can play a role in enhancing the credibility of reforms. These findings provide crucial insights and learning points for other countries within SSA including South Africa and Namibia, and beyond. The findings of Chapter 3 may form a reference point from which nations can borrow in order to protect the viability of agriculture, maintain the efficiency of the land and agriculture credit markets, and to promote trust.

Throughout the thesis, an important sub-theme is the data scarce context in which measurement of socio-economic change in SSA takes place. All the papers in the thesis employ novel approaches in an attempt to contribute to a cure for the "African Statistical Tragedy", a terminology that is widely touched upon in the text. The application of machine learning on Landsat imagery to generate crop cultivation data is not a first, but is significant in the sense that most remote sensing studies analyse imagery for small geographic areas, or test the precision of machine learning algorithms. In this thesis crop cultivation data is generated from remotely sensed satellite images for the entire country in both historical (Chapter 2) and contemporary (Chapter 3) settings and then goes further to adopt quasi-experimental econometric techniques in order to obtain causal effects in Chapter 3. Novel datasets on crop

cultivation enable research that answers questions pertinent to SSA's socio-economic development.

In Chapter 4, the measurement of socio-economic change in SSA in a data scarce context comes full circle. The chapter presents citizen science approaches as an alternative way of extracting useful information from VHR satellite imagery. While this is not the first time this application has been used to analyse land use, the research contributes to understanding the conditions under which the citizen science approach can be employed to detect/classify informal structures (both commercial and residential) with better precision. Traditionally, the understanding is that the consensus of many can improve classification; however, this conventional wisdom does not hold in the examination of informal structures. The more individuals classify an image, the greater the change for consensus to be concentrated in only very small areas; the result is that only a very low proportion of land is appropriately detected in certain classes.

A massive 2005 clean-up operation in Zimbabwe that destroyed informal structures made this experiment possible. This operation enabled the creation of treatment and placebo areas in which “change” and “no change” detection was expected, respectively. By laying bare some of the issues that need consideration in the employment of citizen science to identify informal structures and other land use types, the study makes an important contribution towards the cost effective creation of cadastres by urban councils for security of tenure enforcement.

5.3 Summary of Findings

5.3.1 The complementarity between property rights and market access for crop cultivation in Rhodesia: Evidence from historical satellite data

Udry (2011), bemoaned the absence of strong evidence on the benefits of land titles in the SSA region. In Chapter 2 using the case of Rhodesia, we show that access to markets is a necessary precondition for land titles to result in increased crop cultivation. The chapter uses 1970s crop cultivation data – as the debate on the importance of land titles in SSA intensified in this period. For identification, we relied on Rhodesia's Land Apportionment Act (1930).

The study compares the proportion of land under crop cultivation of African customary tenure wards [TTAs] to that of their fellow counterparts in the NPAs who had land titles. Estimates show that African NPA farmers had more land under crop cultivation as long as they were located close to main cities, railways and main roads. The results remain robust after

controlling for population, as well as climatic and soil variability. These findings emphasise the importance of access to markets for land titles to improve agricultural production in SSA. As was also highlighted in the findings, the proportion of land under cultivation was greater in EAs, which is consistent with the literature that there was population pressure and competition for resources in the customary tenure areas (the TTAs).

5.3.2 Effects of poor land rights enforcement post-redistribution on cultivation and crop quality: Satellite data evidence from Zimbabwe

Lack of property rights enforcement post-agrarian reform has been found to be an impediment to agricultural growth (Deininger et al., 2008). Hence, Chapter 3 investigates this using Zimbabwe's 2000 FTLRP as a case study. This study uses NLD, the proportion of land under crop cultivation and NDVI as proxies for the endogenous variables. In this study, NLD correlates highly with the Zimbabwe Statistical Agency (Zimstat) Census 2011 poverty estimates at the ward level and therefore serves as a proxy for welfare. NDVI is a proxy for crop quality since FTLRP likely affected the extent of application of fertilisers and chemicals to crops given that the newly resettled indigenous farmers were less endowed financially than the previous landowners of European descent.

The spatial empirical strategy relied on treatment groups defined by Rhodesia's Land Apportionment Act (1930) as in Chapter 2. This is because Zimbabwe's land access and racial occupation of the land had remained almost the same as at 2000. Hence, European Areas (EAs) (the areas affected by FTLRP) were considered as the treatment region while the TTAs and NPAs were regarded as the control regions (since indigenous Zimbabwean farmers occupied these). An additional set of estimates considered European Area wards in the Eastern Highlands as the control group and European Areas elsewhere as the treated group, given the indication by Mutangi (2010) that the Eastern Highlands (the mountainous region bordering Mozambique) was mostly spared in the land reform exercise owing to its specialised kind of agriculture (fruit, coffee and other plantations). For identification, the study adopts DID and Spatial Regression Discontinuity Design (S-RDD) [using distance of the ward to the border of the treatment region (EAs)] designs in order to delineate the causal effect.

The specifications control for the following: climate and soil variability; regional imports and exports (to control for exchange rate shocks); agro-ecological fixed effects; and image specific fixed effects (to account for sensor and atmospheric noise at the time of image capture). After adding controls, night-time luminosity did not produce any significant coefficients for the

causal effects, even for the treatment and post-intervention dummies of the DID and RDD estimates. It follows, therefore, that NLD may be not be a viable source of data for measuring socio-economic change in rural areas in developing countries, as these areas are often under-electrified relative to the actual economic activity that takes place. The DID and RDD estimates show a decline in the proportion of land under cultivation and crop quality in European Areas. Hence, the causal effect of the FTLRP is negative. One explanation is government's failure to fully assert property rights for the resettled A1 and A2 farmers. A1 permits, for the smaller farms were inheritable but not marketable, although A2 permits (for the larger farms) had a clause that allowed using the land as collateral. However, Zikhali (2008) notes that the conditions under which the land could be used as collateral were unclear, hence it remained unbankable.

5.3.3 A citizen science approach to classifying urban informality and other urban land use types using satellite imagery

Continuing with the “African Statistical Tragedy” sub-theme, Chapter 4 recognises that urban informality is a widespread phenomenon in SSA and governments could make cost savings through building urban cadastres from VHR satellite imagery. Instead of machine learning, the chapter adopts a crowdsourcing/citizen science approach to classify urban informality and other land use types in Harare, the capital of Zimbabwe. An important aspect that is central to the development of land use maps through crowdsourcing is class noise or simply misspecification or classification imprecision. For urban councils to be able to rely on citizen science to create cadastres (which could be used, in turn, to enforce security of tenure), it is important that these maps are accurate.

Chapter 4 therefore investigates the conditions affecting the accuracy of classification. The chapter relied on the classification efforts of a group of 41 SU students. The students were allotted into two groups, with the first containing 2004 and 2006 images for areas that were affected by a 2005 clean-up operation [Operation Restore Order (ORO)] and the other consisting of imagery over areas that were not affected. The objective of creating the latter group was to have a placebo in which no change detection was expected. Further, these two groups were sub-divided, whereby in one sub-group an individual image was supposed to be classified five times and ten times in the other. The objective was to investigate whether more classification (more consensus) improved classification precision (Linus's Law).

Some students classified more than the required number of images, while some classified less than the required number because the classification online platform did not offer any way of controlling the number of classifications one student could perform, although it was impossible for a student to work on imagery from a different stream. This phenomenon is referred to as participant inequality and it affected the intended setup of the experiment. The chapter was, however, able to answer the objectives since the fractional probit specification controlled for the effect of the number of classifications per image; learning (effect of experience and training); as well as “change/no change” and “few/many” stream/group dummies (main effects).

Descriptively, the consensus classification accuracy rates were very low. The consensus was the intersection of all the user classifications for a particular image and task number (name). Hence, outlier classifications greatly reduced the intersection area – which explains the resultant very low classification accuracy. The fractional probit estimates for the main results (consensus classifications) showed better accuracy for the streams with more volunteers classifying an individual image, whereas year 2006 was found to have worse accuracy compared to 2004. The interaction effects showed that there was worse consensus accuracy for areas that had experienced changes in land use types because the target structures were removed, covered very small areas and were irregularly shaped. An important conclusion is, therefore, that citizen science is fraught with difficulty when users have to identify changes in small informal structures. The individual level fractional probit estimates showed that demographic attributes of volunteers also affect classification accuracy, as does the learning effect. Volunteers should be well-trained and familiar with the environment if they are to obtain useful information from crowdsourcing. As such, citizen science may be best suited to build smaller datasets that train machine learning algorithms which can classify larger geographic areas automatically.

5.4 Research Implications

Borrowing from Rodrik (2006), the arguments proposed and conclusions drawn in the thesis advocate for “need for humility, for policy diversity, for selective and modest reforms and for experimentation”. The conclusions do, however, have important policy implications in as far as moving towards well-defined property rights for poverty alleviation in both rural and urban spaces.

The Zimbabwean case showed that state-led land expropriation suffers the penalty of productive efficiency when property rights of new landowners are not appropriately enforced or they do not support private ownership. While Otsuka (2007) suggests that taxes and subsidies may be a better alternative to land reform, enforcement of property rights and improvements of transport infrastructure (roads and railways) and market access are important issues for consideration at the policy level. It is hoped that this research will help to ensure that other developing countries (particularly those in SSA) who are either facing the need for, or engaged in, the land reform debate may learn something from the case of Zimbabwe. The results of the thesis show that without a clear strategy for securing property rights through land titling, countries risk plunging their economies into crisis – especially if agro-industrial linkages are prominent. Aside from safeguarding property rights, the thesis showed the importance of access to markets as a precondition for land titles to translate to increased crop cultivation. Thus, policymakers may be interested in ensuring that areas which are earmarked for land reform are close to main cities, main roads and railway lines if land titles are to have a positive effect on production.

While the research illustrated the utility of machine learning for land use analysis, the thesis also put forward citizen science/crowdsourcing as a cost effective aid in the development of urban cadastres. Property rights are also important even within an urban setting (and not just for promoting the conditions for agricultural production), especially given that SSA and other developing countries are characterised by sprawling informal commercial establishments and settlements (including backyard structures). These establishments/structures are unregistered, which means that there is no security of tenure. In turn, this results in huge opportunity costs, aside from the direct cost of overburdening existing urban formal amenities and infrastructure. Citizen science could play an important role in aiding urban services management planning and cadastre development, given that some of the factors affecting classification accuracy are known and can be taken into consideration as was investigated in Chapter 4. However, this thesis has highlighted that reliance on non-experts to develop these data sources should proceed with caution – especially in identifying informal structures. This leaves open the role for investigating better methods that involve computerised image recognition.

5.5 Conclusions and Suggestions for Future Research

The thesis reasserts the importance of property rights in the form of land titles within SSA's rural sector. This is important given that the majority of people in the region live in rural areas

and directly depend on agriculture as a source of income. By adopting policies that promote increased agricultural production, SSA and other developing regions can effectively reduce poverty and narrow the gaps in inequality. Central to these policies is the issue of property rights in unlocking farm credit, and promoting and maintaining trust, as well as ensuring efficient functioning of land markets so that a land parcel is traded openly until the most efficient producer owns it. An important caveat here is that asserting modern property rights in SSA is a departure from the traditional/precolonial institutions. Modern property rights are a new (relative to traditional or customary tenure) and colonial concept that is not necessarily amenable to all society. Hence, the move towards modern property rights is expectedly fraught with implementation challenges and friction since it takes land power away from the traditional custodians.

Moreover, the importance of property rights in the form of land titles does not represent an end in itself, given that there are other associated factors that need consideration as well. A key issue is access to markets, as demonstrated in Chapter 2. Where property rights are existing, or they have been asserted/restored, the next step is to ensure there is transportation infrastructure assets in the form of main roads and railways. In short, market access is a necessary precondition for land titles to translate to increased crop cultivation. A potential area on which future researchers may want to focus their attention is an analysis of the same problem using recent data. The 1970s and early 1980s, the period of analysis in Chapter 2, was a remarkably different time than the present. SSA's economies have undergone different kinds of structural transformation (for example a number of economies have rebased to take cognisance of the growth of the services economy – for example, Nigeria rebased its economy to reflect the true contribution of the entertainment industry). However, building infrastructure is a difficult task that takes years; legislative processes that introduce land reforms can take place comparatively quicker. While land titling can improve production, it is often adopted for political reasons without the necessary preconditions in place for it to become successful. These political economy trade-offs weaken the potential for land reforms to achieve their intended goals.

Jayne et al. (2003), show landlessness to be pervasive in SSA. Hence, redistribution is a given, especially in contexts where colonial institutions created imbalances in access to land. Chapter 3 discussed the benefits of land reform, as well as the motivations behind it. Save for political motivation, where elites intend to ring-fence their own interests, there might indeed be a case for land reform from efficiency, equity and poverty alleviation perspectives. The approach of implementing agrarian reform remains a key factor in informing the success of land reform.

Even more important, property rights have no effect if they are not fully asserted in the post-phase. Owing to the absence of property rights enforcement after Zimbabwe's 2000 land reform exercise, the country witnessed reductions in areas under crop production as well as in the quality of the crops as measured by NDVI.

For both Chapters 2 and 3, there are several aspects that future researchers may want to concentrate more on. The first is that FTLRP took place in 2000 and it was not possible to go back in time and create training data that would allow improvement from the binary classification to one in which different crops types are identified. Crop type may be an important confounding attribute on the amount of land under crop cultivation as well as crop quality. This was not accounted for due to the fact that training data for specific crops is difficult to create without the benefit of historical ground-truthing. The training data for crop cultivation was generated based on expert analysis. However, future researchers working on a similar problem set may want to control for crop type if the data is available or if their classification approach can differentiate land under cultivation by crop type. Additionally, the results for Chapters 2 and 3 do not consider the quantum and extent of livestock farming per ward given that this data was unavailable. This is another area that future research may improve upon. Lastly, the study showed that NLD might not be a viable proxy for measuring socio-economic change in rural areas in SSA given the low rates of electricity proliferation; hence, future researchers may restrict their use of NLD to urban areas for SSA.

Chapter 4 of the study shows that the number of times an image is classified, and learning and experience, as well as the demographic attributes of the user such as nationality, gender, race and whether or not they had first-hand experience of the land use type in question are important in determining classification accuracy. Armed with this knowledge and the respective directions that these factors affect classification accuracy, urban councils may begin examining how they can cost effectively create cadastres through citizen science. However, the very low accuracy rates that were obtained for the consensus classifications show the significant level of caution that is required, particularly where classification involves changes in land use management, and structures are small (such as informal structures). Thus, future researchers may be interested in comparing the citizen science classification accuracy generated in this study to the accuracy of the classification of a machine learning algorithm for the same areas in Harare between 2004 and 2006.

Despite the limitations discussed in the preceding paragraphs, the thesis set out and answered important policy questions that are relevant for the socio-economic development in SSA. Poverty reduction and the narrowing of inequality gaps are central ambitions within the region's socio-economic development. In an agro-centric region such as SSA, access to markets and property rights (in both rural and urban areas) are important issues that need sufficient attention, especially at the policy level. Therefore, the thesis took advantage of different land policies/interventions in Zimbabwe in order to investigate how access to markets and land titles, as well as the lack of tenure security enforcement post-land reform affect agricultural activity using the area under crop cultivation as a proxy. In addition, given the importance of tenure security in urban areas, the thesis also investigated the extent to which citizens can be used in the development of urban cadastres. In answering the research objectives, the thesis innovatively borrowed, customised and applied machine learning, citizen science, geo-processing and image analysis skills and techniques from other disciplines to create data and then analysed this data using up to date econometric procedures to effectively measure socio-economic change in SSA.

References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1-19.
- Abdulai, A., Owusu, V. & Goetz, R. (2011). Land tenure differences and investment in land improvement measures: Theoretical and empirical analyses. *Journal of development Economics*, 96(1), 66-78.
- Abu, B. M., Issahaku, H. & Nkegbe, P. K. (2016). Farmgate versus market centre sales: A multi-crop approach. *Agricultural and Food Economics*, 4(1), 21.
- Acemoglu, D., Johnson, S. & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.
- Adams, M. E. (1995). Land reform: New seeds on old ground? (Vol. 6): Overseas Development Institute London.
- Adams, M., Sibanda, S. & Turner, S. D. (1999). *Land Tenure Reform and Rural Livelihoods in Southern Africa*. London: Overseas Development Institute.
- Adjognon, S. G., Liverpool-Tasie, L. S. O. & Reardon, T. A. (2017). Agricultural input credit in Sub-Saharan Africa: Telling myth from facts. *Food Policy*, 67, 93-105.
- Ahmad, S., Kalra, A. & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, 33(1), 69-80.
- Ahmed, U., Ying, L., Bashir, M., Abid, M., Elahi, E. & Iqbal, M. (2016). Access to output market by small farmers: The case of Punjab, Pakistan. *Journal of Animal and Plant Sciences*, 26(3), 2016.
- Albertus, M., Diaz-Cayeros, A., Magaloni, B. & Weingast, B. R. (2012). *Authoritarian Resiliency and Poverty Traps: Land Reform in Mexico*. Stanford, CA: Department of Political Science, Stanford University. Retrieved from: www.stanford.edu/~magaloni/dox/2012landgrowth.pdf
- Albuquerque, J., Herfort, B. & Eckle, M. (2016). The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing*, 8(10), 859.
- Alston, L. J., Libecap, G. D., & Schneider, R. (1996). The determinants and impact of property rights: Land titles on the Brazilian frontier. *The Journal of Law, Economics, and Organization*, 12(1), 25-61.
- Amanor-Wilks, D.-E. (2009). Land, labour and gendered livelihoods in a “peasant” and a “settler” economy. *Feminist Africa 12 Land, Labour and Gendered Livelihoods*.
- Andersson, M. & Green, E. (2016). Development under the surface: The rise of indigenous agriculture as an unintended consequence of settler institutions in Rhodesia, 1896–1962. *Journal of International Development*, 28(1), 127-146.

- Arimah, B. C. & Branch, C. M. (2011). *Slums as expressions of social exclusion: Explaining the prevalence of slums in African countries*. Paper presented at the OECD International Conference on Social Cohesion and Development, Paris.
- Arrighi, G. (1967). *The Political Economy of Rhodesia* (Vol. 16). The Hague: Mouton.
- Arrighi, G. (1970). Labour supplies in historical perspective: a study of the proletarianization of the African peasantry in Rhodesia. *The Journal of Development Studies*, 6(3), 197-234.
- Arsanjani, J. J. & Vaz, E. (2015). An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. *International Journal of Applied Earth Observation and Geoinformation*, 35, 329-337.
- Aryeetey, E. & Udry, C. (2010). Creating property rights: Land banks in Ghana. *American Economic Review*, 100(2), 130-134.
- Ashton, T. S. (1997). *The industrial revolution 1760-1830*. OUP Catalogue.
- Atack, J., Jaremski, M., & Rousseau, P. L. (2014). American banking and the transportation revolution before the Civil War. *The Journal of Economic History*, 74(4), 943-986.
- Austin, R. (1975). *Racism and Apartheid in Southern Africa: Rhodesia: A Book of Data*. Paris: Unesco Press.
- Auty, R. M. (1994). Industrial policy reform in six large newly industrializing countries: The resource curse thesis. *World Development*, 22(1), 11-26.
- Bambio, Y. & Agha, S. B. (2018). Land tenure security and investment: Does strength of land right really matter in rural Burkina Faso? *World Development*, 111, 130-147.
- Banerjee, A., Duflo, E., & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China (No. w17897). *National Bureau of Economic Research*.
- Barracough, S. L. (1999). *Land Reform in Developing Countries: The role of the state and other actors*. Geneva: United Nations Research Institute for Social Development.
- Batini, N., Levine, P., Kim, Y.-B. & Lotti, E. (2010). *Informal labour and credit markets: A survey*. IMF Working Paper, 1-41.
- Becker, K. F. (2004). *The Informal Economy: Fact Finding Study*. Stockholm: Sida.
- Benjamin, N., Mbaye, A. A. & Diop, I. T. (2012). *The informal sector in Francophone Africa: firm size, productivity and institutions*. World Bank Publications.
- Berry, S. (2002). Debating the land question in Africa. *Comparative Studies in Society and History*, 44(4), 638-668.
- Bertrand, M., Duflo, E. & Mullainathan, S. (2002). *How much should we trust differences-in-differences estimates?* National Bureau of Economic Research.
- Besley, T. & Burgess, R. (2000). Land reform, poverty reduction and growth: Evidence from India. *Quarterly journal of Economics*, 389-430.
- Besley, T. & Ghatak, M. (2010). Property rights and economic development. *Handbook of development economics*, 5, 4525-4595

- Bezabih, M., Kohlin, G. & Mannberg, A. (2011). Trust, tenure insecurity and land certification in rural Ethiopia. *The Journal of Socio-Economics*, 40(6), 833-843.
- Bezemer, D. & Headey, D. (2008). Agriculture, development and urban bias. *World Development*, 36(8), 1342-1364.
- Bhattacharya, P. C. (1995). The economics of development: A review article. *Journal of Economic Studies*, 22(2), 59-74.
- Binswanger, H. P., Deininger, K. & Feder, G. (1995). *Power, distortions, revolt and reform in agricultural land relations*. Handbook of development economics, 3, 2659-2772.
- Binswanger-Mkhize, H. P. (2014). From failure to success in South African land reform. *African Journal of Agricultural and Resource Economics*, 9(4), 253-269.
- Birner, R. & Resnick, D. (2010). The political economy of policies for smallholder agriculture. *World Development*, 38(10), 1442-1452.
- Black, D., Galdo, J. & Smith, J. A. (2005). *Evaluating the regression discontinuity design using experimental data*. Unpublished manuscript.
- Bloom, D. E., Sachs, J. D., Collier, P., & Udry, C. (1998). Geography, demography, and economic growth in Africa. *Brookings papers on economic activity*, 1998(2), 207-295.
- Bogart, D. (2009). Nationalizations and the development of transport systems: Cross-Country evidence from railroad networks, 1860-1912. *The Journal of Economic History*, 69(1), 202-237.
- Bond, P. (1999). Political reawakening in Zimbabwe. *Monthly Review*, 50(11), 1.
- Boone, C. (2017). Sons of the soil conflict in Africa: Institutional determinants of ethnic conflict over land. *World Development*, 96, 276-293.
- Boone, C. & Nyeme, L. (2015). Land institutions and political ethnicity in Africa: evidence from Tanzania. *Comparative Politics*, 48(1), 67-86.
- Boserup, E. (1996). Development theory: An analytical framework and selected application. *Population and Development Review*, 505-515.
- Boughton, D., Mather, D., Barrett, C. B., Benfica, R., Abdula, D., Tschirley, D. & Cunguara, B. (2007). Market participation by rural households in a low-income country: An asset-based approach applied to Mozambique. *Faith and economics*, 50(1), 64-101.
- Branca, G., McCarthy, N., Lipper, L. & Jolejole, M. C. (2011). Climate-smart agriculture: a synthesis of empirical evidence of food security and mitigation benefits from improved cropland management. *Mitigation of Climate Change in Agriculture Series*, 3, 1-42.
- Calabresi, G. (1968). Transaction Costs, Resource Allocation and Liability Rules--A Comment. *The Journal of Law and Economics*, 11(1), 67-73.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372-404.

- Cattaneo, M. (2017). Regression Discontinuity Designs in Stata, available online: https://www.stata.com/meeting/columbus15/abstracts/materials/columbus15_cattaneo.pdf, accessed 28-12-2019
- Chamunorwa, A. (2010). Comparative Analysis of Agricultural Productivity between Newly Resettled Farmers and Communal Farmers in Mashonaland East Province.
- Chaney, T. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy*, 126(1), 150-177.
- Chapman, L., Bell, C. & Bell, S. (2017). Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *International Journal of Climatology*, 37(9), 3597-3605.
- Chavunduka, G. (1975). Rural and urban life. *Zambezia*, 4(II).
- Chen, X. & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proc Natl Acad Sci U S A*, 108(21), 8589-8594. doi: 10.1073/pnas.1017031108
- Chigumira, E. (2010). My Land, My Resource: Assessment of the Impact of the Fast Track Land Reform Programme on the Natural Environment, Kadoma District, Zimbabwe.
- Chimhowu, A. (2019). The 'new' African customary land tenure. Characteristic, features and policy implications of a new paradigm. *Land Use Policy*, 81, 897-903.
- Chimhowu, A. & Woodhouse, P. (2006). Customary vs private property rights? Dynamics and trajectories of vernacular land markets in Sub-Saharan Africa. *Journal of Agrarian Change*, 6(3), 346-371.
- Chingozha, T. & von Fintel, D. (2019). The Complementarity between Property Rights and Market Access for Crop Cultivation in Rhodesia: Evidence from Historical Satellite Data. *Economic History of Developing Regions*, 1-24.
- Chitekwe-Biti, B., Mudimu, P., Nyama, G. M. & Jera, T. (2012). Developing an informal settlement upgrading protocol in Zimbabwe—the Epworth story. *Environment and Urbanization*, 24(1), 131-148.
- Christodoulou, D. (1990). *Unpromised Land: Agrarian Reform and Conflict Worldwide*. London: Zed books.
- Christopher, A. (1971). Recent trends in land tenure in Rhodesia 1961-70. *Geography*, 56(2), 140-144.
- Ciamarra, U. P. (2004). Access to land through rental markets: a (counter-) evolution in the World Bank's land policy?, 2004/2(land reform land settlement and cooperatives), 9-21.
- Clarke, D. (1975). Land inequality and income distribution in Rhodesia. *African Studies Review*, 18(1), 1-7.
- Cliffe, L., Alexander, J., Cousins, B., & Gaidzanwa, R. (2011). An overview of fast track land reform in Zimbabwe: editorial introduction. *Journal of Peasant Studies*, 38(5), 907-938.
- Coase, R. H. (1960). The problem of social cost. In *Classic papers in natural resource economics* (pp. 87-137). *Palgrave Macmillan*, London.

- Colin, J.-P. & Woodhouse, P. (2010). Interpreting land markets in Africa. *Africa*, 80(1), 1-13.
- Collier, P. & Dercon, S. (2014). African agriculture in 50 years: Smallholders in a rapidly changing world? *World Development*, 63, 92-101.
- Congedo, L. (2014). *Land Cover Classification of Cropland: A Tutorial Using the Semi-Automatic Classification Plugin for QGIS*. Available online: <https://www.directionsmag.com/article/1428>, accessed 20-11-2017, 1800 hours
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1-45.
- Cook, T. D. & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity design. *Annales d'Economie et de Statistique*, 127-150.
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cotula, L., Toulmin, C. & Quan, J. (2006). *Better Land Access for the Rural Poor: Lessons from Experience and Challenges Ahead*. Rome: IIED/FAO.
- Crosby, A. W. (2004). *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*. Cambridge: Cambridge University Press.
- Davis, B., Di Giuseppe, S. & Zezza, A. (2017). Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy*, 67, 153-174.
- De Janvry, A., Gonzalez-Navarro, M. & Sadoulet, E. (2014). Are land reforms granting complete property rights politically risky? Electoral outcomes of Mexico's certification program. *Journal of development Economics*, 110, 216-225.
- De Janvry, A., Sadoulet, E. & Wolford, W. (2001). Access to land and land policy reforms: UNU World Institute for Development Economics Research, Helsinki.
- De Schutter, O. (2011). How not to think of land-grabbing: three critiques of large-scale investments in farmland. *Journal of Peasant Studies*, 38(2), 249-279.
- De Soto, H. (2000). *The Mystery of Capital: Why capitalism Triumphs in the West and Fails Everywhere Else*. New York: Basic books.
- De Soto, H. (2001). The mystery of capital. *Finance and Development*, 38(1), 29-33.
- De Villiers, B. (2003). *Land Reform: Issues and Challenges*. Johannesburg: Konrad Adenauer Foundation.
- Deininger, K. (1999). *Making negotiated land reform work: Initial experience from Brazil, Colombia and South Africa*. Washington D.C.: The World Bank.
- Deininger, K. W. (2003). *Land Policies for Growth and Poverty Reduction*. Washington D.C.: World Bank Publications.
- Deininger, K. & Castagnini, R. (2006). Incidence and impact of land conflict in Uganda. *Journal of Economic Behavior & Organization*, 60(3), 321-345.

- Deininger, K. & Jin, S. (2006). Tenure security and land-related investment: Evidence from Ethiopia. *European Economic Review*, 50(5), 1245-1277.
- Deininger, K., Ali, D. A., Holden, S. & Zevenbergen, J. (2008). Rural land certification in Ethiopia: Process, initial impact and implications for other African countries. *World Development*, 36(10), 1786-1812.
- Deininger, K., Jin, S., Xia, F. & Huang, J. (2014). Moving off the farm: Land institutions to facilitate structural transformation and agricultural productivity growth in China. *World Development*, 59, 505-520.
- Deininger, K., Savastano, S. & Xia, F. (2017). Smallholders' land access in Sub-Saharan Africa: A new landscape? *Food Policy*, 67, 78-92.
- Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, 78(6), 1863-1903.
- Demsetz, H. (1974). Toward a theory of property rights. In *Classic papers in natural resource economics* (pp. 163-177). *Palgrave Macmillan*, London.
- Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, 59(S1).
- Dhumal, R. K., Yogesh Rajendra, K. & Mehrotra, S. (2013). Classification of Crops from remotely sensed Images: An Overview. *International Journal of Engineering Research and Applications* (IJERA), 3(3), 758-761.
- Do, Q.-T. & Iyer, L. (2008). Land titling and rural transition in Vietnam. *Economic Development and Cultural Change*, 56(3), 531-579.
- Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2), 799-858.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934.
- Dorn, H., Törnros, T. & Zipf, A. (2015). Quality evaluation of VGI using authoritative data—A comparison with land use data in Southern Germany. *ISPRS International Journal of Geo-Information*, 4(3), 1657-1671.
- Duggan, W. R. (1980). The Native Land Husbandry Act of 1951 and the rural African middle class of Rhodesia. *African Affairs*, 79(315), 227-239.
- Eastman, J. R. (2003). *IDRISI Kilimanjaro: Guide to GIS and Image Processing*. Worcester, MA: Clark Labs, Clark University.
- Eastwood, R. & Lipton, M. (2000). Pro-poor Growth and Pro-growth Poverty Reduction: Meaning, Evidence and Policy Implications. *Asian Development Review*, 18(2), 22-58
- Ebener, S., Murray, C., Tandon, A. & Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4(1), 1.
- Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B. & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8), 1652-1660. doi: 10.1016/j.cageo.2009.01.009

- Fairbairn, D. & Al-Bakri, M. (2013). Using geometric properties to evaluate possible integration of authoritative and volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2(2), 349-370.
- Fishlow, A. (1965). *American Railroads and the Transformation of the Ante-bellum Economy* (Vol. 127). Cambridge, MA: *Harvard University Press*.
- Falco, P., Kerr, A., Rankin, N., Sandefur, J. & Teal, F. (2011). The returns to formality and informality in urban Africa. *Labour Economics*, 18, S23-S31. doi: 10.1016/j.labeco.2011.09.002
- FAO. (2016). *Chapter 2: Zimbabwe's natural regions and farming systems*. FAO Corporate Document Repository.
- FAOSTAT. (2015). *Trade Data on Crops and Livestock Products*. Retrieved from: <http://faostat.fao.org/beta/en/?#data/TP>.
- Farrell, J. (1987). Information and the Coase theorem. *Journal of Economic Perspectives*, 1(2), 113-129.
- Federal Department of Trigonometrical and Topographical Surveys, R. a. N. (1963). *Rhodesia: Land Apportionment*. Retrieved from: <http://www.digitalcollections.lib.uct.ac.za/special-collections-maps>
- Fenske, J. (2011). Land tenure and investment incentives: Evidence from West Africa. *Journal of Development Economics*, 95(2), 137-156.
- Fernandes, J. L. R. (2015). *Analysis of Classification Algorithms for Crop Detection using LANDSAT 8 Images*. Universidade de Évora.
- Floyd, B. N. (1962). Land apportionment in Rhodesia. *Geographical Review*, 52(4), 566-582.
- Fonte, Bastin, L., See, L., Foody, G. & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science*, 29(7), 1269-1291.
- Fonte, C., Minghini, M., Patriarca, J., Antoniou, V., See, L. & Skopeliti, A. (2017). Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information*, 6(4), 125.
- Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C. & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS*, 17(6), 847-860.
- Frankema, E., Green, E. & Hillbom, E. (2016). Endogenous processes of colonial settlement. the success and failure of European settler farming in sub-Saharan Africa. *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* (New Series), 34(02), 237-265.
- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F.,... Obersteiner, M. (2009). Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1(3), 345-354.

- Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D.,... Laso-Bayas, J.-C. (2017). A global dataset of crowdsourced land cover and land use reference data. *Scientific Data*, 4, 170075.
- Gallani, S. & Krishnan, R. (2017). Applying the fractional response model to survey research in accounting. School HB.
- Galor, O. & Özak, Ö. (2014). The agricultural origins of time preference: National Bureau of Economic Research.
- Galor, O. & Özak, Ö. (2016). The agricultural origins of time preference: National Bureau of Economic Research.
- Galor, O., Özak, Ö. & Sarid, A. (2016). *Geographical Origins and Economic Consequences of Language Structures*. Brown University Working Paper, 2016.
- Glaeser, E. L. (2013). Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier (an excerpt). *Journal of Economic Sociology*, 14(4), 75-94.
- Garrison, H. (1982). Internal migration in Mexico: a test of the Todaro model. *Food Research Institute Studies*, 18(2), 197-214.
- Gislason, P. O., Benediktsson, J. A. & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300.
- Goldstein, M. & Udry, C. (2008). The profits of power: Land rights and agricultural investment in Ghana. *Journal of political economy*, 116(6), 981-1022.
- Good, K. (1976). Settler colonialism: economic development and class formation. *Journal of Modern African Studies*, 14(4), 597-620.
- Grabowski, R. (1988). The theory of induced institutional innovation: a critique. *World Development*, 16(3), 385-394.
- Green, E. (2016). *The Development of Settler Agriculture in British Africa Revisited: Estimating the Role of Tenant Labour in Rhodesia, C 1900-1960*: African Economic History Network, 29(2016)
- Grier, R. M. (1999). Colonial legacies and economic growth. *Public Choice*, 98(3-4), 317-335.
- Grindle, M. S. (2004). Good enough governance: poverty reduction and reform in developing countries. *Governance*, 17(4), 525-548.
- Guardado, J. (2018). Land tenure, price shocks and insurgency: Evidence from Peru and Colombia. *World Development*, 111, 256-269.
- Hahn, J., Todd, P. & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.
- Haklay, M., Basiouka, S., Antoniou, V. & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315-322.
- Harttgen, K., Klasen, S. & Vollmer, S. (2013). An African growth miracle? Or: what do asset indices tell us about trends in economic performance? *Review of Income and Wealth*, 59(S1).

- Haviland, W. (1953). Tobacco farm organization, costs and land-use in Rhodesia. *South African Journal of Economics*, 21(4), 367-380.
- Hazell, P., Poulton, C., Wiggins, S. & Dorward, A. (2010). The future of small farms: trajectories and policy priorities. *World Development*, 38(10), 1349-1361.
- Hegazy, I. R. & Kaloop, M. R. (2015). Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt. *International Journal of Sustainable Built Environment*, 4(1), 117-124.
- Henderson, J. V., Storeygard, A. & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *Am. Econ. Rev.*, 102(2), 994-1028. doi: 10.1257/aer.102.2.994
- Hentschel, J. (1998). *Combining census and survey data to study spatial dimensions of poverty: A case study of Ecuador* (Vol. 1928): World Bank Publications.
- Herbst, J. (1991). The dilemmas of land policy in Zimbabwe. *Africa Insight*, 21(4), 269-276.
- Herranz-Loncán, A. & Fourie, J. (2017). “For the public benefit”? Railways in the British Cape Colony. *European Review of Economic History*, 22(1), 73-100.
- Holden, S. T. & Ghebru, H. (2016). Land tenure reforms, tenure security and food security in poor agrarian economies: Causal linkages and research gaps. *Global Food Security*, 10, 21-28.
- Hornbeck, R. (2010). Barbed wire: Property rights and agricultural development. *The Quarterly Journal of Economics*, 125(2), 767-810.
- Huang, C., Davis, L. & Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725-749.
- Hughes, A. J. B. (1971). Tribal land tenure-an obstacle to progress?: Special reference to the Tribal Trust Land of Rhodesia. *Africa Insight*, 1(1), 56-73.
- Jayne, T. S. & Jones, S. (1997). Food marketing and pricing policy in Eastern and Southern Africa: A survey. *World Development*, 25(9), 1505-1527.
- Jayne, T. S., Chamberlin, J. & Headey, D. D. (2014). Land pressures, the evolution of farming systems and development strategies in Africa: A synthesis. *Food Policy*, 48, 1-17.
- Jayne, T. S., Yamano, T., Weber, M. T., Tschirley, D., Benfica, R., Chapoto, A. & Zulu, B. (2003). Smallholder income and land distribution in Africa: Implications for poverty reduction strategies. *Food Policy*, 28(3), 253-275.
- Jeanty, P. W. (2012, July). Stata Implementation of the non-parametric spatial heteroskedasticity and autocorrelation consistent covariance matrix estimator. *In Stata conference San Diego* (pp. 26-27).
- Jedwab, R. & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: Evidence from Africa. *Review of Economics and Statistics*, 98(2), 268-284.
- Jedwab, R., Kerby, E. & Moradi, A. (2017). History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya. *The Economic Journal*, 127(603), 1467-1494.

- Jennings, A. & Huggins, G. (1935). Land Apportionment in Rhodesia. *Journal of the Royal African Society*, 34(136), 296-312.
- Jerven, M. (2010). Accounting for the African growth miracle: The official evidence—Botswana 1965–1995. *Journal of Southern African Studies*, 36(1), 73-94.
- Jian, P. & Weifeng, L. (2013). Intercalibration of DMSP-OLS night-time light data by the invariant region method. *International Journal of Remote Sensing*, 34(20), 12.
- Johnson, B. A. & Iizuka, K. (2016). Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. *Applied Geography*, 67, 140-149.
- Johnson, B. A., Iizuka, K., Bragais, M. A., Endo, I. & Magcale-Macandog, D. B. (2017). Employing crowdsourced geographic data and multi-temporal/multi-sensor satellite imagery to monitor land cover change: A case study in an urbanizing region of the Philippines. *Computers, Environment and Urban Systems*, 64, 184-193.
- Jordaan, H., Grové, B. & Backeberg, G. R. (2014). Conceptual framework for value chain analysis for poverty alleviation among smallholder farmers. *Agrekon*, 53(1), 1-25.
- Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1), 127-155.
- Kelly, M. (2019). The Standard Errors of Persistence. London, *Centre for Economic Policy Research*. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=13783
- Keswell, M. & Carter, M. R. (2014). Poverty and land redistribution. *Journal of development Economics*, 110, 250-261.
- Krugman, P. (1995). Increasing returns, imperfect competition and the positive theory of international trade. *Handbook of international economics*, 3, 1243-1277
- Kumbawa, D. (2002). *Informal sector now Zimbabwe's biggest employer*. Harare: Financial Gazette. Available online <http://www.gdrc.org/informal/6-zimbabwe.html>, accessed 20-09-2019, 0800 hours
- Kulkarni, R., Haynes, K., Stough, R., & Riggle, J. (2011). Light based growth indicator (LBGI): exploratory analysis of developing a proxy for local economic growth based on night lights. *Regional Science Policy & Practice*, 3(2), 101-113.
- Lange, M. K. (2004). British colonial legacies and political development. *World Development*, 32(6), 905-922.
- Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods Estimation of Spatial Panels. *Foundations and Trends® in Econometrics*, 4(3), 165-224. doi: 10.1561/08000000014
- Li, X., Ge, L. & Chen, X. (2013). Detecting Zimbabwe's Decadal Economic Decline Using Nighttime Light Imagery. *Remote Sensing*, 5(9), 4551-4570. doi: 10.3390/rs5094551
- Lin, J. Y. (1995). Endowments, technology, and factor markets: A natural experiment of induced institutional innovation from China's rural reform. *American Journal of Agricultural Economics*, 77(2), 231-242.

- López, R. & Valdés, A. (2000). Fighting rural poverty in Latin America: New evidence of the effects of education, demographics and access to land. *Economic Development and Cultural Change*, 49(1), 197-211.
- Lowder, S. K., Scoet, J. & Raney, T. (2016). The number, size and distribution of farms, smallholder farms and family farms worldwide. *World Development*, 87, 16-29.
- Lowe, M. (2014). Night lights and ArcGis: A brief guide. *Available online: <http://economics.mit.edu/files/8945>*, accessed on 22 September 2019.
- Lund, C. (2011). Property and citizenship: conceptually connecting land rights and belonging in Africa. *Africa Spectrum*, 46(3), 71-75.
- Lund, C. & Boone, C. (2013). Introduction: land politics in Africa—constituting authority over territory, property and persons. *Africa*, 83(1), 1-13.
- Machingaidze, V. E. (1991). Agrarian Change from Above: The Rhodesia Native Land Husbandry Act and African Response. *International Journal of African Historical Studies*, 557-588.
- Machová, K., Barcak, F. & Bednár, P. (2006). A bagging method using decision trees in the role of base classifiers. *Acta Polytechnica Hungarica*, 3(2), 121-132.
- Malarvizhi, K., Kumar, S. V. & Porchelvan, P. (2016). Use of high resolution Google Earth satellite imagery in landuse map preparation for urban related applications. *Procedia Technology*, 24, 1835-1842.
- Mandizadza, S. (2010). *The fast track land reform programme and livelihoods in Zimbabwe: A case study of households at Athlone Farm in Murehwa District*. Harare: Institute for Development Studies.
- Maravanyika. (2007). *Black Friday*. Harare: Unknown
- Marimira, S. C. (2010). Institutions, Leadership and Service Delivery in New Resettlement Areas of Zimbabwe. *PLAAS/IDS Working Paper*, 2010
- Markelova, H., Meinzen-Dick, R., Hellin, J., & Dohrn, S. (2009). Collective action for smallholder market access. *Food policy*, 34(1), 1-7.
- Markussen, T. (2008). Property Rights, Productivity and Common Property Resources: Insights from Rural Cambodia¹. *World Development*, 36(11), 2277-2296.
- Markussen, T. & Tarp, F. (2014). Political connections and land-related investment in rural Vietnam. *Journal of Development Economics*, 110, 291-302.
- Marx, B., Stoker, T., & Suri, T. (2013a). There is no free house: Ethnic patronage and property rights in a Kenyan slum.
- Marx, B., Stoker, T., & Suri, T. (2013b). The economics of slums in the developing world. *Journal of Economic perspectives*, 27(4), 187-210.
- Masuku, M., Makhura, M. & Rwelarmira, J. (2001). Factors affecting marketing decisions in the maize supply chain among smallholder farmers in Swaziland. *Agrekon*, 40(4), 698-707.

- Matondi, P. B., & Dekker, M. (2011). Land rights and tenure security in Zimbabwe's post fast track Land Reform Programme. *A Synthesis report for LandAc*. Project ID WS, 320005(3).
- Mbereko, A. (2010). *An assessment of the outcomes of "Fast Track" Land Reform Policy in Zimbabwe on rural livelihoods: The case of Gudo ward (Mazvihwa Communal Area) and Chirere Area (A1 Resettlement Area)*. Harare: Institute of Development Studies.
- McIver, D. K. & Friedl, M. A. (2001). Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), 1959-1968.
- McMillan, Rodrik, D. & Verduzco-Gallo, Í. (2014). Globalization, structural change and productivity growth, with an update on Africa. *World Development*, 63, 11-32.
- Meier, P. (2013). Human computation for disaster response Handbook of human computation (pp. 95-104): Springer.
- Meinzen-Dick, R., Quisumbing, A., Doss, C. & Theis, S. (2017). Women's land rights as a pathway to poverty reduction: Framework and review of available evidence. *Agricultural Systems*, 172, 72-82
- Mellor, J. W. & Johnston, B. F. (1984). The world food equation: Interrelations among development, employment and food consumption. *Journal of Economic Literature*, 22(2), 531-574.
- Mellor, J. W. & Malik, S. J. (2017). The impact of growth in small commercial farm productivity on rural poverty reduction. *World Development*, 91, 1-10.
- Mlambo, A. S. (1994). From dirt tracks to modern highways: Towards a history of roads and road transportation in colonial Zimbabwe, 1890 to World War II. *Zambezia*, 21(2), 147-166.
- Moyana, H. V. (1975). Land and race in Rhodesia. *African Review*, 5(1), 17-41.
- Moyo, P. (2010). Land reform in Zimbabwe and urban livelihoods transformation.
- Moyo, P., Jha, P. & Yeros, P. (2013). The Classical Agrarian Question: Myth, Reality and Relevance Today. *Agrarian South. Journal of Political Economy*, 2(1), 93-119. doi: 10.1177/2277976013477224
- Muchomba, F. M. (2017). Women's land tenure security and household human capital: Evidence from Ethiopia's land certification. *World Development*, 98, 310-324.
- Munslow, B. (1985). Prospects for the socialist transition of agriculture in Zimbabwe. *World Development*, 13(1), 41-58.
- Murisa, T. (2011). Local farmer groups and collective action within fast track land reform in Zimbabwe. *Journal of Peasant Studies*, 38(5), 1145-1166.
- Musemwa, L., & Mushunje, A. (2012). Factors affecting yields of field crops and land utilisation amongst land reform beneficiaries of Mashonaland Central Province in Zimbabwe. *Journal of Development and Agricultural Economics*, 4(4), 109-118.

- Mutangi, G. (2010). The changing patterns of farm labour after the Fast Track Land Reform Programme: The case of Guruve District. *Livelihoods after Land Reform in Zimbabwe Working Paper*, 13.
- Nagendra, H., Munroe, D. K. & Southworth, J. (2004). From pattern to process: landscape fragmentation and the analysis of land use/land cover change. *Agriculture, Ecosystems & Environment*, 101(2), 111-115.
- Naik, A. K. (2009). *Informal sector and informal workers in India*. Paper presented at the seminar measuring the informal economy in developing countries, Kathmandu, Nepal.
- Ncube, M. (2000). Employment, unemployment and the evolution of labour policy in Zimbabwe. *Zambezia*, 27(2), 161-194.
- Ndlela, D. & Robinson, P. (2007). *Distortions to Agricultural Incentives in Zimbabwe*. Washington D.C.: World Bank
- Neis, P. & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet*, 6(1), 76-106.
- Nelson, H. D. (1982). *Area Handbook Series: Zimbabwe*. Washington D.C.: American University.
- Newman, C., Tarp, F. & Van Den Broeck, K. (2015). Property rights and productivity: The case of joint land titling in Vietnam. *Land Economics*, 91(1), 91-105.
- Nielsen, J. (2006). *Participation inequality: Encouraging more users to contribute*. http://www.useit.com/alertbox/participation_inequality.html
- North, D. C., & Thomas, R. P. (1970). An economic theory of the growth of the western world. *The economic history review*, 23(1), 1-17.
- Nunn, N. (2008). The long-term effects of Africa's slave trades. *The Quarterly Journal of Economics*, 123(1), 139-176.
- Nunn, N. & Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7), 3221-3252.
- Obi, A. (2011). Performance of smallholder agriculture under limited mechanization and the fast track land reform program in Zimbabwe. *International Food and Agribusiness Management Review*, 14(1030-2016-82908), 85-104.
- Olson, M. (2008). *The rise and decline of nations: Economic growth, stagflation, and social rigidities*. Yale University Press.
- Otsuka, K. (2007). Efficiency and equity effects of land markets. *Handbook of Agricultural Economics*, 3, 2671-2703.
- Palmer, R. (1971). War and land in Rhodesia. *Transafrican Journal of History*, 1(2), 43-62.
- Palmer, R. (1990). Land reform in Zimbabwe, 1980-1990. *African Affairs*, 89(355), 163-181.
- Papke, L. E. & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619-632.

- Peters, P. E. (2009). Challenges in land tenure and land reform in Africa: Anthropological contributions. *World Development*, 37(8), 1317-1325.
- Pettorelli, N., Gliozzo, G. & Haklay, M. (2016). Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. *Ecology and Society*, 21(3).
- Phimister, I. R. (1974). Peasant production and underdevelopment in Rhodesia, 1890-1914. *African Affairs*, 73(291), 217-228.
- Phimister, I. R. (1988). *An Economic and Social history of Zimbabwe, 1890-1948: Capital Accumulation and Class Struggle*. Longman: London.
- Pinkovskiy, M. & Sala-i-Martin, X. (2014). Lights, Camera,... Income! *Estimating Poverty Using National Accounts, Survey Means and Lights*. National Bureau of Economic Research.
- Place, F. (2009). Land tenure and agricultural productivity in Africa: a comparative analysis of the economics literature and recent policy strategies and reforms. *World Development*, 37(8), 1326-1336.
- Pollak, O. B. (1975). Black farmers and white politics in Rhodesia. *African Affairs*, 74(296), 263-277.
- Potts, D. (2006). 'Restoring Order'? Operation Murambatsvina and the Urban Crisis in Zimbabwe. *Journal of Southern African Studies*, 32(2), 273-291. doi: 10.1080/03057070600656200
- Poulton, C., Kydd, J. & Dorward, A. (2006). Overcoming market constraints on pro-poor agricultural growth in Sub-Saharan Africa. *Development Policy Review*, 24(3), 243-277.
- Punt, E. (1979). *The Development of African Agriculture in Rhodesia with Particular Reference to the Interwar Years*.
- Ranga, D. (2004). The migration effects of the economic structural adjustment programme on a rural community in Zimbabwe. *African Population Studies*, 19(1), 165-185.
- Ravallion, M. & Van De Walle, D. (2006). Land reallocation in an agrarian transition. *The Economic Journal*, 116(514), 924-942.
- Ravallion, M. & Van de Walle, D. (2008). Does rising landlessness signal success or failure for Vietnam's agrarian transition? *Journal of Development Economics*, 87(2), 191-209.
- Ravi, S., Kapoor, M. & Ahluwalia, R. (2012). *The Impact of NREGS on Urbanization in India*. Available at SSRN 2134778.
- Redding, S. J., & Sturm, D. M. (2008). The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, 98(5), 1766-97.
- Rei, D., & Bhattacharya, M. (2008). *The impact of institutions and policy on informal economy in developing countries: an econometric exploration*. Geneva: ILO.
- Reyes, C. M. (2002). Impact of agrarian reform on poverty: *PIDS discussion paper series*. 2002(2)
- Ribas, R. (2016). Multidimensional Regression Discontinuity and Regression Kink Designs with Difference-in-Differences. *2016 Stata Conference, Stata Users group*. Chicago

- Richardson, C. J. (2007). How much did droughts matter? Linking rainfall and GDP growth in Zimbabwe. *African Affairs*, 106(424), 463-478. doi: 10.1093/afraf/adm013
- Riddell, R. C. (1978). The land problem in Rhodesia; alternatives for the future. FAO
- Rifkind, M. (1969). *The Politics of Land in Rhodesia*. Edinburgh: University of Edinburgh.
- Rigg, J. (2006). Land, farming, livelihoods and poverty: Rethinking the links in the rural South. *World Development*, 34(1), 180-202.
- Rigg, J., Salamanca, A. & Parnwell, M. (2012). Joining the dots of agrarian change in Asia: A 25 year view from Thailand. *World Development*, 40(7), 1469-1481.
- Riley, L. (2014). Operation Dongosolo and the geographies of urban poverty in Malawi. *Journal of Southern African Studies*, 40(3), 443-458.
- Robbins, C. R. (1934). Northern Rhodesia; an experiment in the classification of land with the use of aerial photographs. *Journal of Ecology*, 22(1), 88-105.
- Rodrik, D. (2016). An African growth miracle? *Journal of African Economies*, 27(1), 10-27.
- Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C. & Roberts, D. (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment*, 112(5), 2272-2283.
- Rogers, D. J., & Williams, B. G. (1993). 11. Tsetse distribution in Africa seeing the wood and the trees. *Large-scale ecology and conservation biology. the 35th symposium of the British Ecological Society with the Society for Conservation Biology*, University of Southampton, 249-273.
- Rothenberg, A. D., Gaduh, A., Burger, N. E., Chazali, C., Tjandraningsih, I., Radikun, R.,... Weiland, S. (2016). Rethinking Indonesia's informal sector. *World Development*, 80, 96-113.
- Ruttan, V. W., & Hayami, Y. (1984). Toward a theory of induced institutional innovation. *The Journal of development studies*, 20(4), 203-223.
- Rybnikova, N. A. & Portnov, B. A. (2015). Using light-at-night (LAN) satellite data for identifying clusters of economic activities in Europe. *Letters in Spatial and Resource Sciences*, 8(3), 307-334.
- Schwiebert, J. (2018). *A Sample Selection Model for Fractional Response Variables*. Lüneburg: Leuphana Universität Lüneburg, Institut für Volkswirtschaftslehre. Working Paper Series in Economics, No. 382
- Scoones, I., Marongwe, N., Mavedzenge, B., Murimbarimba, F., Mahenehene, J. & Sukume, C. (2011a). Zimbabwe's land reform: challenging the myths. *Journal of Peasant Studies*, 38(5), 967-993.
- Scoones, I., Marongwe, N., Mavedzenge, B., Murimbarimba, F., Mahenehene, J. & Sukume, C. (2011b). *Zimbabwe's Land Reform: A Summary of Findings*. Brighton: IDS.
- See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D.,... Kraxner, F. (2015). Harnessing the power of volunteers, the internet and Google Earth to collect and validate global

- spatial information using Geo-Wiki. *Technological Forecasting and Social Change*, 98, 324-335.
- Seligson, M. A. (1984). Implementing land reform: The case of Costa Rica. *Managing International Development*, 1 (2), 29-46.
- Sender, J. & Johnston, D. (2004). Searching for a weapon of mass production in rural Africa: Unconvincing arguments for land reform. *Journal of Agrarian Change*, 4(1-2), 142-164.
- Senyolo, G., Chaminuka, P., Makhura, M. & Belete, A. (2009). Patterns of access and utilization of output markets by emerging farmers in South Africa: Factor analysis approach. *African Journal of Agricultural Research*, 4(3), 208-214.
- Shao, Y. & Lunetta, R. S. (2012). Comparison of support vector machine, neural network and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 78-87.
- Shapiro, A. F. (2015). Institutions, Informal Labor Markets and Business Cycle Volatility. *Economía*, 16(1), 77-112.
- Shipton, P. & Goheen, M. (1992). Introduction. Understanding African land-holding: Power, wealth and meaning. *Africa*, 62(3), 307-325.
- Sikor, T. & Müller, D. (2009). The limits of state-led land reform: An introduction. *World Development*, 37(8), 1307-1316.
- Sithole, P. (1972). Labour Problems in Rhodesia: A Trade Union Viewpoint. *The Rhodesian Journal of Economics*, 6(4), 9-15
- Sokoloff, K. L. & Engerman, S. L. (2000). Institutions, factor endowments and paths of development in the new world. *Journal of Economic Perspectives*, 14(3), 217-232.
- Srivastava, S., Lobry, S., Tuia, D. & Vargas-Muñoz, J. (2018). *Land-use characterisation using Google Street View pictures and OpenStreetMap*. Paper presented at the Proceedings of the Association of Geographic Information Laboratories in Europe Conference (AGILE).
- Stewart, O., Lubensky, D. & Huerta, J. M. (2010). *Crowdsourcing participation inequality: a SCOUT model for the enterprise domain*. Paper presented at the Proceedings of the ACM SIGKDD Workshop on Human Computation.
- Stocking, M. A. (1978). Relationship of agricultural history and settlement to severe soil erosion in Rhodesia. *Zambezia*, 6(2), 129-145.
- Stoeffler, Q., Alwang, J., Mills, B. & Taruvinga, N. (2016). Multidimensional poverty in crisis: Lessons from Zimbabwe. *The Journal of Development Studies*, 52(3), 428-446.
- Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A. & Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5(4), 256-266.
- Sutton, P. C. & Costanza, R. (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover and ecosystem service valuation. *Ecological Economics*, 41(3), 509-527.

- Tarisayi, K. S. (2013). Land Reform: An Analysis of Definitions, Types and Approaches. *Asian Journal of Agriculture and Rural Development*, 4(3), 2014.
- Thirtle, C., Atkins, J., Bottomley, P., Gonese, N., Govereh, J. & Khatri, Y. (1993). Agricultural productivity in Zimbabwe, 1970-90. *The Economic Journal*, 103(417), 474-480.
- Tibaijuka, K. (2005). *UN Human Special Envoy on Human Settlements Issues in Zimbabwe: Report of the Fact-finding Mission to Zimbabwe to Assess the Impact of Operation Murambatsvina*. Vienna: UN Publications.
- Udry, C. (2011). Land tenure. *The Oxford Companion to the Economics of Africa*, 1.
- Ustuner, M., Sanli, F., Abdikan, S., Esetlili, M. & Kurucu, Y. (2014). Crop type classification using vegetation indices of rapideye imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(7), 195.
- Valdés, A. & Foster, W. (2010). Reflections on the role of agriculture in pro-poor growth. *World Development*, 38(10), 1362-1374.
- Valente, C. (2009). The food (in) security impact of land redistribution in South Africa: microeconomic evidence from national data. *World Development*, 37(9), 1540-1553.
- Vaz, E. & Jokar Arsanjani, J. (2015). Crowdsourced mapping of land use in urban dense environments: An assessment of Toronto. *The Canadian Geographer/Le Géographe Ccanadien*, 59(2), 246-255.
- Verick, S. (2006). *The impact of globalization on the informal sector in Africa*. Economic and Social Policy Division, United Nations Economic and Social Commission for Africa.
- Vitoria, B., Mudimu, G. & Moyo, T. (2012). *Status of Agricultural and Rural Finance in Zimbabwe*. Johannesburg: FinMark Trust.
- Waeterloos, E. & Rutherford, B. (2004). Land reform in Zimbabwe: challenges and opportunities for poverty reduction among commercial farm workers. *World Development*, 32(3), 537-553.
- Warriner, D. (1969). *Land Reform in Principle and Practice*. Oxford, UK: Clarendon Press.
- Weinrich, A. K. H. (1979). *Women and Racial Discrimination in Rhodesia*. United Nations Educational.
- Whillow, J. (1980). Land use, population pressure and rock outcrops in the tribal areas of Zimbabwe-Rhodesia. *Zimbabwe Rhodesia Agricultural Journal*, 77(1), 3-11.
- Wiggins, A. & Crowston, K. (2011). *From conservation to crowdsourcing: A typology of citizen science*. Paper presented at the 2011 44th Hawaii international conference on system sciences.
- Willmott, C. J. & Matsuura, K. (2015). Terrestrial Precipitation and Air Temperature 1900-2014 *Gridded Monthly Time Series*. http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html.
- Willmott, C. & Matsuura, K. (2001). Terrestrial Air Temperature and Precipitation. *Monthly and Annual Time Series (1950-1999)*.

http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html

Wooldridge, J. (2007). *What's New in Econometrics?* Lecture 10 difference-in-differences estimation. NBER Summer Institute. Available at: www.nber.org/WNE/Slides7-31-07/slides_10_diffindiffs.pdf, accessed April 9, 2011.

Xing, H., Meng, Y., Hou, D., Song, J. & Xu, H. (2017). Employing crowdsourced geographic information to classify land cover with spatial clustering and topic model. *Remote Sensing*, 9(6), 602.

Yngstrom, I. (2002). Women, wives and land rights in Africa: Situating gender beyond the household in the debate over land policy and changing tenure systems. *Oxford Development Studies*, 30(1), 21-40.

Yoo, D. & Steckel, R. H. (2010). *Property Rights and Financial Development: The Legacy of Japanese Colonial Institutions*. Cambridge: National Bureau of Economic Research.

Youé, C. (2002). Black squatters on white farms: segregation and agrarian change in Kenya, South Africa and Rhodesia, 1902-1963. *The International History Review*, 24(3), 558-602.

Zarin, A. & Bujang, A. A. (1994). Theory on land reform: An overview. *Bulletin Ukur*, 5(1), 9-14.

Zikhali, P. (2008). *Fast Track Land Reform and Agricultural Productivity in Zimbabwe*. Gothenburg: University of Gothenburg School of Business, Economics and Law

Zimstat. (2015). Zimbabwe Poverty Atlas.

Zingoni, E., Love, D., Magadza, C., Moyce, W. & Musiwa, K. (2005). Effects of a semi-formal urban settlement on groundwater quality: Epworth (Zimbabwe): Case study and groundwater quality zoning. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11-16), 680-688.

Zinyama, L. (1986). Agricultural development policies in the African farming areas of Zimbabwe. *Geography*, 105-115.

Zinyama, L. & Whitlow, R. (1986). Changing patterns of population distribution in Zimbabwe. *GeoJournal*, 13(4), 365-384.

A. APPENDIX A: SUPPLEMENTARY TABLES AND INFORMATION

A.1 Chapter 2 Results Using Non-Linear Estimators

Table A.1 Class Effects of Colonial Land Policy (Fractional Probit Estimates)

	(1)	(2)	City (3)	Road (4)	Rail (5)	(6)	(7)	City (8)	Road (9)	Rail (10)
	<i>Period: 1972-1979</i>					<i>Period: 1980-1984</i>				
European Areas	0.080** (0.037)	0.086** (0.037)	-0.13 (0.266)	0.076 (0.092)	0.20 (0.132)	0.13** (0.057)	0.14** (0.058)	0.17 (0.412)	-0.11 (0.149)	0.079 (0.213)
NPA	-0.052 (0.070)	-0.059 (0.070)	1.79** (0.765)	1.16*** (0.304)	1.62*** (0.260)	0.10 (0.112)	0.097 (0.113)	1.97* (1.029)	0.73* (0.430)	1.74*** (0.464)
ln(dist city)			0.0043 (0.046)	0.017 (0.038)	0.016 (0.038)			0.051 (0.077)	0.032 (0.063)	0.043 (0.063)
ln(dist main road)			0.018 (0.018)	0.024 (0.022)	0.020 (0.018)			-0.045* (0.027)	-0.072** (0.033)	-0.043 (0.027)
ln(dist rail)			-0.0048 (0.024)	-0.0017 (0.024)	0.021 (0.031)			0.064* (0.038)	0.063* (0.038)	0.063 (0.047)
EAs x Distance			0.050 (0.058)	0.0079 (0.029)	-0.027 (0.036)			-0.0029 (0.089)	0.092* (0.048)	0.023 (0.059)
NPAs x Distance			-0.38** (0.157)	-0.37*** (0.091)	-0.44*** (0.071)			-0.39* (0.218)	-0.18 (0.131)	-0.43*** (0.129)
Constant	-0.45* (0.260)	-0.29 (0.337)	-0.33 (0.382)	-0.43 (0.364)	-0.49 (0.368)	-0.87*** (0.174)	-0.034 (0.431)	-0.55 (0.563)	-0.38 (0.524)	-0.54 (0.530)
FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ward Controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
N	2108	2108	2108	2108	2108	696	696	696	696	696

NOTES: Individual ward controls are population, rainfall, and temperature and caloric suitability index. FEs denote region, frame, month and year fixed effects. White heteroscedasticity robust standard errors are reported in parentheses; Base category = Tribal Trust Areas (TTAs). Columns 3 and 8; 4 and 9; 5 and 10 present results from regressions that interacts land class with distance to main city, distance to main road and distance to rail station or siding respectively. * p < 0.10, ** p < 0.05, *** p < 0.01

Table A.2 Class Effects of Colonial Land Policy (Tobit Estimates)

	(1)	(2)	City (3)	Road (4)	Rail (5)	(6)	(7)	City (8)	Road (9)	Rail (10)
	<i>Period: 1972-1979</i>					<i>Period: 1980-1984</i>				
European Areas	0.027** (0.012)	0.030** (0.012)	-0.049 (0.090)	0.028 (0.031)	0.073* (0.044)	0.052** (0.023)	0.057** (0.023)	0.069 (0.160)	-0.042 (0.058)	0.042 (0.085)
NPA	-0.017 (0.023)	-0.019 (0.023)	0.61** (0.262)	0.39*** (0.107)	0.57*** (0.095)	0.042 (0.048)	0.039 (0.045)	0.78** (0.393)	0.28* (0.161)	0.66*** (0.161)
ln(dist city)			0.0071 (0.006)	0.0088 (0.008)	0.0077 (0.006)			-0.018* (0.011)	-0.029** (0.013)	-0.017 (0.011)
ln(dist main road)			-0.0017 (0.008)	-0.00058 (0.008)	0.0076 (0.011)			0.027* (0.016)	0.027* (0.015)	0.028 (0.020)
ln(dist rail)			-0.00071 (0.015)	0.0043 (0.013)	0.0040 (0.013)			0.018 (0.031)	0.010 (0.025)	0.015 (0.025)
EAs x Distance			0.019 (0.020)	0.0023 (0.010)	-0.010 (0.012)			-0.0016 (0.035)	0.036* (0.019)	0.0057 (0.024)
NPAs x Distance			-0.13** (0.054)	-0.12*** (0.031)	-0.15*** (0.025)			-0.15* (0.083)	-0.067 (0.050)	-0.16*** (0.045)
Constant	0.34*** (0.102)	0.38*** (0.124)	0.38*** (0.139)	0.34** (0.133)	0.32** (0.134)	0.17 (0.106)	0.42** (0.167)	0.22 (0.220)	0.30 (0.203)	0.23 (0.206)
FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ward Controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
N	2108	2108	2108	2108	2108	696	696	696	696	696

NOTES: Tobit results account for censoring from below at 0 and from above at 1. Individual ward controls are population, rainfall, and temperature and caloric suitability index. FEs denote region, frame, month and year fixed effects. White heteroscedasticity robust standard errors are reported in parentheses; Base category = Tribal Trust Areas (TTAs). Columns 3 and 8; 4 and 9; 5 and 10 present results from regressions that interacts land class with distance to main city, distance to main road and distance to rail station or siding respectively. * p < 0.10, ** p < 0.05, *** p < 0.01

A.2 Chapter 3 Regression Discontinuity Design (RDD) estimates at 2 and 5km from border

Table A.3 RDD Regression Estimates (2km Buffer from Treatment Region Border)

	<i>Lights</i>			<i>Crops</i>			<i>NDVI</i>		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control Group	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands
Treat	-0.027 (0.810)	-1.92 (1.650)		-0.16 ^{***} (0.057)	-0.16 (0.125)	0.086 (0.245)	-489.4 ^{***} (153.207)	433.6 (285.977)	
Post	-0.34 (0.350)	0.16 (1.034)	-0.32 [*] (0.170)	-0.035 (0.046)	-0.18 [*] (0.102)	-0.25 (0.315)	241.7 ^{***} (84.714)	120.9 (194.117)	32.7 (53.144)
Treat x Post	0.045 (0.411)	-0.49 (1.053)		-0.092 [*] (0.053)	0.056 (0.105)	0.11 (0.317)	-214.5 ^{**} (100.228)	-88.2 (200.963)	
Controls	N	N	N	N	N	N	N	N	N
<i>N</i>	269	204	220	557	426	414	449	341	333
<i>R</i> ²	0.393	0.514	0.589	0.337	0.324	0.312	0.341	0.309	0.321
Treat	1.09 (0.692)	0.71 (1.410)		-0.12 ^{**} (0.062)	-0.20 (0.128)	-0.086 (0.265)	-383.5 ^{**} (161.897)	468.1 [*] (283.660)	
Post	-0.0027 (0.300)	0.68 (0.862)	-0.098 (0.190)	-0.078 (0.048)	-0.19 [*] (0.100)	-0.25 (0.301)	148.1 (90.655)	36.2 (197.152)	-24.5 (66.619)
Treat x Post	0.012 (0.332)	-0.77 (0.872)		-0.10 ^{**} (0.053)	0.024 (0.101)	0.070 (0.301)	-210.3 ^{**} (100.481)	-63.7 (200.619)	
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
<i>N</i>	257	193	209	537	408	396	439	333	325
<i>R</i> ²	0.631	0.699	0.707	0.376	0.373	0.371	0.355	0.321	0.335

Notes*: In (1), (4) and (7) EAs are the Treatment Area while TTLs are the Control Area. In (2), (5) and (8) EAs are the Treatment Area while NPAs are the Control. In (3), (6) and (9), EAs Elsewhere are the Treatment Areas while European Highlands are the Control Area.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4 RDD Regression Estimates (5km Buffer from Treatment Region Border)

	<i>Lights</i>			<i>Crops</i>			<i>NDVI</i>		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control Group	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands	Tribal Trust	Native Purchase	European Highlands
Treat	-0.77* (0.416)	0.75 (0.841)		-0.12*** (0.038)	-0.067 (0.068)	0.20 (0.167)	-172.7* (88.649)	125.5 (139.528)	
Post	-0.32** (0.148)	-0.23 (0.652)	-0.32** (0.134)	-0.10*** (0.015)	-0.16*** (0.041)	-0.25 (0.297)	294.7*** (29.239)	184.8** (72.137)	61.4** (30.035)
Treat x Post	0.047 (0.197)	-0.063 (0.666)		-0.053** (0.022)	0.0035 (0.044)	0.096 (0.298)	-237.7*** (42.738)	-121.0 (77.679)	
Controls	N	N	N	N	N	N	N	N	N
<i>N</i>	996	586	574	2116	1151	1006	1998	1064	923
<i>R</i> ²	0.260	0.310	0.307	0.406	0.287	0.271	0.337	0.243	0.240
Treat	-0.90** (0.398)	-0.53 (0.756)		-0.085** (0.038)	-0.060 (0.070)	0.21 (0.170)	-130.7 (88.470)	165.4 (138.045)	
Post	-0.16 (0.151)	-0.22 (0.579)	-0.084 (0.143)	-0.15*** (0.017)	-0.19*** (0.041)	-0.26 (0.290)	184.8*** (32.337)	94.5 (73.296)	-5.11 (37.090)
Treat x Post	0.052 (0.186)	0.14 (0.584)		-0.069*** (0.022)	-0.0056 (0.043)	0.066 (0.290)	-259.4*** (42.336)	-107.2 (76.531)	
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
<i>N</i>	969	567	555	2049	1114	969	1941	1037	896
<i>R</i> ²	0.356	0.468	0.489	0.434	0.310	0.299	0.374	0.265	0.256

Notes*: In (1), (4) and (7) EAs are the Treatment Area while TTLs are the Control Area. In (2), (5) and (8) EAs are the Treatment Area while NPAs are the Control. In (3), (6) and (9), EAs Elsewhere are the Treatment Areas while European Highlands are the Control Area.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3 Chapter 3 Allocation of different crop exports to Agro-ecological regions

Table A.5 Export crops allocation

Natural region	Crops Types
I	Dairy farming, Tea, Coffee Bananas, Apples
II	Wheat, Maize, Tobacco Cotton, Citrus
III	Barley, Soya beans
IV	Millet, Sorghum
V	Cattle ranching, Oranges

Table A.6 Correlating trade in crops and exchange rates

Crop	Exports		Imports	
	r (quantity)	r (value)	r (quantity)	r (value)
Apples	-0.0858	-0.0938	-0.1082	-0.1082
Bananas	-0.3781	-0.3159	0.2617	0.2617
Barley			0.3108	0.2707
Catte_and_beef	0.6005	0.4448	-0.3881	-0.3676
CitrusJuice	0.4283	0.5721	-0.3574	-0.2358
Coffee	0.6319	0.4672	-0.4446	-0.4141
Cotton	-0.6413	-0.5550	-0.3083	-0.2569
Dairy		0.1063		-0.5111
Maize	0.4424	0.4851	0.2969	0.1646
Millet	0.4326	0.5665	0.3311	0.3737
Oranges	-0.8194	-0.6381	-0.0253	-0.0562
Sorghum	0.2776	0.4726	-0.0572	-0.0572
Soyabeans	0.0182	0.0858	0.5414	0.5506
Tea	-0.3841	-0.3841	-0.3391	-0.3761
Tobacco	0.4603	0.4285	-0.4964	-0.4964
Wheat	0.2522	0.1740	0.3101	0.3101

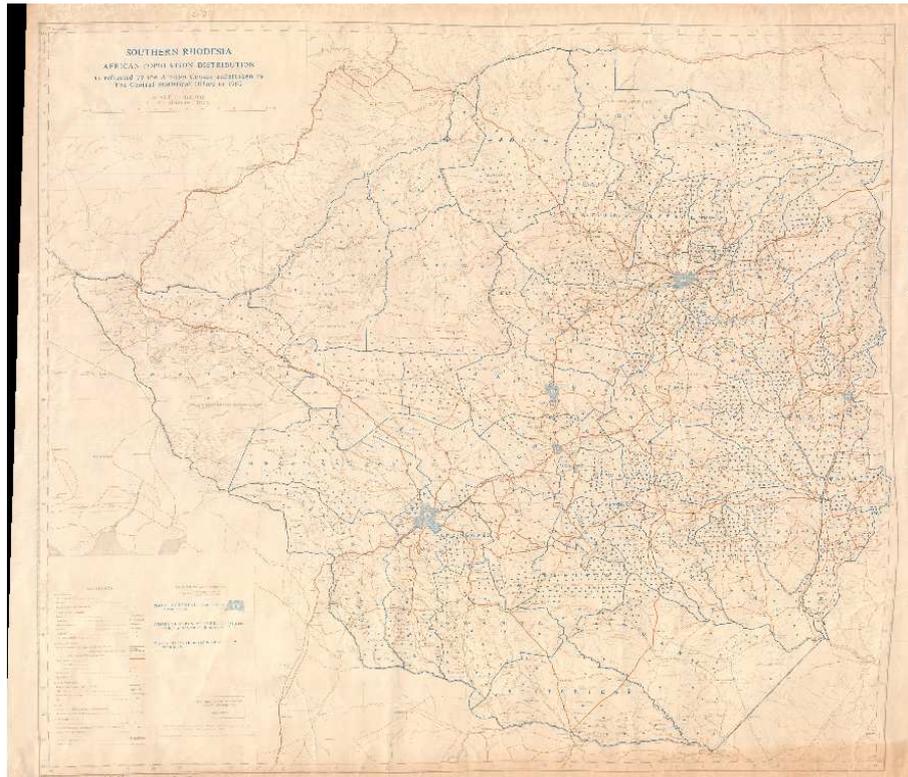


Figure A.1 Southern Rhodesia 1962 Population

Source: Federal Department of Trigonometrical and Topographical Surveys, R. a. N. (1963). Southern Rhodesia: African Population Distribution. from <https://digitalcollections.lib.uct.ac.za/collection/islandora-25215>

Table A.7 European Commercial Agriculture

Land class	Population Count
Tribal Trust Areas (TTAs)	4 163 000
European Areas (EAs)	2 220 000
Native Purchase Areas (NPAs)	297 000
Unassigned Class	226 000

Table A.6 shows the rural population data by land class based on the 1962 Southern Rhodesia census. Native Purchase Areas (NPAs) population as a percentage of total was 6.66%. We then split the data on the African population for 1970 and 1982 as provided by Zinyama and Whitlow (1986) (Table 1) into NPAs and TTAs using that ratio. We then allocate the 1970 and 1982 population figures to wards based on the 1962 population count, and impute figures for the intermediate years using a geometric mean. Using these figures, we repeat our analysis and the results remain robust.

Table A.8 Comparison between EAs and NPAs

Column	(1)	(2)	City(3)	Road(4)	Rail(5)	(6)	(7)	City(8)	Road(9)	Rail(10)
	<i>Period: 1972-1979</i>					<i>Period: 1972-1979</i>				
F-stat	3.20	3.77	4.88	10.74	15.92	0.06	0.13	2.05	2.36	4.7
p-value	0.07*	0.05*	0.027**	0.001***	0.0001***	0.8	0.7	0.15	0.12	0.03**

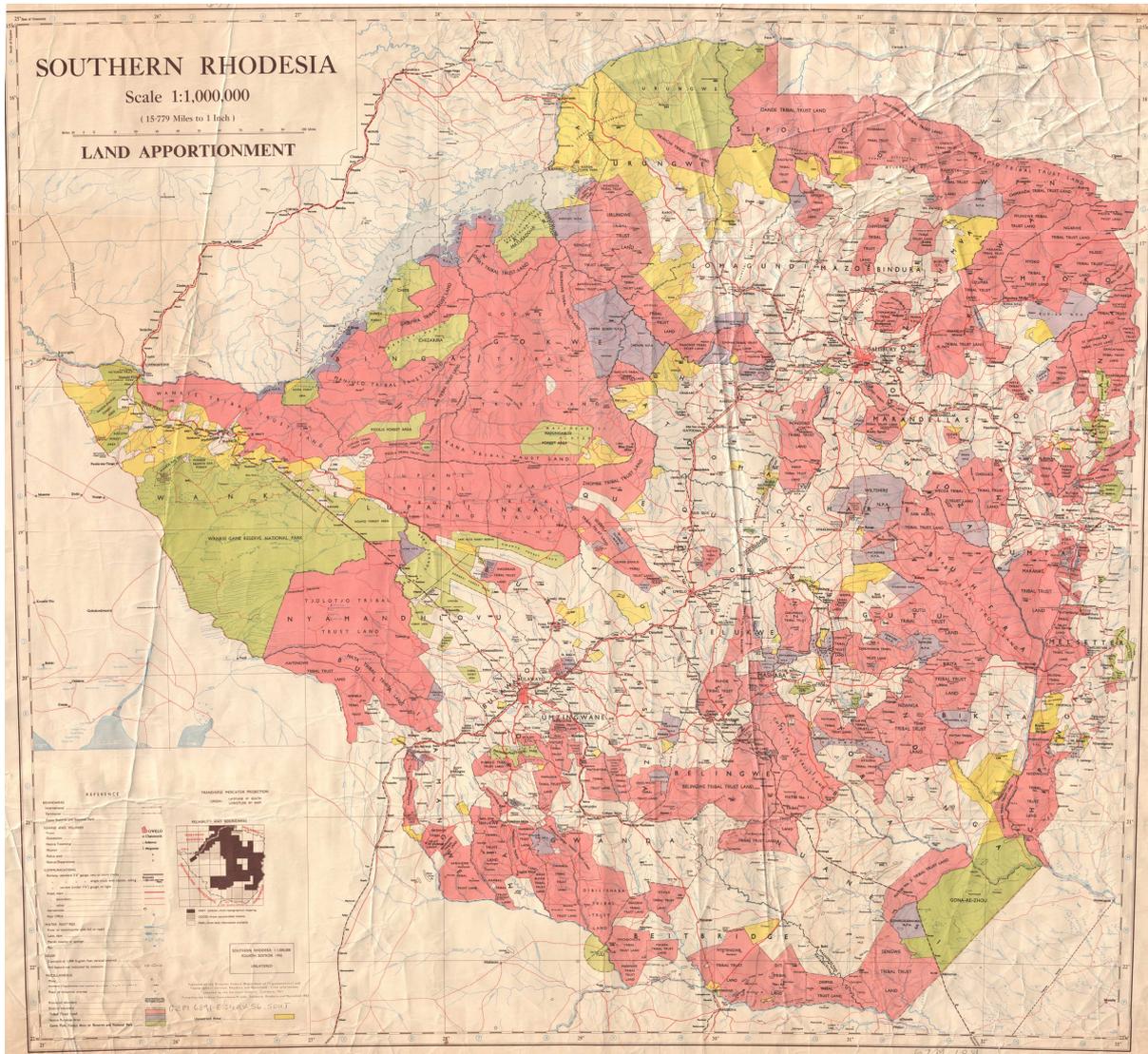


Figure A.2 1930 Land Apportionment Act

Source: Federal Department of Trigonometrical and Topographical Surveys, R. a. N. (1963). Southern Rhodesia: Land Apportionment. from <https://digitalcollections.lib.uct.ac.za/collection/islandora-25210>

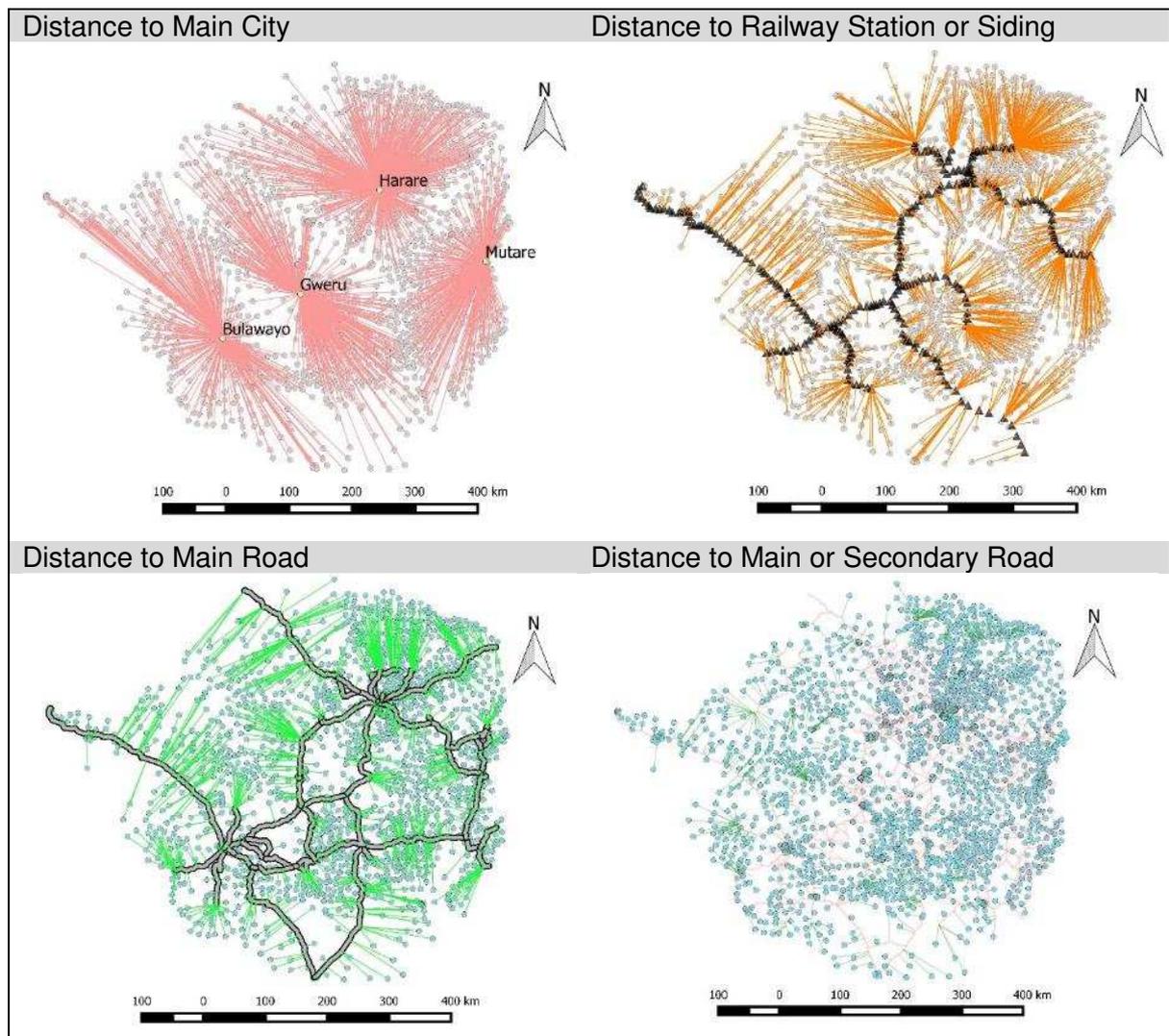


Figure A.3 Calculating Distance variables in QGIS

B. APPENDIX B: DESCRIPTION OF CROP CULTIVATION DATA WORK

B.1 The Data Approach

Chapters 2 and 3 relied on the application of the Support Vector Machine (SVM) Learning algorithm on Landsat imagery to generate crop cultivation data. Additional details on the machine learning approach used in the thesis are presented here.

B.2 Machine Learning

Landsat imagery were downloaded using US Geological Survey (USGS)'s Bulk Application Downloader (BAD) (<https://www.usgs.gov/media/images/earthexplorer-bulk-download-application-bda>). The images for Chapter 2 and 3 are listed in Tables B.1 and B.2 respectively.

Table B.1 List of Landsat Imagery used for Chapter 2 by date

Frame	Dates				
168072	08-Sep-72	25-Mar-73	24-Jun-84		
168073	08-Sep-72	25-Mar-73	26-Apr-79		
168074	03-Aug-72	13-Jan-73	25-Jul-79		
168075	07-Dec-72	30-Jan-73	25-Jul-79		
169072	04-Sep-73	09-Sep-79	02-Aug-84		
169073	08-Dec-72	17-Aug-73	10-Aug-76	18-Aug-84	
169074	17-Aug-73	10-Aug-76			
169075	08-Dec-72	10-Aug-76	16-Jun-78		
170071	10-Sep-72	22-Jun-84	27-Jul-79		
170072	11-Nov-72	09-Mar-73	31-Dec-84		
170073	03-Nov-72	14-Jan-73	11-Aug-76	26-Sep-84	
170074	10-Sep-72	05-Sep-73	08-Jan-76		
170075	03-Nov-72	19-Feb-73	22-Sep-75		
171071	22-Nov-72	06-Sep-73	12-Aug-76		
171072	06-Sep-73				
171073	06-Sep-73	12-Aug-76	29-Jun-84	22-Nov-72	
171074	11-Sep-72	28-Mar-73	13-Jul-75	12-Aug-76	29-Jun-84
172072	23-Nov-72	20-Jun-84			
172073	29-Dec-72	07-Sep-73	06-Sep-75	13-Aug-76	16-Aug-79
172074	29-Dec-72	13-Oct-73	13-Aug-76		
173072	02-Oct-72	05-Feb-73	14-Aug-76	02-Oct-79	
173073	14-Aug-76	18-Nov-84			

Table B.2 List of Landsat Imagery used for Chapter 3 by date

Frame	Dates					
168072	12-Apr-98	18-Nov-03	19-Dec-99	25-Apr-00	30-Apr-01	
168073	06-Dec-03	19-Dec-99	28-Apr-01	25-Apr-00	12-Apr-98	
168074	06-Dec-03	19-Dec-99	29-Dec-00	11-Mar-97	28-Apr-01	
168075	09-Apr-00	19-Dec-99	28-Apr-01	11-Mar-98	04-May-03	10-Feb-02
169072	06-Apr-02	08-Nov-99	09-Apr-03	10-Nov-97	19-Apr-01	
169073	08-Nov-99	10-Nov-97	13-Jan-98	29-Apr-03		
169074	03-Apr-01	10-Nov-00	14-Feb-98	26-Dec-99		
169075	02-Mar-98	09-Apr-03	16-Jan-02	26-Dec-99		
170071	02-Feb-97	05-Feb-98	15-May-02	11-Dec-00		
170072	02-Feb-97	05-Feb-98	12-Mar-02	26-Apr-01		
170073	02-Feb-97	04-Nov-01	05-Feb-98	23-Apr-00	23-Nov-02	
170074	17-Dec-99	15-Mar-03	26-Apr-98	23-Jan-02		
170075	07-Jan-02	23-Apr-97	15-Mar-03	23-Apr-00	26-Apr-98	15-May-99
171071	07-Jan-03	01-Jan-98	20-Apr-02	06-May-99		
171072	01-Apr-98	20-Apr-02	24-Dec-99	16-May-97	16-May-03	
171073	01-Apr-98	04-Apr-02	07-Apr-03	11-Jan-01	16-May-97	
171074	08-Jan-97	24-Dec-99	24-Nov-00	28-Feb-98	30-Jan-02	
171075	01-Apr-01	04-Apr-02	01-Apr-98	07-Apr-03	24-Dec-99	24-Nov-97
172072	05-Apr-00	06-Feb-02	11-Apr-99	20-Dec-01	24-Apr-98	
172073	06-Feb-02	10-Jan-01	21-Apr-97	13-Apr-00		
172074	03-Feb-98	07-Mar-01	13-Apr-00	14-Apr-03	22-Feb-99	21-Jan-02
173072	11-Dec-01	21-Apr-03	23-Feb-97	28-Jan-99		
173073	10-Feb-98	12-Apr-00	13-Feb-99	27-Dec-01	31-Jan-03	

As can be seen in Tables B.1 and B.2, Landsat images are available in Frames (see Figure B.1). After the images were downloaded, they were organised by Frame and dates for preparation before applying the SVM algorithm. Using Congedo (2014)'s Semi-Automatic Classification Plugin, Top of the Atmosphere (TOA) corrections account for the effect of atmospheric noise on the satellite sensor at the point of capturing the image. Using the SCP plugin, training polygons were collected for every Frame. Identification of land under cultivation and natural forest (for our binary classification) relied on expert visual inspection and analysis. Patil and Kumar (2011), asserts that crop leaf plant morphology is used for crop detection since it is inherently different to natural forest. Through visual inspection on Near Infrared (NIR) stacked virtual rasters in Quantum GIS (QGIS), fields and natural forest were identified. The decision rule for identifying land under crops relied on the bright-red regularly shaped blocks of fields since crops reflect more electromagnetic energy to the satellite sensor more in the NIR band. Image downloading targeted the November – March period since this is Zimbabwe's farming

season. Kussul et al., (2012) note problems with discerning summer crops using only optical imagery mainly due to cloud cover. Hence in Chapter 2⁴⁷, there is reliance on imagery falling out of Zimbabwe's summer cropping season. Subsequently, controls for season Fixed Effects (FEs) are included in the regression analysis. Hence the machine learning techniques in Chapter 2 and 3 rely on applying an SVM algorithm on Landsat optical imagery only for a binary cultivation and non-cultivation classification.

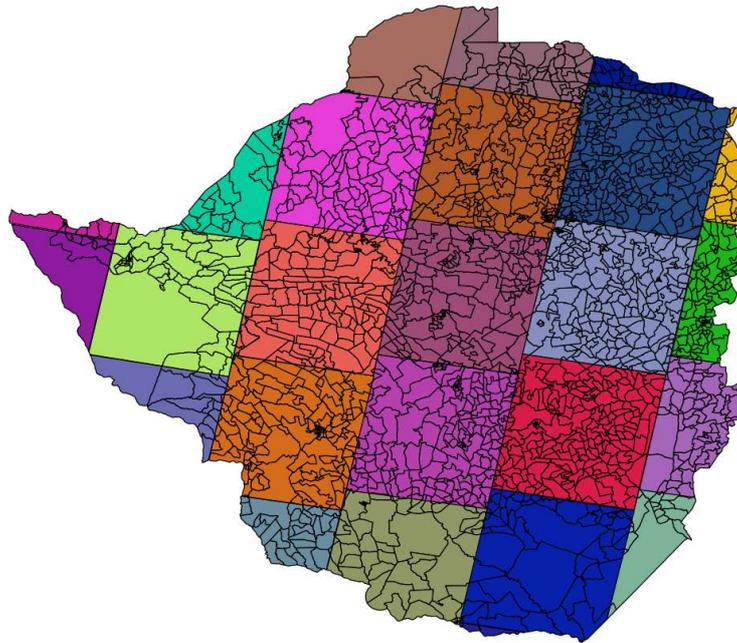


Figure B.1 Landsat Frames over Zimbabwe

I calculated the mean pixel values in the training polygons created in QGIS with the SCP plugin across the different multispectral bands and then saved that information as a shape file in preparation for implementing the SVM classification.

B.3 Classification

Classification followed Fernandes et al., (2015). Bioshop (2006) and Kussul et al., (2012) acknowledge the superiority of the SVM classifier. This is also reflected in the Kappa coefficients in the confusion matrices supplementary text for both Chapters 2 and 3 (<https://tinyurl.com/y64wdmow>), as well by visual inspection of the predicted raster versus the NIR image composites and Google Earth (GE) images. Appendix C also goes a step further in

⁴⁷ The period of analysis for Chapter 2 is 1972 – 1984, and imagery is from the first genres of Landsat satellites and as a result image quality and frequency is not that high for this period.

validating the crop cultivation data against other available macro level datasets for the Chapter 2 analysis period.

Classified pixel binary values are converted to 0 (natural forest) and 1 (land under crops). Given that the Landsat images for the different frames overlap (see Figure B.1), the predicted images are clipped using a common boundary for adjacent frames. The result is that wards falling on the frame boundary are cut as well. The cut area (Figure B.2) segments of such wards are added back together in STATA to obtain crop cultivation proportion values for the entire ward.

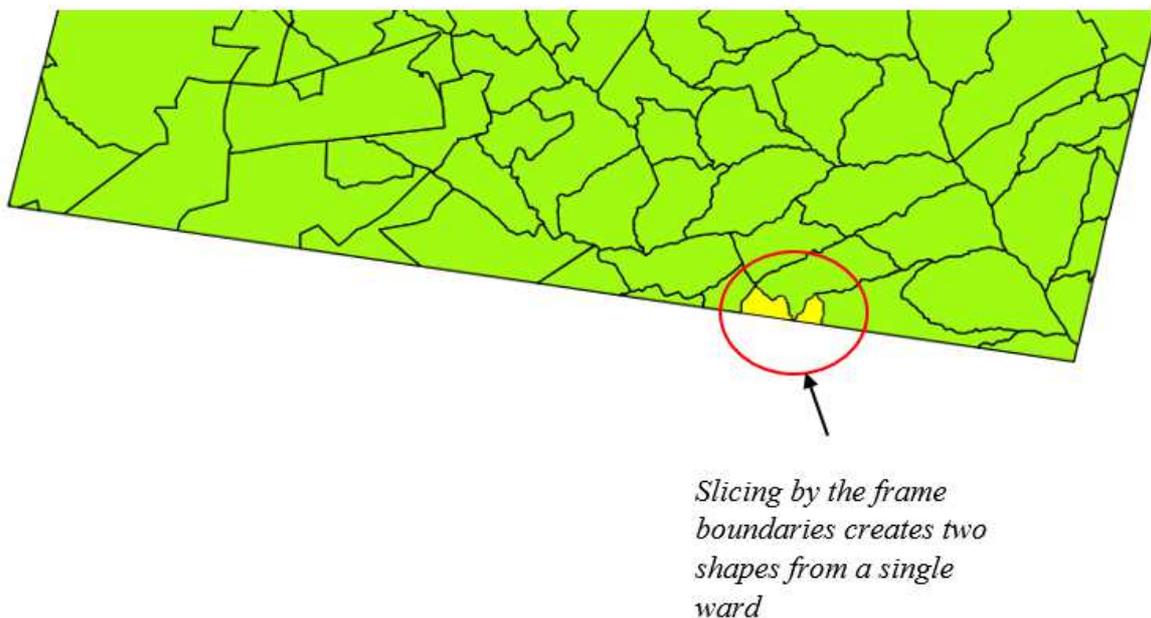


Figure B.2 Cutting Wards at Landsat Boundary

The zonal statistics (zonal mean) of the 0:1 pixel values in conducted in R, and the generated text files for each image are imported into STATA for data preparation pre analysis.

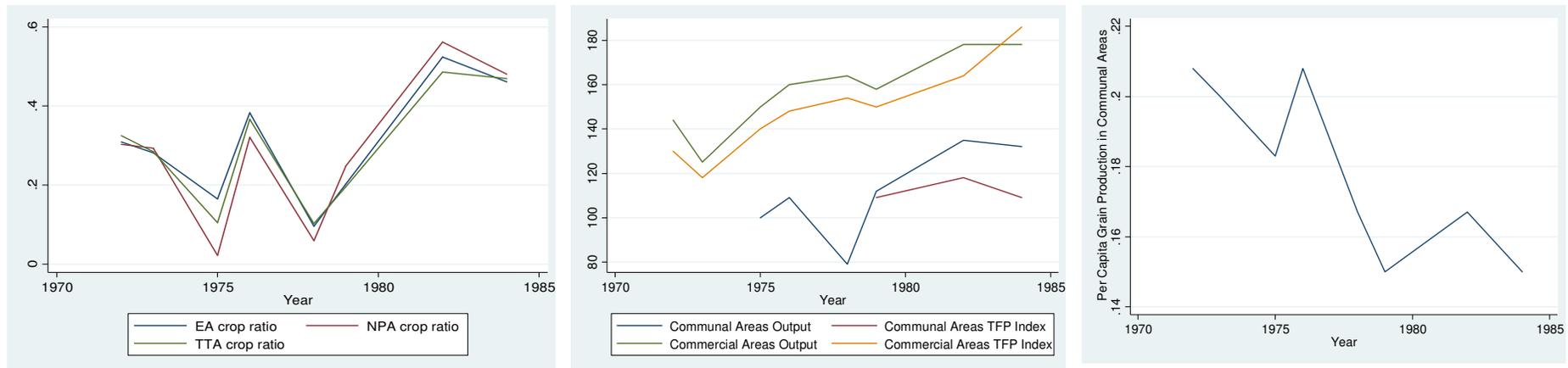
B.4 Processing Normalised Difference Vegetation Index

Chapter 3 relied on Normalised Vegetation Index (NDVI) as a proxy for crop quality. The data was also downloaded using USGS's BDA application. We used National Oceanic and Atmospheric Administration (NOAA) Climate Data Records (CDR) NDVI, which provides daily NDVI raster images which are created from weather satellites (<https://tinyurl.com/y3thxb8y>). MODIS NDVI products were not useful for our analysis because they become available after year 2000, yet our period of analysis begins in 1997. To arrive at annual figures, we averaged the daily NDVI raster images across the farming season. However,

the NDVI product includes natural forest. Hence we multiplied the annualised NDVI raster image with the crop classification raster image [0 for natural forest and 1 for area under cultivation] (after co-registration) by frame to filter out NDVI for natural forest.

C. APPENDIX C: SOME ACCURACY VALIDATION

For Chapters 2 and 3, classification performance for each Landsat path/row frame or footprint are available here: <https://tinyurl.com/y64wdmow>. However, the data validation in this section of the appendices was done using data generated for Chapter 2 only. Figure C.1 shows a comparison of the agriculture data generated by the study to aggregate data in Thirtle et al., (1993) and Jayne and Jones (1997). The scales are different, making this an uneasy task, yet the very high correlations shown in Table C.1 inspire confidence in the data.



a. Ratio of crop to ward area (endogenous variables) created with machine learning from Landsat images. It is shown here for European Areas (EA), Native Purchase Areas (NPAs) and Tribal Trust Areas (TTAs)

b. African communal areas and European commercial areas agriculture aggregate output and Total Factor Productivity (TFP) indices in Thirtle et al., (1993).
* Graph reconstructed using approximate figures from the original.

c. Per capita grain production in the African communal lands from Jayne and Jones (1997).
* Graph reconstructed using approximate figures from the original.

Figure C.1 Comparison of machine-learning generated data to other sources

Table C.1 Time Series Correlations of European Commercial Agriculture with Aggregates from Classified Images

	EAs crop ratio	EA crop hectorage	Commercial Agriculture Output	Commercial Agriculture TFP Index
EAs crop ratio	1			
EAs crop hectorage	0.9645	1		
Commercial Agriculture Output	0.4495	0.4413	1	
Commercial Agriculture TFP Index	0.4527	0.399	0.9448	1

Table C.2 Correlations of Time Series of African Agriculture with Aggregates from Classified Images

	Per Capita Grain Production in Communal Areas	Communal Areas Output	Communal Areas TFP Index	NPA crop ratio	NPA crop hectorage	TTA crop ratio	TTA crop hectorage
Per Capita Grain Production in Communal Areas	1						
Communal Areas Output	0.6003	1					
Communal Areas TFP Index	1	0.6003	1				
NPA crop ratio	0.7004	0.9913	0.7004	1			
NPA crop hectorage	0.0733	0.8417	0.0733	0.7632	1		
TTA crop ratio	0.5419	0.9975	0.5419	0.9794	0.8779	1	
TTA crop hectorage	0.627	0.9994	0.627	0.9952	0.8229	0.9945	1

As already discussed, the individual graphs in Figure C.1 are not superimposed on each other for easier comparison due to the different scales. Rather, correlation coefficients are calculated between the data classified from Landsat imagery through machine and aggregate indicators in Jayne and Jones (1997) and Thirtle et al. (1993). Table C.1 and C.2 show the correlation results for European commercial and African agriculture respectively. In Table C.1 and C.2, EAs crop ratio, EAs crop hectorage, NPA crop ratio, NPA crop hectorage, TTA crop ratio and TTA crop hectorage represent the data created through machine learning from Landsat imagery, whereas the rest represents data from the other sources. The correlations are quite high (as high as 0.9994), which means that the data are reliable.

NIR Composite



SVM Predicted Algorithm



Figure C.2 Comparing Near Infrared composite and predicted raster

Source: Near Infrared Composite (NIR) created in QGIS using Landsat multispectral bands. After identifying the training points, the SVM algorithm is implemented in R to obtain the predicted image in black and white on the right.

D. APPENDIX D: VALIDATION OF DID AND RDD ASSUMPTIONS

This section of the appendices presents the assumptions underpinning the DID and RDD estimation that was relied upon in the study. For DID, we focus on the parallel trends assumption, while for the RDD estimation we explore and discuss the validity of a number of assumptions. The RDD subsection deals with the issues/assumptions of i) discontinuity at border, ii) compound effects, iii) spatial variation in treatment (heterogeneity), iv) bandwidth selection, v) choosing between naïve and chordal (geodetic) distances, vi) non-parametric polynomial estimation asymptotic bias, vii) covariate balance and viii) falsification tests.

D.1 Difference in Difference (DID) – Parallel Trends Assumption

We implanted semi-parametric matching algorithm following Abadie (2005) and results suggest that the parallel trends identifying assumption only holds for the case of NDVI (for the comparison between EAs Elsewhere and EAs Eastern Highlands) as shown in Figure D.1.

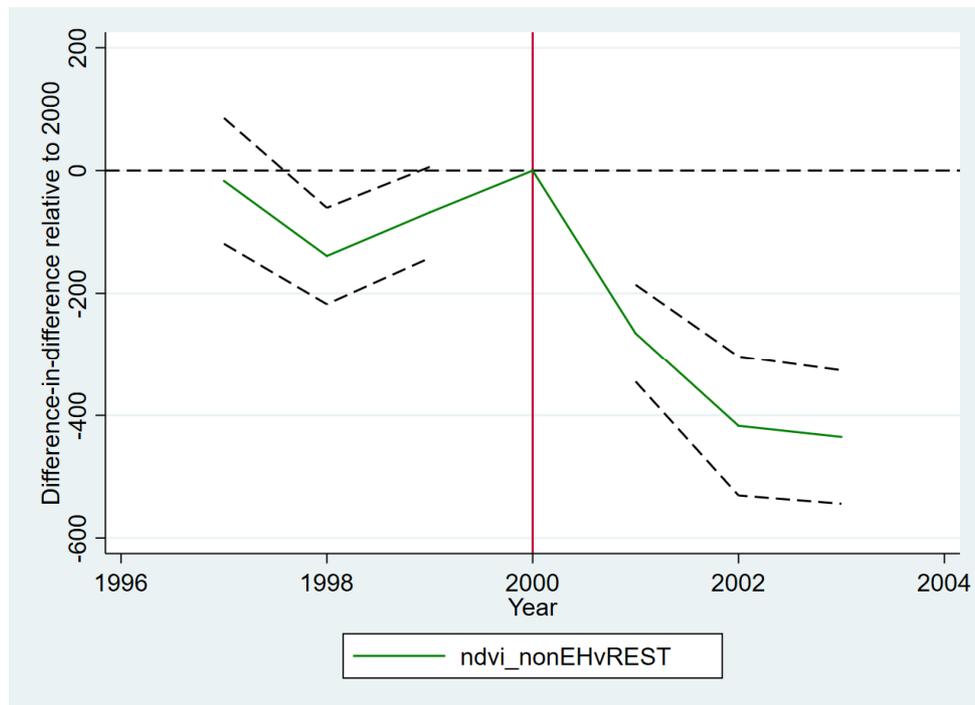


Figure D.1 DID Parallel Trends Assumptions

D.2 RDD Assumptions

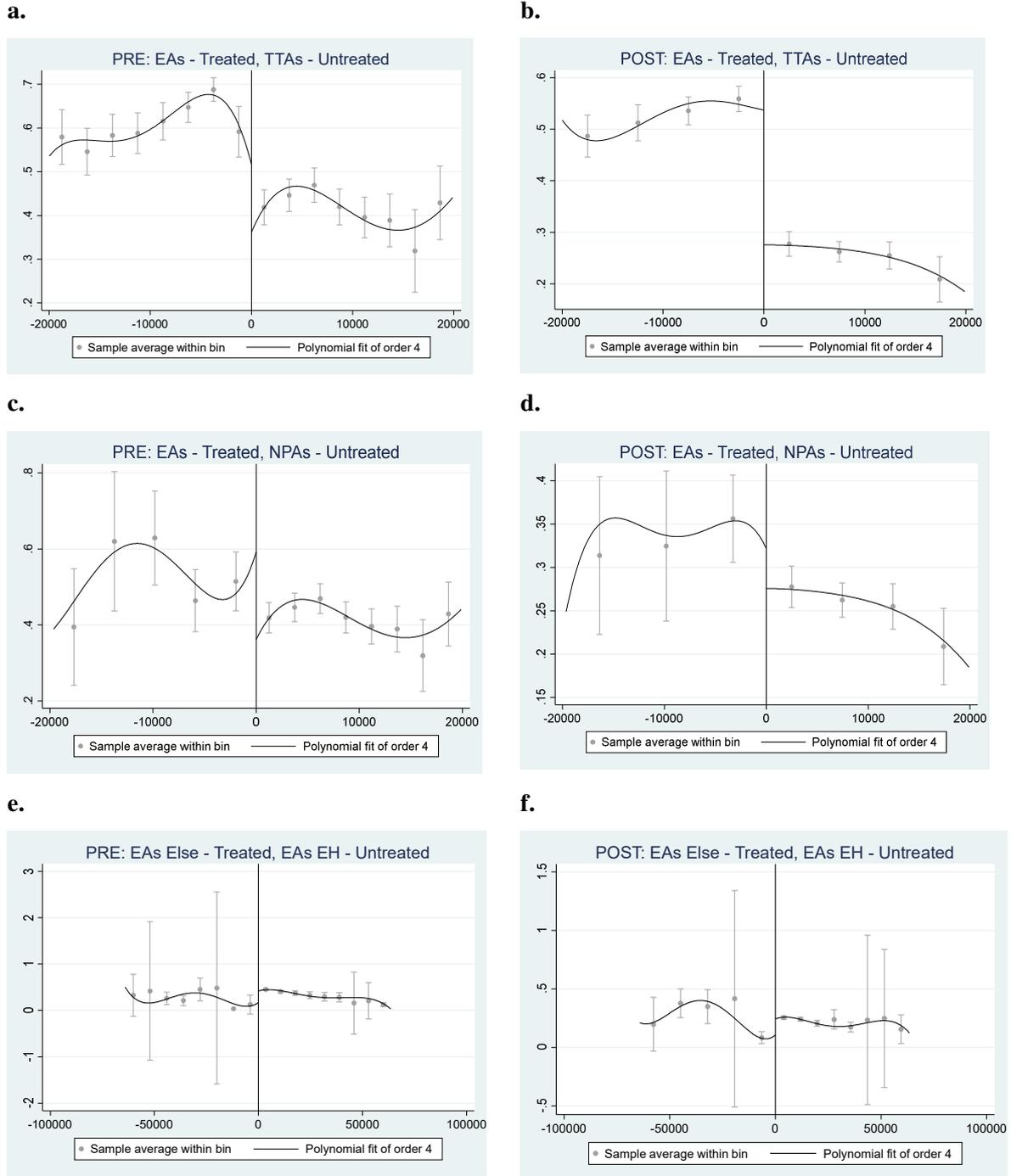
i) Dis-continuity at Border

To check for dis-continuity at border, we construct RD plots (using *rdplot* command) for the three comparison groups in the PRE and POST periods. We use observations within 20km from the border (on both sides) because confidence interval bands expand as distance from the border increases. In Figures D.2 and D.3 panels **a.** to **f.** appear to show that discontinuities at border increased after FTLRP. For Figure D.2 Panel **c.** overlapping Confidence Intervals (CIs) on the left and right side of the border means that the discontinuity is not statistically significant in the PRE phase while the CIs do not overlap in panel **d.** – which shows that the discontinuity is statistically significant after FTLRP. In other panels, for example **a.** and **d.** there is growth in discontinuity in the POST phase. Overall, the plots in Figures D.2 and D.3 show evidence of dis-continuity at the border – which is an important identifying assumption under the RDD design.

ii) Compound Treatments

Keele and Titiunik (2014) point that in Geographic Regression Discontinuity (GRD), there is the risk that there may be several treatments affecting an area because treatment boundaries (for example tenure regime as defined by LAA 1930 in our study) often overlap with other administration boundaries. In Chapter 3, the boundaries of EAs, TTAs and NPAs coincide with ward boundaries (our unit of observation) (*see* Figure D.4). This means that moving from one land class type to another essentially means moving from one ward to another, which may have different idiosyncratic factors affecting the outcome of interest (in our case crop cultivation and crop quality).

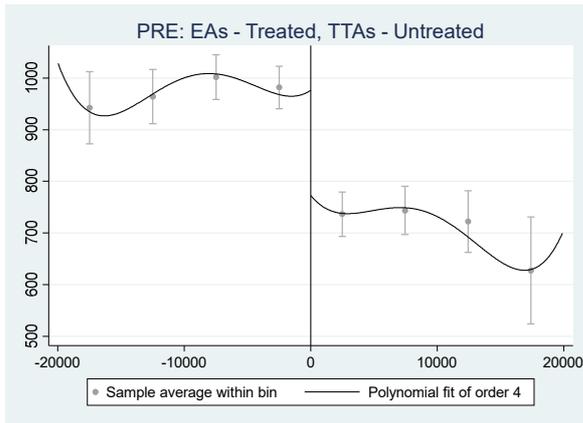
Keele and Titiunik (2014) suggest that one strategy to deal with compound effects is to make inference for geographic areas where treatment assignment differs in the same administrative unit. If we had household level data, we would carry out the causal inference for areas in which the land use type boundary divides a particular ward into treated and untreated sub-regions, yet all other administrative factors would be held constant (as shown in Figure D.4). However, a consequence of the lack of individual farm level data is also the lack of any within ward variation (for both endogenous variable and the covariates) since the ward is our unit of analysis in the small area estimation. Hence distances are calculated from the centre of the ward (the centroid). Figure D.5 shows Ward 10 in Beitbridge District and Ward 7 in Mangwe District, Matebeleland South Province. In Panel B of the image, these wards are split into TTA and EA regions, a scenario ideal for isolating compound treatments if we had household level data.



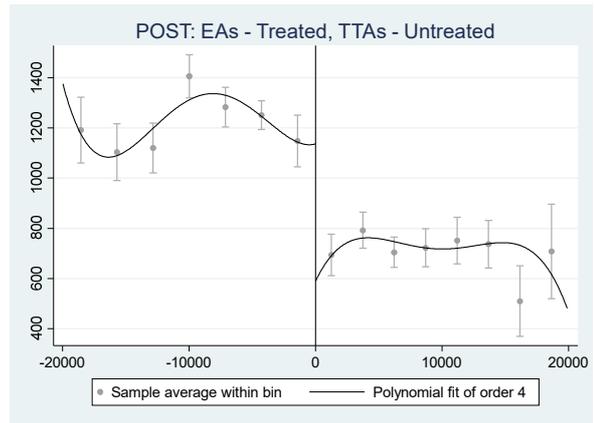
NOTES*: The RD plots are drawn using 20km from border as distance cutoffs. The last row on the figure shows the comparison with EAs EH (treated) and EAs Else (untreated). The distances for EAs EH fall within the 10km threshold because the EH is a relatively small region on the eastern fringes of Zimbabwe.

Figure D.2 RD plots with CI – Crop Cultivation

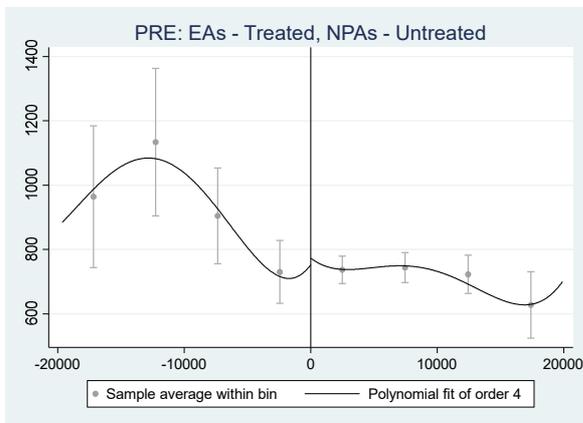
a.



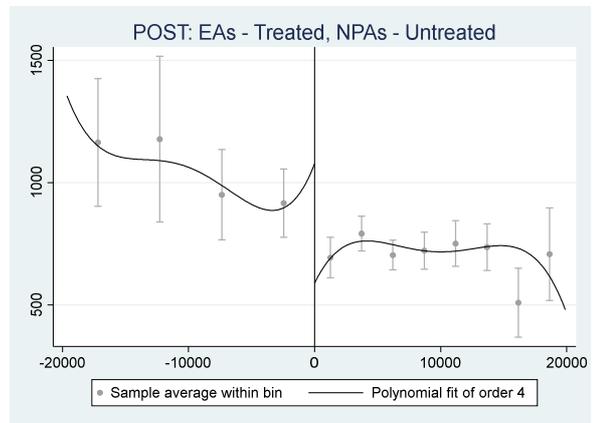
b.



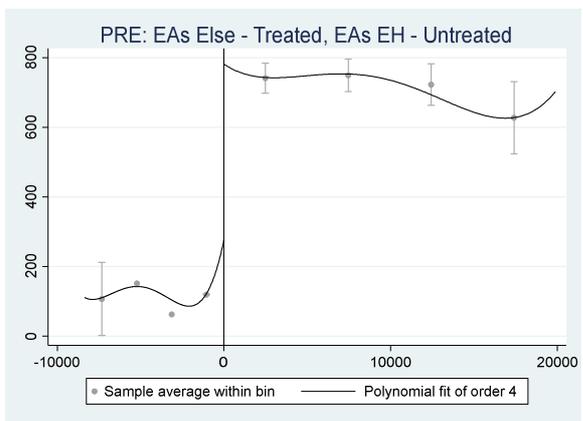
c.



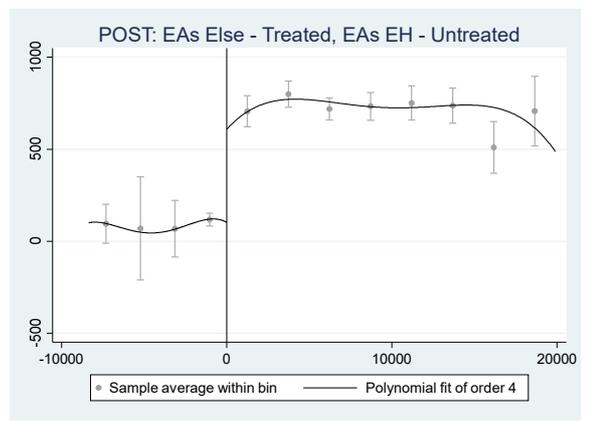
d.



e.



f.



NOTES*: The RD plots are drawn using 20km from border as distance cutoffs. The last row on the figure shows the comparison with EAs EH (treated) and EAs Else (untreated). The distances for EAs EH fall within the 10km threshold because the EH is a relatively small region on the eastern fringes of Zimbabwe.

Figure D.3 RD plots with CI – NDVI

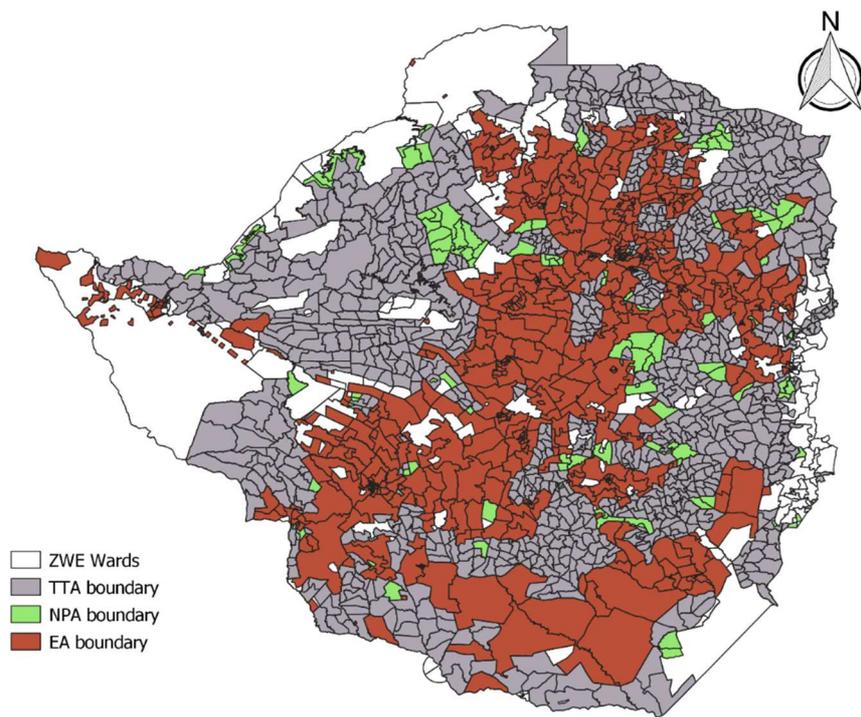


Figure D.4 Ward boundaries overlapping with LAA 1930

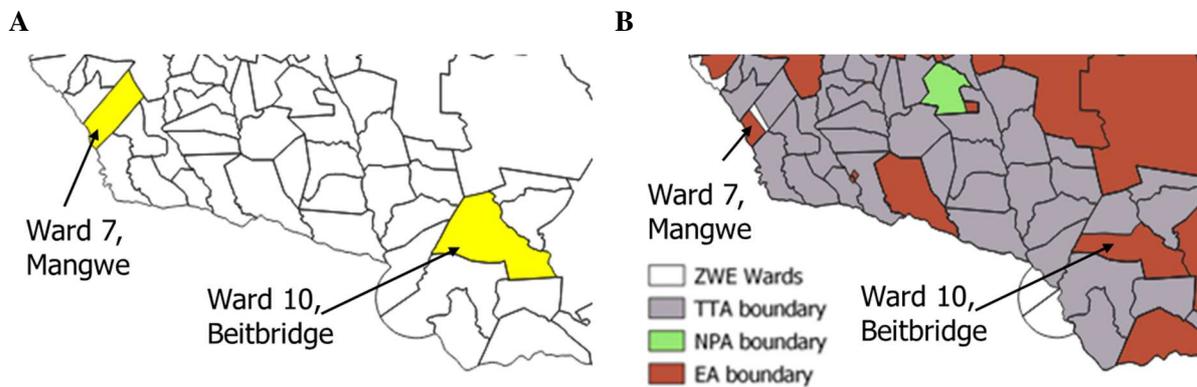


Figure D.5 Ideal conditions for GRD

The nature of our analysis makes it impossible⁴⁸ to rely on within ward variation, therefore we can only test the effect of compounding effects at the district administrative level. As a workaround, we could deal with potential compound effects of multiple treatment by making comparisons across similar wards within the same districts that have both treated and untreated sub-regions. As shown in Figure D.6, most of the districts (if not all) have both treated (EAs) and untreated (either TTAs or NPAs) hence the compounding effects of other treatments that may hold for the Zimbabwean case is accounted for by controlling for district FEs in the RDD estimation.

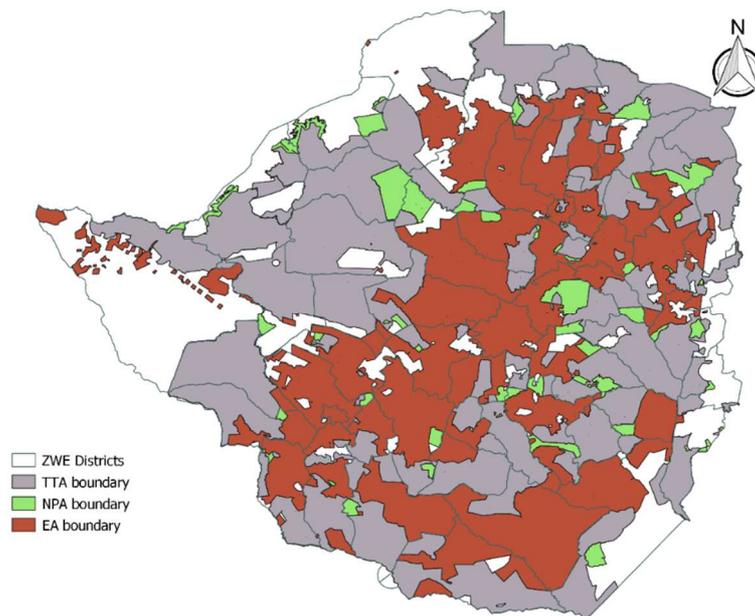


Figure D.6 Controlling for District FEs to deal with compound Effects

iii) Dealing with asymptotic bias of standard errors

Standard confidence intervals are ordinarily premised on the asymptotic distribution of the OLS and robust standard errors, yet these do not hold for the for RDD non-parametric local polynomial estimator, unless distance distances cut-offs are small enough (under smoothing) (Keele and Titianic, 2015). Our strategy of dealing with this is a two pronged one. First, we employ distance cut-offs of 1km, 2km and 5km for our main results, which should be small enough. The second approach is a data driven approach where we estimate effect without

⁴⁸ Our unit analysis is the ward. Distances are measured from the ward centroid. The only ward is the only geo-location unit we have information on. We do not have household level data (and the accompanying geo-locations).

specifying the distance cut-offs, thereby allowing *mdrd* command and the related *rbwselect* to select the distance cutoff (h) and bias correction bandwidth (b) (see Keele and Titiunik, 2015; Rafas, 2016). Standard estimation of the RDD polynomial is implemented in STATA using the ‘*rdrobust*’ package (see Calonico et al., 2017; Cattaneo, 2015), but the *rdrobust* command however suffers implementation challenges for multi-dimensional RDD (where there is geography but also time component). We also rely on *mdrd* package by Ribas (2016) since it is easier to implement. Also our main estimates rely on small bandwidths (1km, 2km and 5km) which should also deal with this bias.

iv) Selection of Cutoffs

In Chapter 4 main estimation, we present results reduced at 2, 5 and 10 km from the border. For robustness, we repeat analysis using the *bwselect* command of the *rdrobust* package. The *bwselect* command is data driven, in that it established bandwidths that are optimal for the dataset. We use *rdrobust* estimations for the for the PRE and POST phases. The ‘*rdrobust*’ results for the full specifications (with all controls) are shown in Table D.1 and they are robust.

Table D.1 Using *rdrobust*

<i>Dep. var</i>	Lights		Crop Cultivation		NDVI	
	PRE	POST	PRE	POST	PRE	POST
	β /S.E	β /S.E	β /S.E	β /S.E	β /S.E	β /S.E
Conventional ¹	-0.02566 (0.0472)	0.01905 (0.0767)	-0.13524*** (0.03201)	-0.25423*** (0.03509)	-113.08** (53.277)	-366.2*** (75.476)
Conventional ²	0.05124 (0.05848)	-0.06095 (0.11878)	-0.13234** (0.05331)	-0.0467 (0.05122)	19.151 (92.297)	-197.54 (173.33)

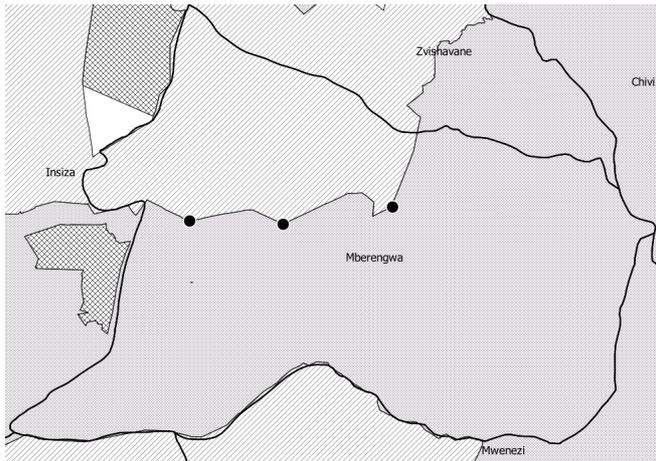
Notes*: Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Conventional¹ = coefficient for EAs (Treated) and TTAs (Control). Conventional² = coefficient for EAs (Treated) and NPAs (Control). Automatic bandwidth selection for Eastern Highlands was unsuccessful due to small distances to the border in the area. Thus, results are only presented for two treatment and control groupings.

v) Demonstrating heterogeneity using Euclidian Distance

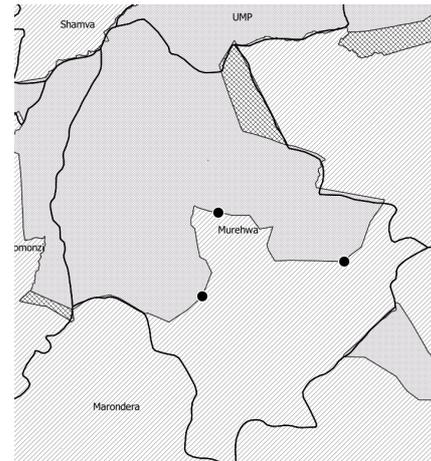
Keele and Titiunik (2015) used chordal distance as opposed to naïve distance because the latter masked important spatial relationships amongst observations. The empirical analysis by Keele and Titiunik (2015) relied on a single school district (West Windsor-Plainsboro) since they needed to deal to compound effects of treatment. As explained in the second section where we

dealt with compound treatments, there is enough heterogeneity by land class (according to LAA 1930) in each and every district, hence we rely on all of Zimbabwe's districts for our analysis and we show the average effect for the entire country. Keele and Titiunic (2015) selected three points on the border (from the set of all border points G), and calculated the chordal distance between each observation in their sample the border point. In this section we use Euclidian distance from the ward centroid to the three arbitrary points on the border between TTAs and EAs for selected provinces. We mimic Keele and Titiunic (2015)'s approach, and the aim is to check the robustness of findings if Euclidian distance is used and also to demonstrate heterogeneity of effect if we had chosen not to compute the average countrywide effect as we do in the main estimates shown in the chapter. We arbitrarily select 3 districts ('*Mberengwa*', '*Murehwa*' and '*Kwekwe*') (see Figure D.7). Analysis at the district level is not possible due to insufficient observations. Therefore, we carry out the robustness at the provincial level to gain more statistical power. Tables D.2 and D.3 presents statistical estimates for Midlands and Mashonaland East Provinces. The estimates for crop cultivation are robust to our main results.

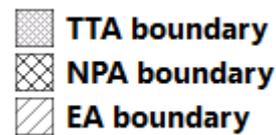
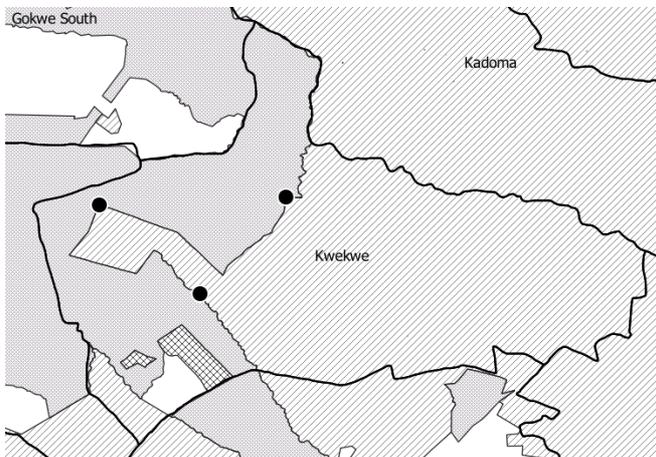
Mberengwa



Murehwa



Kwekwe



NOTES*: Black bold marking is used for District boundaries.

Figure D.7 Mberengwa, Murehwa and Kwekwe Test Areas

The point estimates for both Midlands and Mashonaland East Provinces are consistent with the main estimates in the chapter (negative effect on Lights, Cultivation and NDVI for the three points). But they also show some level of heterogeneity and thus require estimating of the national average effect as we do in the main effects.

Table D.2 Heterogeneity Robustness: EAs (Treated) TTAs (Untreated) - Midlands

	Boundary Point	PRE			POST		
		Lights	Cultivation	NDVI	Lights	Cultivation	NDVI
<i>No Controls</i>	1	0.47 (0.569)	-0.068 (0.178)	515.2 (404.991)	0.65 (0.530)	-0.49*** (0.118)	-756.8 (494.626)
	2	-0.44 (0.618)	0.14 (0.225)	-379.5 (610.317)	-0.54 (0.771)	-0.40*** (0.136)	-353.4 (486.928)
	3	-1.53* (0.847)	-0.22 (0.256)	147.8 (637.373)	-1.16 (1.117)	-0.50*** (0.181)	-477.6 (448.124)
<i>Controls</i>	1	-0.29 (0.431)	-0.15 (0.200)	983.1*** (331.597)	-0.45 (0.456)	-0.27 (0.172)	-332.5 (285.631)
	2	-0.55 (0.640)	0.18 (0.215)	-683.8 (592.161)	0.36 (0.933)	-0.24 (0.174)	-214.9 (492.517)
	3	-2*** (0.497)	-0.42 (0.260)	-19.5 (366.261)	-1.84** (0.738)	-0.55*** (0.119)	-1128.6*** (273.244)

NOTES*: Bandwidths are chosen selected based on CCT MSE minimization approach [see Calonico, Cattaneo and Titiunik (2016)].

Table D.3 Heterogeneity Robustness: EAs (Treated) TTAs (Untreated) – Mashonaland East

	Boundary Point	PRE			POST		
		Lights	Cultivation	NDVI	Lights	Cultivation	NDVI
<i>No Controls</i>	1	-5.2*** (1.695)	-0.12 (0.175)	-651.0* (364.608)	-2.43 (1.586)	-0.33** (0.130)	-407.4 (587.842)
	2	-0.40 (0.422)	-0.19 (0.116)	-24.2 (442.087)	0.094 (0.456)	-0.30** (0.117)	-351.2 (357.328)
	3	-1.19 (0.787)	-0.21 (0.241)	-661.1* (379.218)	-3.6*** (1.306)	-0.42*** (0.143)	-540.5 (566.358)
<i>Controls</i>	1	-4.2*** (1.605)	0.046 (0.198)	-210.8 (353.844)	-0.97 (1.624)	-0.040 (0.077)	-1605.6*** (351.588)
	2	1.60*** (0.411)	-0.11 (0.112)	446.2 (319.200)	3.15*** (0.726)	-0.17* (0.094)	-63.2 (426.989)
	3	-0.28 (0.458)	-0.071 (0.202)	-167.8 (317.441)	-0.56 (1.051)	-0.21*** (0.071)	-453.8** (192.227)

NOTES*: Bandwidths are chosen selected based on CCT MSE minimization approach [see Calonico, Cattaneo and Titiunik (2016)].

vi) Pre-treatment Covariate Balance

Another important assumption of RDD is balance in the pre-treatment covariates. It is only the outcome variable (in this case distance) that is expected to show discontinuity (jump) at the border. We test pre-treatment covariate balance only for the main comparison groups (EAs and TTAs) using the covariates if rainfall, temperature and population PRE and POST. As shown

in Figure D.7, there were no ‘sharp’ discontinuities in the covariates at the border PRE and POST. The plots are also more or less similar between the PRE and POST phases.

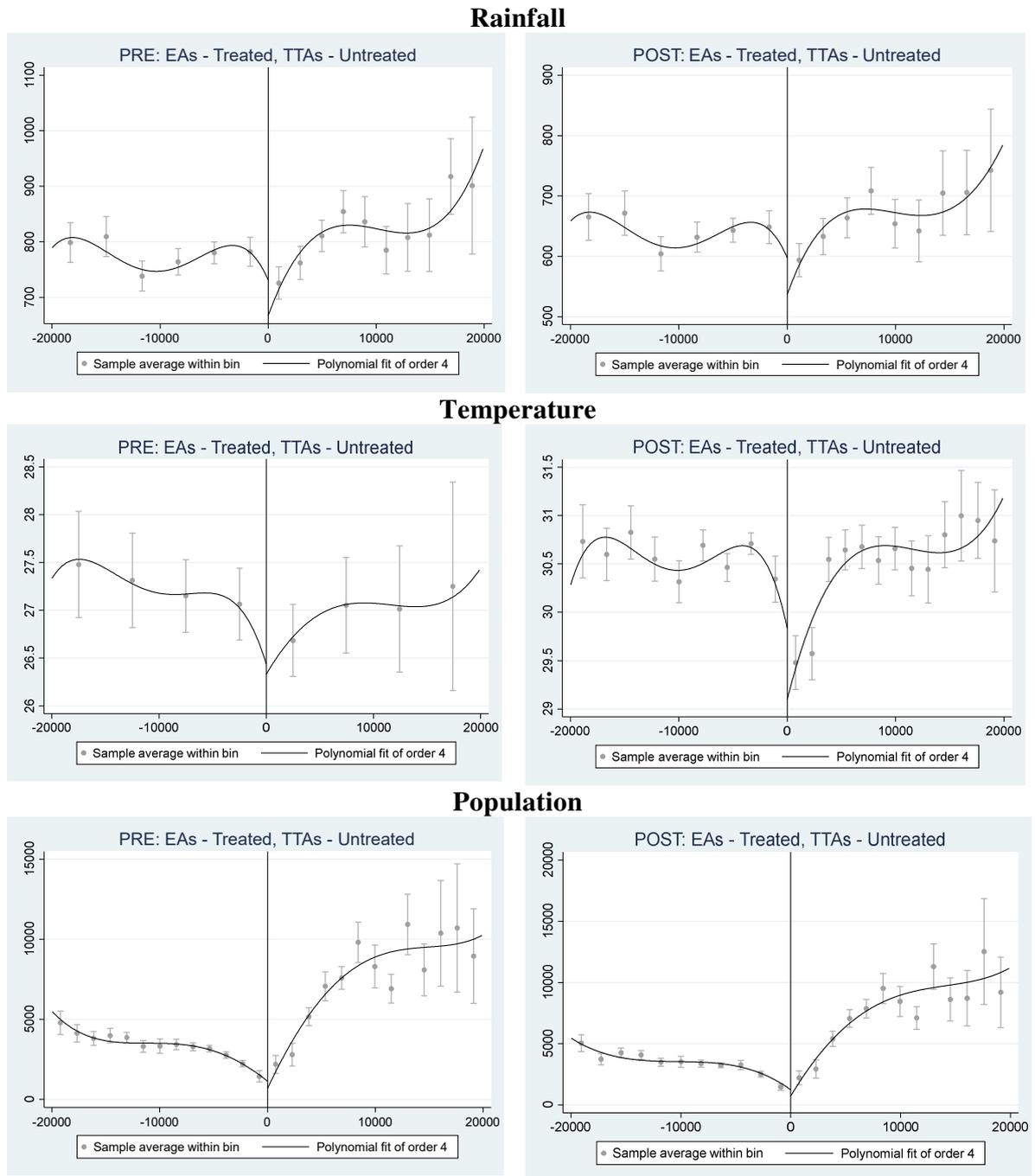


Figure D.8 Pre-treatment covariate balance

E. APPENDIX E: ZOONIVERSE DATA WORK

Chapter 4 employed citizen science to detect informal structures and other urban land use types from Very High Resolution (VHR) images. Citizen science has been used in various fields and the Zooniverse platform (on which this project was hosted) was created primarily for to allow lay astronomers and other enthusiasts to identify new galaxies from night-sky images. In this section, the process of data preparation is outlined. The study recruited 41 Stellenbosch University (SU) students to pilot the classification of informality and other land use types. 180 images split equally between 2004 and 2006 for the same areas: - and for areas affected and not affected by the 2005 clean-up operation were downloaded from Google Earth (GE) and uploaded onto the Zooniverse platform. After classification, the classified polygons were exported from the Zooniverse platform in CSV format and then processed in QGIS mainly using the PyQGIS library.

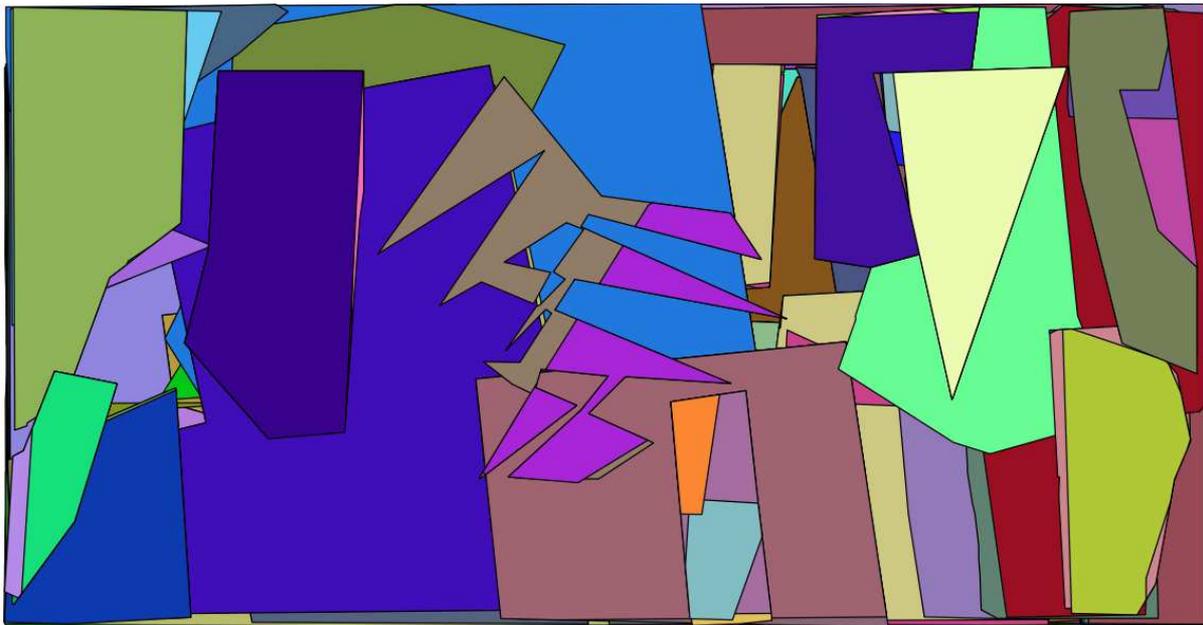


Figure E.1 User classification polygons in QGIS map canvas

The CSV files imported from Zooniverse.org were cleaned using STATA. The STATA script separated classification by image, user/volunteer and task number⁴⁹ and saved them as CSV files importing into QGIS. After importing into QGIS, the vertices of the user polygons drawn during classification on Zooniverse were plotted into polygon shape files. Figure E.1 shows

⁴⁹ Volunteers/students had four classification tasks for each image (although not all classification tasks were identifiable on each image)

several user polygons drawn on QGIS map canvas. These user polygons were then intersected with the reference polygons (from our expert classifications) and again saved as CSV files for importing into STATA. In STATA, the classification accuracy rates were computed using the formulas described in the main thesis body. The process described in the preceding paragraph was iterated for preparing consensus estimates. Instead of intersecting individual user classifications by image and task number with the reference classification, all user classifications for the same image name and task number were intersected with each other to obtain the consensus classification. This consensus classification/polygon was then intersected with the reference polygon. The consensus classification accuracy rates eventually created in STATA were low due to including erroneous outlier classifications in the first step of creating a consensus polygon.

Table E.1 Area Classified by Type

Class	2004 <i>(km²)</i>	2006 <i>(km²)</i>
unassigned class	29,87	21,52
informal business	0,84	0,16
backyard structures	7,31	0,08
formal housing	4,19	15,51
formal business	3,78	4,58
Total Area	46	46