2017

Applications of econometrics and machine learning to development and international economics

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

APPLICATIONS OF ECONOMETRICS AND MACHINE LEARNING TO DEVELOPMENT AND INTERNATIONAL ECONOMICS

by

JONATHAN S. HERSH

B.A., University of Chicago, 2005
A.M., University of Pennsylvania, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader

Marianne Baxter, Ph.D.
Professor of Economics

Second Reader

Sam Bazzi, Ph.D.
Assistant Professor of Economics

Third Reader

Ray Fisman, Ph.D.
Professor of Economics

DEDICATION

This work is dedicated to the late Robert Fogel, who manifested the excitement of research; to the late Albert Maysles, who continues to remind me that compassion and intellect can co-exist; and to my father, who was dedicated to my intellectual development long before I was.

ACKNOWLEDGMENTS

This work was completed in part while in residence at the International Bank for Reconstruction and Development at the World Bank. I am grateful to all of my colleagues there, but in particular to David Newhouse, Nobuo Yoshida, Kristen Himelein, Ana Areias, Ghazala Mansuri, Andrew Whitby, and Bruno Sanchez-Andrade Nuno. I also gratefully acknowledge support from the Hariri Institute at Boston University, and thank Azer Bestavros and John Byers for their advice and guidance.

I am indebted to the economics faculty at Boston University whose time and support helped shape this research. I am grateful to all of the economics faculty there, but those whose work directly influenced this document include Kehinde Ajayi, Stefania Garetto, Adam Guren, Robert King, Dilip Mookherjee, and Pierre Perron. Particular appreciation is owed to my dissertation committee, Marianne Baxter, Sam Bazzi, and Ray Fisman, without whom this work would be much less intellectually satisfying.

APPLICATIONS OF ECONOMETRICS AND MACHINE LEARNING TO DEVELOPMENT AND INTERNATIONAL ECONOMICS

JONATHAN S. HERSH

Boston University Graduate School of Arts and Sciences, 2017

Major Professor: Marianne Baxter, Professor of Economics

ABSTRACT

In the first chapter, I explore whether features derived from high resolution satellite images of Sri Lanka are able to predict poverty or income at local areas. I extract from satellite imagery area specific indicators of economic well-being including the number of cars, type and extent of crops, length and type of roads, roof extent and roof type, building height and number of buildings. Estimated models are able to explain between 60 to 65 percent of the village-specific variation in poverty and average level of log income.

The second chapter investigates the effects of preferential trade programs such as the U.S. African Growth and Opportunity Act (AGOA) on the direction of African countries' exports. While these programs intend to promote African exports, textbook models of trade suggest that such asymmetric tariff reductions could divert African exports from other destinations to the tariff reducing economy. I examine the import patterns of 177 countries and estimate the diversion effect using a triple-difference estimation strategy, which exploits time variation in the product and country coverage of AGOA. I find no

evidence of systematic trade diversion within Africa, but do find evidence of diversion from other industrialized destinations, particularly for apparel products.

In the third chapter I apply three model selection methods – Lasso regularized regression, Bayesian Model Averaging, and Extreme Bound Analysis -- to candidate variables in a gravity models of trade. I use a panel dataset of of 198 countries covering the years 1970 to 2000, and find model selection methods suggest many fewer variables are robust that those suggested by the null hypothesis rejection methodology from ordinary least squares.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGOA .... African Growth and Opportunity Act

ARCH .... Autoregressive Conditional Heteroskedasticity

BU .... Boston University

CNN .... Convolutional Neural Networks

DGP .... Data Generating Process

DS .... Divisional Secretariat

FGT .... Foster-Greer-Torbecke Index

FT .... Fourier Transform

GDP .... Gross Domestic Product

GLCM .... Grey-Level Co-Occurrence Matrix

GLM.. Generalized Linear Models

GN.. Grama Niladhari

GSP .... General System of Preferences

GS2SLS .... Generalized Spatial Two-Stage Least Squares

HIES... Household Income and Expenditure Survey

HOG.. Histogram of Oriented Gradients

HRSF .... High Resolution Spatial Features

HSRI .... High Spatial Resolution Imagery

IMF .... International Monetary Fund

LASSO .... Least Absolute Shrinkage and Selection Operator

LBPM... Local Binary Pattern Moments

LDC ..... Least Developed Countries

LOOCV ..... Leave-One-Out Cross-Validation

LSR ..... Line Support Regions

MAE ..... Mean Absolute Error

MSE ..... Mean Squared Error

NDVI ..... Normalized Differenced Vegetation Index

NMAE ..... Normalized Mean Absolute Error

NRMSE ..... Normalized Root Mean Squared Error

NTL ..... Night Time Lights

OBIA ..... Object-Based Image Analysis

OLS ..... Ordinary Least Squares

ROA ..... Return on Assets

SAR ..... Spatial Autoregression

SURF ..... Speeded Up Robust Features

SVM ..... Support Vector Machine

TFP ..... Total Factor Productivity

USAID ..... United States Agency for International Development

USITC ..... United States International Trade Commission

WTO ..... World Trade Organization

CHAPTER ONE

Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being[1]

Ryan Engstrom[2]

Jonathan Hersh[3]

David Newhouse[4]

---

[1] This project benefited greatly from discussions with Sarah Antos, Ana Areias, Marianne Baxter, Sam Bazzi, Azer Bestavros, Kristen Butcher, John Byers, Francisco Ferreira, Ray Fisman, Alex Guzey, Klaus-Peter Hellwig, Kristen Himelein, Tariq Khokhar, Trevor Monroe, Dilip Mookherjee, Pierre Perron, Bruno Sánchez-Andrade Nuño, Kiwako Sakamoto, David Shor, Benjamin Stewart, Andrew Whitby, Nat Wilcox, Nobuo Yoshida and seminar participants at the Boston University Development Reading Group, Chapman University, World Bank DEC, and the Department of Census and Statistics of Sri Lanka. All remaining errors in this paper remain the sole responsibility of the authors. Sarah Antos, Benjamin Stewart, and Andrew Copenhaver provided assistance with texture feature classification. Object imagery classification was assisted by James Crawford, Jeff Stein, and Nitin Panjwani at Orbital Insight, and Nick Hubing, Jacqlyn Ducharme, and Chris Lowe at Land Info, who also oversaw imagery pre-processing. Hafiz Zainudeen helped validate roof classifications in Colombo. Colleen Ditmars and her team at DigitalGlobe facilitated imagery acquisition, Dung Doan and Dilhanie Deepawansa developed and shared the census-based poverty estimates, and we thank Dr. Amare Satharasinghe for authorizing the use of the Sri Lankan census data. Liang Xu provided research assistance. Zubair Bhatti, Benu Bidani, Christina Malmberg-Calvo, Adarsh Desai, Nelly Obias, Dhusynanth Raju, Martin Rama, and Ana Revenga provided additional support and encouragement. The authors gratefully acknowledge financial support from the Strategic Research Program and World Bank Innovation Labs Big Data for Innovation Challenge Grant, and the Hariri Institute at Boston University. The views expressed here do not necessarily reflect the views of the World Bank Group or its executive board, and should not be interpreted as such.

[2] rengstro@gwu.edu *Department of Geography, George Washington University, 1922 F Street, Washington DC*

[3] jhersh@bu.edu, *Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, and Poverty and Equity Global Practice, World Bank.*

[4] dnewhouse@worldbank.org, *Poverty and Equity Global Practice, World Bank, 1818 H Street, Washington DC*

Abstract

Measuring poverty is important for targeting aid and forming policy in developing economies. This is impeded by the high administrative and labor costs of large-scale surveys. This paper investigates the ability of high spatial resolution satellite images to accurately estimate poverty and economic well-being. We extract both object and texture features from satellite images of Sri Lanka. These data are then used to estimate models of local area poverty and economic well-being. The important features derived from satellite imagery include the number and density of buildings, shadow area (a proxy for building height), number of cars, density and length of roads, type of farmland, and roof material. These variables are used to estimate poverty rates and average log consumption for 1,291 villages (Grama Niladhari Divisions). Predictions from a binomial logit model, using only these satellite features as explanatory variables, explain sixty percent of poverty and average log consumption at the village level. We control for overfitting by using Lasso regularization. Two out of sample applications, extrapolating predictions into adjacent areas and estimating local area poverty using an artificially reduced Census, confirm the out of sample predictive capabilities. We conclude that satellite imagery has the potential to revolutionize poverty measurement, and that surveys should adjust to take advantage of the rise in useful ancillary data.

Keywords: poverty estimation, satellite imagery, machine learning

JEL classification: I32, C50

1 Introduction

Despite the best efforts of national statistics offices and the international development community, small area estimates of poverty and economic welfare remain rare. This is due in part to the lack of available household survey data measuring economic welfare in developing countries. Between 2002 and 2011, as many as 57 countries conducted zero or only one survey capable of producing poverty statistics, and data are all the more scarce in the poorest countries. (Serajuddin et al, 2015). Even where household surveys are conducted regularly, they are typically too small to produce reliable estimates below the district level. Generating welfare estimates for smaller areas requires both a household welfare survey and contemporaneous census data, and the latter is typically available once per decade at best. Furthermore, safety concerns prohibit survey data collection in many conflict areas altogether. Lack of timely information on living standards in small areas impedes the efforts of policymakers and aid organizations to direct scarce resources to the poor, and prevents their constituents to hold them accountable for doing so.

Satellite imagery has generated considerable enthusiasm as a potential supplement to household data that can help fill these severe data gaps. In recent years, private companies such as DigitalGlobe and Airbus have rapidly expanded the coverage and availability of high spatial resolution imagery (HSRI), driving down commercial prices. Planetlabs currently operates more satellites than any organization other than the US and Russian governments, and by mid-2016 will have launched enough satellites to acquire coverage of the entire globe with imagery resolution of 4 to 5 m per pixel on a daily basis.

Continued technological advances will increasingly allow social scientists to benefit from this type of imagery, which has been utilized intensively by the intelligence and military communities for decades.

This paper investigates the ability of object and texture features derived from HSRI (High Spatial Resolution Imagery) to estimate and predict poverty rates at local levels. The area of our study covers 3,500 square kilometers in Sri Lanka, which contain 1,291 villages (Grama Niladhari (GN) divisions). For each village, we extract both object and so-called "texture" feature to use as explanatory variables in poverty prediction models. Object features extracted include the number of cars, number and size of buildings, type of farmland (plantation vs. paddy), the type of roofs, the share of shadow pixels (building height proxy), road extent and road material, along with textural measures. These objects are identified using a combination of deep learning based Convolutional Neural Networks (CNN) and manual "heads up" digitization. These satellite derived features are then matched to household estimates of per capita consumptions imputed into the 2011 Census for the 1,291 GN Divisions.

We investigate four main questions: 1) To what extent can variation in village economic well-being -- poverty rates defined at the 10 and 40th percentiles of national income and average GN consumption -- be explained by high-resolution features? 2) Which features are most strongly correlated with welfare? 3) Can these models predict into geographically adjacent areas?; and 4) How should surveys adjust to meet the rise of useful ancillary data?

We find that: i) satellite features are highly predictive of economic well-being and explain about sixty percent of the variation in village average consumption or poverty; ii) Measures of built-up area and roof type are particularly strong correlates of welfare both in urban and rural areas. Car counts and building height are important correlates in urban areas, while the share of paved roads and agricultural type are strong correlates in rural areas; iii) Out-of-sample predictions are less accurate but tend to preserve rank; and iv) Satellite indicators can substitute for reduced number of households sampled per village. These results suggest that features extracted from high-resolution imagery hold considerable promise for contributing to poverty estimation and survey design.

This paper contributes to a literature exploring how remotely sensed data may be used to assess welfare. The most popular remotely sensed measure for economic applications has been night-time lights (NTL), which measures the intensity of light captured passively by satellite. Strong correlations between NTL and GDP appear at the country level (Henderson et al., 2009) although within a country NTL appears more strongly correlated with density than welfare. The relationship between lights and wages or other measures of income appears weak (Mellander et al., 2013), casting doubt on its reliability as a proxy for small area estimates of welfare. Additionally, NTL is ill-suited for identifying variation in welfare within small areas because of its low spatial resolution. Even the most advanced NTL satellite, VIIRS, has a spatial resolution at nadir of approximately 1.0 squared km.[5]

New developments in computer vision algorithms, however, allow for meaningful

[5] Pixel size can vary depending on the angle of the satellite relative to the ground site.

information to be extracted from daytime imagery. Advances from Deep Learning such as Convolutional Neural Networks (CNN) have the capability to algorithmically classifying objects such as cars, building area, roads, crops and roof type (Krizhevsky, Sutskever, and Hinton, 2012). These objects may be more strongly correlated with local income and wealth than NTL. An alternative approach to analyzing HSRI involves calculating textural and spectral variation in the imagery instead of identifying objects (Graesser et al. 2012, Engstrom et al. 2015, Sandborn and Engstrom 2016). In this approach the spatial and spectral variations in imagery are calculated over a neighborhood (a group of pixels) to characterize the local scale spatial pattern of the objects observed in the imagery. These measures, which we refer to as “texture” or “spectral” measures, capture information about an area that may not be clear from object recognition alone.

This paper also contributes to a literature exploring how supervised learning techniques from machine learning may be applied to unstructured data to reveal information about human welfare (Athey, 2017). Glaeser, Kominers, Luca, and Naik (2015) apply texture-based machine vision classification to images that are captured from Google Street View, trained using subjective ratings of the images on the basis of the perceived safety. They estimate a support vector machine model and show the fitted model can reliably predict block level income in New York City. Jean et al. (2016) employ a transfer learning approach to generate estimates of economic well-being from satellite imagery. They first extract a set of features from the Google imagery, using a convolutional neural network trained to predict NTL. In the second step, these features are then used to predict average consumption at the cluster level, taken from living standard measurement surveys.[7] This two-step approach allows the model to extrapolate relevant features of the imagery using the full range of NTL. This method improves upon the use of NTL alone to predict poverty. The model explains about 42 percent of the variation in village per capita consumption in a pooled sample of all four countries. However, the improvement over NTL alone is modest, as the two-step method raises the R^2 in the pooled estimate from about 0.38 with night-time lights alone to 0.42. This is consistent with the model's substantial predictive power being mostly driven by differences in population density derived from NTL.

In contrast to the aforementioned papers, this paper explores a two stage procedure for generating estimates of economic well-being: first features are extracted from satellite imagery, and then these features are used for supervised learning in estimating economic well-being. We believe this process has several advantages over more opaque prediction methods, including enabling a greater understanding of the performance of particular features in different contexts. For example, our approach makes it straightforward to investigate how well particular features predict poverty rates in urban and rural areas. In addition, this method demonstrates the potential to extract indicators from imagery that are useful for other purposes, such as better understanding road networks, agricultural productivity, and patterns of urbanization. We further distinguish from Jean et al. in that we estimate poverty in very small areas, with an average village size of just over 2 square kilometers, and use poverty and income estimates generated directly from the Census.

[7] It is not clear whether the consumption aggregates have been spatially deflated, and failing to deflate the aggregates would exaggerates the predictive power of population density in explaining variation in welfare.

The paper proceeds as follows: Section 2 summarizes how the data were created and presents brief summary statistics. Section 3 presents the statistical methodology. Section 4 examines the predictive power of high resolution satellite features (HRSF) to estimate poverty in small areas at the village level. Section 5 examines out of sample performance using two applications from estimating local area economic well-being. Section 7 concludes.

2 Data Description

We use a sample site of approximately 3,500 sq. km. in Sri Lanka, shown in figure 1- 1 highlighted in white. We collected high resolution imagery covering 47 Divisional Secretariat (DS) Divisions, the administrative area one level higher than GNs. A fully random sample of DSes was not feasible due to lack of complete coverage across DSes of high resolution imagery. We sampled DSes conditional on HSRI being available, drawing equal areas from urban, rural, and estate sectors.[8] The total sample contains 1,291 GN divisions, with each GN division covering an average area of 2.12 square kilometers.

[8] Sri Lanka is unique in that it classifies sectors as urban, rural, or estate. The Estate sector is a classification by the Department of Census that refers to “plantation areas, which are more than 20 acres in extent and have 10 or more residential laborers.” For the purposes of this study, except for sample stratification, the estate sector is grouped together with the rural sector.

Figure 1- 1: Coverage Area of High Resolution Satellite Imagery

2.1 Derivation of GN Welfare Statistics

Ideally village poverty and income statistics would be generated directly from the 2012/13 Household Income and Expenditure Survey (HIES), a detailed consumption survey that measures the consumption patterns of 16,000 households. However, given the sample size the HIES was insufficient alone to generate consistent poverty estimates. We instead use the Elbers, Lanjouw, and Lanjouw (2003) imputation methodology to impute income estimates into the 2011 Census of Population and Housing, a method that is identical to one used to generate official poverty estimates at the DS Division level (Department of Census and Statistics and World Bank, 2015). For each household in the census, per capita consumption was estimated based on models developed from the HIES, using household indicators that are common to both the Census and the HIES.

We use as a poverty line relative income thresholds from the national income distribution.[9] We use two poverty thresholds – one at 10% of the national income distribution, and another at the 40% level of national income distribution – intended to determine how satellite measures capture both poverty and inequality, respectively. We derive village headcount poverty rates using the standard Foster-Greer-Thorbecke (Foster et al., 1984) methodology

$$povrate_{j,z_k} = \frac{1}{N_j}\sum_{i=2}^{N_j} I(y_i < z_k)$$

where $PovRate_{j,z_k}$ is the poverty headcount rate in GN Division j at the z_k poverty threshold, defined at either the 10 or 40 percent of national income. The variable y_i is per capita household consumption for individual i and N_j is the total population of GN j. We calculated the national poverty threshold level, z_k, as the $k \in \{10,40\}$ percentile of the national predicted per capita consumption in the census. I is an indicator function that takes on one if individual i's per capita household consumption lies below the relative poverty threshold z_k. Since we only consider the headcount rate, this is equivalent to the ratio of the number of individuals below the threshold to the total population in each village.

Figure 1- 2 plots the histograms of poverty and consumption. The lower relative poverty rate (10%) shows considerable skewness. Many villages have zero estimated

---

[9] According to the official poverty line, which was developed in 2002 and subsequently updated for inflation, 6.7 percent of the population was poor in 2012/13.

poverty at this lower threshold, shown as a mass point near zero. At the higher relative poverty threshold, the number of villages with zero estimated poverty is lower and the distribution of poverty rates is closer to normal. Average village consumption (income) in log points is present in the bottom center panel of figure 1- 2, showing a distribution that is possibly bimodal and left skewed.

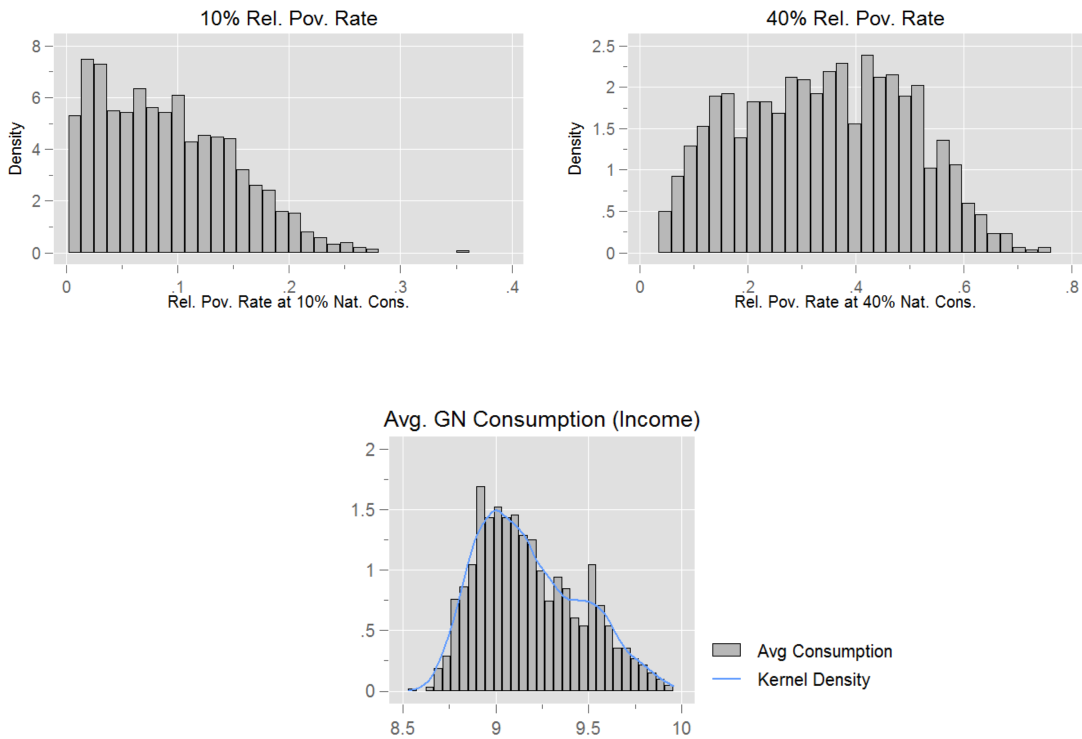


Figure 1- 2: Distribution of Poverty and Average Log GN Consumption

Notes: Histogram of village relative poverty. The two relative poverty rates refer to the fraction of individuals in each village below 10% or 40% of national income. Data is sourced from the poverty estimates imputed into the 2011 Sri Lankan Census.

2.2 Comparison of GN Poverty Rates and Mean NTL Reflectance

A simple visual comparison between mean NTL and village poverty rates illustrates why using NTL as a primary source of information on sub-national welfare is unwise.

Figure 1- 3 presents a panel of three images for the Divisional Secretariat of Seethawaka: mean raw NTL (left), poverty rates derived from the 10% national income threshold (middle), and log of mean population density (right). Comparing the left and middle panels, there is only a small association between villages that have low NTL reflectance and those that are high in poverty. Problems of overglow (Henderson et al., 2012) mean that poor villages adjacent to wealthy ones will be misclassified as non-poor. While NTL tracks the general contours of poverty for the DS – lower poverty areas in the Northwest and higher poverty areas in the Southeast – this coarse association is only of limited use for public policy applications such as poverty targeting or budget allocations.

What NTL does appear to approximate is population density of the underlying GN Divisions, which is consistent with the findings of Mellander et al. (2013). Comparing the right and left panel there appears to be a strong association between high NTL areas and areas with a high population density. We take this to suggest that the information content contained within NTL related to human welfare is limited. While lights at night may indicate gross associations at the lowest levels of poverty, only so much information about human welfare can be learned from the intensity of lights outputted at night. In contrast, HRSI daytime imagery may provide a much richer picture of on the ground welfare conditions.

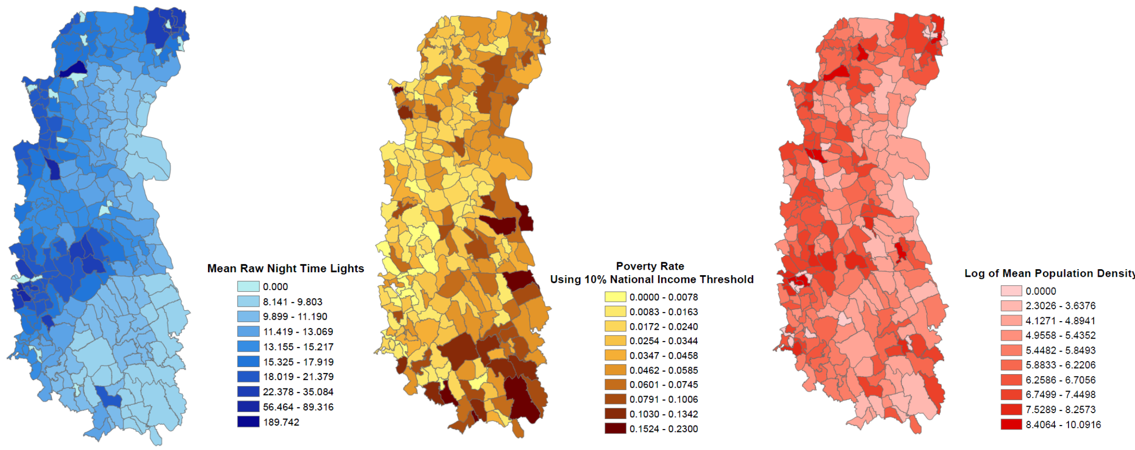


Figure 1- 3: Comparison of Mean Night Time Lights (NTL), Poverty Rate, and Mean Population Density, Seethawaka, Sri Lanka

2.3 Description of High Resolution Spatial Features (HRSF)

The derived high resolution spatial features fall into six broad categories: (1) Agricultural Indicators, (2) Cars, (3) Building Density, (4) Road and Transportation; (5) Roof Type; and (6) Textural and Spectral characteristics. Table 1- 1 present summary statistics for these variables. Under agricultural indicators we collected the fraction of GN agriculture that is paddy (rice cultivation) or plantation (cash crops such as tea), with the excluded category being agriculture not identified as either paddy or plantation. Agriculture in our sample is evenly split between paddy and plantation. We also calculated the fraction of total GN area that is either paddy, plantation, or any agriculture.

Three car related variables were collected – log total number of cars in a GN, total cars divided by road length, and cars divided by area. The average village in our sample has 23 cars. However, there are many outliers with a large number of cars, with the largest GN containing 4,000 cars, and many having none.

Building density variables include the fraction of an area covered by built-up area, and two indicators that capture "shadows" of tall buildings which functions as a proxy for building height. Built up area captures any human settlements – buildings, homes, etc – regardless of use or condition. The shadow variables use the angle of the sun as it shines on a building, and the shadows it displaces, to estimate the height of a building[10]. We calculate the log of total shadows as well as the fraction of shadows covering a valid area. This will act as a crude proxy for building height. Villages with more shadows will have taller buildings on average, indicating higher density of development.

The road variables we collect are the log of total road length, fraction of roads that are paved, and length of airport and length of railroad identified. For roof type, we collect the fraction of roofs in a village that are either clay, aluminum, asbestos, with the omitted category being roofs that are identified as none of the above, the vast majority being gray cement roofs. Roof type can be identified through remote sensing by using hyperspectral imaging, or using reflectance from several contiguous spectral bands. Different roof materials exhibit different spectral properties, particularly in the sub-visible bands of the spectrum. The roofs in our sample are clay (36.5%) aluminum (14.08%), asbestos (7.8%) or gray concrete (41.6%).

We calculate seven separate types of spectral and textural features: Fourier transform, Gabor filter, Histogram of Oriented Gradients (HoG), Line support regions (LSR), Pantex, and Speed-Up Robustness Features (SURF). These features can be considered outputs from some dimension reduction technique, in that they are reduced

---

[10] Valid area refers to areas at the foot of building where shadows may appear.

dimensionality descriptions of a complex 2-D satellite imagery. These are often used in machine vision problems to decompose an image. Like dimension reduction outputs, we may have to squint at the output to determine what these spectral or textural features are capturing. They are intended to capture aspects of a neighborhood that are not so easily identified directly, such as “sluminess” (presence of many irregular building lines) or high density.

Because these measures may be novel to readers without backgrounds in remote sensing, further description may be helpful. We consider Pantex here to be a measure of human settlements. It’s a spatial similarity index, where each cell is compared to adjacent cells in all directions. Forests will have a low Pantex level, since cells in all directions have similar contrast, as will cells with straight roads. Cities dense with many buildings in all directions will have high Pantex values. HOG captures “local intensity gradients or edge directions” (Dalal and Triggs, 2005) and in context here captures intensity of lines of development or agriculture. Local binary patterns (LBPM) captures local spatial patterns and gray scale contrast. SURF detects local features used for characterizing grid patterns, and measures orderliness of building development, the opposite of which is typically referred to as a slum. Areas with right angles, corners, or areas with regular grid patterns, will have larger SURF values relative to areas with chaotic or irregular spacing.

Although technically a spectral characteristic, we consider normalized differenced vegetation index (NDVI) a subset of building development, since its presence at lower resolutions can indicate development such as parks or lawns. We collect NDVI at two scales: 64 meters and 8 meters. These two measures will tell us whether the village contains many large plots of vegetation (64 meter scale) or many small plots of vegetation (8 meters resolution). The latter may be more important for urban areas in distinguishing poor areas from rich, since rich urban areas tend to have more small plots of vegetation used as lawns and parks than poor areas within cities.

For more detail on imagery and the feature extraction process we refer the reader to appendix A, In brief, object-based features were classified through a combination of convolutional neural network (CNN) training and object based image assessment (OBIA). Accuracy varied by feature, but overall class accuracy was above 90% for all features[11]. To illustrate the classification process, we will discuss two training examples for built-up area (building footprint) and cars, shown in figures 1- 4 and 1- 5. Figure 1- 4 shows raw satellite imagery in the right panel, and classified imagery according to the CNN algorithm in the left panel. Areas highlighted in green in are true positive building classifications, where areas classified by the algorithm as buildings were confirmed as such by manual identification. Areas in red are false positives that were erroneously classified by the algorithm as buildings even though no buildings are present. Figure 1- 5 shows raw satellite imagery from Colombo with cars identified by the CNN algorithm are highlighted by blue circles. The classifications occur on the road paths or in parking lots where cars are expected to appear. There are some false negatives – cars in the image not classified as such – particularly in areas where cars are obscured by trees or vegetation. However, the classifier appears accurate enough to distinguish high car areas from low car areas.

[11] See for example the ROC curve for the buildings in figure A1.

Figure 1- 4: Example Developed Area (Buildings) Classification

Notes: above image shows raw (left) and classified (right) for developed area building classifier from raw satellite imagery. Areas in green show are true positive building classifications. Images in red are false positives: erroneously classified areas as buildings.

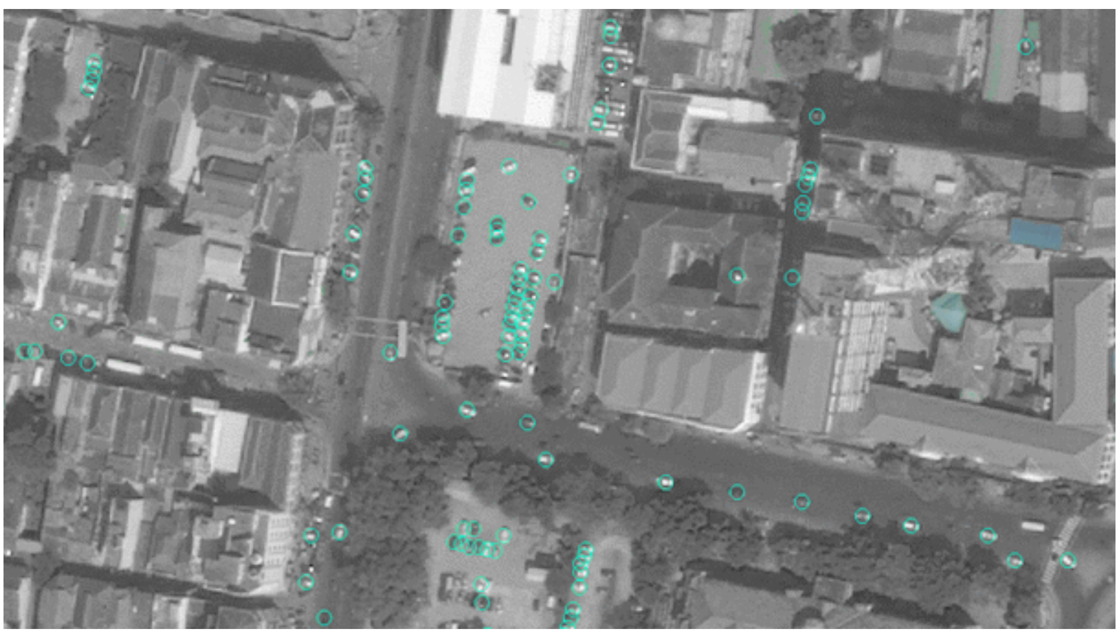


Figure 1- 5: Example Car Classification

Notes: above image shows satellite imagery overlaid with cars identified algorithmically shown in blue.

3 Statistical Methodology

Our methodological aim is to find which high-resolution spatial features are consistent predictors of local poverty rates in absence of any other on-the-ground information on the GN itself. There are many ways in which we could model the poverty rate of each GN. Linear least-squares regression is the simplest method but is unlikely to produce accurate predictions of the true poverty rate given that the support the dependent variable we are trying to model lies on the interval $[0,1]$ and OLS will generate predictions outside of that range. A second and related problem is that poverty data is typically right skewed, with many observations having an estimated poverty rate of zero. Estimation will likely result in heteroskedasticity.[12]

3.1 Baseline GLM Model

We employ as a baseline a binomial logit model, that is a binomial model with a logit link function. This model has several favorable properties, including ease of interpretability of coefficients and estimation. There are several competing models we could have estimated. If the number of zeros is very large, discrete mixture models such as zero-inflated models or hurdle models may perform better. For robustness, we employ a standard OLS model.

Given the list of covariates in table 1, variable choice is not straightforward.

[12] It is possible we could improve the fit by using log transformations of the dependent variable, or more precisely log transforms including a constant term such as $\log(y_j + 1)$, or even an inverse hyperbolic sine transform (IHS). Because we are interested in recovering $E[y|x]$ and not $E[\ln(y)|x]$ or $E[\text{IHS}(y)|x]$ this may introduce retransformation problems when retransforming the data to its original form (Mullahy, 1998).

Estimating a model with the full set of variables in table 1- 1 will likely produce predictions that are overfit, that is that produce good in-sample estimates but have poor out of sample performance (Athey and Imbens, 2015). One attractive method for variable selection among a large selection of covariates is Lasso regularization. The Lasso regression is a regularized regression that estimates a regression model with an added constraint that enforces parsimony (Tibshirani, 1996). The motivation for the shrinkage estimator is that by reducing the parameters of the model, one increases bias (in-sample error) at the expense of lower variance (out-of-sample error).

We can apply regularization to the binomial logit by modifying the log-likelihood such that the estimated likelihood is of the form

$$\beta_{BinLasso} = \underset{\beta}{\operatorname{argmin}} \sum_{j=1}^{N} \{ \underbrace{y_j \log(\pi_j) + (n_j - y_j) \log(1 - \pi_j)}_{Binomial\ log-likelihood} + \underbrace{\lambda \sum_{j=1}^{K} |\beta_j|}_{Shrinkage\ factor} \} \tag{1}$$

Where the poverty rate in a village is given by the logit transform $\pi_j = \exp(x_j^T \beta)$ and $\lambda \geq 0$ is a shrinkage parameter that penalizes the absolute values of the coefficients. As we relax the penalization factor -- that is as $\lambda \to 0$ – the Lasso coefficients converge to the unconstrained estimates: $\beta_{BinLasso} \to \beta_{Bin}$. As $\lambda \to \infty$, we increase the penalty, and $\beta_{BinLasso}$ converges to the zero vector. Lasso regressions are useful as a variable selection methodology because of the sharp L_1 metric, which creates a de-facto variable selection (Varian, 2014) in addition to "shrinking" coefficients in magnitude towards zero. To choose the appropriate value of λ we apply 10-fold cross validation, and choose the value of λ that minimizes root-mean squared error (RMSE) across folds.

So far we haven't discussed inferential standard errors, which are typically absent from Lasso models. Because of the Oracle property of the Lasso estimator (Fan and Li, 2001), we can employ a two step methodology: first estimate a Lasso model over the full candidate set of covariates, x, resulting in $\hat{\beta}_{BinLasso}$ and $x_{selected} \subset x$, the "selected" set of variables from the Lasso estimation where selected indicates Lasso returns a non-zero coefficient. We can then estimate an unconstrained GLM model using only the reduced set of selected coefficients. The Oracle property ensures that inference in the second stage using the reduced set of variables selected in the first stage is consistent with inference were we to use a single stage estimation strategy using only the selected variables present in the true data-generating process.

4 Poverty Validation Using High Resolution Features

Table 1- 2 presents the baseline estimates. The first two columns use the 10% relative poverty rate (poverty incidence) as the dependent variable, the next two columns use the 40% relative poverty rate (inequality) as the dependent variable. Lasso regularization selects 25 out of 31 candidate variables for the 10% poverty rate models, and 19 out of 31 candidate variables for the 40% poverty rate models. T-statistics are presented and reflect clustering at the DS level. We present coefficients as elasticities, where each covariate is evaluated at its mean value. These can be interpreted as a 1% change leads to a coefficient change in estimated poverty rate.

Four of the five agricultural variables are selected for the poverty incidence models, and three out of five are selected for the inequality models. The largest coefficient in these groups is fraction of village devoted to paddy (rice) cultivation, with an estimate of -0.198, indicating that every 1% increase in paddy cultivation leads to a -0.198 reduction in poverty. Only one out of three cars variables – the log number of cars -- is selected in either model. The estimated elasticity is between -0.162 and -0.138. All four of the road variables are selected, with length of roads having large estimated elasticities. All three of the built-up area variables are selected and statistically significant. Fraction of area with buildings is negatively related to poverty, and the building height proxy (shadows) is strongly positively associated with poverty. Two out of three roof type variables are selected, fraction clay, and fraction aluminum, both having positive elasticities, suggesting that poorer households tend to use clay and aluminum as roofing material. At first glance, this is surprising, as previous analysis in Kenya documents that roofs with higher luminosity are associated with lower levels of poverty (Suri et al., 2015). We find the opposite, suggesting that the relationship between roof luminosity and poverty is more complex, and is likely context and region dependent. Of the texture variables, three out of nine are selected for the inequality models (LBPM, LSR, and Gabor), and seven out of nine are selected for the poverty models (NDVI, LBPM, LSR, Gabor, Fourier, SURF). LBPM has the largest elasticity estimate of 1.9. Such a large coefficient suggest that these types of clusters of pixels picked up by LBPM suggests this type of clustering is important for human welfare.

To complement the GLM models, table 1- 3 presents Post-Lasso OLS estimates for poverty (Belloni and Chernuzhukov, 2013), and includes an additional model where average village income (consumption) is the dependent variable. Lasso regularization was again performed against the full set of candidate variables. The estimated models explain an impressive amount of the variation in economic well-being. R-squared values vary from 0.608 for the average village consumption models, 0.61 for the 10% poverty line, and 0.618 for the 40% poverty line. This is to say that a linear model that includes only remotely sensed information explains 61-62 percent of the variation of a village's poverty rate and 60 percent in the variation in that village's income or consumption. Many of the same variables that were selected in the GLM models were selected with Regularized OLS. For the average village consumption models, fraction of roads paved, shadow height, roof type, NDVI, LBPM and agricultural variables are particularly important.

To get a further sense of how these indicators perform in comparison to other remotely sensed indicators, table 1- 4 presents OLS models covering the same sample area using night time lights as the independent variable. The first three columns present poverty and income models. Aggregate night time lights is positive (negative) correlated with income (poverty), however the total explanatory power is low: R^2 values for the three regressions are between 0.1 and 0.147.[13] Models built using high resolution satellite indicators capture more than three times as much variation in poverty or income than NTL. Columns 4-6 of table 1- 4 estimate a model of poverty or income against NTL and including DS fixed effects. Night time lights is no longer significant in any of the specifications,

[13] Even using additional transformations of NTL – squared, cubed, or standard deviation – only increases R^2 values to 0.15.

indicating that within a DS NTL has little correlation with economic well-being.

As a check of the accuracy of the linear models, we plot the predicted economic well-being against true economic well-being in figure 1- 6.[14] Each point represents a village, where the location on the x-axis corresponds to the true poverty rate and location on the y-axis is the model predicted poverty rates. A model that is perfectly able to predict poverty using satellite variables would represent a 45 degree line, starting at the origin and ending at the upper right of the graph. Note that for all of the measure of economic well-being, the predicted true points are roughly straddling the 45 degree line. Our methods are well able to distinguish high poverty areas from low poverty ones, albeit with some noise.

As a further check of the models, figure 1- 7 presents a choropleth map showing the true welfare measures on the left panel, against the predicted welfare measures on the right, for a particular DS, Seethawaka. The top panel shows predicted income from the OLS model against actual income. The model is able to distinguish the poorer eastern areas from the richer western ones. Even poor GNs adjacent to richer ones can be distinguished. Note the scale of the figure, the smallest GNs are less than a half mile across, and yet the HRSF model is able to distinguish with considerable accuracy the variation in average consumption. The middle panel shows predicted and true poverty rates defined at the 10% relative poverty level. Again, the predicted model approximates the true poverty rates with considerable accuracy. The lower poverty regions in the south and north east are replicated in the predicted values. The model tends to under-predict poverty in the lowest poverty areas in the mid-west. This indicates that we may have a better fit using mixture models

[14] These plots are often called “predicted-true” plots.

meant to model observations with many zeros such as zero-inflated poisson or hurdle models.

4.1 Urban and Rural Linear Models

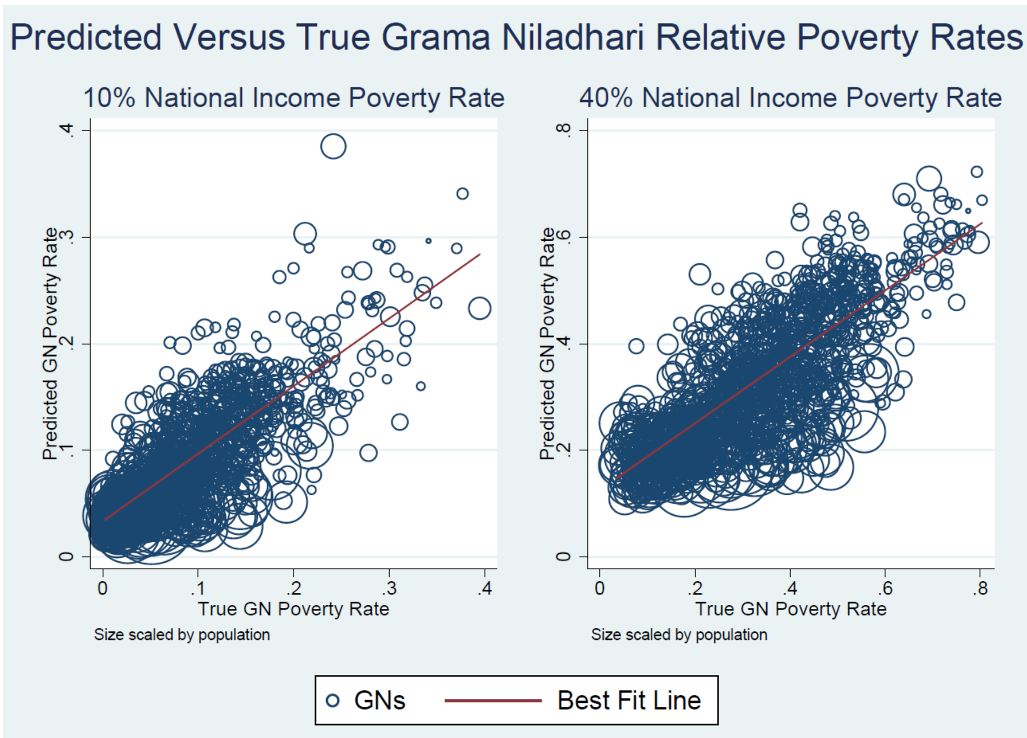


Figure 1- 6: Predicted Versus True Welfare Measures, 10% Relative Poverty Rate (left), 40% Relative Poverty Rate (right)

To consider how the models perform differently in urban versus rural areas table 1-4 shows model estimates estimated separately for 393 urban villages and the 898 rural ones. To parse out the villages into distinct urban and rural groups, we use the official Census definition of urban and rural.[15] Variables were again selected through Lasso estimation.

[15] In principle, a small area's urban or rural characteristic could be estimated via remote sensing using a probability model based on variables such as NDVI, Built-up density, shadow pixels (building height index), agricultural product

The urban model selects fewer variables – 13 of the candidate variables in the urban model are selected versus 16 for the rural model. This suggests that data generating process (DGP) for urban poverty is more sparse than the DGP for the rural poverty. R-squared values are slightly higher in rural areas (0.656) and significantly lower in urban areas (0.445).[16] For the urban model, log number of cars, built-up development, and shadow pixels are important. In rural models, agricultural variables, roof type, shadow pixels, NDVI, Pantex and LBPM are important. Cars is notable in that the presence of cars matters for economic well-being in urban areas and not rural ones.

4.2 What share of the explained variance is accounted for by different features?

The results indicate that features derived from satellite imagery explain a large portion of village income or poverty. However, these results don't address the question of how much variation is explained by which features. To address this issue, we decompose the R^2 using a Shapley decomposition (Shorrocks, 2013; Huettner and Sunder, 2012; Israeli, 2007). This procedure calculates the marginal R^2 of a set of explanatory variables, as the amount by which R^2 declines when removing that set from the set of variables. In other words, for a model with k sets of explanatory variables, the procedure will estimate 2^{k-1} models and average the marginal R^2 obtained for each set of independent variables across all estimated models. Note this is insensitive to the order in which variables are entered into the model

[16] This might be due to the nature of the consumption module in the HIES, which is structured to capture rural income moreso than urban consumption.

Table 1- 6 presents the R^2 decomposition. Building density variables (built-up area, shadow pixels, and NDVI) and roof type collectively explain the majority of the variation in welfare: between 50 to 55 percent of the variation in welfare across the welfare measures is explained by these two variable subgroups. Roof type alone explains between 15% and 17% of the variation. Of the building density variables, built-up area explains the most of the variation in welfare, explaining between 14 and 20 percent. Both shadow and NDVI also add explanatory power, adding between 6.7 – 16.6 and 5.5 – 9.7 respectively. Agricultural land variables add about 5 percent to the explanatory power. Cars explain a little more than that, between 6 and 8 percent. Perhaps this is due to the noise with which we observe cars, in that cars are not always located where their owners live. Although few texture features are individually statistically significant, as a group they contribute 10 to 15 percent of the explained variation of the model. Variables on roads explain about 9 percent of the variation, which in results now shown is largely accounted for by the share of roads that are paved.

Given the prevalence, ease of use and familiarity with night time lights, one might ask how much more explanatory power do the high resolution predictions provide *in addition to* night time lights? Table 1- 7 answers that question, by adding night time lights to the above Shapley decomposition. Within the night time lights category we've included average, squared, cubed, and average standard deviation transformation of NTL to push the information content of night time lights to the limit. Night time lights explains only 8.4 to 12.8 percent of the variance in consumption or poverty according to the decomposition, meaning by there is roughly a 90 percent additional variation in poverty or income that is captured through high resolution satellite predictions. Clearly there is a lot that is missed when using NTL as a singular proxy of economic well-being.

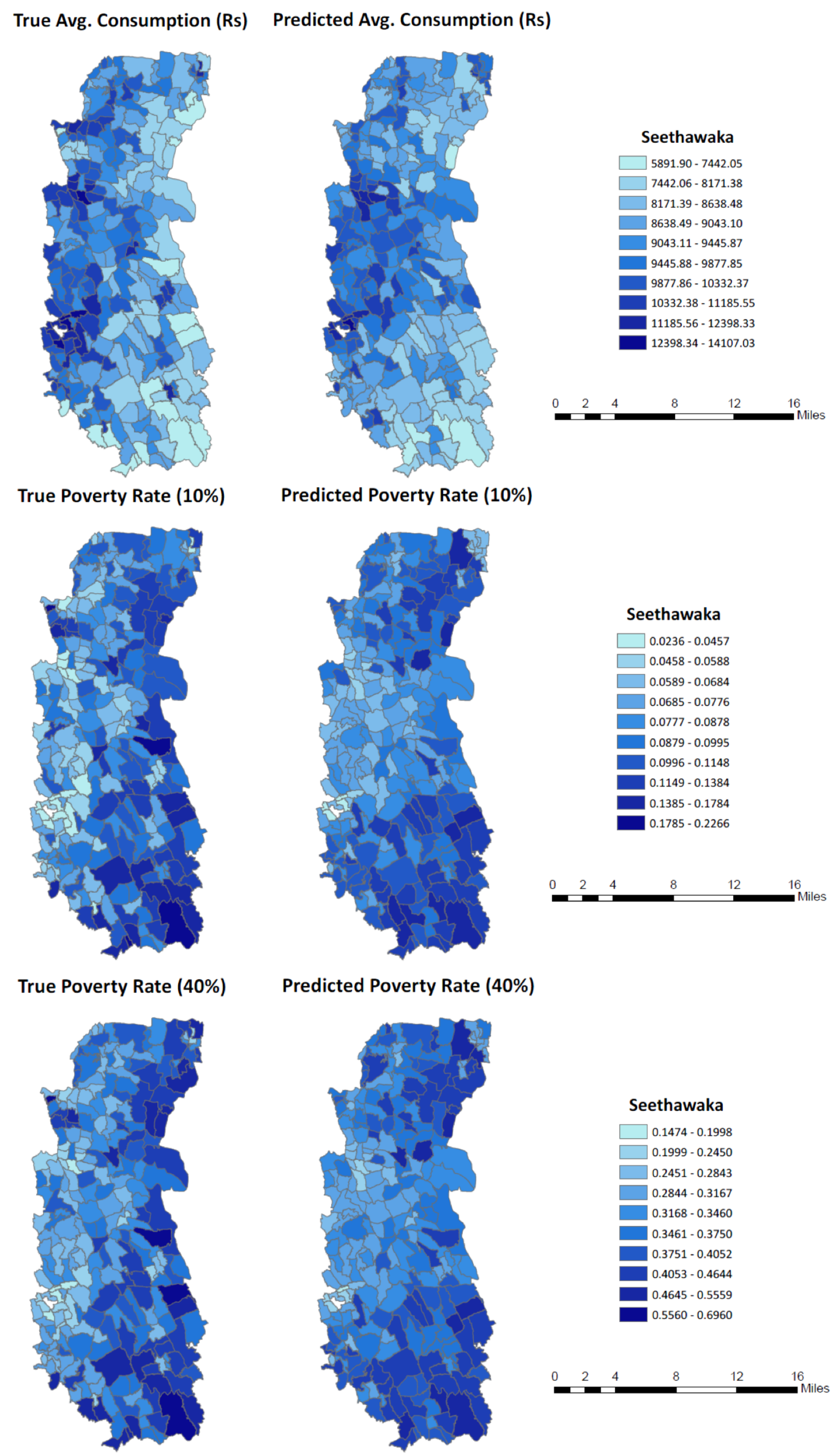


Figure 1- 7: Predicted Versus True Welfare Measures, Binomial Logit Models,

4.3 Correcting for Spatial Autoregression

One unaddressed concern is whether the presence of either spatial autocorrelation or spatial heterogeneity confounds the results. Spatial autocorrelation can occur in the presence of geographic spillovers or interactions (Anselin, 2013), and considering the village-level observations one could develop plausible stories by which poverty is influenced by this mechanism. A Moran's I test for the presence of such disturbances according to Anselin (1996) rejects the null hypothesis that there is no spatial autocorrelation present.

To correct for the spatial autocorrelation we model explicitly the spatial autoregression (SAR) process and allow for SAR disturbances, a so called SARAR model. This is implemented via a generalized spatial two-stage least-squares (GS2SLS) as shown in Drukker et al. (2013). The results presented in table 1- 8 show that after correcting for spatial autocorrelation most high-resolution spatial features remains significant predictors of local area poverty. It appears that although there is some presence of autocorrelation, it is not sufficient to alter the joint significance of the spatial variables.

4.4 Do High Resolution Satellite Features Explain the Poverty Gap?

Measuring the intensity of poverty is another useful metric for understanding the distribution of economic well-being in a country. The Foster-Greer-Thorbecke (Foster et al., 1984) indices are one of the most widely used methods of measuring the intensity of poverty. It measures poverty depth by considering how far the poor are from a given poverty line. We calculate for our sample the FGT_1 metric, which is defined as $FGT_1 =$

$\frac{1}{N}\sum_{i=1}\left(\frac{z-y_j}{z}\right)$, where y_j is an individual's income, and z is the poverty threshold. We compute the average FGT_1 for each village, and use this measure as a dependent variable in a regression where the right hand side includes the features created from high resolution satellite imagery. We consider two different measures of FGT: one measured using z defined at the 10^{th} percentile of national income, and one where it is measured at the 40^{th} percentile. Table 1- 9 presents the results estimated via OLS. The coefficients can be interpreted as a unit change in the distance between the poverty gap and the poverty line for the average village. We find that high resolution features explain the poverty gap well with adjusted R^2 values between 0.588 and 0.609. Many of the same variables that explain headcount poverty explain poverty gap, which is not surprising given their high degree of correlation.

5 Out of Sample Performance with Two Applications

5.1 Poverty Mapping Using Partial Census Sample Size Combined with HRSF

The gold standard for estimating local area poverty is to combine a full Census with a consumption survey. For many areas where we would like to have accurate poverty maps, a full Census is either not available or only partial Census data is available. In this section we test whether a consumption survey combined with partial Census information provides an accurate measure of economic well-being. We conclude, with some reservation, that it does.

For this exercise we produce several simulations of the dependent variable (either income or poverty rate) using a Census with reduced sample size. We compute subsamples of 25% and 50% of villages (GNs) on which to train our high resolution models[18]. We also vary the number of households the sample and subsample that are "surveyed". Income or poverty is measured in each GN using either 25%, 50% or 100% of the actual households in that GN. For example, a 50% GN sample where 25% of the households are surveyed trains a model on half of the villages, where the training data was computed using only a quarter of the households in the village. The estimated actual poverty rate of a village will become less precise the fewer households that are sampled per village. The fewer villages on which to train our models will make the estimates less precise. Increasing both the number of households sampled per village and the number of villages sampled is expensive, and this exercise is intended to estimate the tradeoffs in terms of poverty estimate accuracy of reducing both.

Table 1- 10 presents model performance in this simulation exercise. We present in-sample and out-of-sample R^2, which calculates the coefficient of variation for villages included or excluded from the training sample. First, out of sample R^2 is slightly *higher* than in-sample, indicating the models in the previous section may be underfit. For the 10% relative poverty rate models, out of sample R^2 varies between 0.637 and 0.621 while in-sample R^2 lies between 0.605 and 0.595. Estimated out of sample R^2 is slightly increasing with the number of villages sample. In terms of average error rate of individual predictions,

[18] Villages enter the training or test sets randomly, which differs from the exercise in the following section in which villages enter or exit the training set systematically based on contiguous geography.

we present normalized mean absolute error, which computes mean absolute error expressed as a percentage of average poverty rate or income. Average error does not seem to systematically decline when fewer households per village are sampled. Average normalized error is around 1/3 of the poverty rate for the 10% poverty level, ¼ of the poverty rate for the 40% inequality measure, and on a few percentage points of the average village income, regardless of number of households sampled per village. These results indicate that HRSF can act as a substitute for fewer households sampled per area, potentially saving millions in surveying costs.

5.2 Poverty Extrapolation to Adjacent Areas Using HRSF Models

A strong motivation for using satellite imagery is to extrapolate poverty estimates into areas where survey data on economic well-being does not exist. While most of the data deprivation that characterizes the developing world occurs at the country level, it is also common for surveys to omit selected regions, due to political turmoil, violence, animosity towards the central government, or prohibitive expense. For example, from 2002 through 2009/10, Sri Lanka's Household Income and Expenditure Survey failed to cover certain districts in the North and Eastern part of the country due to civil conflict, and Pakistan's 2012/13 Household Income and Expenditure Survey excluded the Federally Administered Tribal Areas and Jammu and Kashmir.

To assess how well a model "travels" to a different geographic area, we use a form of "leave-one-out cross-validation" (LOOCV), a common method used to infer statistical out of sample performance (Gentle et al., 2012). In standard LOOCV, a model is fit for every observation excluding one, then the estimated relationship is used to predict into the withheld observation. This is repeated for every observation in the dataset until every observation has an associated predicted value. This ensures that for all observations i, the fitted value of $\hat{y}_i$ used to build the model is not influenced by the relationship between x_i and y_i.

Our approach differs from the standard case in that for each estimation we exclude, or "leave out", an entire Divisional Secretariat (DS), an administrative sub-unit at the level immediately below the district. To give a sense of size, our sample contains 47 unique DS divisions. This type of LOOCV is a more stringent test of out of sample performance, but one that more accurately approximates the intended use-case of extrapolating poverty into areas where data are not present. While traditional LOOCV assesses out-of-sample performance relative to a large set of single observations, omitted at random, most cases of incomplete survey coverage omit one or more regions within a country.[19] If the data generating process for poverty is geographically heterogeneous, LOOCV at the DS level will give a more accurate assessment of our methodology in practice.

Our algorithm for adjacent prediction LOOCV is as follows:

1. Estimate a binomial Logit model $\hat{f}_{\{i\}}(x_{-i})$ on all Divisional Secretariats save for holdout DS i.

2. Use the estimated model in step 1 to predicted values for withheld DS $y_i = \hat{f}_{\{i\}}(x_i)$.

[19] A further complication is uncovered regions are not selected at random and are likely differ from the surveyed regions in unobserved ways. This will contribute to prediction error, since extrapolation requires assuming that the model estimated in the surveyed regions applies to the uncovered region as well.

3. Repeat until all DSes have predicted values.

To assess accuracy we use three different metrics to compare predicted poverty rates against the true poverty rates at the GN Division level. 1) Normalized root mean squared error (NRMSE) $= \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\hat{y}_j^{\,CV} - y_j\right)^2}$, where $\hat{y}_j^{\,CV}$ is the cross-validated predicted village economic well-being, y_j is the true economic well-being, and j indexes each GN. 2) Normalized mean absolute error as defined as in the previous section, and 3) Spearman rank correlation[20].

Table 1- 11 shows the simulation results. The adjacent prediction error rates are low for the average consumption models: NRMSE is estimated at 0.083 and NMAE at 0.241. For the poverty rates, adjacent prediction error is higher. At the 10% poverty threshold NRMSE is 0.559, and NMAE is 0.404. For the 40% poverty threshold NRMSE is 0.363 and MAE is 0.276. These error rates may not be sufficient for calculating official statistics, but they may be sufficient for generating rank ordering of villages by poverty or income. The correlation between the predicted and the true values confirms this. Spearman's ρ is estimated at between 0.68 and 0.7 for the three models. What to make of these results? Predicting into adjacent areas given this modeling strategy is insufficient for generating official statistics. It is possible that the choice of model influences out of sample behavior. Models with better non-linear properties, or models that can more easily handle

[20] Spearman's rank correlation is a non-parametric method that measures only the correlation in the monotonic rank ordering between two variables, in this case the predicted and true poverty rates. This metric indicates whether the GN divisions can successfully be ranked on the basis of their predicted poverty rates. A rank correlation coefficient of 1 would indicate that models using only HRSF can create a perfect rank ordering of village economic well-being using the more stringent DS leave one out cross validation.

heterogeneity – such as random forests, SVM, ensemble methods, or deep learning – will likely perform better predicting into adjacent areas.

6 Conclusion

How should surveys be designed in an era of Big Data? We believe these results show that ancillary data, particularly from high resolution satellite imagery, can act as a strong substitute for certain types of Census data. Traditionally, small area poverty estimates have been derived by combining a Census with a consumption survey. Given the prohibitive cost of conducting surveys sufficiently large to provide accurate statistics for small areas, measures of economic well-being are computed relatively infrequently, at best once every three years and often with a lag much larger than that. The welfare consequences of too infrequent measures of poverty and inequality are unknown but possibly large, given the many uses of accurate measures of economic well-being, from impact evaluation, to budget allocation to social transfers. If indicators derived from imagery explain a sufficiently large portion of variation in welfare, they could potentially serve as viable substitutes for Census data, which would greatly increase the frequency of small area poverty estimates.

How well do indicators derived from satellite imagery predict poverty and which indicators are most important? We investigate these questions using a sample of 1,291 villages in Sri Lanka, linking measures of economic well-being with features derived from high resolution satellite imagery. The results indicate that the correlation between satellite derived indicators and economic well-being is remarkably strong. GLM explain 61-61 percent in the variation in poverty, and 61 percent of the variation in average log income. In both rural and urban areas, variables measuring building density and built-up area are the strongest predictors of variation in poverty. This includes a built-up area measure, vegetation indexes, roof type, and number of shadow pixels, a proxy for building height. As expected, the extent and lushness of vegetation is negatively correlated with incomes in rural areas, and positively correlated with incomes in urban areas, suggesting that vegetation and gardens are a luxury in urban areas.

The analysis also included several mathematical transformations indicating the "texture" of the image, which have been utilized for optical recognition purposes and other imagery processing applications. As a whole, these are correlated with poverty and even moreso with inequality. They generally appear to capture contrasts and sharp edges, which characterize wealthier areas, although additional analysis is needed to better understand which features explain variation in welfare, at what scale they should be calculated and what specific characteristics of the built environment they are measuring. The major advantage of these "texture" features is they provide insight into landscape variability without having to define objects, as they are simply characterizing the variability in the imagery. Because of this, these variables are simple and straight forward to calculate and require no training data to create unlike the other remote sensing variables utilized in this research.

While these results are very encouraging, additional analysis suggests caution when extrapolating predictions into geographically adjacent areas. The normalized error rates from 1/4 to 1/2 of the poverty rates, depending on the incidence of poverty. A likely impediment to extrapolation is geographic heterogeneity. Another factor is time differences at which satellites images were taken, contributing to overall independent variable noise. This could impact selected indicators such as car counts, which vary according to the day of the week the imagery was obtained. Indeed, the correlation between car counts and poverty in urban areas is negative and statistically significant when controlling for image dummies, but not in the full sample. Measures of agriculture also exhibit considerable seasonal variation which could also confound extrapolation to adjacent areas. This suggests that some indicators may be more biased than others when extrapolating across space, and that the size date of the image is an important consideration when considering spatial extrapolation using satellite-based indicators.

These findings raise a host of questions for further work. The most immediate of these is whether satellite indicators can substitute for Census data in different contexts. Does the strong correlation between satellite-based indicators and economic well-being extend to income measured directly from an expenditure survey? Because of the size of the consumption survey at our disposal, we could not test this directly. The village poverty rates used were generated from models that explained 40 to 50 percent of the variation in measured consumption. The strength of the results reported above is very encouraging. However, it is not clear that satellite-based indicators can legitimately substitute for census data to generate small area estimates if, for example, they can only explain one third of the variation in measured per capita consumption. Second, it is important to better understand the extent to which these results generalize to different ecological environments, such as Africa, the Middle East, and other parts of Asia. There is no guarantee that the predictive power of building density, agriculture, and spatial features documented above, for example, will hold in all environments.

A second line of research could explore whether changes in satellite imagery could be used to forecast changes in economic well-being across space and time. The ability to "now-cast" measures of economic well-being using contemporary changes in satellite imagery applied to old measures of poverty could be the panacea for a "data deprived" world. More research is needed to understand whether this is possible and if so, which satellite indicators reliably forecast changes in poverty in which contexts. Secondly, more research is needed before predicting into adjacent areas not covered by surveys is a reality. More flexible modeling specifications will likely improve adjacent area predictions. In general, the inevitable increase in the availability of imagery and feature identification algorithms, along with the encouraging results from this study, suggest that satellite imagery will become an increasingly valuable tool to help governments and stakeholders better understand the spatial nature of poverty.

References

Afzal, M., Hersh, J., and Newhouse, D. (2016). “Building a better model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka”. Mimeo, World Bank.
Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324).

Anselin, Luc. *Spatial econometrics: methods and models.* Vol. 4. Springer Science & Business Media, 2013.

Anselin, Luc, et al. "Simple diagnostic tests for spatial dependence." *Regional science and urban economics* 26.1 (1996): 77-104.

Athey, S., & Imbens, G. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. arXiv preprint arXiv:1504.01132.

H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded Up Robust Features. Lecture Notes in Computer Science, 3951:404–417, 2006

Belloni, Alexandre and Chernozhukov, V. (2013). “Least squares after model selection in high-dimensional sparse models” *Bernoulli*. 19(2).

Besley, T., & Ghatak, M. (2006). “Public goods and economic development”. *Understanding Poverty*. (pp. 285-302). Oxford: Oxford University Press.

N. Dalal, and B. Triggs, “Histograms of oriented gradients for human detection,” in Computer Vision and Pattern Recognition (CVPR), San Diego, CA, 2005, pp. 886-893.

Department of Census and Statistics and World Bank, 2015 “The Spatial Distribution of Poverty in Sri Lanka”, available at: http://www.statistics.gov.lk/poverty/SpatialDistributionOfPoverty2012_13.pdf

Donaldson D., and Storeygard A. “Big Grids: Applications of Remote Sensing in Economics”, *forthcoming, JEP.*

Drukker, David M., Ingmar R. Prucha, and Rafal Raciborski. "Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances." University of Maryland, Department of Economics (2011).

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. Econometrica, 71(1), 355-364.

Elbers, Chris, Peter F. Lanjouw, and Phillippe G. Leite. "Brazil within Brazil: Testing the poverty map methodology in Minas Gerais." World Bank Policy Research Working Paper Series, Vol (2008).

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogrammetric Engineering and Remote Sensing, 63(6), 727-734.

Engstrom, R., Ashcroft, E., Jewell, H., & Rain, D. (2011, April). Using remotely sensed data to map variability in health and wealth indicators in Accra, Ghana. In Urban Remote Sensing Event (JURSE), 2011 Joint (pp. 145-148). IEEE.

Engstrom, R., Sandborn, A., Yu, Q. Burgdorfer, J., Stow, D., Weeks, J., and Graesser, J. (2015) Mapping Slums Using Spatial Features in Accra, Ghana. Joint Urban and Remote Sensing Event Proceedings (JURSE), Lausanne, Switzerland, 10.1109/JURSE.2015.7120494

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348-1360.

Foster, James; Joel Greer; Erik Thorbecke (1984). "A class of decomposable poverty measures". *Econometrica*. 3. 52: 761–766.

Gabor, D. (1946). Theory of Communication. Journal of the Optical Society of America-A, 2 (2), 1455-1471.

Gentle, J. E., Härdle, W. K., & Mori, Y. (Eds.). (2012). Handbook of computational statistics: concepts and methods. Springer Science & Business Media.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2015). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life (No. w21778). National Bureau of Economic Research.

J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no.4, pp. 1164-1176, Jul, 2012.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. The American Economic Review, 102(2), 994-1028.

Huettner, Frank, and Marco Sunder. "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values." Electronic Journal of Statistics 6 (2012): 1239-1250.

Israeli, Osnat. "A Shapley-based decomposition of the R-square of a linear regression." The Journal of Economic Inequality 5.2 (2007): 199-212.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790-794.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. The American Economic Review, 105(5), 491-495.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

Marx, B., Stoker, T. M., & Suri, T. (2013). The Political Economy of Ethnicity and Property Rights in Slums: Evidence from Kenya.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. The American economic review, 102(4), 1508.

Mooney, D. F., Larson, J. A., Roberts, R. K., & English, B. C. (2009). Economics of the variable rate technology investment decision for agricultural sprayers. In Southern agricultural economics association annual meeting, Atlanta, Georgia, January.

Mullahy, J. (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of health economics*, 17(3), 247-281.

Mullainathan, S. (2014, August). Bugbears or legitimate threats?: (social) scientists' criticisms of machine learning?. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 4-4). ACM.

Newhouse, David; Suarez Becerra, Pablo; Doan, Dung. 2016. Sri Lanka Poverty and Welfare: Recent Progress and Remaining Challenges. World Bank, Washington, DC. © World Bank. https://www.openknowledge.worldbank.org/handle/10986/23794 License: CC BY 3.0 IGO.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20-36.

M. Pesaresi, A. Gerhardinger, and F. Kayitakire, “A robust built-up area presence index by anisotropic rotation-invariant textural measure,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 1, no. 3, pp. 180-192, Oct, 2008.

Sandborn, A. and Engstrom, R (In Press) Determining the Relationship Between Census Data and Spatial Features Derived From High Resolution Imagery in Accra, Ghana. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS) Special Issue on Urban Remote Sensing

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). Data deprivation: another deprivation to end. World Bank Policy Research Working Paper, (7252).

Shorrocks, Anthony F. "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value." Journal of Economic Inequality (2013): 1-28.

S. W. Smith, The scientist and engineer’s guide to digital signal processing. San Diego, CA: California Technical Publishing, 1997.

Tucker CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8: 127-150.

Varian, H. R. (2014). Big data: New tricks for econometrics. The Journal of Economic Perspectives, 28(2), 3-27.

L. Wang, and D. He, “Texture classification using texture spectrum,” Pattern Recognition, vol. 23, no. 8, pp. 905-910, 1990.

Watmough, G. R., Atkinson, P. M., Saikia, A., & Hutton, C. W. (2016). Understanding the Evidence Base for Poverty–Environment Relationships using Remotely Sensed Satellite Data: An Example from Assam, India. World Development, 78, 188-203.

Wong, T. H., Mansor, S. B., Mispan, M. R., Ahmad, N., & Sulaiman, W. N. A. (2003, May). Feature extraction based on object oriented analysis. In *Proceedings of ATC 2003 Conference* (Vol. 2021).

W. P. Yu, G. W. Chu, and M. J. Chung, “A robust line extraction method by unsupervised line clustering,” Pattern Recognition, vol. 32, no. 4, pp. 529-546, Apr, 1999.

Tables

Table 1- 1: Village (Grama Niladhari) Summary Statistics

	Mean	**Sd**	**Min**	**Max**
Economic Well-Being				
Avg Consumption in Rs	10274.2	3052.7	4881.9	21077
Rel. Pov. Rate at 10% Nat. Cons.	0.0903	0.066	0.0023	0.39
Rel. Pov. Rate at 40% Nat. Cons.	0.332	0.16	0.035	0.8
Geographic Descriptors				
log Area (square meters)	14.73	1.01	12.1	18
= 1 if urban	0.304	0.46	0	1
province==[1] Western	0.587	0.49	0	1
province==[3] Southern	0.255	0.44	0	1
province==[6] North-Western	0.0643	0.25	0	1
province==[7] North-Central	0.0155	0.12	0	1
province==[8] UVA	0.0782	0.27	0	1
Agricultural variables				
% of GN area that is agriculture	0.168	0.15	0	0.94
% of GN agriculture that is paddy	44.4	37.5	0	100
% of GN agriculture that is plantation	46.38	37.8	0	100
% of Total GN area that is paddy	8.629	10.9	0	74.7
% of Total GN area that is plantation	8.168	11	0	94.1
Cars				
log number of cars	3.123	1.44	0	8.3
Total cars divided by total road length	0.00556	0.01	0	0.17
Total cars divided by total GN Area	3.77E-05	0.00007	0	0.00093

Building Density				
% of area with buildings	7.817	6.82	0.13	33.9
% shadows (building height) covering valid area	6.509	6.01	0.31	34.9
ln shadow pixels (building height)	12.96	1.04	7.31	17.6
Road variables				
log of Sum of length of roads	9.445	0.94	1.47	13.1
fraction of roads paved	38.3	28.7	0	100
ln length airport roads	0.013	0.33	0	9.25
ln length railroads	1.098	2.67	0	10.8
Roof type				
Fraction of total roofs that are clay	36.5	22	0	100
Fraction of total roofs that are aluminum	14.08	7.06	0	71.9
Fraction of total roofs are asbestos	7.766	11.3	0	71.2
Textural and spectral characteristics				
Vegetation Index (NDVI), mean, scale 64	0.427	0.21	0	0.86
Vegetation Index (NDVI), mean, scale 8	0.566	0.24	0	0.99
Pantex (human settlements), mean	0.627	0.54	0.02	2.94
Histogram of Oriented Gradients (scale 64m), mean	3509.4	2070.3	129.1	10381
Linear Binary Pattern Moments (scale 32m), mean	49.5	1.1	18.1	49.5
Line support regions (scale 8m), mean	0.00836	0.004	-2E-07	0.035
Gabor filter (scale 64m), mean	0.469	0.28	0.014	1.3
Fourier transform, mean	84.34	17.8	4.51	113.4
SURF (scale 16m), mean	12.06	7.77	0.13	31.6
Observations		1291		

Table 1- 2: Estimated GLM Models of Local Area Poverty using High Res Features

	10% Relative Poverty Rate (“Poverty”)		**40% Relative Poverty Rate (“Inequality”)**	
	Mfx: elasticity at means	t	Mfx: elasticity at means	t
Dep Var: Fraction of village below 10/40% of national income				
log Area (square km)	1.893**	[3.04]	0.268	[0.75]
= 1 if urban	-0.152***	[-8.68]	-0.0517***	[-4.71]
% of GN area that is agriculture	-0.0387	[-1.64]		
% of GN agriculture that is paddy	-0.198***	[-4.54]	-0.131***	[-4.71]
% of GN agriculture that is plantation	-0.137***	[-3.41]	-0.0976***	[-3.73]
% of Total GN area that is paddy	-0.0113	[-0.61]	-0.0249**	[-2.98]
log number of cars	-0.162***	[-4.59]	-0.138***	[-5.33]
log of Sum of length of roads	-1.178**	[-3.25]	-0.755***	[-3.52]
fraction of roads paved	-0.140***	[-5.79]	-0.0935***	[-5.58]
ln length airport roads	-0.0013	[-1.55]	-0.00061	[-1.37]
ln length railroads	0.0187**	[2.69]	0.00990*	[2.11]
% of area with buildings	-0.304***	[-5.12]	-0.255***	[-7.05]
% shadows (building height) covering valid area	0.220***	[5.22]	0.148***	[5.28]
ln shadow pixels (building height)	1.401***	[4.50]	1.401***	[6.17]
Fraction of total roofs that are clay	0.307***	[8.80]	0.200***	[9.39]
Fraction of total roofs that are aluminum	0.186***	[6.21]	0.103***	[4.77]
log of Total count of buildings in GN	-0.545**	[-3.29]	-0.362**	[-2.82]

Vegetation Index (NDVI), mean, scale 64	0.251***	[6.35]	0.184***	[8.47]
Vegetation Index (NDVI), mean, scale 8	-0.341***	[-4.38]		
Linear Binary Pattern Moments (scale 32m)	1.969***	[4.37]	1.618***	[5.04]
Line support regions (scale 8m), mean	-0.0595	[-1.37]	0.268	[0.75]
Gabor filter (scale 64m) mean	-0.219**	[-3.06]	-0.0517***	[-4.71]
Fourier transform, mean	1.359***	[6.27]		
SURF (scale 16m), mean	-0.230**	[-3.06]	-0.131***	[-4.71]
Observations	1291		1291	
Log-Likelihood	-272.0925		-527.8597	
AIC	594.185		1095.719	
BIC	723.2643		1198.983	

Notes: Unit of observation is Grama Niladhari (GN) division. Models are clustered at the DS level. Variables were selected using Lasso regularization from the candidate set of variables shown in table 1. Marginal effects presented as elasticities evaluated at mean of independent variable. * p<0.05, ** p<0.01, *** p<0.001

Table 1- 3: OLS Models of Local Area Poverty Rates on High-Res Spatial Features

	10% Poverty Rate		**40% Poverty Rate**		**Average Village Income**	
	coef	t	coef	t	coef	t
log Area (square meters)		[2.52]	0.0093	[0.60]	-0.0079	[-0.31]
= 1 if urban	-0.023	[-1.80]	-0.037	[-1.06]	0.08	[1.18]
% of GN area that is agriculture	coef	[-1.04]	-0.017	[-0.27]		
% of GN agriculture that is paddy	0.020*	[-2.97]	-0.00087**	[-2.97]	0.0014**	[2.92]
% of GN agriculture that is plantation	-0.00021**	[-2.84]	-0.00059*	[-2.66]	0.0012**	[2.72]
% of Total GN area that is paddy	-0.00019	[-0.58]	-0.00083	[-1.10]	0.0016*	[2.10]
Total cars divided by total road length	-0.31	[-1.17]				
Total cars divided by total GN Area	29.6	[0.54]				
log number of cars	-0.0059	[-0.89]	-0.015	[-1.39]	0.024	[1.60]
log of Sum of length of roads	-0.020***	[-3.64]	-0.027*	[-2.32]	0.033	[1.67]
fraction of roads paved	-0.00035***	[-4.24]	-0.00079**	[-3.24]	0.0014**	[3.06]
ln length airport roads	-0.0051	[-1.45]			0.022	[1.52]
ln length railroads	0.00098	[1.31]			-0.0046	[-1.26]
% of area with buildings	-0.0027*	[-2.31]	-0.0093*	[-2.34]	0.020*	[2.56]
% shadows (building height)	0.0022*	[2.04]	0.0064*	[2.18]	-0.013*	[-2.27]

ln shadow pixels (building height)	0.016*	[2.51]	0.039*	[2.64]	-0.047	[-1.95]
Fraction of total roofs that are clay	0.00077**	[3.35]	0.0017**	[3.25]	-0.0027**	[-3.15]
Fraction of total roofs that are aluminum	0.00091***	[3.63]	0.0022**	[3.15]	-0.0040**	[-3.15]
Fraction of total roofs are asbestos	-0.00033	[-1.08]				
log of Total count of buildings in GN	-0.0090**	[-2.71]	-0.019*	[-2.05]	0.029	[1.70]
Vegetation Index (NDVI), mean, scale 64	0.061*	[2.20]	0.14**	[2.94]	-0.21**	[-2.93]
Vegetation Index (NDVI), mean, scale 8	-0.064**	[-2.80]				
Linear Binary Pattern Moments (scale 32m) mean	0.0021**	[2.91]	0.0090***	[5.53]	-0.017***	[-5.92]
Line support regions (scale 8m), mean	-0.66	[-0.87]				
Gabor filter (scale 64m) mean	-0.052	[-1.53]				
Fourier transform, mean	0.0017**	[3.42]				
SURF (scale 16m), mean	-0.0014	[-0.94]	-0.001	[-0.59]	0.0034	[1.06]
Constant	-0.32**	[-3.03]	-0.31	[-1.43]	10.1***	[29.9]
Observations	1291		1291		1291	
R-sq	0.610351		0.618038		0.608118	
R-sq Adj.	0.602022		0.612633		0.60226	

Notes: Unit of observation is Grama Niladhari (GN) division. Variables were selected using Lasso regularization from the candidate set of variables shown in table 1. * p<0.05, ** p<0.01, *** p<0.001

Table 1- 4: Linear Model Estimates Night Lights on Small Area Poverty/Average GN Consumption

	(1)	(2)	(3)	(4)	(5)	(6)
	10% Pov. Rate	**40% Pov. Rate**	**Avg. Income**	**10% Pov. Rate**	**40% Pov. Rate**	**Avg. Income**
Night Lights 2012	-0.583***	-1.546**	2.922**	-0.0383	-0.0898	0.186
	(-3.53)	(-3.38)	(3.32)	(-0.79)	(-0.67)	(0.64)
Observations	1291	1291	1291	1291	1291	1291
R-sq	0.109	0.131	0.147	0.000868	0.000842	0.00103
R-sq Adj.	0.108	0.130	0.146	0.0000932	0.0000671	0.000258
R-sq within				0.000868	0.000842	0.00103
R-sq between				0.372	0.448	0.527
R-sq overall				0.109	0.131	0.147
Divisional Secretariat FEs	No	No	No	Yes	Yes	Yes

Unit of observation is Grama Niladhari (GN) an area 4 administrative boundary.
All models include a regression constant which is omitted from the table.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1- 5: Urban and Rural Models of Local Area Average Income on High-Res Spatial Features

	Rural		Urban	
	coef	t	coef	t
Dependent variable: Average log GN Consumption				
% of GN area that is agriculture	0.12*	[2.34]		
% of GN agriculture that is paddy	0.00076**	[3.11]	0.0002	[0.36]
% of Total GN area that is plantation			-0.0058**	[-3.20]
log number of cars	0.019	[1.27]	0.085***	[5.73]
log Area (square meters)	-0.033	[-1.43]		
log of Sum of length of roads	0.029+	[1.93]		
fraction of roads paved	0.0012**	[3.44]	0.0014+	[2.06]
ln length airport roads	0.044***	[6.59]		
ln length railroads			-0.0052	[-1.50]
% of area with buildings			0.028***	[6.07]
% shadows (building height) covering valid area			-0.015**	[-2.87]
ln shadow pixels (building height)	-0.057**	[-3.23]		
Fraction of total roofs that are clay	-0.0041***	[-6.70]	0.0026	[1.41]
Fraction of total roofs that are aluminum	-0.0051***	[-5.63]	-0.0033+	[-1.84]

Fraction of total roofs are asbestos	-0.0017*	[-2.05]		
log of Total count of buildings in GN	0.040**	[3.53]	0.031	[0.77]
Vegetation Index (NDVI), mean, scale 64	-0.27***	[-4.68]	0.28	[1.65]
Pantex (human settlements), mean	0.18***	[3.73]		
Linear Binary Pattern Moments (scale 32m), mean	-0.013***	[-10.7]		
Line support regions (scale 8m), mean			-1.4	[-0.34]
Fourier transform, mean			-0.0042+	[-1.96]
Constant	10.6***	[36.8]	8.89***	[27.6]
Observations	898		393	
R-sq	0.656151		0.446433	
R-sq Adj.	0.650303		0.427445	

Notes: Unit of observation is Grama Niladhari (GN). Variable selection was performed via Lasso regularization from the candidate set of variables shown in table 1. + p<0.10, * p<0.05, ** p<0.01, *** p<0.00

Table 1- 6: Shapley Decomposition of Share of Variance Explained (R2) by High Resolution Spatial Feature Subgroup

	Average log predicted per capita consumption in GN	10% Poverty Rate	40% Poverty Rate
Urban	10.7	11	10.1
Agricultural land variables	5.1	5.9	5.7
Cars	6.2	7.9	6.8
Building density variables	38.6	38.9	32.9
Of which: Built-up area	20	13.7	16.5
Shadow	13.1	16.6	6.7
NDVI	5.5	8.6	9.7
Road variables	9.8	9.1	9.9
Roof Type	15	17.3	16.1
Texture variables	14.6	9.7	12.9
Observations	1291	1291	1291
R-sq	0.62	0.587	0.622
RMSE	0.1779	0.042	0.0988752
F-stat	81.127	72.05378	83.498
Log Likelihood	409.537	2250.214	1168.524

Notes: Agricultural variables include fraction agriculture plantation, fraction agriculture paddy, and fraction of GN area that is plantation. Car variables include log of car count, and cars per total road length. Building density variables include log of developed area, shadow count (building height proxy), fraction of GN developed, fraction covered by shadow, NDVI at scales 64 and 8. Road variables include log of unpaved road length, log of paved roads narrower than 5m, log of paved roads 5m+, log of airport roads, log of railroad length, and fraction of roads paved. Roof variables include count of roofs by type: clay, aluminum, asbestos, grey cement, and fraction of roofs of same type. Texture variables include Fourier series, Gabor, histogram of oriented gradients, Local Binary Pattern Moments mean and standard deviation, line support regions, and SURF.

Table 1- 7: Shapley Decomposition of Share of Variance Explained (R2) by High Resolution Spatial Feature Subgroup in Addition to Night Time Lights

	Average log predicted per capita consumption in GN	**10% Poverty Rate**	**40% Poverty Rate**
Urban	10	10.4	9.4
Agricultural land variables	4.2	4.6	4.6
Cars	5.7	7.3	6.2
Building density variables	31.3	35.7	33.5
Of which: Built-up area	18.7	15.7	16.7
Shadow	7.2	11.4	10.3
NDVI	5.4	8.6	6.5
Road variables	8.9	8.2	8.7
Roof Type	13.4	15.9	14.45
Texture variables	13.8	8.4	12.2
Night Time Lights (avg, sq, cube, std)	12.8	8.4	10.8
Variance in income/poverty explained by high resolution spatial features in addition to night time lights	87.2	91.6	89.2
Observations	1291	1291	1291
R-sq	0.637	0.61	0.64
RMSE	0.1733	0.04165	0.09623
F-stat	73.761	65.807	75.975
Log Likelihood	446.381	2287.176	1206.085

Notes: Night time lights category includes the following transformations of night time lights: average, squared, cubed, and standard deviation. Agricultural variables include fraction agriculture plantation, fraction agriculture paddy, and fraction of GN area that is plantation. Car variables include log of car count, and cars per total road length. Building density variables include log of developed area, shadow count (building height proxy), fraction of GN developed, fraction covered by shadow, NDVI at scales 64 and 8. Road variables include log of unpaved road length, log of paved roads narrower than 5m, log of paved roads 5m+, log of airport roads, log of railroad length, and fraction of roads paved. Roof variables include count of roofs by type: clay, aluminum, asbestos, grey cement, and fraction of roofs of same type. Texture variables include Fourier series, Gabor, histogram of oriented gradients, Local Binary Pattern Moments mean and standard deviation, line support regions, and SURF.

Table 1- 8: MLE Estimation Correcting for Spatial Autoregression

	Average log Village Consumption (Income)	
	coef	t
log Area (square meters)	-0.046***	[-4.01]
= 1 if urban	0.048+	[1.96]
% of GN area that is agriculture	0.022	[0.42]
% of GN agriculture that is paddy	0.00046+	[1.74]
% of GN agriculture that is plantation	0.00076**	[3.09]
% of Total GN area that is paddy	0.00057	[0.79]
Total cars divided by total road length	-0.93	[-1.20]
Total cars divided by total GN Area	401.4*	[2.28]
log number of cars	0.020***	[3.57]
% of area with buildings	0.0083***	[4.19]
log of Total count of buildings in GN	0.012	[1.23]
Vegetation Index (NDVI), mean, scale 64	0.071	[1.54]
Vegetation Index (NDVI), mean, scale 8	-0.042	[-0.67]
log of Sum of length of roads	0.029**	[2.70]
fraction of roads paved	0.0012***	[6.00]
ln length airport roads	0.0052	[1.50]
ln length railroads	-0.00092	[-0.48]
Fraction of total roofs that are clay	-0.0025***	[-5.83]
Fraction of total roofs that are aluminum	-0.0034***	[-4.92]
Fraction of total roofs are asbestos	0.0014*	[2.26]
Linear Binary Pattern Moments (scale 32m), mean	-0.0080***	[-3.38]
Line support regions (scale 8m), mean	-1.25	[-0.71]
Gabor filter (scale 64m) mean	-0.053	[-0.92]

Fourier transform, mean	-0.0030***	[-3.61]
SURF (scale 16m), mean	0.0052*	[2.24]
Constant	9.74***	[51.6]
Observations	1287	

Notes: Standard errors have been corrected according to Conley (1999, 2008), with model estimation via GMM. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table 1- 9: Estimating Poverty Gap Using High Res Features

	Poverty Gap (FGT1 - 10%)		Poverty Gap (FGT1 - 40%)	
	coef	t	coef	t
log Area (square km)	0.0060**	[2.84]	0.0063	[1.02]
= 1 if urban	-0.0063+	[-2.00]	-0.013	[-1.05]
% of GN area that is agriculture	-0.0081	[-1.29]	-0.018	[-0.76]
% of GN agriculture that is paddy	-0.000087**	[-3.24]	-0.00033**	[-3.10]
% of GN agriculture that is plantation	-0.000053**	[-2.91]	-0.00021*	[-2.63]
% of Total GN area that is paddy	-2.3E-05	[-0.29]	-0.00025	[-0.88]
Total cars divided by total road length	-0.09	[-1.32]		
Total cars divided by total GN Area	9.55	[0.72]		
log number of cars	-0.0014	[-0.83]	-0.0058	[-1.24]
log of Sum of length of roads	-0.0049**	[-2.97]	-0.011*	[-2.48]
fraction of roads paved	-0.000077**	[-3.37]	-0.00023*	[-2.67]
ln length airport roads	-0.00027	[-0.89]		
ln length railroads	0.00026	[1.35]		
% of area with buildings	-0.00062*	[-2.16]	-0.0028*	[-2.04]
% shadows (building height) covering valid area	0.00053+	[1.76]	0.0017	[1.54]
ln shadow pixels (building height)	0.0037*	[2.19]	0.016*	[2.68]
Fraction of total roofs that are clay	0.00020**	[2.96]	0.00070**	[3.12]
Fraction of total roofs that are aluminum	0.00024**	[3.31]	0.00084**	[3.19]
Fraction of total roofs are asbestos	-9.1E-05	[-1.14]		
log of Total count of buildings in GN	-0.0022*	[-2.62]	-0.0073*	[-2.09]
Vegetation Index (NDVI), mean, scale 64	0.017*	[2.33]	0.056**	[2.88]
Vegetation Index (NDVI), mean, scale 8	-0.019**	[-2.95]		

Linear Binary Pattern Moments (scale 32m)	0.00048*	[2.55]	0.0029***	[4.87]
Line support regions (scale 8m), mean	-0.27	[-1.39]		
Gabor filter (scale 64m) mean	-0.016+	[-1.78]		
Fourier transform, mean	0.00046**	[3.44]		
SURF (scale 16m), mean	-0.00025	[-0.67]	-0.0001	[-0.15]
Constant	-0.093**	[-3.41]	-0.17+	[-2.00]
Observations	1234		1234	
R-sq	0.5884		0.6097	
R-sq Adj.	0.5792		0.6039	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table 1- 10: Model Performance Using Simulated Reduced Census Sampling

Sample of villages (GNs) for model training	Sample of households within GNs	In-Sample R^2	Out of Sample R^2	Normalized Mean Absolute Error (NMAE)
Using 10% National Poverty Line As Dependent Variable				
100	100%	0.6325	-	0.3272
	50%	0.6316	-	0.3271
	25%	0.6287	-	0.3283
50	100%	0.6058	0.6379	0.3355
	50%	0.6061	0.6378	0.3348
	25%	0.6092	0.6441	0.3342
25	100%	0.5995	0.6139	0.3393
	50%	0.5990	0.6167	0.3395
	25%	0.5948	0.6212	0.3392
Using 40% National Poverty Line As Dependent Variable				
100	100%	0.6215	-	0.2300
	50%	0.6213	-	0.2300
	25%	0.6209	-	0.2303
50	100%	0.6097	0.6104	0.2348
	50%	0.6097	0.6118	0.2346
	25%	0.6098	0.6154	0.2340
25	100%	0.5947	0.6132	0.2363
	50%	0.5939	0.6152	0.2357

	25%	0.5923	0.6178	0.2346
Using Income (Average Consumption) as Dependent Variable				
	100%	0.6081	-	0.014778
100	50%	0.6080	-	0.014783
	25%	0.6077	-	0.014782
	100%	0.5989	0.5964	0.01499
50	50%	0.5987	0.5982	0.01498
	25%	0.5984	0.6014	0.01493
	100%	0.5880	0.5943	0.01524
25	50%	0.5874	0.5954	0.01520
	25%	0.5866	0.5966	0.01515

Notes: This table simulates estimation error when using a reduced Census size. “Sample of villages” refers to the percentage of the villages within the Census used to train the model. “Sample of households within GN” refers to the number of households within the sampled GNs used to calculate the income or poverty rate statistic. In-sample (out-of-sample) R^2 reports the coefficient of determination for the data used in the training (test) sample. Normalized mean absolute error (NMAE) reports the mean average error rate divided by the average income/poverty rate, such that the statistic gives the average absolute error expressed as a percentage of the income/poverty rate.

Table 1- 11: Estimated Poverty Extrapolation Performance, Using DS Leave-One-Out Cross-Validation (LOOCV)

	Average predicted per capita consumption in GN	**10% Poverty Rate**	**40% Poverty Rate**
Normalized Root Mean Squared Error (NRMSE)	.0836147	.5596012	.3631206
Normalized Mean Absolute Error (NMAE)	.0241663	.404652	.2766029
Spearman Rank Correlation Between Predicted and True Poverty Rates	.6982542	.6942492	.6792735

Notes: Table presents out of sample estimates for of extrapolated poverty rate prediction into withheld Divisional Secretariat (DS) administrative districts. We estimate 47 models, each time withholding one of 47 DS units to reserve as an out of sample test. Using the relationship between poverty and satellite variables estimated using the training data, we predict into the withheld DS – so called "leave-one-out" cross-validation (LOOCV).

Appendix A: Description of Imagery and Extraction of Object and Texture Features from High Resolution Satellite Features

Details on Satellite Imagery

The satellite imagery consists of 55 unique "scenes" purchased from Digital Globe, covering areas specified in our sample area.[21] Each "scene" is an individual image captured by a particular sensor at a particular time. Images were acquired by three different sensors: Worldview 2, GeoEye 1, and Quickbird 2. These sensors have a spatial resolution of $0.46m^2$, $0.41m^2$, and $0.61m^2$, respectively in the panchromatic band and $1.84m^2$, $1.65m^2$, $2.4m^2$ respectively in the multi-spectral bands. Pre-processing of imagery included pan-sharpening, ortho-rectification, and image mosaicking.

Details on Extraction of Object Based Features

Object features were classified using the assistance of two technical partners: Orbital Insight and LandInfo. Orbital Insight produced object classification for three variables: The share of the GN division that is built-up (i.e. consists of buildings), the number of cars in the GN, and the share of pixels in the GN that were identified as shadow pixels, which is a proxy for the gross floor area, or height, of buildings. The classification method used by Orbital Insight is similar to Krizhevsky, Sutskever, and Hinton (2012) which utilizes convolutional neural networks (CNN) to build object predictions from raw

[21] Particular thanks to Digital Globe is due for their "Seeing the World" program, which offered very high spatial resolution imagery at reduced rates for non-commercial purposes.

imagery. LandInfo classified the remaining objects, which included roof type, paved and unpaved roads of different widths, railroads, and the type of agriculture. Landinfo used a combination the Trimble eCognition and Erdas Imagine software platforms to classify objects, except for roads, which were classified using visual interpretation.

The CNN classification algorithm used by Orbital Insight involved four steps:

1. Ingestion/Tiling
2. Model Development
3. Classifying All Pixels Using the Trained Model
4. Aggregating Prediction Results to GN Division level

The tiling stage split the large images into many small images or tiles, in order to make the modeling computationally scalable, as each tile could be distributed to a different GPU core for greater efficiency. In the model development stage, the classification model was trained and tuned. Model building began by manually classifying or labeling a sub-sample of the imagery as a positive or negative value for a given object using a crowdsourced campaign. The classified data was split into an 80% training and a 20% testing set, where the training set was used to build the model. This allowed sample prediction metrics, presented below, to be calculated using the withheld test set. Training was run for 60,000 iterations using the Nesterov solver method, a variant of stochastic gradient descent.

To get a sense of the accuracy, Figure 1-A 1 shows the receiver operator characteristics, or ROC curve, summarizing the classification accuracy of the developed area building classifier. The ROC curve presents the true positive rate on the y-axis, against the false positive rate on the x axis as the discriminant threshold is varied. A classifier that is no better than random would correspond to the 45 degree line, while a perfect classifier would correspond to the point (0,1) in the ROC space. Improvements in classifier accuracy are shown as the curve moves up and to the left. The ROC curve suggests the classification algorithm is highly accurate, corresponding to a 90% accuracy rate overall. Once the model was full trained, the team applied the trained model to the full set of imagery. The results were then summarized at the GN level.[22]

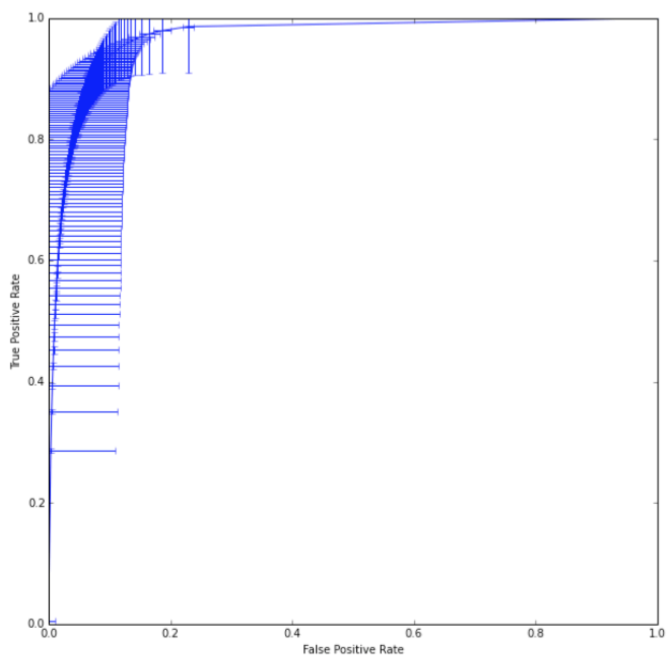


Figure 1-A 1: ROC Curve for Developed Area (Buildings) Classifier

LandInfo extracted roof type, paved and unpaved roads of different widths, railroads, and the type of agriculture, using a combination of an object-based image

---

[22] The GN shapefile was in turn based on the GN shapefile provided by DCS, modified to correct manually identified errors.

analysis (OBIA) methodology utilizing the Trimble eCognition software platform (Benz et al., 2004) and visual interpretation. OBIA is a common classification methodology applied to HSRI because it is designed to map objects that are larger than the pixel size (Blaschke, 2010). For this OBIA analysis, the first step was to segment the image into polygons that represent the features of interest based on spectral and spatial homogeneity of those features. The segments were created based on the degree of homogeneity using a range of scale, shape, and other parameters. Once an image was segmented, the objects themselves were then classified, using Erdas Imagine Software. Using this methodology, all of the roofs of individual buildings within the imagery were extracted and the type of roof was classified as asbestos, aluminum, clay, or grey roofs, in part based on validation of selected roofs in Colombo. The roof types of a few selected buildings were validated by direct observation. Roads and agriculture were extracted using a combination of eCognition and manual visual interpretation. Visual interpretation was required because the roads were covered by trees and in shadow, making automated detection difficult. In addition, the imagery was not always taken during the growing season, which along with the small, irregular shapes of the agricultural plots made automated extraction impractical.

Figure 1-A 2 shows an example OBIA classification for roads, railroads, and road width. The road network in the village has been mapped in detail, showing not just major roads, but minor roads. Not shown are the road type, whether the road is paved or not. These features combine to indicate the extent to which a given area is accessible by road. Figure 1-A 3 shows an example roof type classification, where roof type is distinguished not just by material but by material shading.

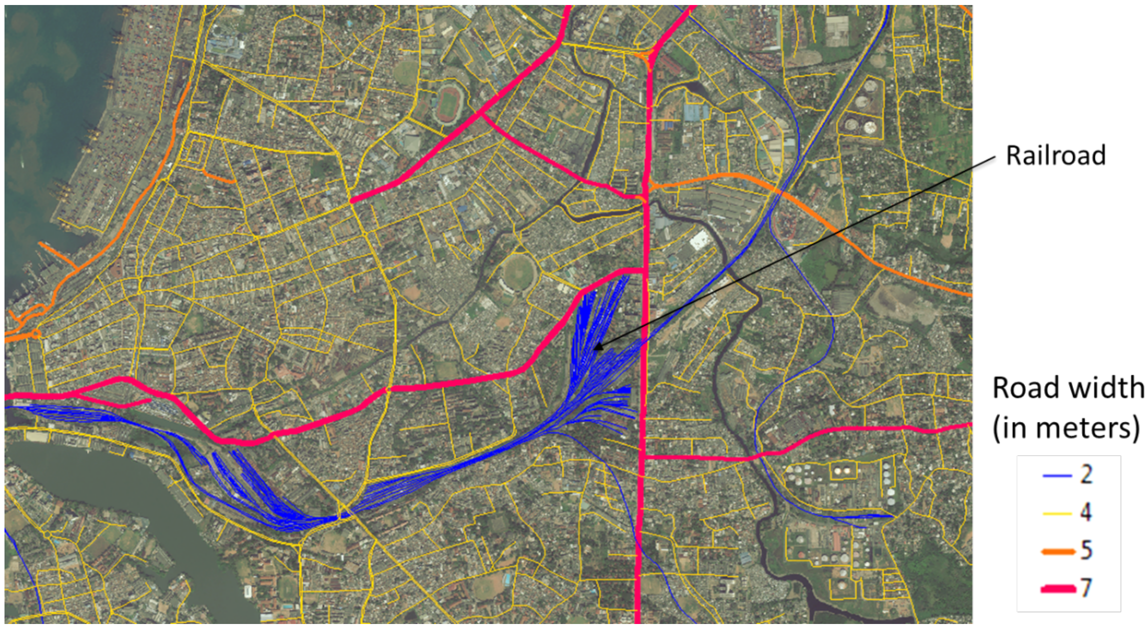


Figure 1-A 2: Example Roads and Railroads Classification

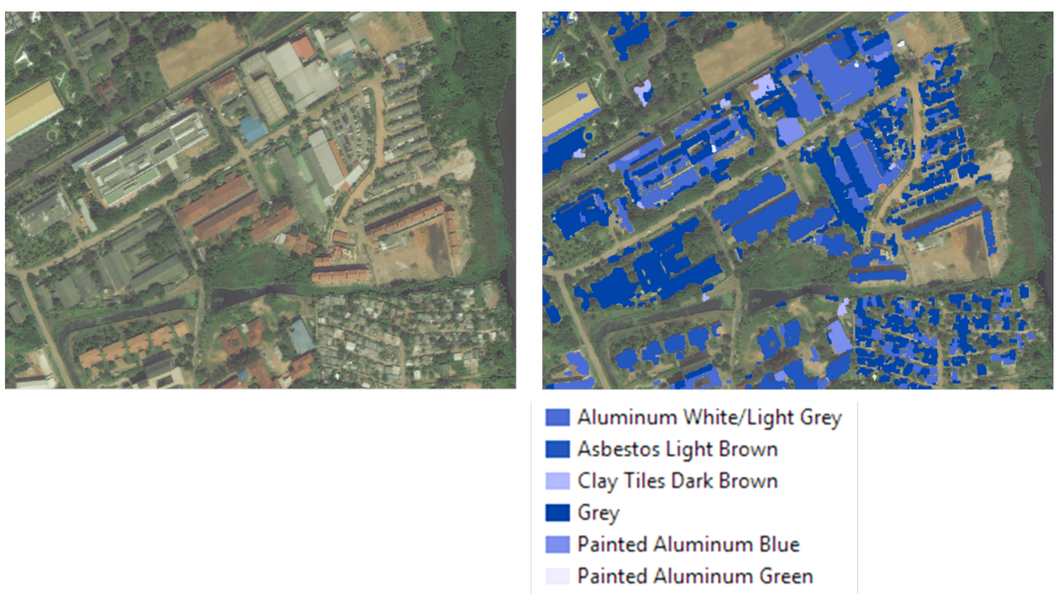


Figure 1-A 3: Example Roof Type Classification

Details on Extraction of Spectral and Textural features

Textural and spectral features were created using a block size of eight pixels and scales of 8, 16, 32, and 64 meters using a methodology similar to Graesser et al. (2012). This resulted in an output image comprised of 165 bands at a spatial resolution of 12.8 or 16m depending on native, multispectral resolution of each of the sensors. The seven textural features calculated for this study were:

1. Histogram of oriented gradients (HOG), which captures edge orientations and sorts them into a histogram (Dalal and Triggs 2005).
2. PanTex, which is a built-up presence index derived from the grey-level co-occurrence matrix (GLCM) (Pesaresi et al. 2008).
3. Line support regions (LSR), which characterize line attributes (Yu et al. 1999)
4. Local binary patterns moments (LBPM), which define contiguous regions of pixel groups and sorts them into a histogram (Wang and He, 1990).
5. Fourier transform (FT) which examines pattern frequency across an image (Smith 1997).
6. Gabor, a linear Gaussian filter used for edge detection (Gabor 1946)
7. Speeded Up Robust Features (SURF), an algorithm that extracts key points (i.e., edges and corners) from an image through pyramidal Gaussian based decomposition (Bay et al., 2006).
8. The Normalized Difference Vegetation Index (NDVI), the most widely used vegetation index that provides information about the presence and abundance of vegetation (Tucker 1979).

Additional spectral features calculated were simply the means of the four individual bands, Blue, Green, and Near Infrared. Once the spatial and spectral features were calculated, the mean, standard deviation, and sum were determined for each GN Division. Previous research has indicated that these features are correlated with census data that indicate poverty such as slum conditions, population density, solid waste collection, unimproved sanitation (Sandborn and Engstrom, 2016) and to map informal and slum areas within cities (Graesser et al. 2012, Engstrom et al. 2015).

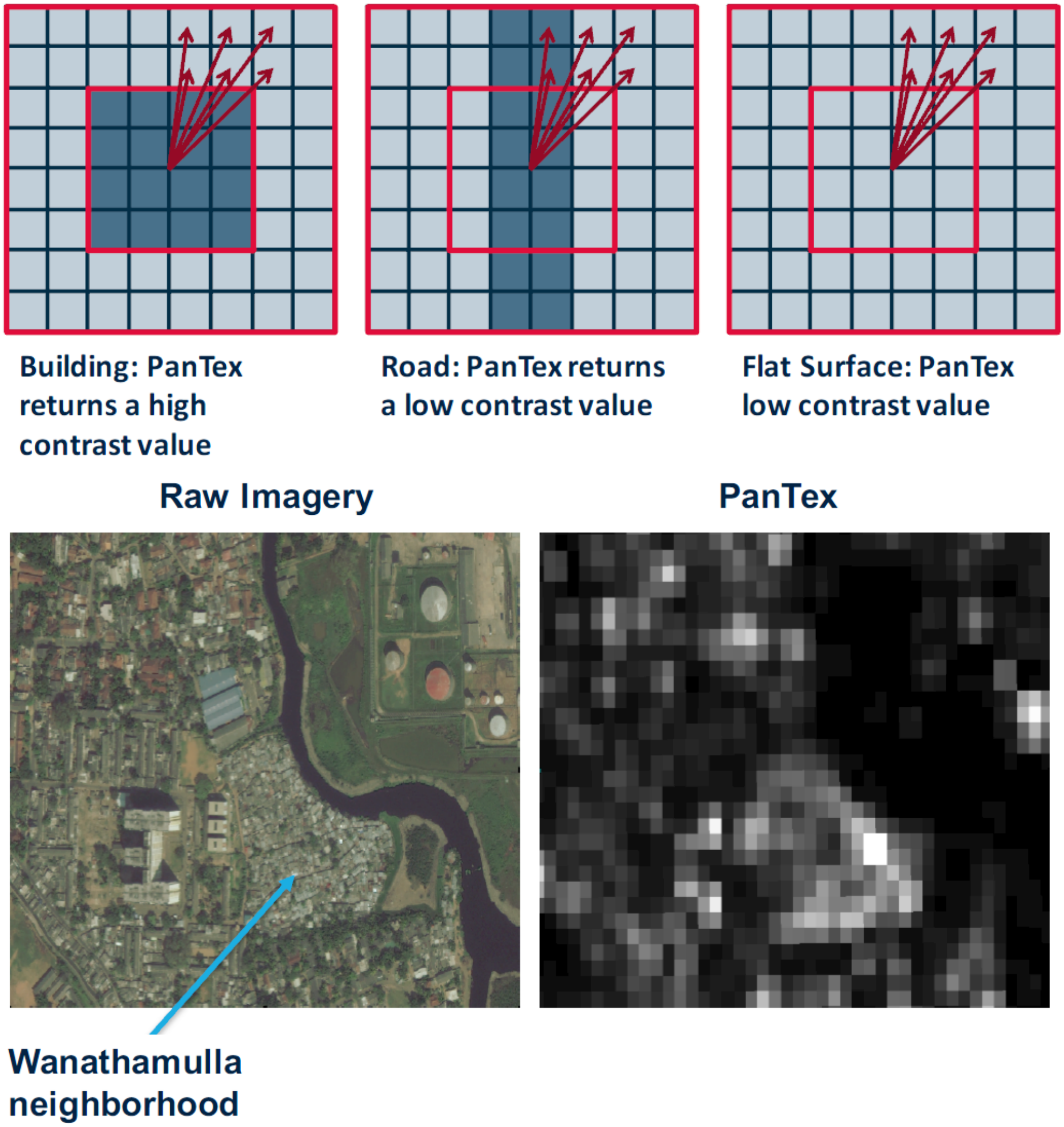


Figure 1-A 4: Pantex Classification Algorithm Description (top) and Example Classification (bottom right) Applied to Raw Imagery (bottom left)

Sampled divisional secretariats include Ambagamuwa, Ambalantota, Ambanpola, Bandaragama, Biyagama, Bulathsinhala, Colombo, Dehiwala, Devinuwara, Dodangoda, Doluwa, Dompe, Galle Four Gravets, Hali Ela, Hambantota, Homagama, Horana, Ingiriya, Kaduwela, Kalutara, Kamburupitiya, Katana, Kattankudy, Kelaniya, Kesbewa, Kirinda

Puhulwella, Kolonnawa, Kotapola, Kothmale, Kurunegala, Madurawala, Maharagama, Malimbada, Manmunai North, Matara Four Gravets, Moratuwa, Nagoda, Negambo, Nuwara Eliya, Nuwaragam Palatha East, Padukka, Panadura, Panvila, Puttalam, Rathmalana, Rattota, Seethawaka, Sri Jayawardanapura Kotte, Thihagoda, Thimbirigasyaya, Tissamaharama, dapalatha, Udunuwara, Ukuwela, and Uva Paranagama.

Appendix B: Information on Sri Lankan Administrative Divisions

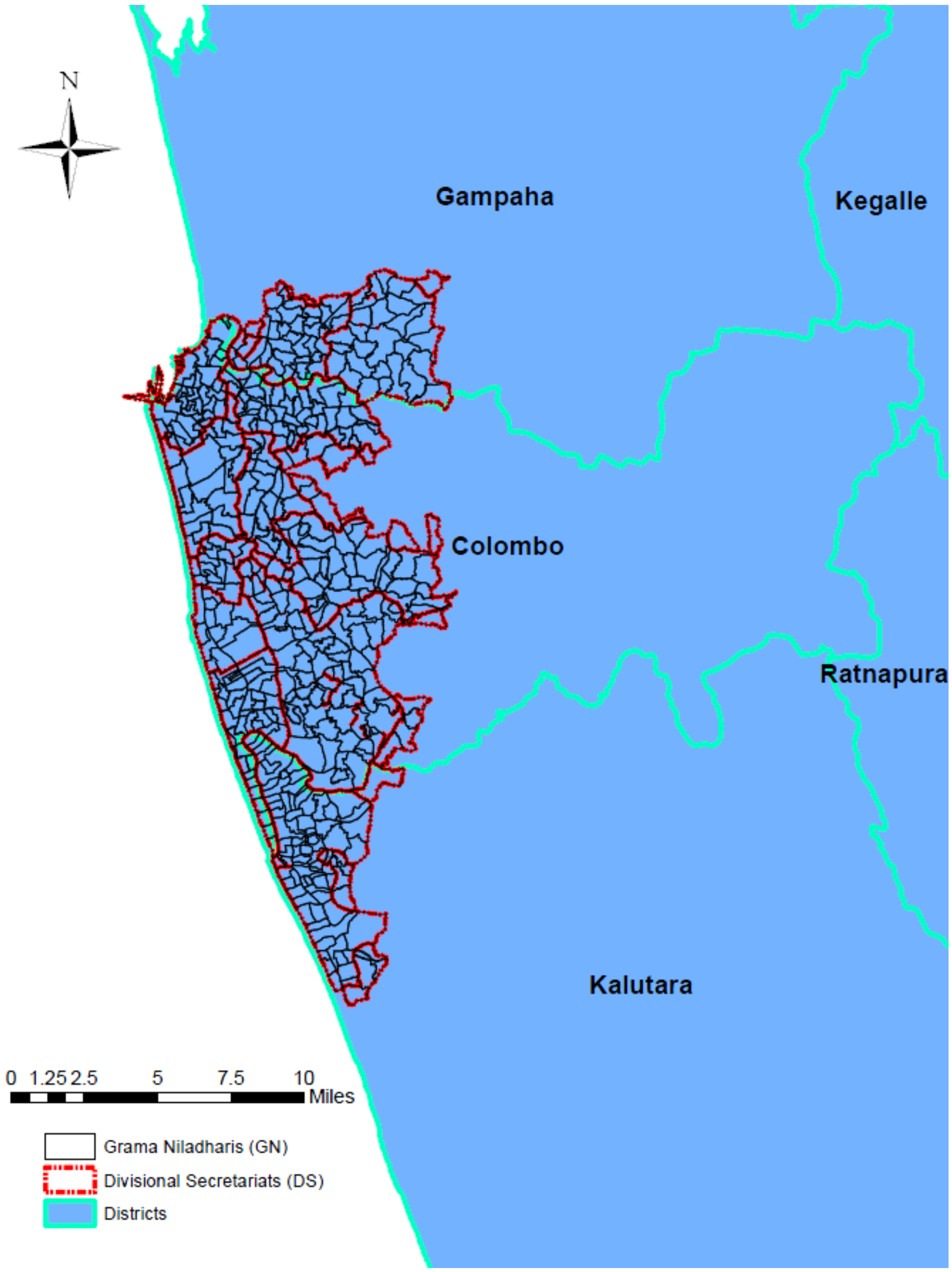


Figure 1-B 1: Graphical Depiction of Relative Size of Districts, Divisional Secretariats, and Grama Niladharis

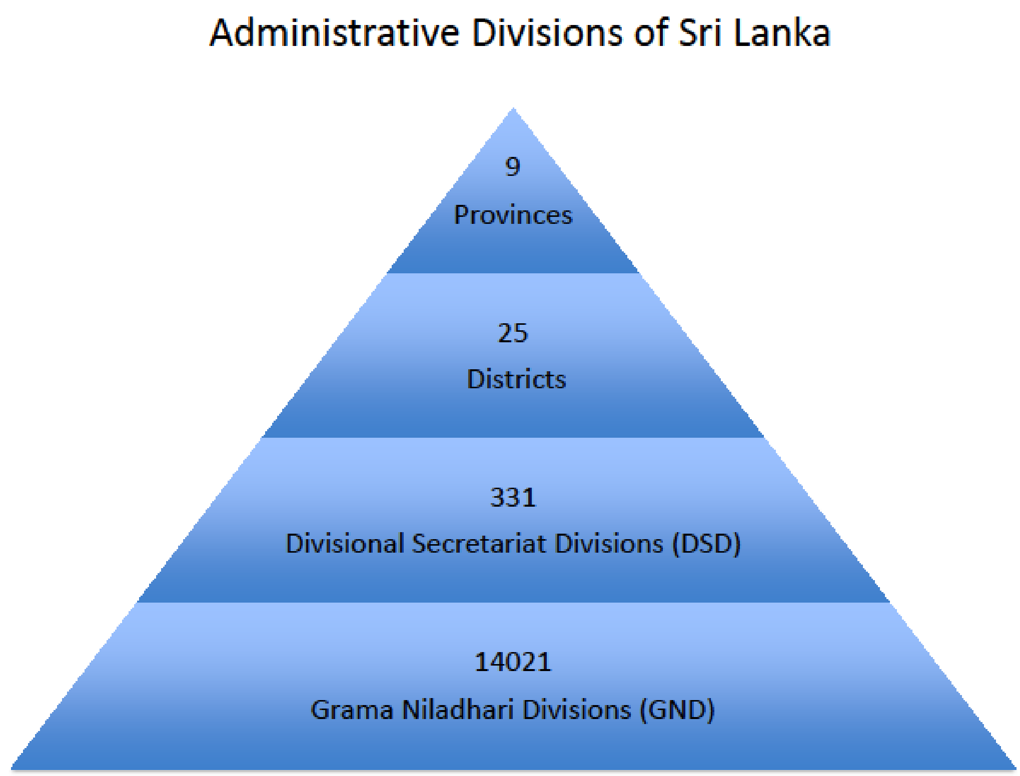


Figure 1-B 2: Count of Administrative Divisions of Sri Lanka by Type

CHAPTER TWO

Preferential Tariffs and African Trade[23]

Klaus-Peter Hellwig
khellwig@imf.org
International Monetary Fund

Jonathan Hersh
jhersh@bu.edu
Boston University

Abstract

This paper investigates the effects of preferential trade programs such as the U.S. African Growth and Opportunity Act (AGOA) on the direction of African countries' exports. While these programs intend to promote African exports, textbook models of trade suggest that such asymmetric tariff reductions could additionally divert African exports from other destinations to the tariff reducing economy. We examine the import patterns of 177 countries and estimate the diversion effect using a triple-difference estimation strategy, which exploits time variation in the product and country coverage of AGOA. We find no evidence of systematic trade diversion within Africa, whereas diversion from other industrialized destinations to the US was significant, in particular for apparel products. At the same time we show that, more than diverting trade, AGOA had positive spillovers on

[23] We thank Marianne Baxter, Samuel Bazzi, Stefania Garetto, Adam Guren, Christian Henn, Siddharth Kothari, Dilip Mookherjee, Megha Mukim, David Newhouse, Alberto Osnago, Bob Rijkers, and seminar participants at Boston University for the many helpful comments and discussions. Yun Liu and Guangzhi Yi provided excellent research assistance. All errors are our own.

the product composition of trade, which suggests that the product coverage of preferential trade agreements can influence structural change in Africa.

Keywords: African Trade, Trade Diversion, Regional Trade Integration, Preferential Trade Agreements

JEL Classification: F14, F15, O24

1 Introduction

The African Growth and Opportunity Act (AGOA) is a cornerstone of the United States' economic and political relations with Sub-Saharan Africa. Since its inception in late 2000, AGOA has offered substantial tariff reductions on a wide range of products originating from African beneficiary countries, with the intention of promoting export-led growth in the region. Other large economies, including the E.U. and China, established similar trade programs. Such unilateral agreements were enacted without any of the reciprocity requirements common to bilateral or multilateral trade agreements. This paper studies the impact of those programs on the direction of Africa's exports. While the existing literature suggests that AGOA has served its principal purpose of promoting Africa's exports and thereby enhancing the continent's integration with the world economy, our central question is to what extent this growth reflects a redirection of trade flows. To study this question, we exploit the variation in the timing and product coverage of tariff reductions.

We find considerable heterogeneity in the impact of AGOA. While some trade got diverted, mainly from other industrialized destinations to the US, we show that AGOA also had positive spillovers, encouraging more African exporters to enter new markets. The response of bilateral trade flows to AGOA also varies widely across products, with diversion being particularly strong for apparel items.

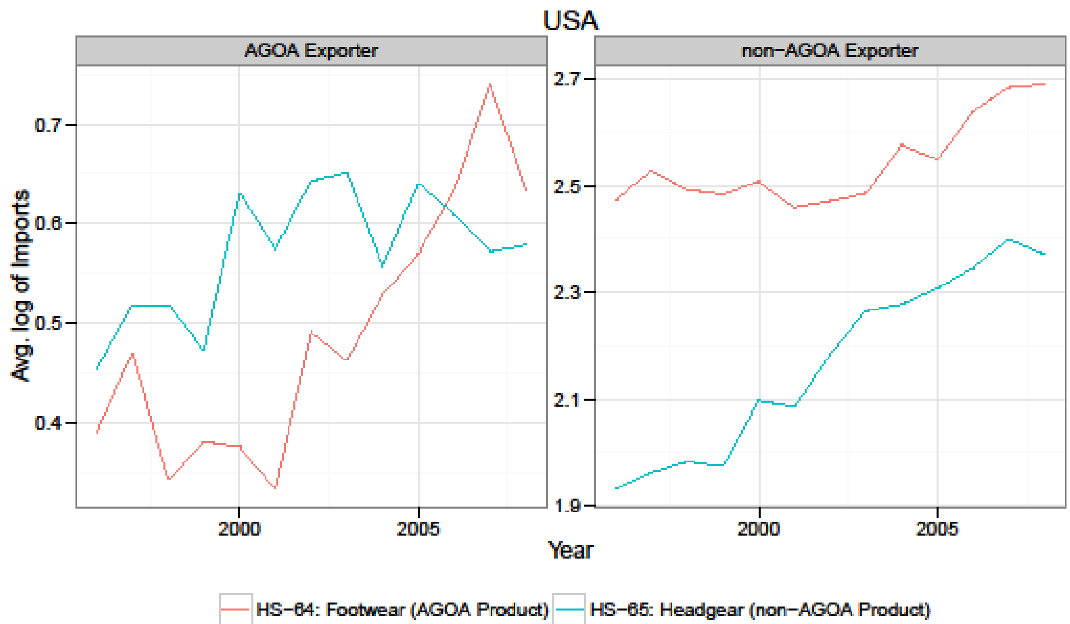
To illustrate our approach and main results, let us begin with a simple example. Consider the US imports of two product groups, footwear and headgear.[24] Seventy-five percent of product lines in footwear were covered under AGOA, whereas zero percent of headgear products received such treatment.[25] Figure 2- 1a compares average US imports of footwear and headgear, showing differences between AGOA eligible exporters in the left panel, and ineligible exporters in the right panel. For non-beneficiaries, there appears to be no systematic difference across product groups. However, for AGOA beneficiaries, there is a strong increase in the footwear exports from around 0.35 log points in 2000 to more than 0.7 log points in 2007, whereas no strong increase in headgear exports is seen. The figure is just one example of how, as intended, the program helped to promote exports of AGOA products from AGOA beneficiaries to the US.

But what if we look at imports of the same product groups into *Kenya*? For geographic reasons, Kenya is an important trading partner to several AGOA beneficiaries. The left panel of Figure 2- 1b compares Kenyan imports of footwear and headgear from AGOA exporters before and after the implementation of AGOA. Imports of footwear products averaged of 1.6 log points in 1998, but steadily declined over the early 2000s, reaching 0.4 log points in 2008. By contrast, no systematic decline is seen for headgear,

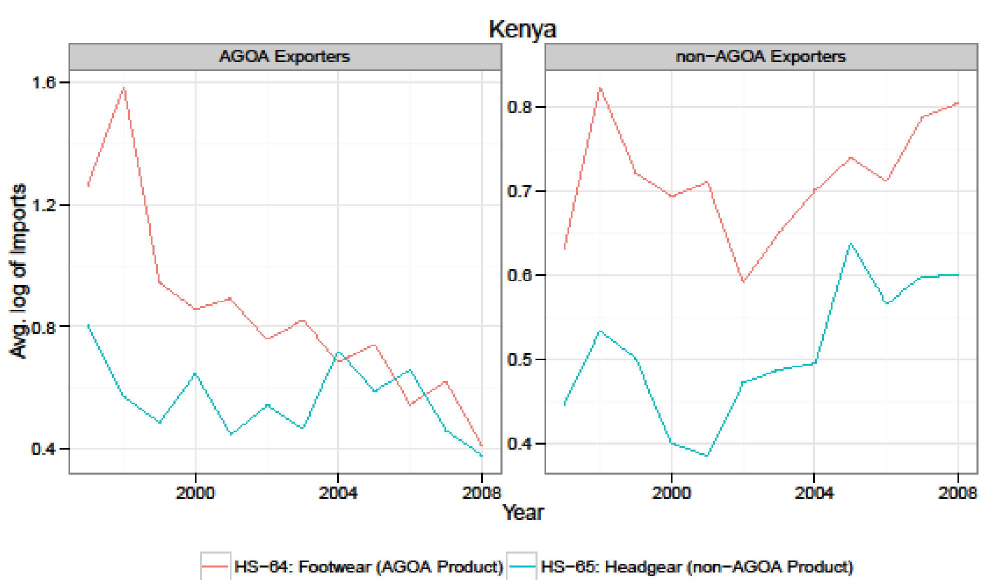
[24] Product groups are identified by their 2-digit HS codes (64, and 65 respectively)
[25] We define treatment here and throughout the paper as a product being eligible for an AGOA preferential tariff rate.

the non-treated product group. Similarly, when looking at Kenyan imports from non-AGOA exporters, we see no systematic difference between imports of headgear and footwear, suggesting the decline in footwear imports into Kenya from AGOA-beneficiaries is due to some characteristic specific to AGOA product treatment. We interpret the process described by Figure 1b as *trade diversion*, that is, evidence that AGOA designated products were diverted from SSA countries to other destinations.

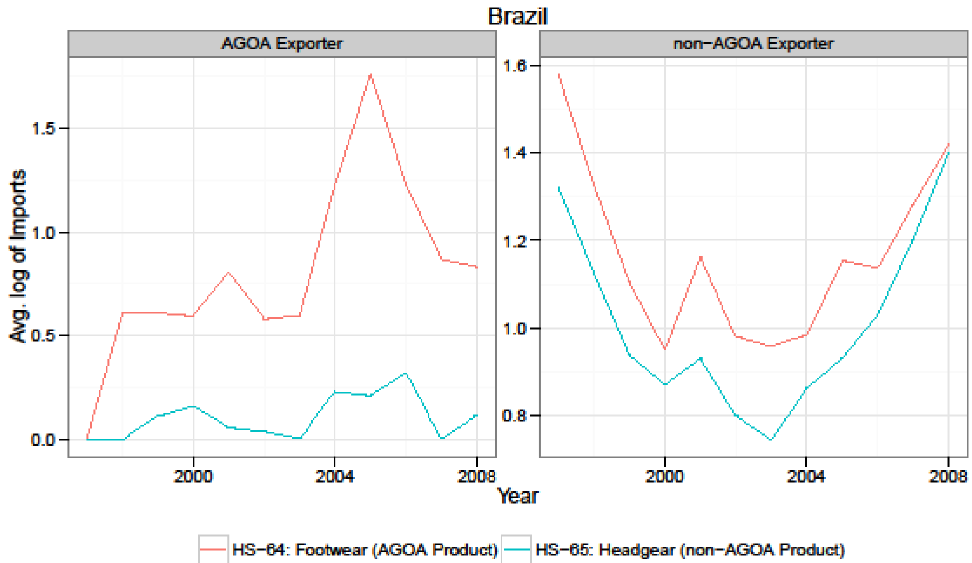
Figure 2- 1c looks at imports into Brazil. Brazil had no comparable program to promote exports from SSA, in particular no program that favored footwear over headgear. Despite this, we see increases in footwear imports from AGOA exporters following the implementation of AGOA. Looking at non-AGOA exporters in the right panel, we see no systematic difference between imports of footwear and headgear between non-AGOA and AGOA exporters. We interpret this response of African exports in footwear as positive spillovers.



(a) U.S. Imports



(b) Kenyan Imports



(c) Brazilian Imports

Figure 2- 1: Imports from AGOA and non-AGOA exporters for high-AGOA treated product category (footwear) and low-AGOA treated category (headgear), before and after AGOA implementation in late 2000.

Existing research and the policy discussion on AGOA have focused on the narrative of Figure 2- 1a. For example, Frazer and van Biesebroeck (2010) estimate the impact of AGOA on US imports using a triple-difference estimation strategy.[26] Controlling for several sets of fixed effects to ensure that their result is not driven by product-specific demand shocks or macroeconomic conditions in exporting countries, they find that, on average, AGOA led to a 13.5 % increase in imports of AGOA products from beneficiary countries. We follow their idea of exploiting the variation in the country and product coverage of AGOA to identify its effect on trade patterns. But rather than looking at the

[26] Note that Figure 2- 3a depicts the triple between exporters, across products, and over time.

origin of U.S. imports, we focus on the *destination of African exports*. Using a similar triple-difference methodology, we ask whether Figure 2- 1b and 1c are representative for a wider range of products.

The use of triple-difference regressions has proven useful in estimating the effects of preferential trade programs. Yet, we are also mindful of the potential problems of applying what is essentially a micro-econometric approach to a macroeconomic question. Beyond the mere diversion of trade from one destination to another, a general equilibrium model of trade would suggest that an asymmetric tariff reduction by a large export market will lead to a change in the production structure of African economies: Input factors are withdrawn from high-tariff sectors and used by sectors that benefited from the tariff reduction. This mechanism could be exacerbated by the presence of increasing returns in production or exporting.[27] To the empiricist such spillovers raise the question to what extent the left panel in Figure 2- 1a should be interpreted as a *relative increase* in African exports of footwear and to what degree it shows a *relative decline* in African exports of headgear. In other words, from a theoretical point of view, it is hard to make the case that products such as headgear can serve as a proper control group. To address this concern, we consider an alternative specification which controls for a countries total exports of a given product, which allows us to study the *composition* of exports by destination in the spirit of a quadruple-differences regression. For example, rather than asking "did exports of footwear from Ethiopia to Kenya increase?", we ask "did Kenya's share in Ethiopia's

[27] It is noteworthy that several African countries responded to AGOA with domestic industrial policies targeting sectors that benefited from AGOA to leverage its impact.

footwear export increase?". This approach allows us to some extent to separate *what* countries export from *where* they ship it to.

To be clear, claims regarding the welfare implications of these effects are beyond the scope of this paper. When it comes to trade diversion, it is difficult to assess whether AGOA exacerbated or reduced existing distortions in apparel trade. And the enhanced access to the US market has attracted inflows of foreign direct investment to Africa which, through transfers of technology and know-how can generate important spillovers. However, our results indicate that the impact of tariff reductions on aggregate export growth, especially for apparel, are smaller than suggested by the growth US-African trade. This finding complements the work of Rotunno et al. (2013) who show that a significant amount of US textile imports under AGOA are trans-shipments due to the generous rules of origin.

The spillovers generated by tariff reductions also have important policy implications. McMillan and Rodrick (2011) find evidence that structural change has been TFP reducing for many low income and emerging countries, and they attribute this structural change with the product composition of exports. A growing body of literature, sparked by Hausman et al. (2007), also suggests that the composition of exports matters for the pace of economic development.

This paper speaks to several literatures. Firstly, we contribute to the debate regarding the estimated impact of AGOA. By showing the AGOA had unintended consequences impacting intra-African trade, we suggest that current estimates of the estimated impact of AGOA may be overstated. This paper also contributes to the literature on preferential tariff agreements in general, showing that even unilateral programs may have large general equilibrium effects. Finally, we contribute to the literature on regional trade integration, by shedding light on the extent to which preferential trade programs may impact regional integration among both beneficiary and non-beneficiary countries.

The rest of the paper is organized as follows: Section 2 provides background into the creation of AGOA and gives a brief review of estimated impacts from this program, along with other research focused on trade diversion. Section 3 presents our methodology and gives an overview of the data sources used. Section 4 presents our main results.

2 Background

Since the implementation of its GSP program in 1975, the United States has offered substantial tariff reductions to most low-income and emerging economies, including most African countries, covering up to half of all product categories. The GSP program was expanded in 1997 to offer duty free access for additional products imported from eligible least developed countries (GSP for LDCs). In October 2000, the GSP and GSP for LDCs were complemented by the African Growth and Opportunity Act (AGOA), which was signed into law on May 18th, 2000[28], and strengthened through three separate acts of congress between 2000 and 2015. It stipulates a nonreciprocal trade agreement between the United States and many Sub-Saharan African countries. The key benefits of the AGOA entail an extension of the enhanced market access specified in the GSP for LDCs to non-LDCs (see Figure 2- 4b). Hence, after the reductions of worldwide trade barriers in the

[28] Title I, Trade and Development Act of 2000; P.L. 106–200

1990s had gradually eroded the advantage of reduced tariffs, AGOA helped restoring the average discount in (ad-valorem equivalent) tariff rates for non-LDCs such as South Africa or Mauritius to around 2 percentage points (see Figure 2-5). An additional provision was created for textiles and apparel, knowing as the AGOA Textile program, which allows eligible countries to ship qualifying apparel duty-free and quota-free to the US. Eligible products are allowed to use inputs originating from outside of AGOA countries, but must meet the Rules of Origin requirement[29]. In addition to reduced preferential tariff rates, the U.S. provided technical assistance for AGOA eligible countries through US Agency for International Development (USAID) to assist countries in using the benefits of the initiative. Initially AGOA was set to expire in 2008, but it was extended in 2004 and 2015.

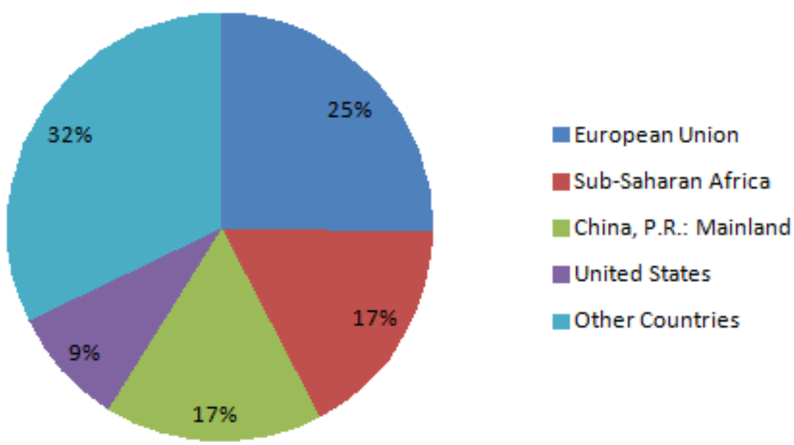


Figure 2- 2: Composition of aggregate export from sub-Saharan Africa (38 countries) by destination, 2013. (Source: IMF, DOTS)

[29] Nevertheless, Rotunno et al. (2013) find that while the Multifiber Agreement was in effect (2001-2005) Chinese firms utilized the AGOA provision to “quota hop” for textile exports facing quota restrictions. They estimate that this form of transshipment accounts for 22% of Africa’s apparel exports from 2001-2008.

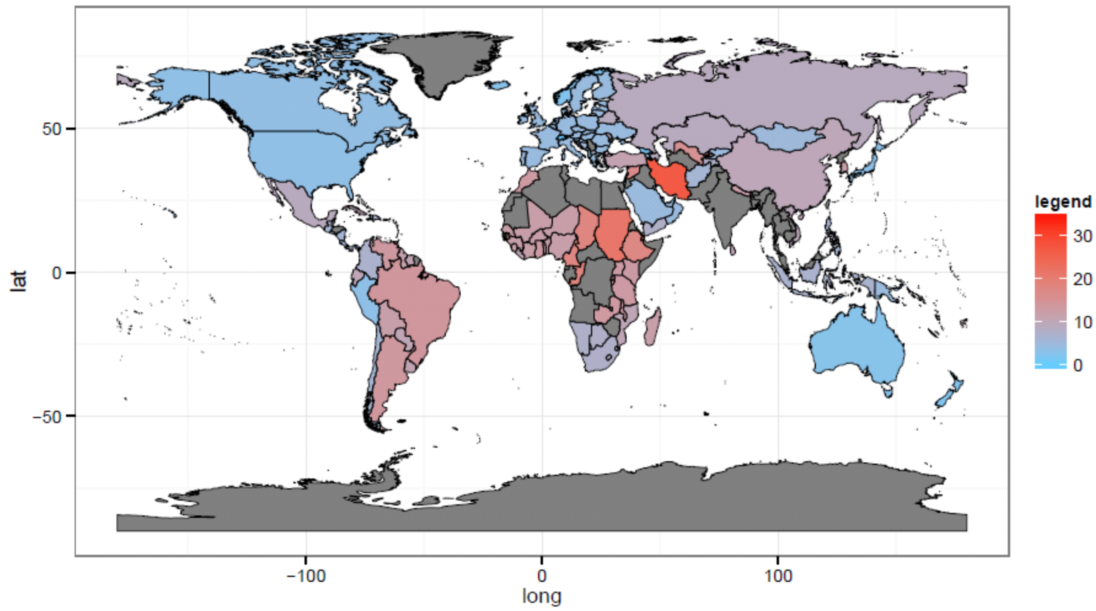


Figure 2- 3: Average MFN tariff rate, 2013 or latest available. (Source: World Bank, WDI

At the time of the law's signing, 34 countries were eligible to receive AGOA benefits, and as of 2015 the number of eligible AGOA countries stands at 40[30]. Country eligibility for the program is subject to a yearly review, and while the program was intended to be inclusive of Sub-Saharan African countries, country eligibility is dependent on three criteria: (1) making progress towards, or having established market based economic reforms, rule of law, elimination of trade barriers and barriers to US investment, protection of worker's rights, increasing availability of health care and the elimination of child labor; (2) not engaging in activities detrimental to US security interests; and (3) not engaging in

[30] Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Chad, Comoros, Republic of Congo, Cote d'Ivoire, Djibouti, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, South Africa, Tanzania, Togo, Uganda, and Zambia.

violations of human rights or harboring or promoting terrorist groups.

Figure 2- 5 presents a map describing AGOA eligibility. Several SSA countries were never eligible for AGOA (Somalia, Sudan, and Zimbabwe. On occasion countries previously receiving AGOA benefits have been deemed ineligible for violating any of the above conditions. Notable examples include Madagascar following undemocratic changes in government in 2009, Cote d'Ivoire in 2005, and Guinea and Niger in 2009, and Guinea-Bissau in 2012 Mauritania, Niger, Guinea, and Cote d'Ivoire had AGOA benefits revoked then subsequently restored, while Madagascar, Eritrea, DR Congo, Central African Republic, Mali, and the Gambia had benefits revoked and not restored as of the time of this writing.

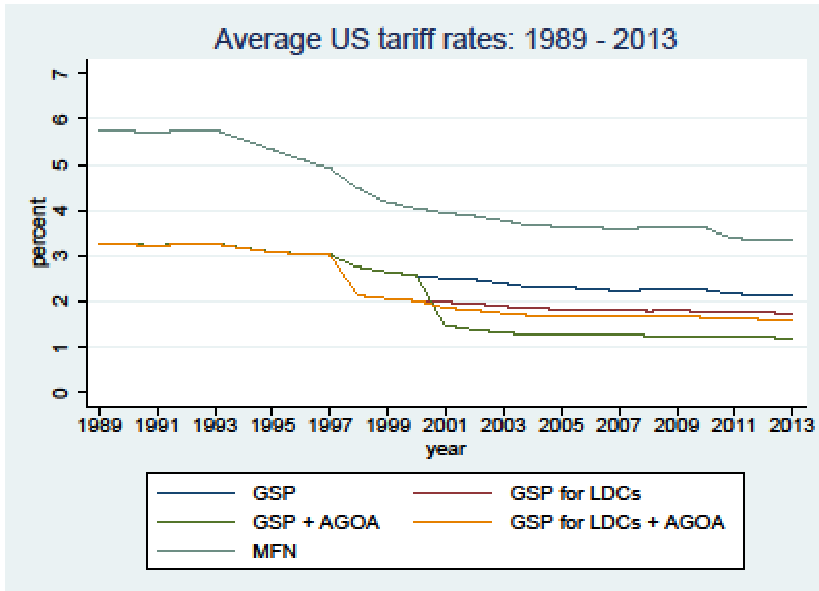
Estimated Impacts of AGOA in the Literature

Estimated impacts of AGOA in the literature have been mixed, but the consensus appears to be that it had a positive causal impact on SSA exports. On a utilization basis, the program has been a large success. Figure 2- 6 compares imports into the US by selected preferential tariff program.[31] Usage of AGOA peaked in 2008, with over $28 billion of imports coming into the US through AGOA designated lines from AGOA beneficiaries. In comparison to other preferential trade programs, AGOA is the largest US preferential trade program on an import value basis, apart from NAFTA. Of course, utilization is not an adequate measure of program success, as one could argue the US would have seen these level of imports from AGOA beneficiaries even without the program being enacted.

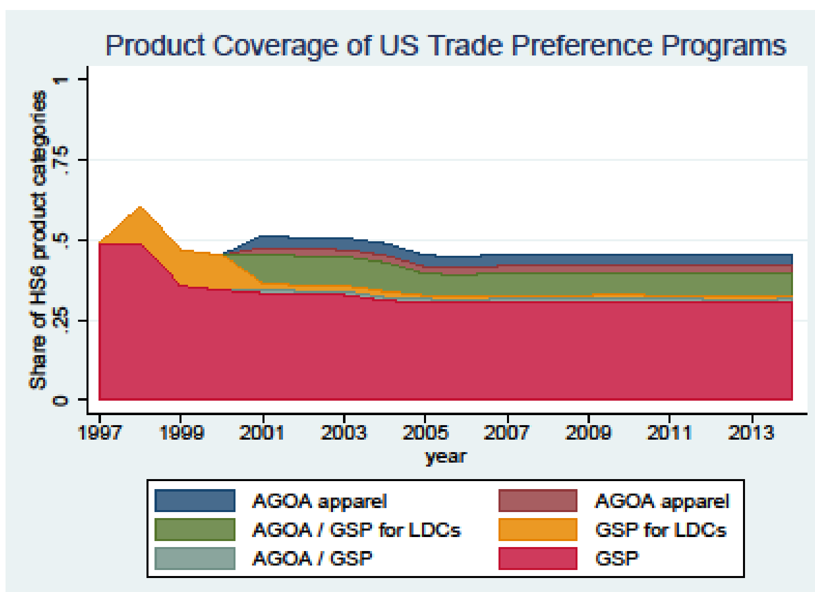
To estimate a causal impact of AGOA researchers have employed a variety of

[31] Source: http://dataweb.usitc.gov/

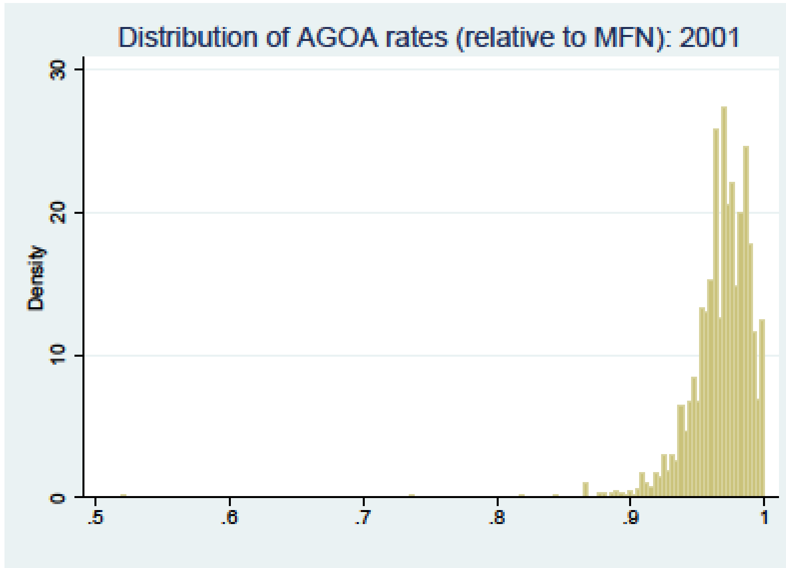
methods. Nouve and Staaz (2003) use a modified gravity equation to estimate whether AGOA increased agricultural exports to the US, and find an insignificant effect of AGOA treatment. More recently, Frazer and van Biesebroeck (2010) estimate the impact of AGOA on all African exports to the US using a triple difference specification. We will pay particular attention to their estimator, as it functions as the starting point for our analysis. The authors find that AGOA product designation has a positive impact on US imports from AGOA designated countries, with an estimated DDD coefficient of 0.127 that is significant at the 5% level. This translates to a marginal impact of AGOA designation on a product of 13.5%, meaning that AGOA designation causally increased imports from beneficiary countries by 13.5% relative to non-designated products. While the authors themselves do not make welfare claims as to the impact of increased African exports to the US, many subsequent researchers have interpreted these findings as such. For example, Frankel (2010) states “Mauritius was one of the first two countries [with Kenya] to be approved for AGOA, which has proven successful.”



(a) United States: Unweighted mean of ad valorem equivalent tariff rates over 6-digit products by country group



(b) United States: Coverage of various preferential trade programs. Most AGOA products were already duty free under the GSP for LDC program. Products exclusively covered by AGOA are predominantly textiles, but also include fresh-cut flowers and vegetables.



(c) United States: preference margin $(1 + \tau_{AGOA}1 + \tau_{MFN})$ for AGOA eligible countries and eligible (6-digit) products, 2001

Figure 2- 4

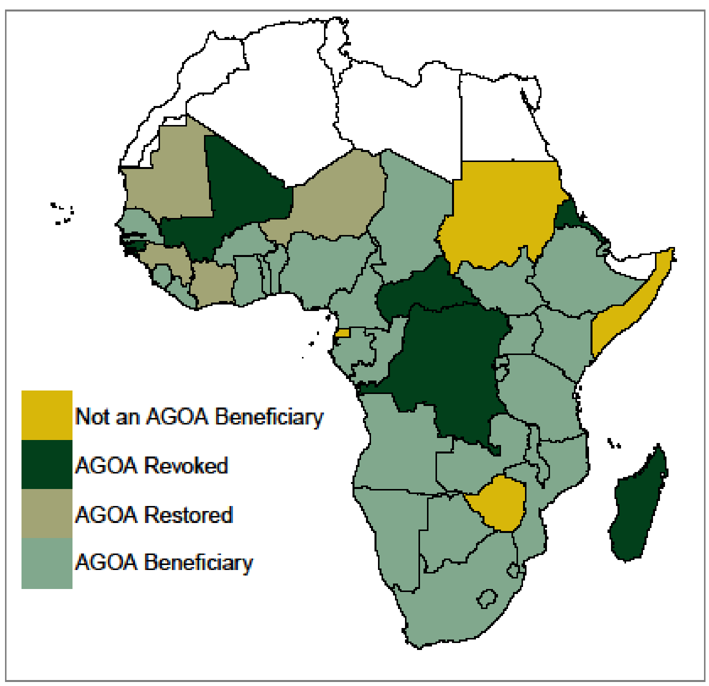


Figure 2- 5: AGOA Eligible Countries

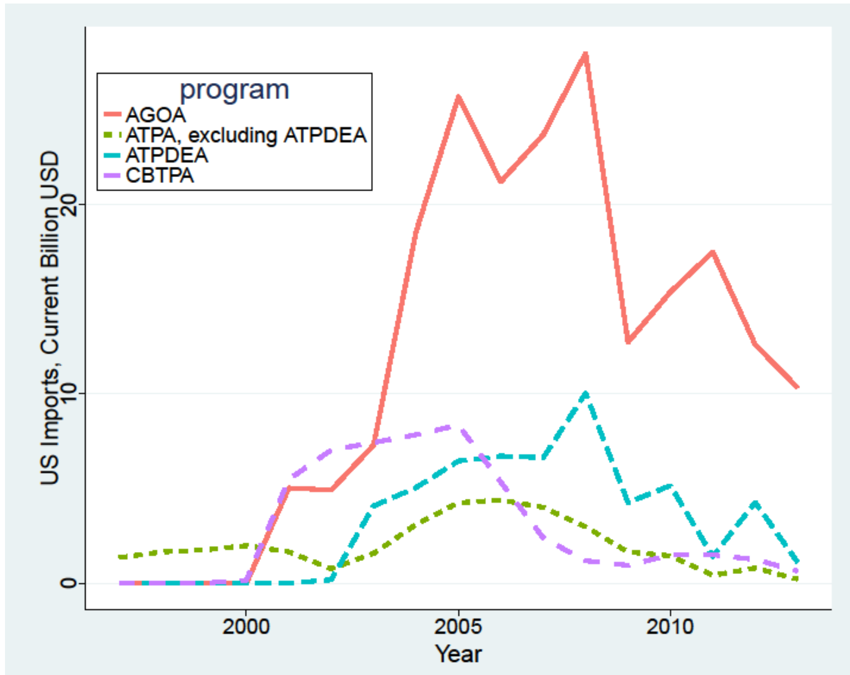


Figure 2- 6: US Imports by Selected Preferential Tariff Agreement, Source: US International Trade Commission

This paper adds to a number of papers which have examined trade diversion resulting from unilateral and multilateral trade programs. Concerns about the diversionary effects of PTAs have a long history going back to Viner (1950)[32], but despite the proliferation of such agreements there has been relatively little attention paid to the effects on trade diversion of these programs. Baldwin and Murray (1977) offered a thorough treatment when analyzing trade diversion and trade creation under the general system of preference (GSP), the formalized system of exemption from WTO/GATT tariff restrictions. Baldwin and Murray propose an estimator to calculate for donor countries the

[32] For a good overview of the history of these concerns, and of PTAs in general, see Bhagwati's *Termites in the Trading System: How Preferential Agreements Undermine Free Trade* (2008).

trade diversion from non-beneficiary countries, which differs from our approach in that we focus on the trade diversion for non-donor countries. On the empirical end, Krueger (1999) estimates whether there was significant trade diversion caused by NAFTA among US, Canada and Mexican imports. The author finds that "the evidence seems to indicate that those commodity categories in which Mexican exports to the U.S. grew most rapidly were also those categories in which it grew most rapidly with the rest of the world" and concludes that NAFTA was likely trade creating and not trade diverting. Clausing (2001) examines the Canada-US Free Trade Agreement and finds little evidence that the act led to trade diversion. In Africa, Holden and McMillan (2006) estimate the impact of the Southern African Development Community (SADC) on imports into South Africa. The authors find no impact of SADC on the composition of imports into South Africa, leading them conclude there was no trade diversion from this program.

3 Methodology and Data

Our starting point is Frazer and van Biesebroeck's (2010) idea to exploit variation in AGOA product assignment and country AGOA eligibility, along with the timing and implementation of the program. Their estimator can intuitively be thought of as the difference between two difference-in-difference (DD) estimators, comparing the difference-in-difference impact of AGOA between AGOA beneficiary countries and non-AGOA beneficiaries. Equation (2) below present the intuitive form of their DDD estimator

(1)

$$DDD = \underbrace{[(lnIMP_{03}^{AP} - lnIMP_{99}^{AP}) - (lnIMP_{03}^{NP} - lnIMP_{99}^{NP})]}_{AGOA\ Country} - \underbrace{[(lnIMP_{03}^{AP} - lnIMP_{99}^{AP}) - (lnIMP_{03}^{NP} - lnIMP_{99}^{NP})]}_{Non-AGOA\ Country}$$

where *AP* indicates an AGOA-designated product, *NP* indicates a non-AGOA product, and *03* and *99* are two arbitrarily chosen years (2003, and 1999) after and before AGOA is implemented respectively. Their specification compares the growth of imports into the US of products at the six digit level from African countries, both designated AGOA beneficiaries and those not designated, and the growth of imports of of AGOA products with that of non-AGOA products. Note that this is exactly what is shown in Figure 2- 1. The first term on the right hand side of equation **(1)** gives the change in AGOA versus non-AGOA products for AGOA beneficiaries, shown in the left panels of Figure 2- 1. The second term on the right hand side of equation **(1)**, the difference in growth between AGOA and non-AGOA products for non-AGOA countries is shown in the right panel.

Frazer and van Biesebroeck (2010) implement the estimator **(1)** as follows:

(2)

$$\ln(z_{o,p,t}) = \beta_{US}^{AGOA} * AGOAproduct_{pt} * AGOAcountry_{ot} + \beta_{US}^{AGOAapparel} * AGOAapparelproduct_{pt} * AGOAapparelcountry_{ot} + \tilde{\mu}_{ot} + \tilde{\gamma}_{pt} + \tilde{\delta}_{op} + \epsilon_{opt}$$

where $z_{o,p,t}$ is the value of exports of product p from exporter o to the US in year t, $AGOAproduct_{pt} \in \{0, 1\}$ indicates whether a product is designated as AGOA in year t,

$AGOAcountry_{ot} \in \{0, 1\}$ indicates whether an exporter is given AGOA status in year t, and $\tilde{\mu}_{ot}$, $\tilde{\gamma}_{pt}$, and $\tilde{\delta}_{op}$ are exporter/year, product/year, and exporter/product fixed effects. Here the main coefficient of interest is β_{US}^{AGOA} which gives the triple-difference estimate of the impact of AGOA product designation on exports from AGOA beneficiary countries to the US. This coefficient is specific to the US. Under Frazer and van Biesebroeck's preferred specification, the estimates of β_{US}^{AGOA} and $\beta_{US}^{AGOAapparel}$ are 0.127 and 0.426 and highly significant, which is interpreted as evidence for a sizable effect of AGOA on export growth.

This interpretation of the sign and magnitude of β_{US}^{AGOA} as evidence of export growth rests on the assumption that, as in an experiment with clearly designated treatment and control groups, each exporter/product pair is an independent observation. Products and countries not covered under AGOA serve as a control group to establish the counterfactual. Trade theory, however, teaches us that this empirical strategy, common to studies of microeconomic phenomena, may have its pitfalls when it comes to evaluating macro policies, if these policies affect relative prices and thereby change production and exporting decisions for *all* sectors. If firms that benefit from AGOA increase their exports, they hire more capital and labor which may otherwise have been employed by other firms. Relative to the counterfactual, these other firms thus use less labor and reduce their exports. The estimator in **(2)** cannot distinguish between the increase of AGOA trade and the decrease of non-AGOA trade and could therefore overestimate the impact of AGOA on exports to the US.

The second assumption behind the above interpretation of β_{US}^{AGOA} is that AGOA did not trigger any diversion of trade from other destinations to the US. To test this assumption, we extend our analysis to all origin/destination/product/year combinations. To do so, we generalize **(2)** to

(3)

$$\ln(x_{o,d,p,t}) = \beta_d^{AGOA} * AGOAproduct_{pt} * AGOAcountry_{ot} + \beta_d^{AGOAapparel} * AGOAapparelproduct_{pt} * AGOAapparelcountry_{ot} + \tilde{\mu}_{odt} + \tilde{\gamma}_{dpt} + \tilde{\delta}_{odp} + \epsilon_{odpt}$$

where x denotes the value of exports and d is the destination country (so that $z_{o,p,t} \equiv x_{o,US,p,t}$). We estimate equation **(3)** separately for each importing country d in our sample. If AGOA was a microeconomic experiment that only affected trade flows between the US and its beneficiaries in treated products, then $\beta_d^{AGOA} = 0$ for all $d \neq US$. If AGOA affects the production structure and leads to more exports of benefiting firms and less exports of other firms, then β_d^{AGOA} can be positive for some $d \neq US$. And if AGOA triggers a diversion of covered products from destination d to the US, then β_d^{AGOA} can be negative.

Note that, while we interpret $\beta_d^{AGOA} < 0$ as trade diversion, $\beta_d^{AGOA} > 0$ does not imply the absence of such diversion. This is because AGOA may trigger both, trade diversion *and* a change in the product mix towards AGOA products, whereas β_d^{AGOA} can only capture the net impact of these two forces. We therefore consider an alternative specification in which we control for $ln(\tilde{x}_{opt})$, where $\tilde{x}_{opt}$ are country o's total exports of a given product p. An alternative would be to\strikeout off\uuline off\uwave off add an origin/product/time fixed effect, and jointly estimate all coefficients for all importers, which would amount to a quadruple-difference equation. Doing so, however, would require considerably larger computational resources. Note that, if we control for covariates at the *origin/product/time* level, the coefficients β obtain a different interpretation. Instead of measuring the relative effect of policies on the *\uuline default\uwave defaultvalue*\strikeout off\uuline off\uwave off of trade flows between country *o* and *d*, they measure the effect on the share of country *d* in country *o*'s exports of product *p*, since $ln(x_{o,d,p,t}\tilde{x}_{opt}) = ln(x_{o,d,p,t}) - ln(\tilde{x}_{opt})$.

Data

We use import data from UN Comtrade's *International Trade Statistics Database* as the main source of data on trade flows between countries.[33] This database provides yearly trade flows between countries disaggregated at the six-digit product level, covering most countries. Whenever available, we use data from 1996 to 2014. Product codes are harmonized to correspond to the 1996 HS nomenclature.[34] Not all countries report import data throughout our period of interest. Among African countries in particular, data coverage often begins later, sometimes as late as the early 2000s. Out of a total of 45 African countries in the UN Comtrade ITSD, 12 have trade flow series which begin after 2001.[35] Our triple difference empirical methodology requires variation in timing of AGOA designated countries. Since AGOA starts in 2001, we would need adequate import data covering the pre-treatment period.[36]This is not problematic for estimating equation **(3)**, as

[33] http://comtrade.un.org/
[34] During the period of study, from 1996-2014, HS product coding changed multiple times. We use the concordance tables published under http://wits.worldbank.org/product_concordance.html
[35] These countries are: Benin, Congo, Comoros, Egypt, Ghana, Guinea, Guinea-Bissau, Libya, Lesotho, Morocco, Swaziland, and Seychelles.
[36] This isn't strictly necessary, as we do have variation in AGOA beneficiary countries post-2001.

we can exclude countries with insufficient coverage. However, when computing a country o's total exports $\tilde{x}_{opt}$ of product p, we need to add up over all importers, i.e.,

$$\tilde{x}_{opt} = \sum\nolimits_{d} x_{odpt}$$

When computing these totals, we only include data from those importers who reported in every year between 1998 and 2010. This ensures that time variation in $\tilde{x}_{opt}$ is not caused by the entry of new reporting countries into the sample.

A data description for each country is presented in Table 2- A.1 in the appendix. The second column gives the first year when our import panel begins. The third column gives the year import data ends for a given country, which for most countries is at the year 2013. Several countries included in our main sample end earlier, such as Grenada which ends in 2009, and Syria which stops in 2010. The fourth column gives the number of positive flows of import lines observed in our data over the period in our data. This varies widely, based on the size of the importer along with their import structure. For example, we observe 2,238,378 positive import flows into the US at the six digit product level, while for the relatively smaller economy of Paraguay we observe 386,954 positive import flows. When estimating our specifications, we follow convention and include instances of zero trade flow between countries. To keep those observations with zero, we follow the convention in the development literature and use the inverse hyperbolic sine (IHS) transformation instead of taking logs.[37]

[37] The IHS is defined as $\log\left(x + \sqrt{x^2 + 1}\right)$. This can be interpreted the same as the standard logarithmic transformation. An alternative would be to transform variables with zeros using $log(x + 1)$, though the former is preferred for small parametrization of x. For more see Burbidge et. al (1988) or MacKinnon and Mageer (1990).

Data on preferential trade agreements is obtained from a variety of sources. For AGOA, we use information from the US International Trade Commission (USITC).[38] For non-apparel, tariff policy is defined at the HS-8 product level, while our trade flows data is available only at the HS-6 level, requiring some aggregation to make the two series compatible. We follow Frazer and van Biesebroeck (2010) and, within each 6-digit product, compute the share of HS-8 products that receive preferential tariff treatment. These shares are weighted using the total US imports between 1998 and 2000 for each product.[39] For apparel products, we do not have a list of covered products. However, from the USITC's trade data, we can recover information on apparel products that were imported under AGOA after 2000. This method allows us to identify AGOA apparel products, but we fail to identify AGOA apparel products which were not imported into the US from any country. By omitting these products, we are likely to overestimate the impact of AGOA for apparel. Information on country coverage is also available online.[40]

4 Results

4.1 Exports to the US

We estimate equation **(3)** separately for each importing country that reports sufficient data to Comtrade. We begin with an analysis of imports into the US and China to revisit the results of Frazer and van Biesebroeck (2010) who estimate the effects of

[38] http://dataweb.usitc.gov
[39] data on US imports at the HS-8 level are also available from the USITC
[40] http://agoa.info/

AGOA between 1998 and 2006. Table 2- 1 presents our estimates. The first column of estimates corresponds to Frazer and van Biesebroeck (2010), although with a longer sample period. The effect of AGOA on the value of US imports remains highly significant and positive, although for non-apparel products the effect is smaller than in the earlier estimates. This decline in the effect of AGOA is not surprising, given that MFN tariffs have declined considerably, thus eroding preferential tariff margins.

Note that, for non-apparel products, the effect of AGOA on US imports becomes insignificant when we control for $ln(\tilde{x}_{opt})$ in the second through fourth column, which suggests that, on average, for these products there was no significant diversion to the US. To the contrary, it appears that the spillovers to other countries were strong enough to keep the share of the US in African exports stable. On the other hand, for apparel, we have a coefficient of 0.187 – evidence for considerable trade diversion. The table also shows that the estimated effects on trade shares are robust to different ways of computing $\tilde{x}_{opt}$, a country *o*'s aggregate exports of product *p* in year *t*.

4.2 Preferential tariffs and trade flows

We now extend the analysis and estimate equation **(3)** for each available importer, but *without* controlling for $ln(\tilde{x}_{opt}^{1997-2013})$. The results for all 177 countries are presented in Table 2- A.2 in the appendix, but we prefer to use the graphical representation in Figure 2- 7 in which each map plots the effect of an individual preferential tariff program (i.e., one map per column of Table 2- A.2).

For non-apparel products, AGOA had a positive impact not just for trade in AGOA products with the US, but also for African trade in these products with most other countries (Figure 2- 7a). An exception are several European destinations and Japan where we estimate a negative coefficient. For African countries, with the exception of Guinea-Bissau, imports of AGOA products from AGOA countries have increased relative to non-AGOA products.

For apparel products (Figure 2- 7b), AGOA had a markedly different impact with negative coefficients for most countries. The relative drop in AGOA versus non-AGOA products is most pronounced for China, where it amounted to more than 50 percent. Interestingly, we find very little effect of AGOA on apparel trade flows within Africa (with the exception of South Africa).

Table 2- 1: Regression results (equation (3)) for the US

	Dependent variable: $ln(x_{odpt})$			
AGOA	0.036**	-0.001	0	0
	(9.47)	(-0.68)	(-0.44)	(0.22)
AGOA (apparel)	0.522**	0.177**	0.189**	0.182**
	(27.77)	(32.27)	(39.07)	(37.85)
$ln(\tilde{x}_{opt}^{1998-2010})$		0.784**		
		(4070.36)		
$ln(\tilde{x}_{opt}^{1997-2013})$			0.785**	
			(4778.16)	
$ln(\tilde{x}_{opt}^{all})$				0.775**

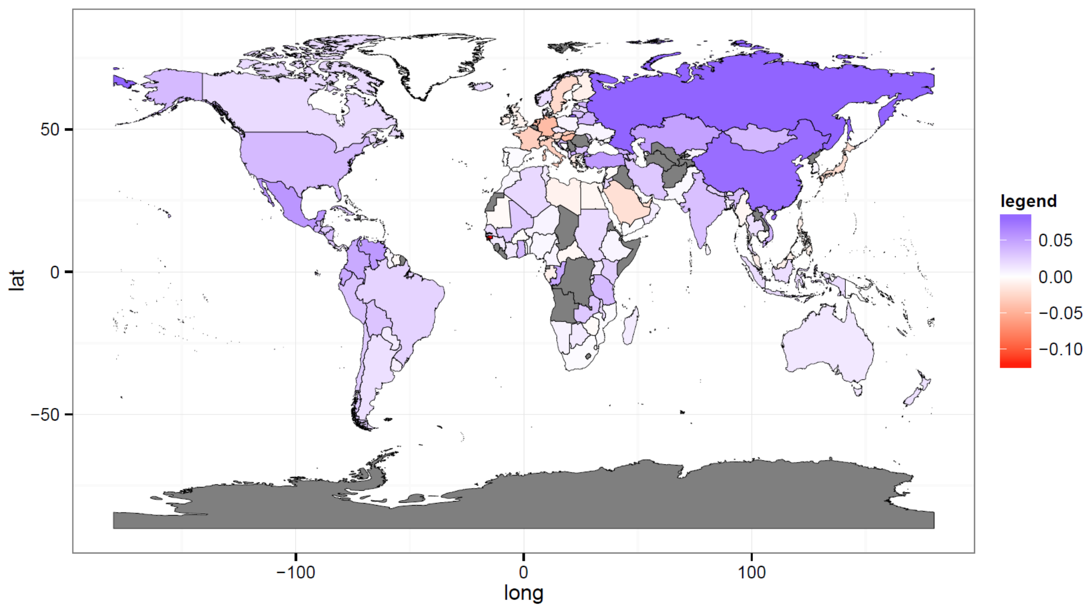
				(4932.93)
country/product FE	yes	yes	yes	yes
product/year FE	yes	yes	yes	yes
country/year FE	yes	yes	yes	yes
Observations	20287080	14651780	19160020	20287080

Notes: t-statistics in parentheses; *significant at 5%; **significant at 1%; standard errors robust to heteroskedasticity. $\tilde{x}_{opt}^{1998-2010}$: aggregate exports of product *p* to countries reporting in every year from 1998-2010; $\tilde{x}_{opt}^{1997-2013}$: aggregate exports of product *p* to countries reporting in every year from 1997-2013; $\tilde{x}_{opt}^{all}$: aggregate of all reported exports of product *p*.

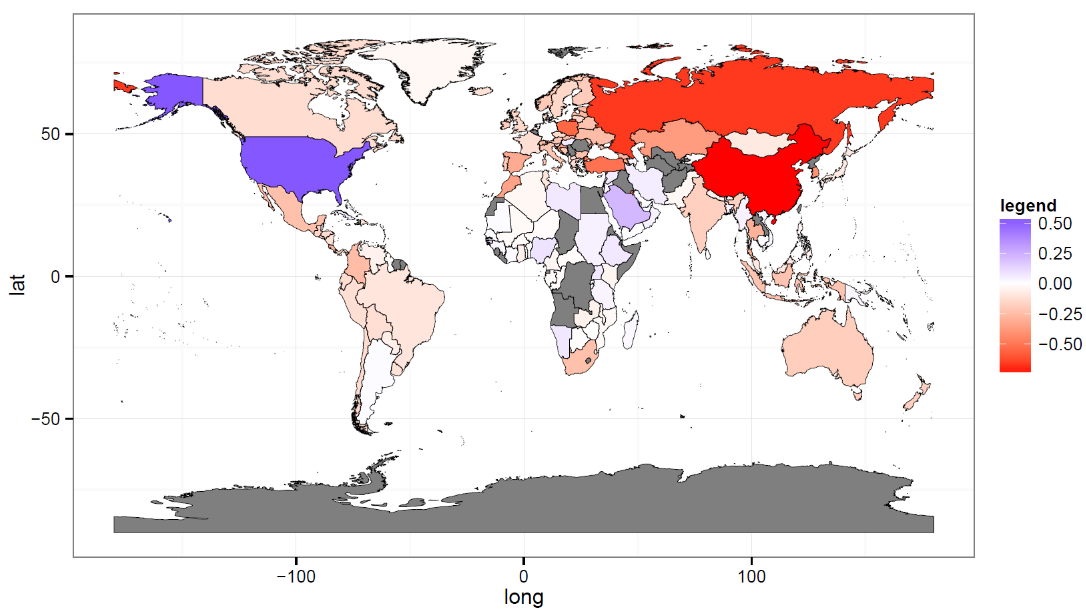
4.3 Preferential tariffs and trade shares

When controlling for a country of origin's aggregate exports of a product $ln(\tilde{x}_{opt}^{1997-2013})$, we find that for most countries in Africa the coefficients on non-apparel AGOA products (see Figure 2- 10a and Table 2- A.3) are smaller than those in Figure 2-7a. Interpreted through the lens of our discussion in Section 3, this is not surprising and indicates that, on average, AGOA affected both the *bilateral* trade of product *p* between country *o* and country *d* as well as country *o*'s *aggregate* exports of product *p*, where the latter effect was positive on average for AGOA products. Moreover, we find that diversion effects were insignificant for most African countries.

For advanced economies such as Japan and several European countries, the results are less straightforward to interpret. While the signs of coefficients remain negative when controlling for the country of origin's total exports, the coefficients are closer to zero. This is surprising, since we now control for any positive spillovers. The same is true for apparel products (see Figure 2- 10b), where we see more positive coefficients than in Figure 2- 7b.



(a) AGOA (non-apparel)

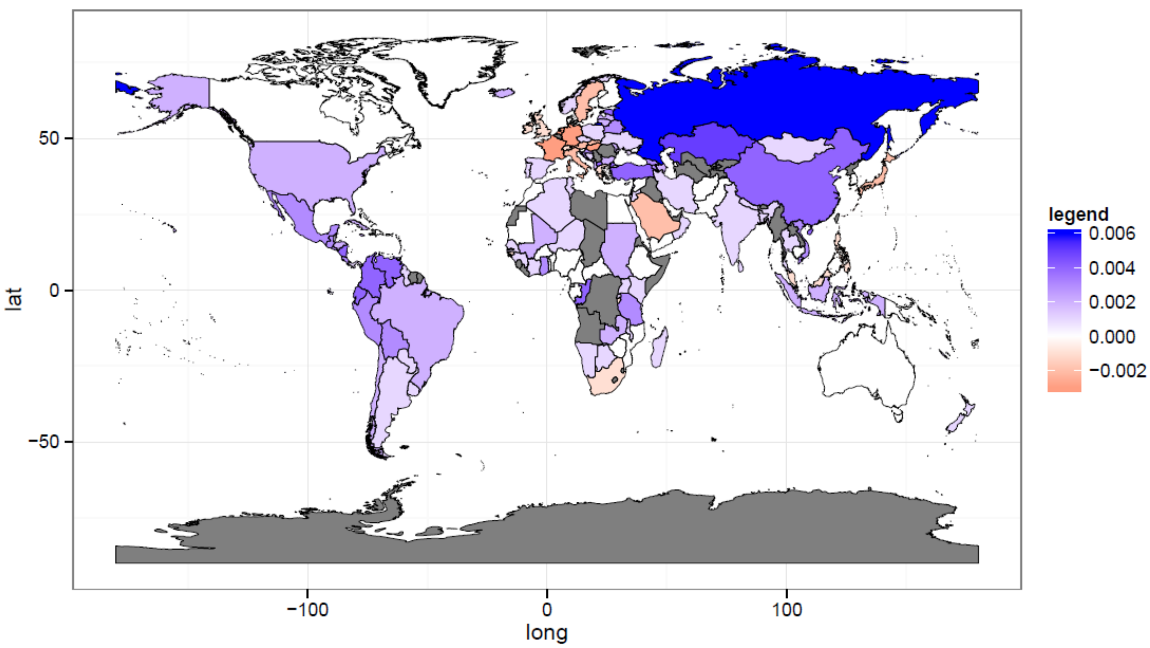


(b) AGOA (apparel)

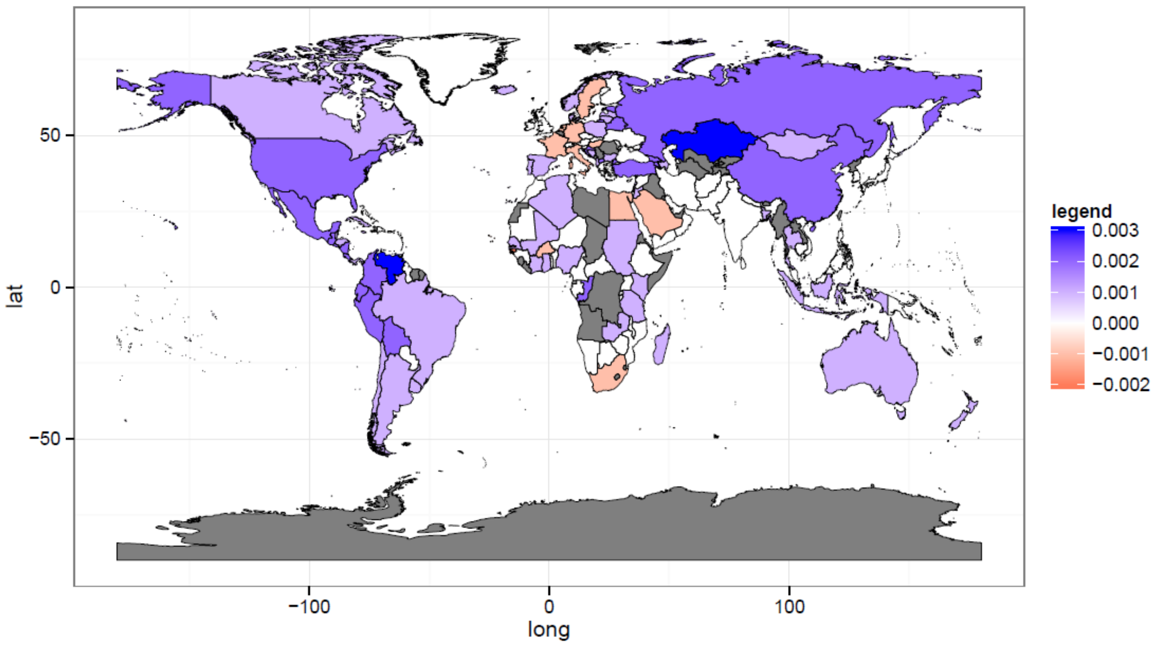
Figure 2- 7: Coefficient estimates: preferential tariffs and trade flows

5 Conclusion

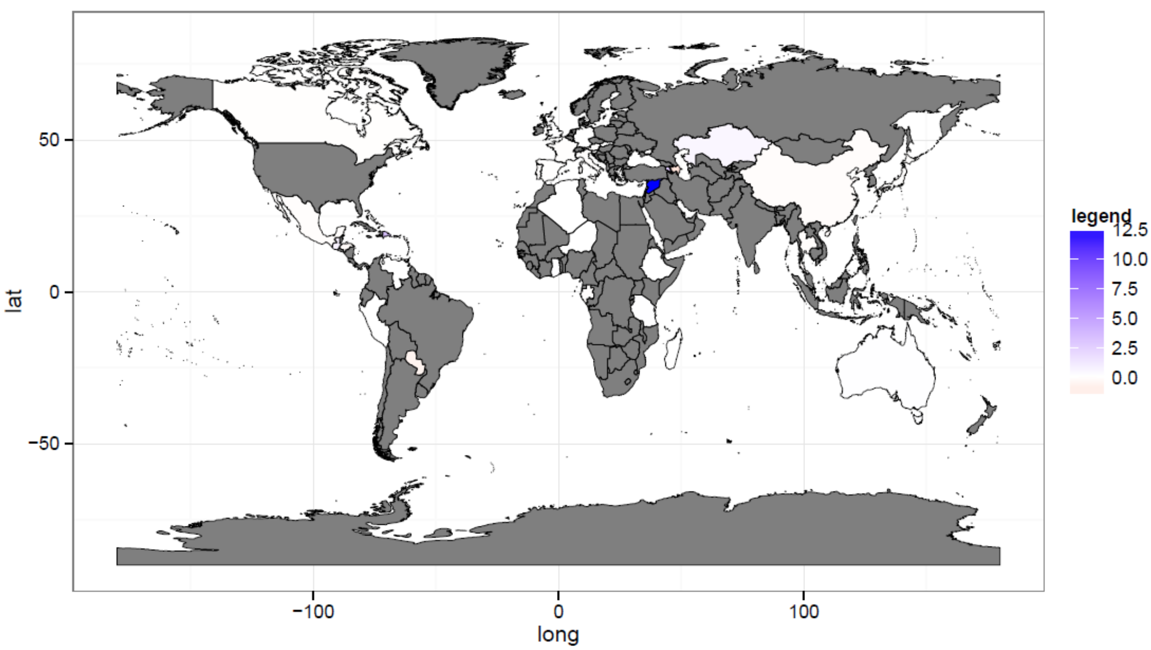
The analysis in this paper has shown that, as suggested by textbook models of trade, the asymmetric tariff reductions granted through AGOA have led to some trade diversion. However, this redirection of trade was of little consequence for intra-African trade. We have also shown that AGOA had a significant impact on the product composition of African exports. By taking these spillovers into account, policy makers in advanced economies can influence the speed and direction of structural change in Africa.



(a) positive exports

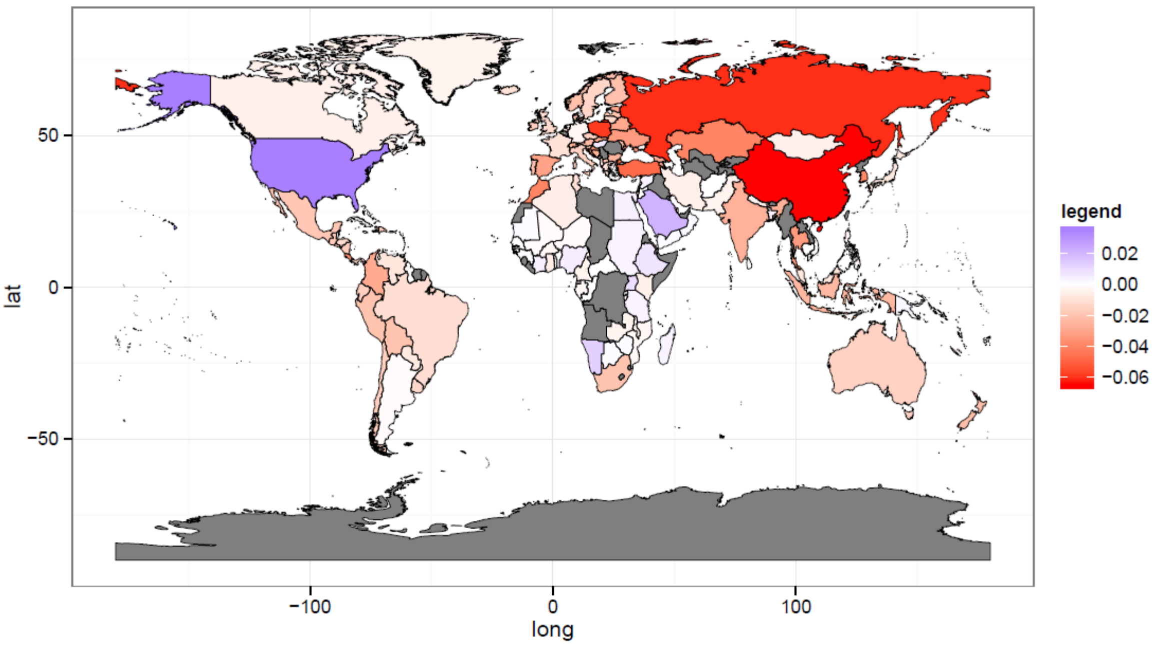


(b) entry

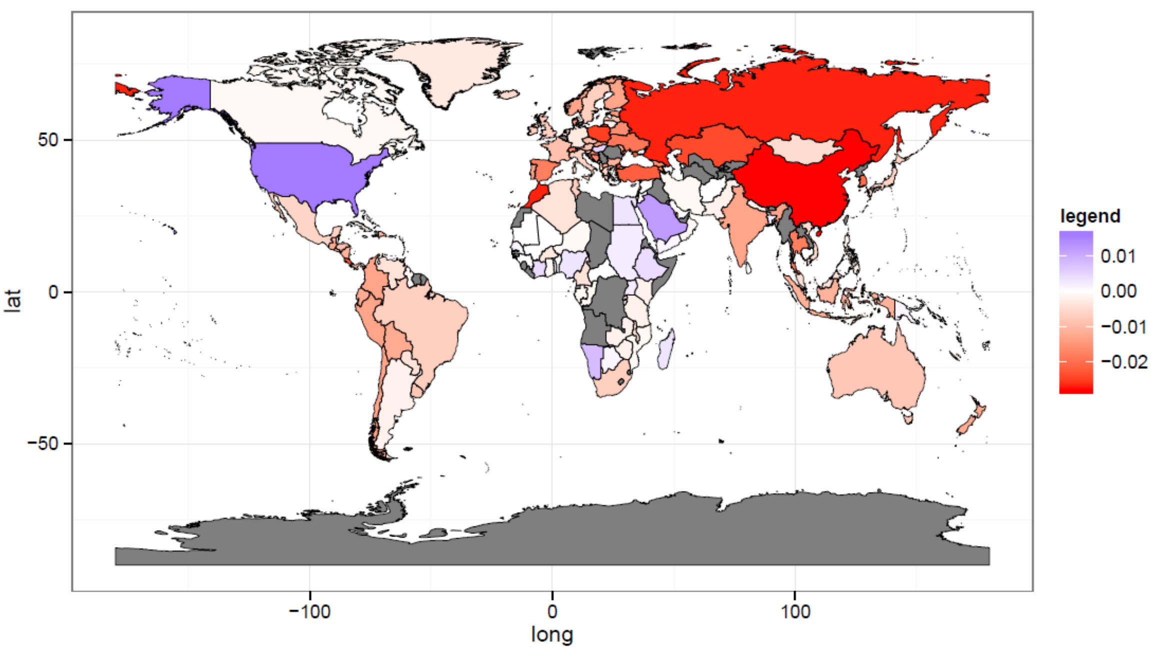


(c) exit

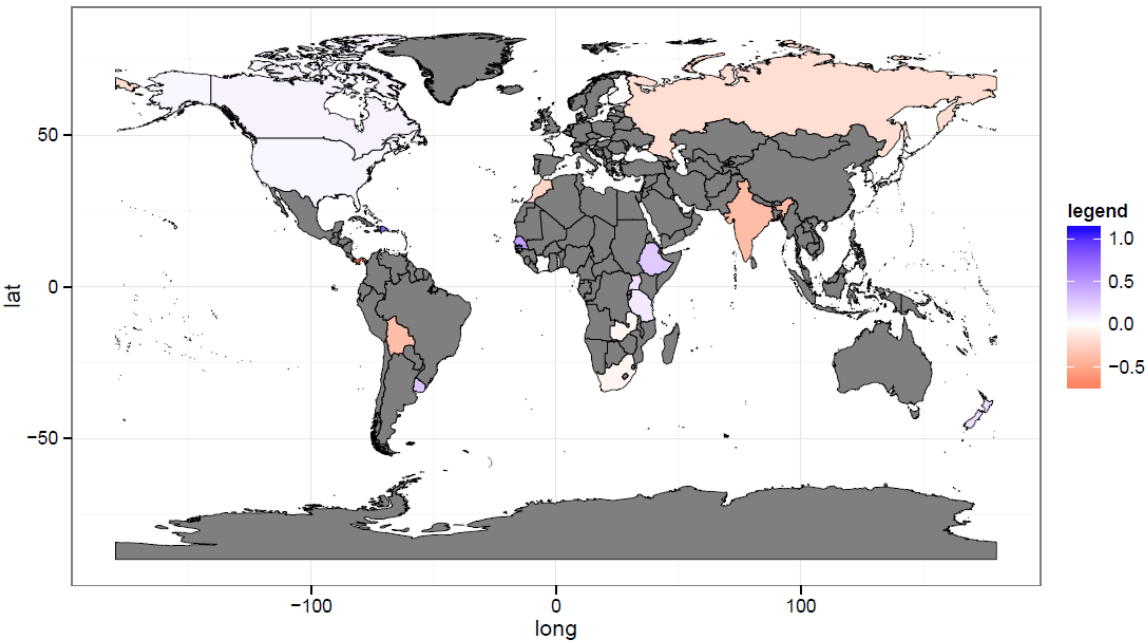
Figure 2- 8: Coefficient estimates: preferential tariffs and the extensive margin of trade (non-apparel)



(a) positive exports

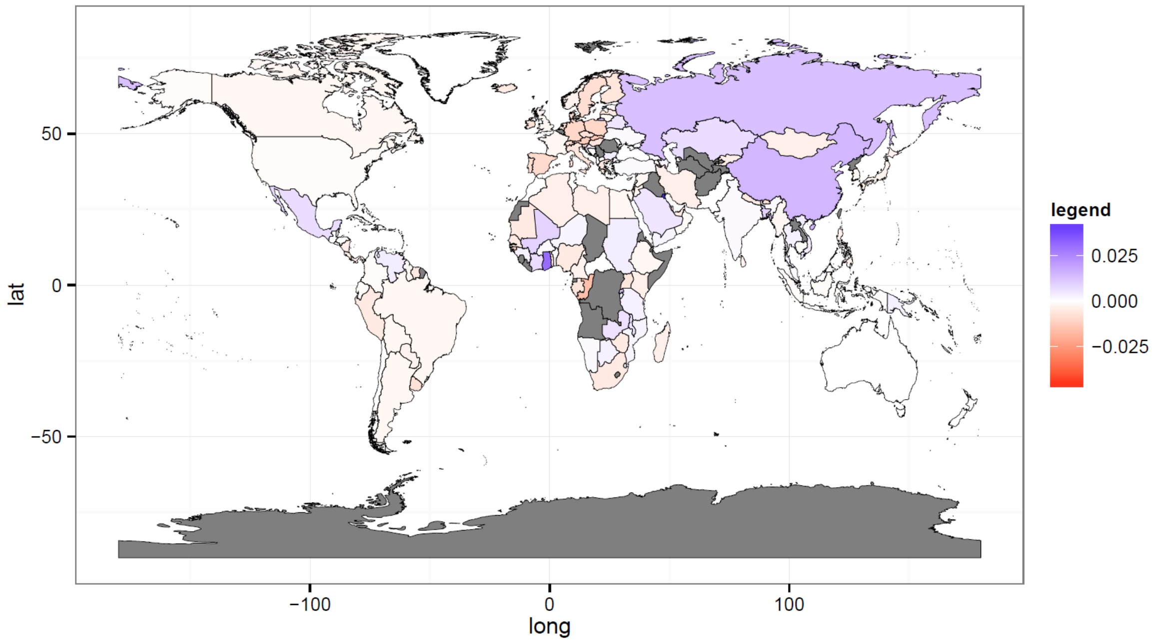


(b) entry

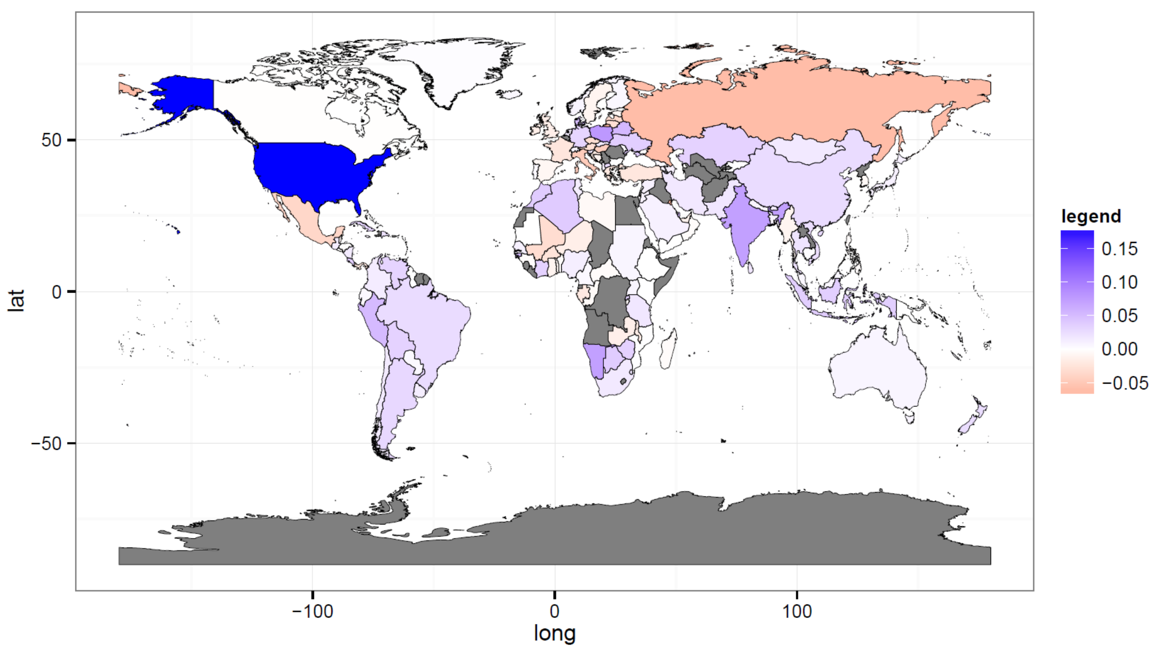


(c) exit

Figure 2- 9: Coefficient estimates: preferential tariffs and the extensive margin of trade (apparel)



(a) AGOA (non-apparel)



(b) AGOA (apparel)

Figure 2- 10: Coefficient estimates: preferential tariffs and trade shares

References

Baldwin, R. E., & Murray, T. (1977). "MFN tariff reductions and developing country trade benefits under the GSP." *The Economic Journal*, 30-46.

Bhagwati, J. (2008). *Termites in the trading system: How preferential agreements undermine free trade.* Oxford University Press. Chicago.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. Journal of the American Statistical Association, 83(401), 123-127.

Clausing, K. A. (2001). "Trade creation and trade diversion in the Canada-United States free trade agreement." *Canadian Journal of Economics*, 677-696.

De Loecker, Jan. 2013. "Detecting Learning by Exporting." American Economic Journal: Microeconomics, 5(3): 1-21.

Edwards, L., & Lawrence, R. (2014). "AGOA rules: the intended and unintended consequences of special fabric provisions." *African Successes: Modernization and Development.* University of Chicago Press.

Frankel, J. (2014). "Mauritius: African success story." *African Successes: Sustainable Growth.* University of Chicago Press. Chicago

Frazer, G., & Van Biesebroeck, J. (2010). "Trade growth under the African growth and opportunity act." *The Review of Economics and Statistics,* 92(1), 128-144.

Halvorsen, Robert and Raymond Palmquist *The American Economic Review* Vol. 70, No. 3 (Jun., 1980), pp. 474-475

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106(26), 10570-10575.

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

Holden, M., & McMillan, L. (2006). "Do free trade agreements create trade for South Africa?" *Economic Research Southern Africa.*

Krueger, A. O. (1999). "Trade creation and trade diversion under NAFTA" (No. w7429). National Bureau of Economic Research.

MacKinnon, J. G., & Magee, L. (1990). Transforming the dependent variable in regression models. International Economic Review, 315-339.

McMillan, Margaret S., and Dani Rodrik. (2011) “Globalization, structural change and productivity growth.” (No. w17143). National Bureau of Economic Research.

Nouve, Kofi, and John Staatz. “The African Growth and Opportunity Act and the Latent Agricultural Export Response in Sub-Saharan Africa.” Annual Meeting of the American Agricultural Economics Association, Montreal, Quebec, Canada, July. N.p., 2003. 27–30.

Rotunno, Lorenzo, Pierre-Louis Vézina, and Zheng Wang. "The rise and fall of (Chinese) African apparel exports." Journal of Development Economics 105 (2013): 152-163.

Sandrey, R. (2006) “Trade Creation and Trade Diversion resulting from SACU trading agreements.” Working Paper No 11. Stellenbosch:

Viner, J. (1950). The custom unions issue. Carnegie Endowment for World Peace. New York.

A Tables

Table 2-A 1: Import Data Summary Statistics, by Importer

Country	Min: Year	Max: Year	N: $x_{odpt} > 0$
Aruba	2000	2013	88337
Albania	1996	2013	387233
Argentina	1996	2013	901756
Armenia	1997	2013	349844
Antigua and Barbuda	1999	2013	143078
Australia	1996	2013	1464686
Austria	1996	2013	1674500
Azerbaijan	1999	2013	307512
Belgium	1999	2013	1405185
Benin	1998	2013	153831
Burkina Faso	2001	2013	149616
Bulgaria	1996	2013	974817
Bahamas	1997	2013	159289
Belarus	1998	2013	657971
Belize	1998	2013	185893
Bolivia	1997	2013	505677
Brazil	1997	2014	1138396

Barbados	2000	2013	349160
Botswana	2000	2013	211843
Central African Republic	1997	2013	35175
Canada	1996	2013	2021374
Switzerland	1996	2013	1615157
Chile	1997	2013	963086
China	1996	2013	1690627
Cote d’Ivoire	2001	2013	366888
Cameroon	2000	2012	337909
Colombia	1996	2013	895067
Cabo Verde	1997	2013	214964
Costa Rica	1997	2013	762977
Cyprus	1996	2013	726140
Czech Republic	1996	2013	1516918
Germany	1996	2013	2460386
Dominica	1999	2012	133812
Denmark	1996	2013	1296273
Dominican Republic	2001	2013	516297
Algeria	1996	2013	728515
Ecuador	1996	2013	656641

Spain	1996	2013	1801144
Estonia	1996	2013	960900
Ethiopia	2001	2013	407171
Finland	1996	2013	1219590
France	1996	2013	2185653
Faroe Islands	1996	2009	306809
Gabon	1997	2009	189002
United Kingdom	1996	2013	2096269
Georgia	1998	2013	448345
Gambia	1996	2013	144801
Greece	1996	2013	1051610
Grenada	2000	2009	119604
Greenland	1996	2013	153090
Guatemala	1997	2013	626631
Guyana	1997	2013	214522
Hong Kong	1996	2013	1075948
Honduras	1997	2012	392197
Croatia	1997	2013	1054499
Hungary	1996	2013	1089324
Indonesia	1996	2013	1093136

India	1996	2013	1320410
Ireland	1996	2013	1005005
Iran	1997	2011	304526
Iceland	1997	2013	760963
Israel	1996	2013	958322
Italy	1996	2013	1994868
Jamaica	1998	2013	354970
Jordan	1998	2013	472091
Japan	1996	2013	1252957
Kazakhstan	1998	2013	736762
Kenya	1997	2013	510046
Kyrgyzstan	2000	2013	211553
Cambodia	2000	2013	183280
Saint Kitts and Nevis	1999	2011	127378
Korea, Republic of	1996	2013	1290909
Lebanon	1997	2013	779545
Sri Lanka	1999	2013	641312
Lithuania	1997	2013	894155
Luxembourg	1999	2013	550454
Latvia	1997	2013	774038

Macao	1996	2012	288342
Moldova, Republic of	2000	2013	450495
Madagascar	1996	2013	415744
Maldives	1997	2013	339384
Mexico	1996	2013	1423617
Macedonia	1996	2013	662964
Mali	1998	2012	177161
Malta	1996	2013	515263
Mongolia	1997	2013	156513
Mozambique	2001	2013	270940
Mauritania	2000	2013	79865
Montserrat	1999	2013	48677
Mauritius	1997	2013	573311
Malawi	1999	2013	223183
Malaysia	1997	2013	1054520
Mayotte	2000	2009	134972
Namibia	2000	2013	214878
New Caledonia	2001	2013	400449
Niger	1998	2013	192014
Nigeria	1999	2013	395627

Nicaragua	1997	2013	456238
Netherlands	1996	2013	1659269
Norway	1996	2014	1522331
Nepal	1998	2013	118332
New Zealand	1996	2013	1147968
Oman	2000	2013	392582
Panama	1998	2013	449291
Peru	1998	2013	763166
Philippines	2000	2013	639572
Papua New Guinea	2001	2012	125592
Poland	1996	2013	1258822
Portugal	1996	2013	1081196
Paraguay	1998	2014	386954
French Polynesia	1996	2013	443496
Qatar	2000	2013	465732
Romania	1997	2013	1181402
Russian Federation	1997	2013	1517826
Rwanda	2001	2013	184522
Saudi Arabia	1999	2013	849153
Sudan	1999	2011	246707

Senegal	1996	2013	380408
Serbia	2000	2013	754673
Singapore	1997	2013	1403708
El Salvador	1997	2013	530798
Sao Tome and Principe	1999	2013	50638
Slovakia	1997	2013	1078971
Slovenia	1996	2013	1213759
Sweden	1996	2013	1422442
Syrian Arab Republic	2001	2010	124107
Turks and Caicos	1999	2012	19184
Togo	1998	2013	109412
Thailand	1999	2013	1184924
Trinidad and Tobago	1999	2010	295975
Tunisia	2000	2013	577592
Turkey	1996	2013	1275049
Tanzania	1997	2013	601849
Uganda	1996	2013	456473
Ukraine	2001	2013	813656
Uruguay	1997	2013	552950
United States	1996	2013	2238378

Saint Vincent	1998	2012	198737
Venezuela	1996	2013	829753
Viet Nam	2000	2013	650556
Samoa	2001	2013	85310
South Africa	1997	2013	1509265
Zambia	1997	2013	367290
Zimbabwe	2001	2013	263962

Notes: Data are from UN Comtrade. "N:$x_{odpt} > 0$" indicates the total number of positive observations the six-digit product level by an importer.

Table 2-A 2: Results: Effect of preferential tariffs on trade flows (by importing country and program).

Importing country	**AGOA**	**AGOA (apparel)**	**Observations**
Albania	0.02**	-0.064**	18290232
	(12.736)	(-23.182)	
Algeria	0.02**	-0.029**	16961274
	(8.694)	(-6.478)	
Andorra	-0.001	-0.132**	6705972
	(-0.367)	(-18.257)	
Antigua and Barbuda	0.005**	0.002	8659910
	(2.579)	(0.396)	
Argentina	0.016**	0.011*	17994636
	(7.306)	(2.154)	
Armenia	0.031**	-0.281**	14789376
	(17.659)	(-56.222)	
Aruba	-0.004	-	1659600
	(-0.153)		
Australia	0.012**	-0.187**	20216196
	(4.204)	(-20.878)	
Austria	-0.023**	-0.219**	21067542

	(-8.765)	(-20.931)	
Azerbaijan	0.025**	0.027**	13510080
	(8.78)	(9.568)	
Bahamas, The	-0.001	0.026**	9771396
	(-0.462)	(6.251)	
Bahrain	-0.001	-0.114**	10032260
	(-0.426)	(-7.587)	
Bangladesh	0.008**	0.033**	11009600
	(3.299)	(5.697)	
Barbados	0.007**	0.026**	15027320
	(4.747)	(5.689)	
Belarus	0.039**	-0.227**	17551536
	(19.574)	(-49.49)	
Belgium-Luxembourg	-0.035**	-0.169**	21502638
	(-11.086)	(-16.927)	
Belize	0.009**	-0.019**	13283328
	(6.789)	(-7.039)	
Benin	0.004	-0.004	15501600
	(1.12)	(-0.856)	
Bhutan	0.007	-0.057**	3178170

	(1.342)	(-5.413)	
Bolivia	0.031**	-0.12**	18868725
	(20.876)	(-37.211)	
Bosnia and Herzegovina	0.003	-0.209**	12110461
	(1.887)	(-12.274)	
Botswana	0.011**	-0.007	13002444
	(4.235)	(-0.668)	
Brazil	0.023**	-0.097**	19475098
	(9.506)	(-17.527)	
Brunei	0.005	-0.038**	3903012
	(1.465)	(-4.64)	
Bulgaria	0.03**	-0.229**	20948400
	(15.157)	(-50.817)	
Burkina Faso	-0.001	-0.012	9235776
	(-0.071)	(-1.376)	
Burundi	-0.001	-0.003	8446482
	(-0.513)	(-0.534)	
Cambodia	-0.001	0.01**	12949104
	(-0.359)	(5.199)	
Cameroon	0.005	-0.031**	12915240

	(1.815)	(-4.95)	
Canada	0.019**	-0.129**	20363382
	(6.298)	(-10.543)	
Cape Verde	-0.009**	-0.009*	14524392
	(-4.924)	(-2.313)	
Central African Republic	-0.004	-0.009*	5974500
	(-1.405)	(-2.291)	
Chile	0.027**	-0.132**	17379236
	(11.611)	(-19.728)	
China	0.078**	-0.727**	19526472
	(25.214)	(-89.313)	
Colombia	0.045**	-0.251**	20389824
	(24.349)	(-51.558)	
Congo, Rep.	0.045**	-0.011**	5070338
	(3.008)	(-2.749)	
Costa Rica	0.029**	-0.282**	19375461
	(17.513)	(-64.342)	
Cote d'Ivoire	0.011*	0.014	12690612
	(2.427)	(1.66)	
Croatia	0.035**	-0.336**	19909890

	(17.062)	(-47.325)	
Cuba	0.027**	0.05**	7552192
	(8.637)	(10.158)	
Cyprus	-0.01**	0.04**	19269684
	(-5.4)	(6.836)	
Czech Republic	0.01**	-0.139**	21469086
	(4.105)	(-17.768)	
Denmark	0.008**	-0.149**	20997468
	(3.673)	(-19.556)	
Dominica	0.003	-0.006*	8477105
	(1.789)	(-2.272)	
Dominican Republic	0.004**	-0.117**	14962402
	(2.813)	(-16.719)	
East Timor	0.003	-	1085472
	(0.325)		
Ecuador	0.046**	-0.126**	19985706
	(26.298)	(-36.532)	
Egypt, Arab Rep.	-0.005	-	5229642
	(-1.335)		
El Salvador	0.044**	-0.151**	16796000

	(24.476)	(-44.229)	
Estonia	0.037**	-0.251**	20691504
	(19.785)	(-46.375)	
Ethiopia(excludes Eritrea)	0.005*	0.071**	19632620
	(2.183)	(15.3)	
Faeroe Islands	0.005**	-0.076**	14619150
	(3.6)	(-20.71)	
Fiji	-0.001	0.001	11923728
	(-0.726)	(0.149)	
Finland	-0.009**	-0.195**	20942712
	(-4.294)	(-27.584)	
France	-0.031**	-0.118**	21305592
	(-7.165)	(-9.188)	
French Polynesia	0.015**	-0.05**	17763480
	(11.065)	(-11.261)	
Gabon	-0.008*	-0.012*	10305750
	(-2.152)	(-1.96)	
Gambia, The	-0.003	0.012**	15161580
	(-1.254)	(4.926)	
Georgia	0.041**	-0.263**	13969440

	(19.575)	(-60.221)	
Germany	-0.043**	0.002	21309750
	(-11.538)	(0.192)	
Ghana	0.037**	-0.046**	10704280
	(4.069)	(-3.641)	
Greece	-0.011**	-0.122**	20603520
	(-4.557)	(-22.947)	
Greenland	0	-0.026**	14771160
	(0.033)	(-9.421)	
Grenada	0.011**	-0.01**	7527790
	(5.365)	(-3.196)	
Guatemala	0.038**	-0.152**	18663399
	(22.212)	(-46.057)	
Guinea	0.011**	0.021**	7783680
	(3.304)	(4.188)	
Guinea-Bissau	-0.124*	0.116**	213759
	(-2.421)	(3.873)	
Guyana	0.005**	-0.003	16664760
	(4.124)	(-1.402)	
Honduras	0.027**	-0.132**	13976550

	(13.29)	(-40.287)	
Hong Kong, China	-0.014**	-0.246**	17962560
	(-5.232)	(-33.292)	
Hungary	-0.043**	0.058**	19502208
	(-18.859)	(10.172)	
Iceland	0.017**	-0.11**	14475602
	(7.973)	(-13.625)	
India	0.033**	-0.178**	19922868
	(10.447)	(-38.368)	
Indonesia	0.017**	-0.22**	20639808
	(6.592)	(-47.347)	
Iran, Islamic Rep.	0.029**	0.054**	9091620
	(7.755)	(17.64)	
Ireland	-0.009**	-0.182**	21243222
	(-3.814)	(-18.983)	
Israel	0.016**	-0.109**	18418752
	(6.17)	(-19.941)	
Italy	-0.03**	-0.215**	20987856
	(-8.781)	(-20.097)	
Jamaica	0.023**	-0.007	14040640

	(11.039)	(-1.838)	
Japan	-0.021**	-0.072**	19836576
	(-8.116)	(-7.276)	
Jordan	0.022**	0.021**	12364448
	(7.749)	(4.02)	
Kazakhstan	0.05**	-0.368**	16779870
	(21.608)	(-52.11)	
Kenya	0.025**	-0.037**	15985800
	(7.622)	(-5.527)	
Kiribati	0.003	0.042**	2245470
	(0.347)	(4.607)	
Korea, Rep.	0.004	-0.397**	21176100
	(1.415)	(-51.018)	
Kuwait	0.007	-0.467**	4060640
	(0.951)	(-3.973)	
Kyrgyz Republic	0.005**	-0.007*	10882704
	(2.613)	(-2.237)	
Latvia	0.019**	-0.261**	15961776
	(8.9)	(-57.217)	
Lebanon	0.006**	-0.187**	19436610

	(3.066)	(-21.148)	
Libya	-0.008	0.068**	1787208
	(-0.765)	(4.039)	
Lithuania	0.029**	-0.142**	17389776
	(13.511)	(-32.706)	
Macao	-0.017**	-0.237**	14653711
	(-10.582)	(-56.034)	
Macedonia, FYR	0.043**	-0.135**	19594080
	(26.301)	(-35.471)	
Madagascar	0.01**	0.02**	18800460
	(4.025)	(3.96)	
Malawi	0.016**	-0.028**	14079912
	(4.592)	(-3.276)	
Malaysia	-0.009**	-0.08**	19734926
	(-3.457)	(-19.095)	
Maldives	-0.031**	0.087**	6041545
	(-3.939)	(9.784)	
Mali	0.028**	-0.018**	12114760
	(6.085)	(-2.766)	
Malta	-0.008**	0.023**	18320850

	(-4.937)	(4.757)	
Mauritania	-0.003	0.013**	11063808
	(-1.207)	(3.471)	
Mauritius	0.003	-0.006	18527960
	(1.245)	(-0.812)	
Mayotte	0	0.031**	8125110
	(0.019)	(3.289)	
Mexico	0.056**	-0.235**	20447316
	(23.221)	(-28.173)	
Micronesia, Fed. Sts.	-0.004	0.022**	5302530
	(-1.63)	(4.799)	
Moldova	0.02**	-0.178**	14536368
	(10.746)	(-34.437)	
Mongolia	0.038**	-0.078**	6852329
	(14.548)	(-15.165)	
Montenegro	-0.003	-0.146**	7905744
	(-1.566)	(-3.893)	
Montserrat	0.005*	0.009**	4439968
	(2.428)	(3.038)	
Morocco	0.004	-0.343**	12065508

	(1.683)	(-21.56)	
Mozambique	0.002	-0.021**	13541775
	(1.152)	(-2.996)	
Myanmar	-0.005	0.037**	910320
	(-0.74)	(2.903)	
Namibia	0.006*	0.072**	14216328
	(2.239)	(8.13)	
Nepal	0.024**	-0.05**	8645148
	(12.792)	(-13.424)	
Netherlands	-0.041**	-0.04**	21167820
	(-12.614)	(-4.237)	
Netherlands Antilles	0	-0.073**	3836700
	(-0.032)	(-6.792)	
New Caledonia	0.006**	-0.073**	13477620
	(4.543)	(-9.309)	
New Zealand	0.015**	-0.171**	20766636
	(7.082)	(-26.9)	
Nicaragua	0.036**	-0.111**	18478575
	(25.12)	(-38.415)	
Niger	0.006	-0.009	15009072

	(1.889)	(-1.386)	
Nigeria	0.004	0.08**	14610401
	(0.99)	(15.573)	
Norway	0.018**	-0.202**	20144520
	(7.568)	(-25.718)	
Occ.Pal.Terr.	-0.011*	-0.009*	4430706
	(-2.268)	(-2.449)	
Oman	0.009**	0.039**	11911900
	(3.489)	(7.395)	
Pakistan	0.003	-0.026**	12329328
	(1.26)	(-4.143)	
Palau	-0.002	0.007	3306940
	(-0.754)	(0.518)	
Panama	0.013**	-0.162**	10401105
	(3.783)	(-18.422)	
Papua New Guinea	0	0.027**	5757288
	(0.154)	(6.123)	
Paraguay	0.01**	-0.061**	13815360
	(5.804)	(-17.878)	
Peru	0.032**	-0.121**	18446736

	(15.336)	(-25.167)	
Philippines	-0.009**	-0.035**	15185450
	(-3.406)	(-8.93)	
Poland	0.005	-0.546**	21035340
	(1.839)	(-68.353)	
Portugal	0.001	-0.202**	21134700
	(0.473)	(-27.519)	
Qatar	0.019**	-0.21**	8998770
	(4.731)	(-17.222)	
Romania	0.003	-0.334**	19874530
	(1.569)	(-63.893)	
Russian Federation	0.083**	-0.663**	19713336
	(29.34)	(-73.866)	
Rwanda	0.014**	0.011	12269904
	(3.386)	(1.308)	
Samoa	0.003**	-0.001	6984055
	(2.721)	(-0.174)	
Sao Tome and Principe	0.004	0.009	7050000
	(0.91)	(1.809)	
Saudi Arabia	-0.02**	0.221**	16011735

	(-5.557)	(26.173)	
Senegal	0.018**	0.011**	16736184
	(7.358)	(3.298)	
Seychelles	-0.012*	-0.032**	6429680
	(-2.507)	(-3.034)	
Singapore	-0.012**	-0.179**	19394144
	(-4.29)	(-19.43)	
Slovak Republic	-0.028**	-0.24**	20161490
	(-13.714)	(-38.652)	
Slovenia	-0.008**	-0.158**	21130956
	(-4.034)	(-22.067)	
South Africa	0.002	-0.237**	19968370
	(0.638)	(-16.921)	
Spain	0.002	-0.313**	21188520
	(0.649)	(-30.655)	
Sri Lanka	0.005*	0.012**	16549830
	(2.482)	(3.824)	
St. Kitts and Nevis	0.007**	-0.004	10182952
	(5.401)	(-1.675)	
St. Lucia	0.013**	-0.014**	6953040

	(5.437)	(-3.462)	
St. Vincent and the Grenadines	0.007**	0.002	14165010
	(5.969)	(0.686)	
Sudan	0.019**	0.042**	15206730
	(8.398)	(15.602)	
Suriname	-0.003	-	1175490
	(-0.17)		
Swaziland	-0.002	-0.034*	3347224
	(-0.367)	(-2.465)	
Sweden	-0.024**	-0.147**	20991960
	(-10.292)	(-19.387)	
Switzerland	-0.011**	-0.256**	20996064
	(-3.865)	(-28.399)	
Syrian Arab Republic	-0.002	0.061**	7532330
	(-0.926)	(22.153)	
Tanzania	0.036**	0.033**	19300950
	(10.431)	(3.847)	
Thailand	0.016**	-0.288**	17595000
	(5.15)	(-39.945)	
Togo	0	0.001	13931925

	(-0.138)	(0.296)	
Tonga	-0.036**	-	2487550
	(-3.462)		
Trinidad and Tobago	0.018**	-0.019**	12171324
	(7.764)	(-5.247)	
Tunisia	0.01**	-0.089**	13938358
	(4.341)	(-12.782)	
Turkey	0.053**	-0.512**	21056040
	(21.194)	(-65.97)	
Turks and Caicos Isl.	-0.005	-0.005	1603264
	(-1.44)	(-1.426)	
Uganda	0.009**	0.068**	20023956
	(3.086)	(9.204)	
Ukraine	0.005**	-0.249**	14784510
	(2.851)	(-28.027)	
United Arab Emirates	-0.02**	0.02	2660310
	(-2.706)	(0.92)	
United Kingdom	-0.007	-0.14**	21389472
	(-1.941)	(-10.588)	
United States	0.036**	0.522**	20287080

	(9.466)	(27.768)	
Uruguay	0.008**	-0.108**	17068748
	(4.262)	(-24.281)	
Vanuatu	0	0.001	2754000
	(-0.059)	(0.107)	
Venezuela	0.054**	-0.053**	20259720
	(27.413)	(-13.817)	
Vietnam	0.039**	0.029**	15094884
	(14.859)	(6.424)	
Yemen	0.001	0.006*	9632950
	(0.422)	(2.027)	
Yugoslavia	0.036**	-0.375**	14819922
	(12.826)	(-43.084)	
Zambia	0.029**	-0.049**	17466820
	(9.05)	(-5.476)	
Zimbabwe	-0.002	0.004	12151872
	(-1.216)	(0.533)	

Note: t-statistics in parentheses; *significant at 5%; **significant at 1%; standard errors robust to heteroskedasticity. Dependent variable is the log of import value at the six digit product category. Each row represents the results from a separate regression estimating the impact of tariff programs on a country's imports, where the impact of the program is estimated via a triple difference specification comparing differences in product coverage, country coverage, and timing of the program; Fixed effects at the exporter-product, product-year, and exporter-year level are included in all specifications.

Table 2-A 3: Results: Effect of preferential tariffs on trade shares (by importing country and program).

Importing country	AGOA	AGOA (apparel)	$ln(\tilde{x}_{opt}{}^{1998-2010})$	Observations
Albania	0	0.01**	0.525**	13209612
	(-0.51)	(8.775)	(1439.053)	
Algeria	-0.004**	0.039**	0.565**	12249809
	(-4.362)	(22.098)	(1748.06)	
Andorra	-0.004**	-0.007**	0.448**	5215756
	(-4.154)	(-2.738)	(756.543)	
Antigua and Barbuda	0.001	0.013**	0.425**	6061937
	(0.931)	(4.882)	(675.379)	
Argentina	-0.002*	0.029**	0.548**	12996126
	(-2.245)	(14.376)	(1713.927)	
Armenia	-0.003**	0.01**	0.479**	11092032
	(-3.22)	(6.902)	(1322.838)	
Aruba	-0.049**	-	0.46**	829800
	(-3.992)		(438.22)	
Australia	0	0.007*	0.651**	14600586
	(-0.001)	(2.462)	(2973.481)	
Austria	-0.012**	-0.028**	0.606**	15215447
	(-11.363)	(-6.657)	(2390.843)	
Azerbaijan	0.002	0.009**	0.522**	10808064

	(1.787)	(9.001)	(1292.917)	
Bahamas, The	-0.001	0.005**	0.469**	7472244
	(-0.834)	(3.008)	(699.193)	
Bahrain	-0.001	-0.012**	0.546**	9029034
	(-0.834)	(-4.474)	(1625.385)	
Bangladesh	0.009**	0.008*	0.519**	9908640
	(4.166)	(2.044)	(1120.009)	
Barbados	0.002**	0.019**	0.438**	11807180
	(2.678)	(7.222)	(1089.93)	
Belarus	0.005**	0.053**	0.549**	14260623
	(7.53)	(40.663)	(1968.206)	
Belgium-Luxembourg	-0.016**	0.035**	0.663**	15529683
	(-14.637)	(11.296)	(2785.501)	
Belize	-0.001*	-0.006**	0.447**	10792704
	(-2.122)	(-4.169)	(803.747)	
Benin	0.001	-0.002	0.526**	12595050
	(0.567)	(-1.099)	(1000.444)	
Bhutan	-0.008**	0.014**	0.431**	2471910
	(-3.484)	(4.182)	(396.661)	
Bolivia	-0.002**	0.033**	0.496**	14429025
	(-3.511)	(24.938)	(1440.682)	
Bosnia and Herzegovina	0	-0.01*	0.477**	8807608

	(-0.312)	(-2.572)	(1232.788)	
Botswana	0.003	0.038**	0.455**	10216206
	(1.087)	(4.908)	(777.22)	
Brazil	-0.002*	0.026**	0.604**	14892722
	(-2.234)	(12.147)	(2019.517)	
Brunei	-0.006*	0.004	0.463**	2602008
	(-2.073)	(1.121)	(641.283)	
Bulgaria	0.004**	0.001	0.572**	15129400
	(5.481)	(0.686)	(2399.762)	
Burkina Faso	0.001	-0.025**	0.512**	7556544
	(0.083)	(-3.363)	(969.2)	
Burundi	0.001	-0.005	0.492**	6142896
	(0.833)	(-1.445)	(608.329)	
Cambodia	-0.001	0.002	0.502**	10174296
	(-1.254)	(1.7)	(922.835)	
Cameroon	-0.003	0.005	0.48**	10928280
	(-1.352)	(1.349)	(1307.212)	
Canada	-0.002	-0.001	0.597**	14706887
	(-1.459)	(-0.107)	(2392.33)	
Cape Verde	-0.004**	0.01**	0.425**	11106888
	(-3.962)	(4.466)	(879.891)	
Central African Republic	0	-0.003	0.525**	4381300

	(-0.048)	(-0.822)	(564.081)	
Chile	0	0.015**	0.573**	13290004
	(0.036)	(6.561)	(2068.721)	
China	0.015**	0.026**	0.637**	14102452
	(12.711)	(8.286)	(2376.64)	
Colombia	-0.001	0.014**	0.564**	14725984
	(-1)	(6.956)	(1834.385)	
Congo, Rep.	-0.016	-0.004	0.499**	2897336
	(-1.909)	(-1.778)	(506.486)	
Costa Rica	-0.004**	0.035**	0.516**	14816529
	(-6.448)	(19.811)	(1595.459)	
Cote d'Ivoire	0.007*	0.033**	0.485**	9517959
	(2.125)	(5.518)	(1193.788)	
Croatia	0.004**	-0.019**	0.562**	15225210
	(5.023)	(-8.468)	(2463.965)	
Cuba	0.001	0.029**	0.54**	7552192
	(0.84)	(13.091)	(1127.541)	
Cyprus	-0.008**	0.009**	0.532**	13916994
	(-11.458)	(4.211)	(1959.406)	
Czech Republic	-0.007**	-0.03**	0.585**	15505451
	(-7.584)	(-9.746)	(2304.639)	
Denmark	-0.008**	0.049**	0.598**	15164838

	(-8.944)	(17.361)	(2222.143)	
Dominica	0.001	0.003*	0.413**	7825020
	(0.624)	(2.363)	(710.655)	
Dominican Republic	0.001	0.02**	0.502**	11509540
	(1.142)	(5.114)	(1250.786)	
East Timor	-	-	0.36**	723648
			(179.803)	
Ecuador	-0.001	0.005**	0.561**	14434121
	(-0.839)	(3.558)	(1721.662)	
Egypt, Arab Rep.	-0.002	-	0.623**	2614821
	(-0.423)		(902.847)	
El Salvador	0	-0.001	0.52**	12844000
	(-0.468)	(-0.708)	(1334.304)	
Estonia	-0.002*	-0.007**	0.53**	14943864
	(-2.415)	(-3.281)	(2014.581)	
Ethiopia(excludes Eritrea)	-0.002	0	0.533**	15013180
	(-1.929)	(-0.211)	(1488.028)	
Faeroe Islands	-0.002**	-0.004**	0.419**	12530700
	(-3.646)	(-3.013)	(1283.603)	
Fiji	0	0	0.484**	8942796
	(-0.435)	(0.226)	(861.912)	
Finland	-0.006**	0.006*	0.585**	15125292

	(-8.236)	(2.11)	(2230.221)	
France	-0.002	-0.023**	0.744**	15387372
	(-1.315)	(-5.795)	(3850.275)	
French Polynesia	-0.004**	0.001	0.468**	12829180
	(-8.358)	(0.364)	(1562.706)	
Gabon	-0.006*	-0.02**	0.529**	9513000
	(-2.33)	(-4.658)	(1312.225)	
Gambia, The	-0.001	-0.003**	0.479**	10950030
	(-0.458)	(-2.911)	(921.236)	
Georgia	0	0.004**	0.504**	11350170
	(-0.136)	(2.684)	(1445.138)	
Germany	-0.01**	0.031**	0.755**	15390375
	(-8.528)	(8.77)	(3864.219)	
Ghana	0.031**	-0.01	0.481**	7492996
	(3.211)	(-1.091)	(1053.762)	
Greece	-0.005**	-0.02**	0.637**	14880320
	(-6.256)	(-10.843)	(2511.673)	
Greenland	0	0.002**	0.449**	10668060
	(-0.656)	(2.853)	(946.147)	
Grenada	-0.002*	-0.001	0.428**	7527790
	(-2.058)	(-0.774)	(772.672)	
Guatemala	0	0.012**	0.567**	14272011

	(-0.381)	(9.036)	(1680.694)	
Guinea	0.003	-0.002	0.529**	7783680
	(1.664)	(-0.57)	(944.548)	
Guinea-Bissau	-0.01	0.05**	0.486**	213759
	(-0.456)	(3.202)	(146.536)	
Guyana	-0.002**	0	0.503**	12743640
	(-3.397)	(-0.111)	(982.634)	
Honduras	-0.001	-0.001	0.521**	11181240
	(-0.656)	(-0.616)	(1260.064)	
Hong Kong, China	-0.001	-0.015**	0.668**	12972960
	(-0.732)	(-5.06)	(2341.711)	
Hungary	-0.006**	-0.044**	0.64**	14084928
	(-7.722)	(-25.018)	(3085.042)	
Iceland	-0.005**	0.005	0.492**	11069578
	(-5.756)	(1.848)	(1945.298)	
India	0.001	0.072**	0.648**	14388738
	(0.935)	(39.663)	(2415.874)	
Indonesia	-0.001	0.036**	0.594**	14906528
	(-0.742)	(19.768)	(1868.602)	
Iran, Islamic Rep.	-0.004*	0.017**	0.669**	7576350
	(-2.03)	(13.681)	(1736.414)	
Ireland	-0.007**	-0.015**	0.602**	15342327

	(-7.725)	(-3.941)	(2363.099)	
Israel	-0.003**	0.003*	0.636**	13302432
	(-3.837)	(1.962)	(2854.68)	
Italy	-0.006**	-0.047**	0.731**	15157896
	(-6.164)	(-15.012)	(3527.362)	
Jamaica	-0.002**	0.008**	0.52**	11408020
	(-3.156)	(6.232)	(1282.247)	
Japan	-0.003**	0.01**	0.696**	14326416
	(-3.464)	(4.078)	(3185.766)	
Jordan	-0.003**	0.009**	0.599**	10046114
	(-3.421)	(6.214)	(2128.689)	
Kazakhstan	0.007**	0.033**	0.531**	13423896
	(6.627)	(20.187)	(1827.922)	
Kenya	-0.004**	0	0.554**	13854360
	(-2.735)	(-0.032)	(1723.522)	
Kiribati	-0.012	0.009*	0.407**	1571829
	(-0.492)	(2.029)	(379.718)	
Korea, Rep.	-0.002*	-0.005*	0.623**	15293850
	(-2.203)	(-2.059)	(2237.105)	
Kuwait	0.043*	-0.069**	0.574**	3045480
	(2.349)	(-3.942)	(989.843)	
Kyrgyz Republic	-0.004**	0	0.494**	8550696

	(-4.678)	(-0.428)	(1068.613)	
Latvia	-0.005**	-0.033**	0.563**	12206064
	(-6.474)	(-20.881)	(2069.963)	
Lebanon	-0.003**	0	0.565**	14863290
	(-4.24)	(0.065)	(2164.378)	
Libya	-0.004	-0.005	0.626**	1787208
	(-0.683)	(-0.783)	(898.562)	
Lithuania	-0.001	-0.029**	0.557**	13298064
	(-1.113)	(-15.586)	(2042.484)	
Macao	-0.004**	-0.019**	0.499**	11205779
	(-7.434)	(-15.447)	(1328.742)	
Macedonia, FYR	0	0.027**	0.492**	14151280
	(0.577)	(18.995)	(1590.781)	
Madagascar	-0.004**	-0.003	0.477**	13578110
	(-3.245)	(-1.413)	(1321.831)	
Malawi	0.005	-0.017**	0.487**	12068496
	(1.849)	(-2.913)	(952.713)	
Malaysia	0.001	0.006**	0.672**	15091414
	(1.689)	(5.466)	(2768.313)	
Maldives	-0.023**	0.026**	0.445**	4620005
	(-6.704)	(7.692)	(1068.532)	
Mali	0.009**	-0.031**	0.544**	10384080

	(2.602)	(-6.601)	(1095.435)	
Malta	-0.006**	0.014**	0.521**	13231725
	(-11.87)	(10.298)	(1894.002)	
Mauritania	-0.005*	0	0.566**	8692992
	(-2.373)	(-0.006)	(813.599)	
Mauritius	-0.006**	0.025**	0.516**	14168440
	(-5.055)	(10.275)	(1832.962)	
Mayotte	-0.002	0.01	0.42**	8125110
	(-0.805)	(1.73)	(941.223)	
Mexico	0.007**	-0.037**	0.612**	14767506
	(7.322)	(-12.843)	(2253.95)	
Micronesia, Fed. Sts.	-0.009**	0.005	0.421**	3711771
	(-3.876)	(1.794)	(516.314)	
Moldova	0	0.002	0.465**	11421432
	(0.2)	(1.527)	(1382.634)	
Mongolia	-0.004**	0.016**	0.479**	5606451
	(-4.646)	(13.321)	(967.864)	
Montenegro	-0.003	0.026	0.457**	4941090
	(-1.926)	(1.542)	(857.987)	
Montserrat	-0.001	0.001	0.39**	3415360
	(-0.442)	(0.519)	(451.977)	
Morocco	0	0.038**	0.541**	9049131

	(0.101)	(8.451)	(1412.313)	
Mozambique	0.002	0.001	0.485**	10416750
	(0.963)	(0.257)	(970.289)	
Myanmar	-0.002	-0.009	0.557**	910320
	(-0.493)	(-1.756)	(352.06)	
Namibia	0.001	0.072**	0.458**	11169972
	(0.246)	(9.88)	(796.887)	
Nepal	-0.005**	0.007**	0.558**	5763432
	(-5.046)	(4.284)	(780.514)	
Netherlands	-0.006**	0.012**	0.71**	15287870
	(-5.845)	(4.71)	(3580.934)	
Netherlands Antilles	-0.01**	-0.003	0.479**	3836700
	(-3.128)	(-0.499)	(586.766)	
New Caledonia	-0.001	0.005**	0.453**	10367400
	(-1.147)	(2.834)	(1329.605)	
New Zealand	0	0.025**	0.543**	14998126
	(-0.599)	(9.741)	(2068.673)	
Nicaragua	-0.004**	0.015**	0.48**	14130675
	(-5.573)	(12.97)	(1269.962)	
Niger	0.002	-0.015**	0.489**	12194871
	(1.21)	(-4.772)	(889.156)	
Nigeria	-0.005	0.012**	0.656**	11238770

	(-1.51)	(4.845)	(1692.838)	
Norway	-0.003**	0.007**	0.613**	14548820
	(-3.317)	(2.605)	(2720.099)	
Occ.Pal.Terr	-0.009	-0.004*	0.469**	2531832
	(-1.538)	(-2.076)	(473.741)	
Oman	-0.001	-0.002	0.591**	9359350
	(-0.376)	(-1.037)	(1686.82)	
Pakistan	0	0.018**	0.566**	8966784
	(0.218)	(4.829)	(1419.624)	
Palau	-0.003	-0.003	0.352**	1889680
	(-1.052)	(-0.15)	(312.724)	
Panama	-0.004**	-0.02**	0.581**	8320884
	(-3.251)	(-6.746)	(1698.675)	
Papua New Guinea	0.003	0.001	0.509**	3838192
	(0.661)	(0.435)	(559.324)	
Paraguay	-0.003**	0.004**	0.49**	11224980
	(-4.382)	(3.304)	(1232.762)	
Peru	-0.005**	0.052**	0.532**	14987973
	(-6.16)	(24.134)	(1740.359)	
Philippines	-0.003	-0.003	0.602**	11931425
	(-1.668)	(-1.885)	(1911.587)	
Poland	-0.01**	0.078**	0.635**	15192190

	(-11.262)	(28.962)	(2650.559)	
Portugal	-0.008**	0.005	0.583**	15263950
	(-8.524)	(1.763)	(2024.268)	
Qatar	-0.006**	0.051**	0.602**	8180700
	(-3.36)	(16.568)	(2011.868)	
Romania	-0.001	0.02**	0.594**	15198170
	(-1.713)	(10.652)	(2377.829)	
Russian Federation	0.013**	-0.06**	0.617**	15074904
	(12.204)	(-19.568)	(2323.407)	
Rwanda	0.013**	0.031**	0.47**	9202428
	(3.384)	(4.978)	(825.577)	
Samoa	0.001	0.005**	0.449**	5372350
	(1.104)	(3.409)	(500.799)	
Sao Tome and Principe	0.001	-0.001	0.455**	5640000
	(0.371)	(-0.378)	(419.374)	
Saudi Arabia	0.005*	0.01*	0.641**	12809388
	(2.187)	(2.397)	(2521.433)	
Senegal	-0.005**	0.01**	0.524**	12087244
	(-4.61)	(8.228)	(1491.771)	
Seychelles	-0.005	-0.031**	0.44**	5625970
	(-1.671)	(-5.535)	(820.545)	
Singapore	-0.001	0.015**	0.632**	14830816

	(-0.903)	(4.737)	(2326.242)	
Slovak Republic	-0.013**	0.014**	0.551**	15417610
	(-16.836)	(6.503)	(2032.479)	
Slovenia	-0.009**	0.016**	0.533**	15261246
	(-9.87)	(5.798)	(1971.637)	
South Africa	-0.005**	0.016**	0.57**	15269930
	(-3.164)	(2.63)	(2090.242)	
Spain	-0.009**	-0.007	0.664**	15302820
	(-8.231)	(-1.933)	(2750.324)	
Sri Lanka	-0.003*	0.018**	0.514**	13239864
	(-2.571)	(10.078)	(1430.83)	
St. Kitts and Nevis	0.001	-0.001	0.414**	9399648
	(1.817)	(-0.451)	(694.69)	
St. Lucia	0	0	0.431**	6953040
	(-0.094)	(0.058)	(765.8)	
St. Vincent and the Grenadines	-0.001	0.008**	0.428**	12276342
	(-1.551)	(7.523)	(877.772)	
Sudan	0.004*	0.007**	0.585**	13034340
	(2.376)	(4.448)	(1203.033)	
Suriname	-0.005	-	0.528**	783660
	(-0.437)		(377.238)	

Swaziland	0.003	0.02*	0.616**	3347224
	(0.747)	(2.325)	(223.31)	
Sweden	-0.008**	-0.009**	0.633**	15160860
	(-9.243)	(-3.565)	(2592.083)	
Switzerland	-0.007**	0.007*	0.611**	15163824
	(-6.561)	(2.42)	(2487.646)	
Syrian Arab Republic	-0.002*	0.01**	0.633**	7532330
	(-2.201)	(8.545)	(1281.422)	
Tanzania	0.003*	0.018**	0.518**	14759550
	(1.996)	(4.634)	(1526.257)	
Thailand	0.003	0.02**	0.572**	14076000
	(1.368)	(7.802)	(1894.962)	
Togo	-0.005	-0.007	0.525**	11145540
	(-1.669)	(-1.588)	(969.378)	
Tonga	-0.035**	-	0.402**	1492530
	(-3.511)		(295.677)	
Trinidad and Tobago	0	0.01**	0.51**	12171324
	(0.278)	(5.697)	(1332.839)	
Tunisia	0	0.012**	0.538**	10951567
	(0.375)	(4.36)	(1549.235)	
Turkey	0.001	-0.022**	0.644**	15207140
	(1.369)	(-9.219)	(2677.406)	

Turks and Caicos Isl.	0.001	0.001	0.513**	1202448
	(0.662)	(1.174)	(123.618)	
Uganda	-0.006**	0.007	0.501**	14461746
	(-3.511)	(1.71)	(1290.708)	
Ukraine	0.001	0.037**	0.552**	11372700
	(1.141)	(14.317)	(1656.058)	
United Arab Emirates	-0.005	-0.012	0.638**	2660310
	(-1.128)	(-1.372)	(1079.49)	
United Kingdom	-0.002*	-0.014**	0.746**	15447952
	(-1.986)	(-3.541)	(3744.628)	
United States	-0.001	0.177**	0.784**	14651780
	(-0.682)	(32.269)	(4070.357)	
Uruguay	-0.007**	0.012**	0.499**	13052572
	(-9.819)	(7.406)	(1498.31)	
Vanuatu	-0.007	0.007	0.393**	2203200
	(-0.928)	(1.914)	(388.893)	
Venezuela	0.004**	0.032**	0.602**	14632020
	(5.891)	(19.7)	(1874.813)	
Vietnam	0.004**	0.024**	0.627**	11860266
	(2.864)	(10.891)	(1994.349)	
Yemen	0.002	0	0.552**	6743065
	(1.636)	(0.052)	(974.207)	

Yugoslavia	0.001	0.017**	0.541**	11399940
	(0.605)	(5.337)	(1843.61)	
Zambia	0.006**	-0.014**	0.516**	13356980
	(3.407)	(-3.131)	(1178.928)	
Zimbabwe	-0.005**	0.034**	0.469**	9113904

Note: t-statistics in parentheses; *significant at 5%; **significant at 1%; standard errors robust to heteroskedasticity. Dependent variable is the log of import value at the six digit product category. Each row represents the results from a separate regression estimating the impact of tariff programs on a country's imports, where the impact of the program is estimated via a triple difference specification comparing differences in product coverage, country coverage, and timing of the program; Fixed effects at the exporter-product, product-year, and exporter-year level are included in all specifications.

CHAPTER THREE

Robust Determinants of Bilateral Trade

Marianne Baxter[41]

Boston University and NBER

Jonathan Hersh[42]

Boston University

Abstract

What are the policies and country-level conditions which best explain bilateral trade flows between countries? As databases expand, an increasing number of possible explanatory variables are proposed that influence bilateral trade without a clear indication of which variables are robustly important across contexts, time periods, and which are not sensitive to inclusion of other control variables. To shed light on this problem, we apply three model selection methods – Lasso regularized regression, Bayesian Model Averaging, and Extreme Bound Analysis – to candidate variables in a gravity models of trade. Using a panel of 198 countries covering the years 1970 to 2000, we find model selection methods

[41] mbaxter@bu.edu, Department of Economics, 270 Bay State Road, Boston, MA 02215.
[42] jhersh@bu.edu, Department of Economics, 270 Bay State Road, Boston, MA 02215.

suggest many fewer variables are robust that those suggested by the null hypothesis rejection methodology from ordinary least squares.

Keywords: Bilateral trade flows, gravity model, machine learning

JEL classification: F10, F14

1. Introduction

In 1962, Tinbergen proposed that the flow of trade between two countries should be proportional to the size of the countries' economies and inversely proportional to their distance. In reference to Newton's law of universal gravitation, he dubbed this relationship as "gravity." Subsequently, countless empirical studies found gravity to be a robust relationship across a broad range of contexts and time periods. The model was so popular that it led Anderson (1979) to state that the gravity model was "the most successful empirical trade device of the last twenty-five years."

This model was placed on firm theoretical ground through the work of Anderson and van Wincoop (2003) and the large literature that followed At the same time, there has been a revival of interest in empirical 'gravity' models that are motivated by policy questions such as "do currency unions matter?" or "do trade agreements/customs unions increase trade?" both of which remain relevant today. The problem of variable selection for empirical analysis of the gravity model is growing more complex as the availability of machine-readable databases expand along with statistical and computational methods for handling large datasets. It is tempting to include every possible empirical determinant of trade in the gravity model, although there are well known problems with "overfitting."[44] Further, variables that improve fit in-sample may not predict well when applied out of sample or to other datasets. If the empirical gravity model is to be used to inform policy it must balance in-sample and out-of-sample performance.

Our goal in this paper is to evaluate the robustness of commonly included measures of trade frictions, policy decisions, and country characteristics in determining the extent of bilateral trade. We use a panel of 198 countries from 1970 to 2000, and apply a standard empirical form of the gravity equation. To evaluate the robustness of variables included in the gravity equation we use three empirical methods: Bayesian Model Averaging; Lasso; and Extreme Bound Analysis[45] for the purposes of variable selection. Through these methods we learn which variables *should* be included in the gravity equation, that is which ones robustly predict trade flows.

The rest of the paper proceeds as follows. Section 2 reviews the Anderson-van Wincoop (2003) model of the gravity equation, which has become the standard workhorse model used in empirical implementation, and summarizes relevant empirical research. Section 3 presents background on the three empirical methods used for variable selection. Section 4 introduces the data used in our analysis and section 5 contains the results of applying the three approaches to model selection to bilateral trade data. Section 6 concludes.

---

[44] Even, curiously, data on Eurovision scores (Felbermayr and Toubal, 2009).

[45] These methods are not entirely new in economics, with Varian (2014), Belloni and Chernozhukov (2013) advocating for the use of Lasso, and Fernandez, Ley, and Steel (2001) employing Bayesian Model Averaging in the context of cross-country growth regressions.

2. Methodology

Anderson (1979) was the first researcher to present theoretical foundations which rationalized the gravity model. His model rested on the assumption that each country produces a single, imperfectly substitutable good. Anderson and Van Wincoop (2003) extended the single-good framework of Anderson (1979) to an arbitrary number of goods. We use their model with some modifications to the specification of trade costs.

2.1 Model Description

The consumer's objective is:

$$(1) \qquad \max_{c_{ij}} \left\{ \left(\sum_i \beta_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right\} \quad s.t. \; \sum_i p_{ij} c_{ij} = y_j$$

where c_{ij} is the consumption in region j of goods from exporting region i, p_{ij} is the price of region i goods for region j consumers, and y_j is nominal income in region j. The parameter β_i is a scale distribution parameter and σ is the elasticity of substitution between all goods. Trade costs enter the model through costs passed from exporter to importer. That is, t_{ij} is the trade cost factor between i and j, and given an exporter's supply price of p_i we can model the importer's supply price as $p_{ij} = p_i t_{ij}$. Next we let x_{ij} be the nominal value of exports from i to j. Since these exports are eventually consumed in region j we must have $x_{ij} = p_{ij} c_{ij}$. Finally, since each country consumes the value of its income, $y_i = \sum_j x_{ij}$. Maximization of equation (1) yields:

$$(2) \qquad x_{ij} = \left(\frac{\beta_i p_i t_{ij}}{P_j} \right)^{(1-\sigma)} y_j$$

where $P_j = \left[\sum_i \left(\beta_i p_i t_{ij}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$ is the consumer price index in region *j*. In words the optimization condition states that exports from region i to j are related to the exporter's supply price and the trade cost factor divided by the destination country's CPI.

Imposing market clearing gives us

$$(3) y_i = \sum_j x_{ij} = \sum_j \left(\frac{\beta_i t_{ij}}{P_j}\right)^{(1-\sigma)} = (\beta_i p_i)^{1-\sigma} \sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} y_{ij} \ \forall\, i$$

Let $Y^W = \sum_j y_j$ denote world income. Summing over all countries gives the more tractable expression:

$$(4) \qquad x_{ij} = \frac{y_i y_j}{Y^W}\left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma}$$

where

$$(5) \qquad \Pi_i = \left(\sum_i \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{y_j}{Y^W}\right)^{1/1-\sigma}$$

Anderson and van Wincoop (2003) propose one normalization that provides a solution to the set of equations (4) and (5) which is $\Pi_i = P_i$. They note however that this is not innocuous[46]. Nevertheless, with this normalization we arrive at the most widely used form of the gravity equation:

$$(6) \qquad x_{ij} = \frac{y_i y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma}$$

[46] The general solution is of the form $P_i = \lambda \overline{P}_i$ and $\Pi_i = \overline{\Pi}_i/\lambda$. The normalization is not innocuous in that in practice these multilateral resistance terms are estimated via country fixed effects. This is appropriate for cross-sectional estimation, but not for panel estimation. For more on panel estimation with the gravity model see Baier and Bergstrand (2007) and Egger and Nelson (2011).

The "gravity" elements--a negative relationship between trade and distance and a positive relationship between trade and GDP--are evident in equation (6). Trade flows between regions i and j are positively related to the product of the countries' GDP. Because t_{ij} is almost always parameterized to include distance, trade flows are inversely related to the distance between the two regions. The elements P_i and P_j are the "multilateral resistance" terms for regions i and j respectively. Though they are only indexed by i and j, note that both of them include the sum of all other regions' trade costs.

2.2 Empirical Specification

The estimating equation for the linear standard gravity equation is derived by taking the natural log of equation (6) and adding an error term:

(7) $$\ln(T_{ij}) = k + \ln(y_i) + \ln(y_j) + (1-\sigma)t_{ij} + (1-\sigma)\mathrm{P}_i + (1-\sigma)\mathrm{P}_j + \epsilon_{ij}$$

where k is a constant term, and T_{ij} are the bilateral flows from region i to j (*i*'s exports to *j* plus *j*'s exports to *i*). The multilateral resistance terms P_{ij} can be accounted for using country-level fixed effects. The more troubling parameter is t_{ij} , the trade cost specification, which is not observed. In Anderson and van Wincoop (2003) the authors specify the trade cost as a function of bilateral distance and whether the trade flows occur between two different countries: $t_{ij} = b_{ij}d_{ij}^{p}$.

In the related empirical literature, researchers have specified t_{ij} to include (i) measures of cultural or ethnic closeness, such as having a common language or a common legal system; (ii) geographic considerations that affect the ease of moving goods, such as sharing a border, being landlocked, being an island; (iii) membership in the WTO and/or other regional trading groups; (iv) a host of policies that bear on exchange rate stability; the ease of currency convertibility, banking and exchange rate crises. Specifying t_{ij} correctly is important for the out-of-sample performance. Including too few variables and one runs the risk of under-fitting the model, and not capturing enough variation in trade flows. Including too many variables and one risk of over-fitting, that is fitting noise rather than signal. Here is where methods from machine learning may be useful. Many of these were created specifically to solve the problem of building models that perform well out of sample, given a choice of too many predictors that can be reliably estimated using OLS (Varian, 2014). Methods such as Lasso regularization have been used successfully for model selection to select linear covariates from a large set of candidate coefficients (Belloni and Chernozhukov, 2013; Afzal, et al., 2015). Tree-based algorithms such as random forest or gradient boosted trees have been used successfully for pattern discovery[47], and may successfully reveal important patterns here.

[47] See Bajari et al., 2015; Athey and Wager, 2015 for examples in economics

3. Statistical Methods

3.1 Extreme Bound Analysis

Extreme Bound Analysis (EBA) was proposed by Leamer (1983, 1985) to address model uncertainty. EBA attempts to find which variables, in the set of candidate variables $\boldsymbol{X}$, are associated with an outcome variable yet robust to the inclusion of different control variables. As summarized by Leamer (2008):

> *"Extreme bounds analysis is a global sensitivity analysis that applies to the choice of variables in a linear regression. Rather than a discrete search over models that include or exclude subsets of the variables, this sensitivity analysis answers the question: how extreme can the estimates be if any linear homogenous restrictions on a selected subset of the coefficients are allowed?"*

The robustness of each coefficient is determined by whether the coefficient remains statistically significant and of the same sign in a reasonable number of estimated models. More formally, let $\boldsymbol{F}$ be the set of control variables that remain fixed in every model specification, which we call the "fixed" variables. The set $\boldsymbol{X}$ contains the variables that are the focus of the sensitivity analysis, which we refer to as the uncertain set of variables. Finally, let $\boldsymbol{U}_j \subseteq \mathbf{X}$ be the subset of variables which we use as control variables for a given specification j. Let $\mathrm{x} \in \mathbf{X}$ be a singular control variable that we are focusing on with model j. The model we use to estimate the robustness of x has the form

$$y = \beta_{0,j} + \beta_j x + \gamma_j \boldsymbol{F} + \Lambda_{\mathrm{j}} \boldsymbol{U}_j + \epsilon \tag{8}$$

where j indexes the regression models. We estimate this regression for each of the M

possible models depending on the combinations of $\boldsymbol{U_j}$ that are possible. The number of elements in $\boldsymbol{U_j}$ is typically limited to three variables in the literature (see Levine and Renelt (1992)) though the number of elements to be included for each specification is in theory limited only by the size of **X**. In the present application, this specification is estimated for all subsets of **X** with the exception any subsets including x itself. This process yields a distribution of coefficient estimates and associated standard errors, which are used to estimate empirical confidence intervals at some desired level of significance. The "extreme bound" for the coefficient of variable x is given by [*a,b*], where a is the lowest value in any confidence interval and b is the highest value in any confidence interval. The variable x is robust if [*a,b*] does not contain the value zero. The variable is "fragile" if [*a,b*] contains zero.

3.2 Lasso Regression

The Lasso regression is a member of the family of regularized regressions which estimates a regression model with an added constraint that forces parsimony in the coefficient estimate (Tibshirani, 1996). These estimators are referred to as "shrinkage" estimators, so named because relative to OLS their coefficients are biased towards zero. The motivation for shrinking coefficients towards zero comes from the bias-variance tradeoff; by adding more parameters one can easily reduce within-sample error or bias. This comes at the expense of a larger estimator variance or out-of-sample error. Lasso regressions and other shrinkage estimators attempt to strike a balance between in-sample and out-of-sample error.

Formally, the Lasso estimator, β_{lasso} solves the minimization problem

$$(9)\quad \beta_{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \underbrace{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{K} x_{ij}\,\beta_j\right)^2}_{OLS\ Sum\ of\ squared\ residuals} \quad + \quad \underbrace{\lambda \sum_{j=1}^{K}\left|\beta_j\right|}_{Shrinkage\ factor} \right\}$$

where $\lambda \geq 0$ is a parameter that represents a penalty associated with the sum of the absolute values of the coefficients. The Lasso estimator adjusts all parameter estimates by the same absolute amount unless this adjustment would cause the parameter to change sign, in which case, the parameter is set to zero (Friedman, Hastie, and Tibshirani, 2001). Note that as $\lambda \to 0$, the parameter penalization decreases and $\beta_{lasso} \to \beta_{OLS}$. As $\lambda \to \infty,$ variables are penalized more stringently and β_{lasso} converges to the zero vector.

Optimal λ^* is selected through cross-validation and comparing root mean squared error (RMSE) from a vector of possible $\hat{\lambda}$. The cross-validation algorithm for selecting λ^* is as follows: Sample data are split into K equally sized subsamples, or 'folds,' of equal size. Model estimation is performed using (K-1) of the folds and the resulting estimates are used to forecast or fit the data on the withheld fold. The average root mean squared error of this forecast is a function of λ. Thus, the ex-post optimal λ^* is the value of λ that minimizes this RMSE. In practice, however, researchers often choose the value of λ that corresponds to the more restrictive model associated with a one-standard-deviation increase in the cross-validation RMSE.[48] We follow this practice in our analysis.

Belloni and Chernozhukov (2013) discuss the Lasso estimator in the context of statistical inference. The original Lasso method (Tibshirani, 1996) was developed from the

[48] See Krstajic et al (2014) for more discussion on the selection of λ. The value is set to the RMSE minimizing value plus one standard error for the purposes of choosing "the simplest model whose accuracy is comparable with the best model." That is, while a more complicated model may perform better, the more parsimonious model performs comparable well enough to the more complicated one.

standpoint of prediction, which differs from the usual approach of applied economists who are concerned with parameter inference. Belloni and Chernozhukov develop an estimator they call the "post-Lasso" estimator, which they show to perform at least as well as the Lasso estimator with slightly less bias. The post-Lasso estimator is a two-step procedure. In step 1, a Lasso model is estimated over a large set of possible control variables. The variables which have non-zero coefficients are selected and retained for use in the second step. In step 2, an OLS model is fit using only the subset of variables that had non-zero coefficients in the first step. This method leans on the strengths of each approach: Lasso is useful for variable selection but presents biased estimates of coefficients.[49] OLS is unbiased and efficient, though cannot handle models with large number of covariates. In our analysis we present both the Lasso and the Post-Lasso estimates.

3.3 Bayesian Model Averaging

Bayesian model averaging (BMA) is an intuitive approach to model uncertainty where Bayes Rule is applied to the model and data, from which one can construct posterior parameter estimates. One advantage of BMA is that as in typical Bayesian estimation procedures, the output is a posterior distribution of possible parameter estimates, which can be more revealing than the point estimates returned by other methods. BMA is estimated as follows: First, all permutations of a linear regression model are estimated using the set of explanatory variables, $\boldsymbol{X}$. We refer to $\boldsymbol{X}_{\alpha}$ as a particular subset of variables

[49] Since Lasso estimation may shrink fixed effect coefficients, which should be included in any unbiased estimate, we employ a two step procedure. In the first step, we use Frisch-Waugh-Lovell (FWL) theorem (Frisch and Waugh, 1933; Lovell, 1963) to transform the dependent and independent variables to control for the level fixed effects. In the second step, the desired estimator is used on the FWL- transformed data series.

in $\boldsymbol{X}$, and in that sense, each distinct $\boldsymbol{X}_\alpha$ is a separate model. In the second step, a posterior parameter vector is constructed using a weighted average of all parameters estimated from the set of estimated models in step 1. Because some models explain the data better than others, posterior parameters are a weighted function of parameter estimates using posterior model probabilities (PMP), which describe how well a given model (with an associated $\boldsymbol{X}_\alpha$) explain the data. If the dimension of $\boldsymbol{X}$ is K –that is we can choose from K possible explanatory variables to fit our model–this implies that BMA needs to estimate 2^K possible models to estimate every possible explanatory variable combination, a considerable computational undertaking. In practice, restricting estimation to a sample of the 2^K possible model computations reduces model estimation to a manageable size.

To give more structure to the problem, consider the problem of estimating a model of the form

$$y = \beta_{0,\alpha} + X_\alpha \beta_\alpha + \epsilon_\alpha \tag{10}$$

where we must choose which set of variables $X_\alpha \in \boldsymbol{X}$ should be included in a given regression. Using Bayes Rule, we can calculate the posterior model probability–a measure of the reasonableness of the coefficients used–as

$$p(M_\alpha | y, X) = \frac{p(y|M_\alpha, X)p(M_\alpha)}{p(y|M_\alpha)} \tag{11}$$

Where $p(M_\alpha | y, X)$ is the probability of the model given the data, or the posterior model probability; $p(y|M_\alpha, X)$ is the probability of the outcome variable given the model and the set of covariates and $p(M_\alpha)$ is the unconditional probability of the particular specification of the model, M_α. Using an application of the law of total probability we can rewrite the posterior model probability as

(12) $$p(M_\alpha|y,X) = \frac{p(y|M_\alpha,X)p(M_\alpha)}{\sum_{j=1}^{2^K} p(y|M_j,X)p(M_j)}$$

This leads to an expression for the model weighted posterior distribution for any estimator, β_k as

(13) $$p(\beta_k|y,X) = \sum_{\alpha=1}^{2^K} p(\beta_k|M_\alpha,y,X)p(M_\alpha|X,y)$$

The equation above shows that given the posterior model probability (PMP) we can estimate $p(\beta_k|y,X)$ −the probability that any estimator is included in the true model. The left hand side of equation is referred to as the posterior inclusion probability (PIP) and is reported a number between zero and one. The PIP reflects our relative confidence that the true model contains any particular variable. For example, if a variable has a PIP value of 1.0 this indicates that 100% of the weighted models include the variable β_k as a regressor, giving us relative confidence that the true model contains this variable. From the posterior distributions we also recover the posterior mean–the posterior average of the coefficient–and the posterior standard deviation, which give us the weighted average and the weighted standard deviation of the coefficient estimates across estimated model.

4. Data

Our panel covers the sample period of 1970 – 2000. Country coverage varies based on data availability. Regarding variables considered, we include a large set of candidate variables that the literature has suggested as measures of trade frictions, making attempts to use the data sources most commonly employed. However, when faced with a choice we decide in favor of variables that are measured over the entire sample period. Summary statistics for the data used are presented in Table 3- 1.

4.1 List of Data Sources

Measurement of Trade Intensity

Real Bilateral Trade Flows For measures of trade intensity, we use the NBER-UN dataset of bilateral trade flows as described in Feenstra et al. (2005). The NBER-UN dataset offers several advantages: a long panel from 1962-2000, trade statistics covering all reported trading partners as collected by the United Nations, and construction using the more reliable import statistics when these are available. In the few cases where import statistics are not available, Feenstra, et al. use export measures as reported by the trading partner. Since these data primarily use import statistics, the trade intensity data measure CIF trade flows.

Gravity Variables

Distance The defining features of a gravity equation are a positive relationship between trading partners' size and trade intensity and an inverse relationship between distance and trade intensity. Several measures of geographical distance have been proposed and used, with no consensus in the literature as to which one is preferred.[50] We consider four measures of distance, with a goal that the variable selection methods will provide evidence for which measure best explains trade volumes. Each distance measure is provided by the

[50] See Disdier and Head (2008) for an illuminating meta-analysis on distance in gravity models.

Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) (Mayer, and Zignago, 2011). The first distance measure is the natural log of distance between most populated cities -- the most standard distance measure employed in the empirical gravity literature. The second is the natural log of distance between capitals. The next two measures were developed by Head and Mayer (2002). These measures are (i) the natural log of weighted distance and (ii) the natural log of CES-weighted distance. They calculate weighted distance as

(14) $$d_{ij} = \left[\sum_{k\in i}\left(\frac{pop_k}{pop_i}\right)\sum_{l\in j}\left(\frac{pop_l}{pop_j}\right)d_{kl}^{\theta}\right]^{1/\theta}$$

where pop_k measures the population of area k in country i. For simple weighted distance, θ is set to 1. However, with CES weighted distance, θ is set to -1, which intentionally corresponds to the most frequently measured elasticity between trade and distance. Intuitively, these weighted distance metrics measure distance along the dimensions that matter: since good will eventually need to travel to where demand is located, these measures attempt to account for different dispersions in population densities.

Product of GDPs Data on country GDP are from the Penn World Tables, version 7.1 (Heston, Summers and Aten 2012) and are expressed in constant US dollars. Our GDP product variable is constructed as the average of the logs of the two partners' levels of real GDP.

Geographical Determinants of Trade

Contiguous, Island, and Landlocked For geographical determinants of trade, we look at three widely used variables in the literature: (i) an indicator variable for whether the trading partners are contiguous–that is they are adjacent to each other; (ii) an indicator for whether either trading partner is an island, and (iii) an indicator for whether either trading partner is landlocked.

Proxies for Cultural Distance

Next we consider variables which proxy for, or are directly related to, cultural distance between trading partners. These determinants of trade rest on the affinity principle: countries find it advantageous to trade with countries that similar to themselves. Some of these variables, such as language, can be thought of informative of reduced trade costs through easier contracting, or reduced transaction costs. Other cultural variables might proxy for shared demand systems across populations.

Share Official Language, 9%+ Speak Language A language indicator is often included in gravity models of trade. One way to justify its inclusion is in reducing contracting and coordinating costs between trading partners, what we refer to as the ***direct*** mechanism of reducing trade costs. Because language is a specialized skill, it is not necessary for the majority of population to speak the language in order to exploit this channel of reduced trade costs. Therefore, we consider the indicator for whether 9% of the population share a common language as a test for the direct reduction of trade costs through language. There is some support in the literature for broadening the scope of this common covariate. Melitz and Toubal (2014) estimate a model which adds linguistic proximity, shared native language, and spoken language to the usual official-language indicator variable. They find that the inclusion of these variables results in trade impacts twice as large as with official language alone. The second language measurement we consider is whether the trading partners share an official language. This variable captures a sense of shared cultural background between trading partners, either through similarity of culture, or through shared historical past, during which one would have had much time to develop trading linkages.

Religious Distance We consider religious distance as a proxy for shared culture between trading partners. Lewer and van de Berg (2007) construct a series of indicator variables for shared majority religion and find trading partners who share religion have more trade. We take a slightly different approach and construct a continuous metric of similarity of religion that we define as religious distance. We parameterize religious distance as the Euclidean distance between the percent of the population in 16 different religious groups in the two countries, where these groups are defined by the World Religion Dataset[51]. A distance of 0 indicates that the trading partners have populations which have

[51] The World Religion Dataset is available at and gives the percent of population in each of the 16 different major religious groups for 192 countries covering 1945-2010.
http://www.thearda.com/Archive/Files/Downloads/WRDNATL_DL2.asp

identical population fractions for each religious group; higher values of the distance variable indicate less religious similarity.

Factor Endowments

Human Capital, Physical Capital, and Arable Land To measure the factor endowment of human capital we use the Barro-Lee (2013) statistics on the average years of schooling for the population over the age of 15. For physical capital, we define factor endowment as a measure of physical capital per worker. Using data from the Penn World Tables, we calculate the value of capital stock measured at the current PPP exchange rate, divided by the number of employed persons in the economy. Finally, for the factor endowment for arable land is defined as arable land per worker, using data from the World Development Indicators. For each measure of factor endowments (human capital; physical capital; arable land), the factor intensity for the trading pairs i and j is defined as

$$f_t^{\{ij\}} = \ln(F_{it} * F_{jt}) \tag{15}$$

where F_{it} is the endowment of for country i in period t. Countries with similar factor endowments will have larger factor intensity measures.

Impediments to Flows of Goods and Capital

WTO/GATT Membership, Regional Trade Agreements Multilateral agreements, such as the World Trade Organization (WTO) and the General Agreement on Tariffs and Trade (GATT), are one method by which countries can commit to lower impediments to the flow of goods and capital. These trade organizations are tasked with the goal of increasing world trade, thus it is natural to posit that membership in these organizations has a positive effect on trade volumes. Rose (2004) estimated the effect of WTO/GATT membership on trade, parameterizing membership as an indicator for whether either trading partner are included in a trade agreement, and an indicator for whether both partners are in the WTO/GATT. He finds positive effects of membership. Baier and Bergstrand (2006) use a panel framework to attempt to address the endogeneity of membership within a free trade agreement, and find that the trade gains from membership are even larger than those found by Rose. Subramanian and Wei (2007) find that WTO membership has a strong heterogeneous impact on trade, with effects largest when both trading partners are members, and further find that sectors which did not liberalize experienced no trade gains to WTO membership. We use WTO/GATT membership published by *Centre d'Etudes Prospectives et d'Informacions Internationales* (CEPIIWe also include an indicator for membership in a regional trade agreement, the data for which is courtesy of de Sousa (2012).

Common Currency A shared currency between trading partners is widely believed to encourage trade. This was one of the main justifications for the introduction of the Euro in 1999. We use de Sousa's (2012) formulation of shared currency, which parameterizes the variable as equal to 1 if trading partners are part of an explicit or implied currency union. In an explicit currency union, the currency of one country circulates as legal tender in the second country. An implied currency union exists when one country maintains an explicit peg at a fixed rate of their currency to another country's currency. Note this does not include any other type of peg besides a fixed and maintained peg. The effect of a common currency on international trade has received more scrutiny than any other variable. Several influential papers using data from 2000 and earlier (i.e., not including the Euro zone) found that countries with a common currency enjoyed a level of trade from 110% higher to 577% higher, compared with countries that did not share a common currency.[52] In their meta-analysis of studies that estimated this parameter, Rose and Stanley (2005) consider 34 separate studies that overall present 754 estimates of the common currency effect. They find that the mean estimate implies a 136% increase in trade, while the median estimate implies a 70% increase in trade.

<u>Capital Openness</u> The degree to which capital can flow freely between countries may also affect trade. We utilize Chinn and Ito's (2007) index for financial liberalization, which is itself based on the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions* (AREAER). Their index is based on the series of binary indicator variables provided in the financial transactions of the AREAER, for a five year window in which the capital controls were not in effect. They define $share_{i,t} = \frac{\sum_{k,t\in \boldsymbol{K}} k_{i,t}}{card(\boldsymbol{K_t})}$ where $\boldsymbol{K_t}$ is the set of possible capital controls in year t. Their capital openness variable is the first standardized principal component of this share variable. We consider their openness index to be a

[52] In roughly chronological order, these are Rose (2001): 235% higher; Rose and van Wincoop (2001), 136% to 297% higher; Frankel and Rose (2002) 371% higher; Glick and Rose (2002) 110% higher; and Barro and Tenreyo (2007), 577% higher.

parsimonious dimension reduction of a series of measures of capital controls projected into one dimension.

Exchange Rate Measurements *Exchange Rate Volatility, and Exchange Rate Regimes: Fixed Exchange Rate, Crawling Peg, or Moving Band* Nominal exchange rate volatility has been shown to affect trade flows in models where firms set prices in advance (Broda and Romalis, 2003) and it has been tested empirically quite broadly. Nonetheless, Anderson and Van Wincoop (2004) in their review of the literature remark that there is "substantial consensus that the impact of exchange rate volatility on trade is very small at best, with even the sign uncertain (pg. 719.)" There are various ways to parameterize exchange rate volatility; we model exchange rate volatility as the residuals derived from an $ARCH(1,3)$ process for yearly bilateral exchange rates.

To measure the type of exchange rate regime, we use the IMF "coarse" classifications, as reported by Reinhart and Rogoff (2004) and subsequent updates. We define indicators for three exchange rate classifications: fixed exchange rate, crawling peg, and a moving band exchange rate regime. Each classification has two types: whether either trading partner employs this exchange rate regime, and whether both trading partner has this arrangement.

Crises Episodes *Indicators for Debt, Banking and Currency Crises* Debt, banking and currency crises disrupt and depress economic activity in general, therefore it's likely that crises episodes have a large impact on trade flows between countries. We use the IMF's

Systematic Banking Crises Database–which also contains data on currency and debt crises–as developed by Laeven and Valencia (2008) and updated in Laeven and Valencia (2012). The authors define a banking crisis to exist if two conditions are met:

> "1) Significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations)
>
> 2) Significant banking policy intervention measures in response to significant losses in the banking system." (p. 4)

The authors record 147 banking crises since 1970. Their definition of a currency crisis is based on Frankel and Rose (1996). A currency crisis is defined as a nominal depreciation of the currency versus the US dollar of at least 30%. The authors find 218 events which qualify as currency crises during the time period of 1970-2011. The Laeven and Valencia definition of a debt crisis is based on information from Beim and Calomiris (2001), World Bank (2002), IMF reports and other agencies, and Sturzenegger and Zettelmeyer (2006). The authors find 66 events that qualify as sovereign debt crises during the period 1970-2011. Crises events are rare by definition. Since the effect of crisis on trade may not be immediate, we define an indicator variable taking the value one if either trading partner experienced a crisis within the previous three-year window and zero otherwise.

5. Results

The main results are presented in Tables 3-2 and 3-3. All specifications include year, exporter, and importer fixed effects. Table 3-2 presents the coefficient estimates for OLS, Lasso and Post-Lasso, while Table 3-3 presents the results for Bayesian Model

Averaging and Extreme Bound Analysis. We consider a variable to be Lasso robust if it remains non-zero after Lasso Bayesian shrinkage, BMA robust if it has a posterior inclusion probability (PIP) of greater than 0.5, and EBA robust if the estimate upper bound and lower bound do not contain zero.

Table 3-2 compares OLS and Lasso estimates. Of the 31 covariates in the candidate set of variables (excluding the intercept), OLS regression finds that 26 of these variables are statistically significant at the 5% level. However, only 17 are Lasso robust. Table 3-3 shows that 18 are BMA robust and 19 are EBA robust.[53] Together, the three methods suggest a more parsimonious model of the determinants of bilateral trade than would be implied by standard application of OLS.

Figure 3-1 presents histograms of the distribution of estimated regression coefficients for the EBA method, where the vertical red line shows where zero sits in the distribution. These distributions can be highly informative; some variables, such as the landlocked indicator show a tight coefficient distribution, indicating across almost all reasonable specifications we can expect the coefficient estimate to lie within this range. Other variables, such as regional trade agreement, show a bimodal coefficient distribution, suggesting this parameter has heterogeneous effects that may depend on other included covariates.

Figure 3-2 presents the posterior model probabilities from BMA, showing the most likely model specifications with their accompanying probabilities. The model with the

[53] We consider a variable Lasso robust if it remains non-zero after Lasso Bayesian shrinkage, BMA robust if it has a posterior inclusion probability (PIP) of greater than 0.5, and EBA robust if the estimate upper bound and lower bound do not contain zero.

highest estimated posterior model probability has a 17% posterior probability, and includes 19 variables (excluding the variables of banking crisis, crawling peg, moving band, fixed exchange rate, official common language, and currency crisis). The second most likely model has a 16% posterior probability, and includes all the variables in the previous model and includes a crawling peg indicator. The third most likely model 15% posterior model probability and includes all of the variables of the previous model, but includes an indicator for banking crises. Together, these three models have a cumulative probability of 48%. Note that after 70% cumulative model probability the models appear to fragment, with many models having small fractional posterior probabilities.

Figure 3-3 shows the shrinkage path of the Lasso coefficients. The y-axis presents the standardized coefficient value as the value of the shrinkage parameter, lambda, varies. The OLS solution corresponds to the left-most position on the x-axis. As the lambda parameter increases, and we move to the right on the x-axis, variables are shrunk towards zero. For any given value of lambda some coefficients will be estimated to be zero, thus for each value of lambda positive y-values correspond to variables selected via Lasso. The shrinkage path – that is the order in which variables are shrunk to zero – is informative of which variables have the largest explanatory power. For example, for very large values of lambda, log of weighted distance remains while many other variables have been shrunk to zero.

Gravity Variables

Estimation via Lasso we find two out of the four candidate variables were not shrunk to zero: weighted distance and CES distance, with coefficients of -0.994 and -0.133 respectively. In comparison, OLS considers distance and weighted distance highly significant, with the log of distance having a puzzling positive coefficient. If we add the Lasso robust coefficients we get a combined elasticity of distance on trade of -1.127. This is slightly larger than the average elasticity of -0.907 as found in Disdier and Head's meta-analysis of 1,467 gravity models. Disdier and Head find that papers using earlier data tend to have smaller coefficients, and an average coefficient size of 0.9, 0.96, and 0.95 in decades 1970s, 1980s and 1990s respectively. Because BMA and EBA methods are less robust to the inclusion of highly correlated variables, we chose only one of the two Lasso robust distance measures to test using BMA and EBA. Using Bayesian model averaging we estimate a coefficient on CES weighted distance of -1.113. The posterior inclusion probability (PIP) is 1, meaning that 100% of the weighted posterior models included distance in the final model. The posterior standard deviation of the estimated coefficient is 0.007, indicating a small amount of variation across models. Finally using EBA, we find an upper bound and lower bound range of (-1.34 to -1.1), within the range of significance suggested by EBA. Lasso shrinks the product of GDPs to an estimated 0.559 from the OLS estimate of 0.729. Using Post-Lasso, the estimate rises to 0.636. BMA shows the product of GDPs to be a tightly estimated 0.739, very close to the OLS estimate. This variable is also robust under EBA with an estimated range of (0.75, 0.99).

Geographical Determinants

All of the geographical determinants variables are robust according to the three methods used. OLS estimates the contiguous dummy's coefficient as 0.614, while Lasso gives an estimate of 0.431. BMA agrees with the Lasso estimate, giving a mean estimate of 0.405 which is on the lower end of EBA's estimate range from 0.37 to 0.69.

Proxies for Cultural Distance

Both language variables considered remain non-zero after Lasso shrinkage. The coefficient on official language is estimated at 0.066 and the coefficient on 9% speak is estimated at 0.296. Egger and Lassmann (2012), in a meta-analysis of 701 coefficients culled from 81 published articles, find an average coefficient of 0.49, considerably smaller than our estimate, even when one combines the two different language estimates. Our Post-Lasso estimate, is estimated at 0.047 for official language and at 0.327 for 9%+ population, also smaller than the literature average. The estimate from BMA show a posterior inclusion probability of 0.031, meaning only 3.1% of the weighted posterior models included this variable. EBA, however, shows robustness of official language, with an estimated coefficient range between 0.19 and 0.78. This variable's counterpart, 9%+ population*,* however, is robust according to BMA with an estimated PIP of 1, and according to EBA which shows an estimated range of 0.21 and 0.76.

Both former colony and common colonizer appear strongly robust in the Lasso regression model, with coefficients of 0.738 and 0.319 respectively. The post-Lasso estimate highly significant with slightly larger coefficients of 0.867 and 0.394. This estimate is roughly on par with the coefficient estimated by Frankel and Rose (2002), and with the coefficient estimate of 0.45 in Glick and Rose (2002). The BMA and EBA show similar robustness of these covariates. Former colony is robust according to BMA with an estimated PIP of 1 and an estimated post-mean of 0.866. EBA gives the upper bound and lower bound range of (0.86,1.36). Common colonizer has an estimated PIP of 1 and a post mean estimate of 0.413, roughly similar to the post-Lasso estimate. The EBA estimate shows robustness with a rather large estimated range of (0.35 to 0.77). Though no meta-analysis exists for this coefficient, our estimate for former colony seems smaller than the coefficient on this covariate estimated previously, such as in Rose (2004) who estimates a coefficient of 1.28 for post-1970, or Rose and van Wincoop (2001) who find a coefficient of 1.74.

The estimated Lasso coefficient on common legal origin is 0.25, and has a post-Lasso coefficient of 0.286, which is nearly identical to the OLS estimate of 0.282. This estimate is similar in magnitude to others in the literature, such as 0.306 estimated by Head, Mayer and Ries (2010) or 0.410 estimated Felbermayr and Toubal (2009). The post-mean estimate from BMA is very similar in magnitude, estimated at 0.287, with an accompanying PIP of 1. The EBA range of (0.27,0.51) indicates this variable is considered robust according to that method. Religious distance has an OLS estimate of -0.209, and using Lasso we get a coefficient of -0.174. Given that this is an index, it's hard to interpret the magnitude of this coefficient, but given that the standard deviation of this index varies is 0.69 in our sample, moving one standard deviation of religiously dissimilarity is predicted to decrease aggregate trade flows between partners by 11%[54]. The Post-Lasso coefficient estimate is -0.215 and is significant at the canonical levels. This variable is also

[54] Assuming a marginal effect from 0 to 0.69.

robust according to BMA and EBA. BMA gives a posterior inclusion probability (PMP) of 1 with a post-mean coefficient of -0.216 and a standard deviation of 0.009. EBA further finds this variable robust and gives a range of (-0.3, -0.17). In comparison to other work in the literature, our findings suggest a stronger effect of religious similarity than previous estimates (Linders et al., 2005) who estimate a coefficient of 0.22 for (binary) religious similarity between trading partners. Some of this difference may be coming from the continuous versus discrete parameterization of this variable, however when taken at face value our estimate implies a larger response to religious similarity and trade.

Factor Endowments

All three factor endowment variables – human capital, physical capital, and arable land -- are statistically significant using OLS with coefficients of 0.742, -0.129, and -0.231 respectively. Lasso selects only human capital and arable land, with coefficients of 0.389 and -0.15. Using BMA, all three variables have PIP of 1, and EBA finds all three robust. Human capital shows a BMA post-mean of 0.75, and an EBA range from (0.78, 1.37), consistent with the OLS coefficient estimate. Physical capital shows a BMA post-mean of -0.132 and an estimated EBA range of (-0.18, -0.09). Finally arable land shows a BMA post-mean of -0.229 and an estimated EBA range of (-0.41, -0.27).

Impediments to the Flows of Goods and Capital

WTO/GATT and common currency indicator are robust across all three methods. Exchange rate volatility is robust according to BMA, but not when using Lasso or EBA.

Regional trade agreements and capital openness are not significant in any of the specifications. For capital openness, While OLS estimates a coefficient of -0.00298, Lasso estimation estimates a zero coefficient. This result is mirrored in the results for BMA, which estimates a zero PIP, and EBA, which estimates a range of coefficient values of (-0.21,0.04), which is not robust according to the method. This was a surprising result, as the degree to which capital can flow freely seems to a priori affect real trade flows.

WTO/GATT membership looks strongly robust across methods. OLS estimates a significant coefficient of 0.335, Lasso estimates a coefficient of 0.171, and BMA estimates post-mean of 0.348. This compares to the coefficient estimated by Rose (2004) in column 4 of table 1 of 0.15, which is closer to the Lasso result than the OLS estimates. Regarding common currency, OLS estimates a coefficient of 0.448, which Lasso shrinks to 0.109. The BMA post-mean is 0.353 with a PIP of 1, and EBA estimates a range of (0.11, 0.84). Frankel and Rose (2002) estimate coefficients on currency union membership that range from 1.36 to 1.55, which are substantially larger than our estimates.

Exchange Rate Measurements

Using OLS, all of the exchange rate variables are statistically significant from zero. However, there is large agreement across methods, showing only either crawling peg being robust. Lasso selects only either crawling peg indicator as robust and the rest are set to zero. Using BMA, exchange rate volatility, and either crawling peg have PIPs of 1 with estimated coefficients of -0.012 and 0.153 respectively. Both fixed exchange rate has a PIP of 0.893 with an estimated post-mean of 0.071 and either moving band has a PIP of 0.048. Using EBA only either fixed exchange rate and either crawling peg are robust, with estimated ranges of (-0.14, -0.02) and (0.06, 0.18) respectively. The differences in statistical robustness between OLS and the other methods are stark. We can only speculate as to the reason for the differences. Our hypothesis is that OLS may be fitting a significant amount of noise that it interprets as signal, which the other methods do not.

Crisis Measurements

Estimating using OLS we see that of the three crisis episodes considered—debt, banking, or currency—only banking crisis is significant at the standard levels, showing a positive coefficient of 0.0513. However, when estimating via Lasso and applying Bayesian shrinkage this variable is estimated at zero and thus is not considered robust according to Lasso. Using Bayesian Model Averaging, we see that the presence of a debt crisis has a PIP of 0, indicating no probability of inclusion in the true model. The PIP of banking crisis is on the cusp of robustness, showing a value of 0.481 and a post-mean of 0.021. Currency crisis shows a near-zero PIP of 0.006 and a post-mean indistinguishable from zero. Estimating via EBA, we see that neither debt crisis nor currency crisis are robust according to EBA. However, banking crisis is, showing a range of estimated coefficient values of (0.02,0.1). The positive coefficient on banking crisis is slightly puzzling, and given the window of this variable of 3 years, this may indicate that we are picking up the "rebound" period when trade returns to trend after a crisis.

To be sure, this is not to say that crisis episodes considered here do not necessarily have an impact on trade. When estimating trade flows *without* using year fixed effects (available from the authors by request) banking and currency crises are consistently negative and robust. Our results do not preclude the possibility that all worldwide trade is depressed during periods of banking and currency crises. That is to say, it is possible from viewing these results that all countries lose out during banking and currency crisis episodes, not just those that experience the crises themselves. There appears to be some support for this thesis, as shown in Shelburne (2010) who looks at trade decline during the global financial crisis from 2007-2010. How much is worldwide trade depressed? Our results indicate quite a lot. EBA shows the coefficient on banking crisis varies from -0.52 to -0.3, even when controlling for exporter and importer fixed effects. This translates to a marginal effect of -40% to -25.9% per trading partner, which indicates almost implausibly large aggregate declines. For currency crisis, the coefficient varies from -0.44 to -0.23 indicating marginal effects on trade of -35.6% to -20.5%.

6. Conclusion

How do the three variable selection methods refine the set of variables that should define the workhorse empirical gravity model? First, our result reject the robustness of roughly a fifth of the variables in the candidate set for which OLS does not reject the null hypothesis. Second, the set of robust variables is remarkably consistent across the three model selection methodologies.

Table 3-3 shows relative agreement across methods. Very few variables appear highly significant using one method while not very significant in others. In particular Lasso and EBA show very similar results qualitatively, differing in parameter inclusion significance for only 4 variables.

Table 3-4 show the results across methods. A mark in the table indicates that the variable is robust according to that particular method. Specifically, for lasso, a mark indicates a non-zero coefficient. For EBA, a mark indicates that the upper and lower bound do not include zero. For Bayesian Model Averaging, a mark indicates a 50% or great PIP. Overall, our results show that model selection methods that balance "fit" and "prediction" are straightforward to employ; give a consistent set of results, at least in the context of the gravity model; and that these methods represent the best current solution to the problem of variable selection in potentially a wide variety of contexts.

References

Anderson, J. (1979) “A Theoretical Foundation to the Gravity Equation”. American Economic Review, Vol 69, No 1, pp. 106-116

Anderson, J. and van Wincoop, J. (2003) “Gravity with Gravitas: A Solution to the Border Puzzle,” American Economic Review. vol. 93(1), pages 170-192, March.

Baier, S. and Bergstrand, J, (2007) “Do Free Trade Agreements Actually Increase Members' International Trade?” Journal of International Economics, vol. 71(1), pages 72-95.

Baldwin R. and Taglioni, D. “Gravity for Dummies and Dummies for Gravity Equations.” (No. w12516). National Bureau of Economic Research.

Barro, R. J. and Lee, J. W. (2013). “A New Data Set of Educational Attainment in the World, 1950-2010.” Journal of Development Economics, 104, 184-198.

Beim D. and Calomiris, C. (2001). “Emerging Financial Markets.” Irwin Professional Pub.

Belloni, A and Chernozhukov, V. (2013) “Least Squares After Model Selection in High Dimensional Sparse Models”, ArXiv 2009, Bernoulli 2013.

Belloni, A. Chernozhukov, V. and Hansen C. (2014). “Inference on Treatment Effects after Selection among High-Dimensional Controls.” The Review of Economic Studies. 81(2), 608-650.

Broda, Ch. and Romalis. J. (2009). “The Welfare Implications of Rising Price Dispersion.” University of Chicago mimeo.

Bun and Klaassen (2007) “The Euro Effect on Trade.” Oxford Bulletin of Economics and Statistics 64:9 (2007), 473-496.

Carrere, C. (2006). “Revisiting the Effects of Regional Trade Agreements on Trade Flows with Proper Specification of the Gravity Model.” European Economic Review, 50(2), 223-247.

Chaney, T. (2013). “The Gravity Equation in International Trade: An Explanation.” NBER Working Paper No. 19285.

Chinn M. and Ito H. (2008). “A New Measure of Financial Openness.” Journal of Comparative Policy Analysis. 10(3), 309-322.

De Haan, J. and Sturm, J. E. (200)0. "On the Relationship Between Economic Freedom and Economic Growth." European Journal of Political Economy, 16(2), 215-241.

De Sousa, J. "The Currency Union Effect on Trade is Decreasing Over Time." Economics Letters 117(3), 917-920.

Egger P. and Nelson, D. (2011) "How Bad is Anti-Dumping? Evidence from Panel Data", Review of Economics and Statistics, MIT Press, vol 93(4), pages 1374-1390, November.

Egger, Peter H., and Andrea Lassmann. "The Language Effect in International Trade: A Meta-Analysis." Economics Letters 116.2 (2012): 221–224.

Feenstra, R. C. Lipsey, R. E., Deng H., Ma, A. C., Mo, H. (2005). World Trade Flows: 1962-2000 (No w11040). National Bureau of Economic Research.

Felbermayr, Gabriel J., and Farid Toubal. "Cultural Proximity and Trade." European Economic Review 54.2 (2010): 279–293. ScienceDirect. Web. 23 Apr. 2015.

Fernandez, C, Ley, E. and Steel, M. (2001) "Model Uncertainty in Cross-Country Growth Regressions", Journal of Applied Econometrics, Jon Wiley & Sons, Ltd., vol. 16(5) pges 563-576.

Frankel J. and Rose A. (2002) "An Estimate of the Effect of Common Currencies on Trade and Income." The Quarterly Journal of Economics. MIT Press, vol 117(2), pages 437-466.

Frankel. K. Stein, E. and Wei S. J. (1995). "Trading Blocks and the Americas: The Natural, the Unnatural, and the Super-Natural." Journal of Development Economics, 47(1), 61-95.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. Econometrica: Journal of the Econometric Society, 387-401.

Glick, R. and Rose, A. (2002). "Does a Currency Union Affect Trade? The Time Series Evidence," European Economic Review, Elsevier, vol. 46(6) pages 1125-1151.

Granger, C. and H. Uhlig (1990) "Reasonable Extreme Bounds Analysis," Journal of Econometrics" 44 (1990), 159-170.

Head, K. and Mayer, T (2004). "The Empirics of Agglomeration and Trade." Handbook of Regional and Urban Economics, 4, 2609-2669.

Head, Keith, Thierry Mayer, and John Ries. “The Erosion of Colonial Trade Linkages after Independence.” Journal of International Economics 81.1 (2010): 1–14. ScienceDirect.

Heston, A., Summers, R. and Aten, B. (2012). “Penn World Table Version 7.1” Philadelphia: University of Pennsylvania.

Krstajic, Damjan et al. “Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models.” Journal of Cheminformatics 6.1 (2014): 10. PMC. Web. 22 Apr. 2015.

Leamer, E. E. (1983). “Let's Take the Con Out of Econometrics.” The American Economic Review, 31-34.

Leamer, E. E. (1985). “Sensitivity Analyses Would Help.” The American Economic Review, 308-313.

Leamer, E. E. (2008) “Extreme Bounds Analysis.” The New Palgrave Dictionary of Economics, Second Edition, 2008.

Levine, R. and Renelt D. (1992). “A Sensitivity Analysis of Cross-Country Growth Regressions.” The American Economic Review. 942-963.

Lewer, J and Van den Berg (2007). “Religion and International Trade: Does the Sharing of a Religious Culture Facilitate the Formation of Trade Networks?” American Journal of Economics and Sociology, 66(4) 765-794.

Limao, N. and Anthony Venables. (2001). “Infrastructure Geographical Disadvantage, Transport Costs, and Trade.” The World Bank Economic Review 15(3), 451-479.

Linders, Gert-Jan et al. Cultural and Institutional Determinants of Bilateral Trade Flows. Rochester, NY: Social Science Research Network, 2005. papers.ssrn.com. Web. 29 Apr. 2015.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association, 58(304), 993-1010.

Martinez-Zarzoso, I. and Marquez-Ramos L. (2008). “The Effect of Trade Facilitation on Sectoral Trade.” The BE Journal of Economic Analysis and Policy. 8(1).

Mayer, T. and Zignago, S. (2011). “Notes on CEPII's Distances Measures: The GeoDist Database.”

McCallum, J. (1995). “National Borders Matter: Canada-US Regional Trade Patterns. The American Economic Review, 615-623.

Melitz J. and Toubal, F. (2014) “Native Language, Spoken Language, Translation and Trade.” Journal of International Economics. Volume 93, Issue 2, pages 351-363.

Rose. Al. K. (2004). “Do WTO Members Have More Liberal Trade Policy?” Journal of International Economics, 63(2), 209-235.

Sala-i-Martin. X. X. (1997). “I Just Ran Two Million Regressions.” The American Economic Review, 178-183.

Shelburne, Robert. (2010). “The Global Financial Crisis and Its Impact on Trade: The World and the European Emerging Economies.” UNECE Discussion Paper No. 2010.2.

Silva J. S. and Tenreyro S. (2006). “The Log of Gravity.” Review of Economics and Statistics, 88(4), 641-658.

Sturzenegger, F. and Zettelmeyer, J. (2006). “Debt Defaults and Lessons from a Decade of Crises.” MIT Press.

Subramanian, Arvind & Wei, Shang-Jin, 2007. "The WTO promotes trade, strongly but unevenly," Journal of International Economics, Elsevier, vol. 72(1), pages 151-175, May.

Tibshirani, R. (1996). “Regression Shrinkage and Selection Via the Lasso”. Journal of the Royal Statistical Society. Series B (Methodological). 267-288.

Valencia, F and Laeven, L (2008). Systematic Banking Crises: A New Database.” IMF

Valencia, F and Laeven, L (2012). Systematic Banking Crises Database: An Update.” IMF

Varian. H. (2014) “Big Data: New Tricks for Econometrics”. Journal of Economic Perspectives. Volume 28, Number 2, Spring 2014, pp. 3-27(25)

Tables

Table 3- 1: Summary Statistics

	mean	*sd*	*min*	*max*
Ln of Real Bilateral Trade Flows	17.0	2.97	8.14	26.4
ln Dist	8.70	0.78	4.09	9.89
ln of Dist between Capitals	8.70	0.78	4.09	9.89
ln of Weighted Distance	8.71	0.77	4.74	9.89
ln of CES Weighted Distance	8.70	0.79	4.66	9.89
Product of GDPs	22.8	2.29	14.4	31.6
Contiguous	0.025	0.16	0	1
Either Island	0.042	0.20	0	1
Either Landlocked	0.19	0.39	0	1
Share Official Language	0.16	0.37	0	1
9%+ Speak Language	0.18	0.39	0	1
Former Colony	0.034	0.18	0	1
Common Colonizer	0.060	0.24	0	1
Common Legal Origin	0.37	0.48	0	1
Religious Distance	0.70	0.69	0	2.14
Human Capital (product)	1.54	0.40	0.099	2.48
Physical Capital (product)	21.1	1.77	13.4	24.9
Arable Land (product)	-3.21	1.83	-14.3	2.13
WTO/GATT	0.96	0.19	0	1
Regional Trade Agreement	0.051	0.22	0	1

Common Currency	0.0046	0.068	0	1
Capital Openness	0.21	0.26	0	1
Exchange Rate Volatility	0.35	1.80	0.035	65.0
Either Fixed Exchange Rate	0.49	0.50	0	1
Both Fixed Exchange Rate	0.095	0.29	0	1
Either Crawling Peg Exch Rate	0.53	0.50	0	1
Both Crawling Peg Exchange Rate	0.090	0.29	0	1
Either Moving Band Exch Rate	0.39	0.49	0	1
Both Moving Band Exchange Rate	0.045	0.21	0	1
Debt Crisis 3yr Window	0.056	0.23	0	1
Banking Crisis 3yr Window	0.15	0.35	0	1
Currency Crisis 3yr Window	0.19	0.39	0	1

Table 3- 2: OLS, Lasso, and Post-Lasso Models

	(1)	(2)	(3)
Variable	OLS	Lasso	Post Lasso
ln Dist	0.500***	0	
ln of Dist between Capitals	0.201	0	
ln of Weighted Dist	-1.875***	-0.994	-1.690***
ln of CES Weighted Dist	0.0205	-0.133	0.534***
ln of Product of GDPs	0.729***	0.559	0.636***
Contiguous	0.614***	0.431	0.601***
Either Island	0.475***	0.337	0.476***
Either Landlocked	-0.531***	-0.211	-0.520***
Share Official Language	0.0646*	0.066	0.0470
9%+ Speak Language	0.311***	0.296	0.327***
Former Colony	0.855***	0.738	0.867***
Common Colonizer	0.396***	0.319	0.394***
Common Legal Origin	0.282***	0.25	0.286***
Religious Distance	-0.209***	-0.174	-0.215***
Human Capital (product)	0.742***	0.389	0.707***
Physical Capital (product)	-0.129***	0	
Arable Land (product)	-0.231***	-0.15	-0.249***
WTO/GATT	0.335***	0.171	0.353***
Regional Trade Agreement	-0.00572	0	
Common Currency	0.448***	0.109	0.481***

Capital Openness	-0.00298	0	
Exchange Rate Volatility	-0.0120***	0	
Either Fixed Exch Rate	0.0414**	0	
Both Fixed Exch Rate	0.0949***	0	
Either Crawling Peg Exch Rate	0.177***	0.073	0.130***
Both Crawling Peg Exch Rate	0.0814***	0	
Either Moving Band Exch Rate	0.0758***	0	
Both Moving Band Exch Rate	0.0871***	0	
Debt Crisis 3yr Window	-0.0159	0	
Banking Crisis 3yr Window	0.0513***	0	
Currency Crisis 3yr Window	-0.0142	0	
Constant	0.0668***	0.064	0.0667***
Observations	152,213	152,213	152,213

Notes: Dependent variable is the log of real bilateral trade flows for all regression specifications. All regressions include year, exporter, and importer fixed effects. For OLS, t statistics are presented in parentheses. Robust standard errors. T-statistics are hidden in this version of the table. *p<0.05, ** p <0.01, *** p<0.001

Table 3- 3: Bayesian Model Averaging and Extreme Bound Analysis Baseline Results

	Bayesian Model Averaging			Extreme Bound Analysis	
	PIP	Post Mean	Post SD	UB to LB	Robust
ln of CES Weighted Dist	1	-1.113	0.007	(-1.34, -1.1)	y
Product of GDPs	1	0.736	0.016	(0.75, 0.99)	y
Contiguous	1	0.405	0.031	(0.37, 0.69)	y
Either Island	1	0.488	0.026	(0.46, 0.77)	y
Either Landlocked	1	-0.514	0.047	(-0.67, -0.38)	y
Share Official Language	0.031	0.002	0.011	(0.19, 0.78)	y
9%+ Speak Language	1	0.364	0.017	(0.21, 0.76)	y
Former Colony	1	0.866	0.027	(0.83, 1.36)	y
Common Colonizer	1	0.413	0.023	(0.35, 0.77)	y
Common Legal Origin	1	0.287	0.011	(0.27, 0.51)	y
Religious Distance	1	-0.216	0.009	(-0.3, -0.17)	y
Human Capital (product)	1	0.75	0.068	(0.78, 1.37)	y
Physical Capital (product)	1	-0.132	0.012	(-0.18, -0.09)	y
Arable Land (product)	1	-0.229	0.018	(-0.41, -0.27)	y
WTO/GATT	1	0.348	0.025	(0.12, 0.41)	y
Regional Trade Agreement	0	0	0	(-0.18, 0.07)	n
Common Currency	1	0.353	0.062	(0.11, 0.84)	y
Capital Openness	0	0	0	(-0.21, 0.04)	n
Exchange Rate Volatility	1	-0.012	0.002	(-0.01, 0)	n

Either Fixed Exch Rate	0.051	0.002	0.01	(-0.14, -0.02)	y
Both Fixed Exch Rate	0.893	0.071	0.031	(-0.1, 0.1)	n
Either Crawling Peg ER	1	0.153	0.015	(0.06, 0.18)	y
Both Crawling Peg ER	0.449	0.027	0.032	(-0.09, 0.07)	n
Either Moving Band ER	0.924	0.048	0.02	(-0.02, 0.11)	n
Both Moving Band ER	0.174	0.013	0.029	(-0.07, 0.14)	n
Debt Crisis 3yr Window	0	0	0	(-0.08, 0.04)	n
Banking Crisis 3yr Window	0.481	0.021	0.024	(0.02, 0.1)	y
Currency Crisis 3yr Window	0.006	0	0.002	(-0.06, 0.03)	n

Dependent variable is real bilateral trade flows between trading partners. All specifications include exporter, importer, year fixed effects. PIP is the "posterior inclusion probability" and reflects our relative confidence that the true model contains any particular regressor. Post Mean is the weighted average over the posterior estimates of the regressor. Post SD is the standard deviation of coefficient's posterior distribution. LB refers to highest value of the parameter in all of the models estimated, UB refers to the highest value of the parameter estimated. Leamer considers an estimate "robust" if its highest and lowest estimated value does not include zero.

Table 3- 4: Summary of Variable Robustness Across Methods

	OLS	Lasso	BMA	EBA
ln Dist	▪		▪	▪
ln of Dist between Capitals			▪	▪
ln of Weighted Distance	▪	▪	▪	▪
ln of CES Weighted Distance		▪	▪	▪
Product of GDPs	▪	▪	▪	▪
Contiguous	▪	▪	▪	▪
Either Island	▪	▪	▪	▪
Either Landlocked	▪	▪	▪	▪
Share Official Language	▪	▪		▪
9%+ Speak Language	▪	▪	▪	▪
Former Colony	▪	▪	▪	▪
Common Colonizer	▪	▪	▪	▪
Common Legal Origin	▪	▪	▪	▪
Religious Distance	▪	▪	▪	▪
Human Capital (product)	▪	▪	▪	▪
Physical Capital (product)	▪		▪	▪
Arable Land (product)	▪	▪	▪	▪
WTO/GATT	▪	▪	▪	▪
Regional Trade Agreement				
Common Currency	▪	▪	▪	▪

Capital Openness				
Exchange Rate Volatility	▪		▪	
Either Fixed Exchange Rate	▪			▪
Both Fixed Exchange Rate	▪		▪	
Either Crawling Peg Exch Rate	▪	▪	▪	▪
Both Crawling Peg Exchange Rate	▪			
Either Moving Band Exch Rate	▪		▪	
Both Moving Band Exchange Rate	▪			
Debt Crisis 3yr Window				
Banking Crisis 3yr Window	▪		▪	▪
Currency Crisis 3yr Window				

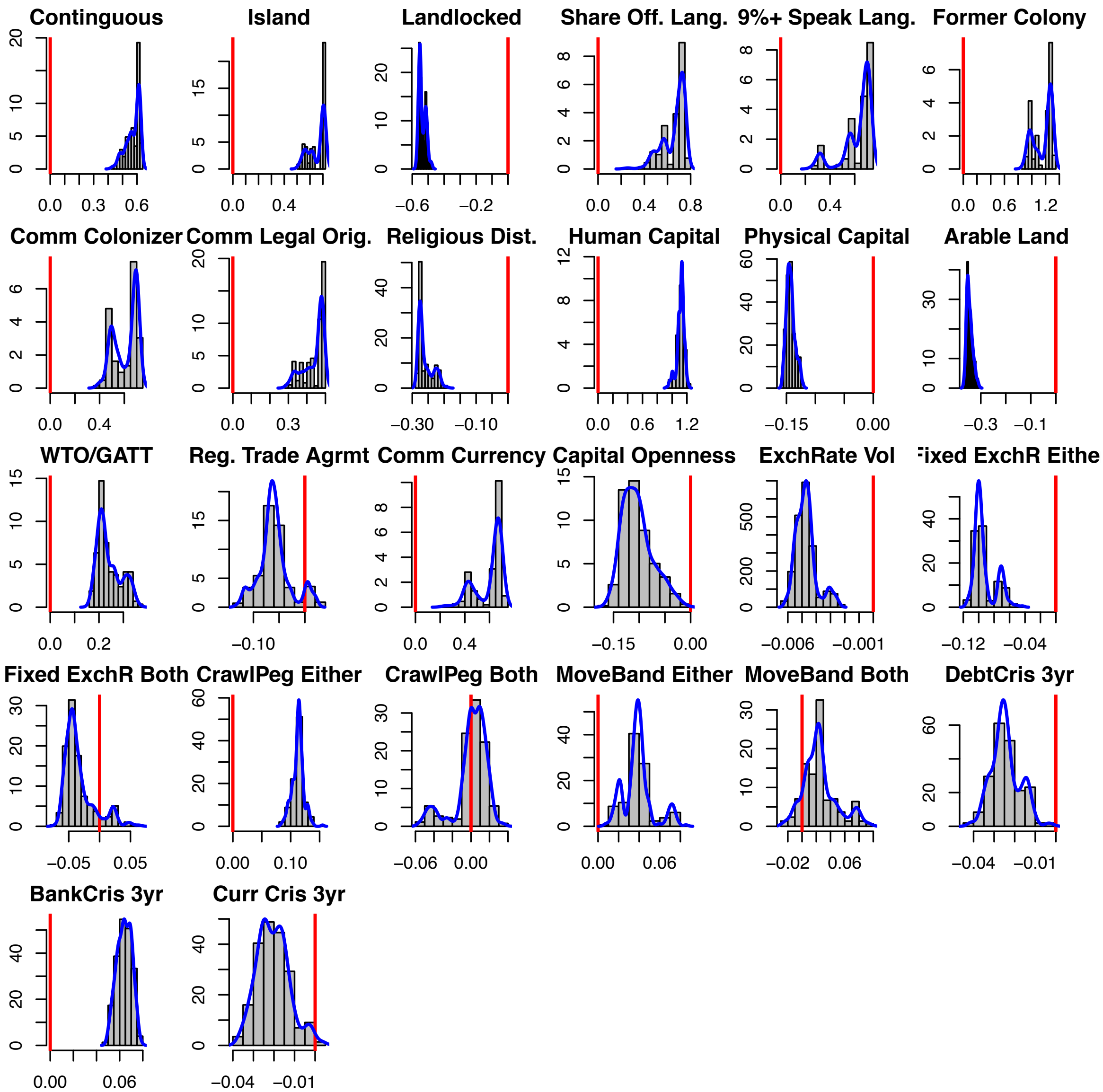


Figure 3- 1: Distributions of parameter estimates generated by Extreme Bounds Analysis

Figure shows histograms of coefficient probability densities from Extreme Bounds Analysis estimation. The vertical red line shows where zero lies on the x-axis. Blue lines show the kernel density smoothed histograms.

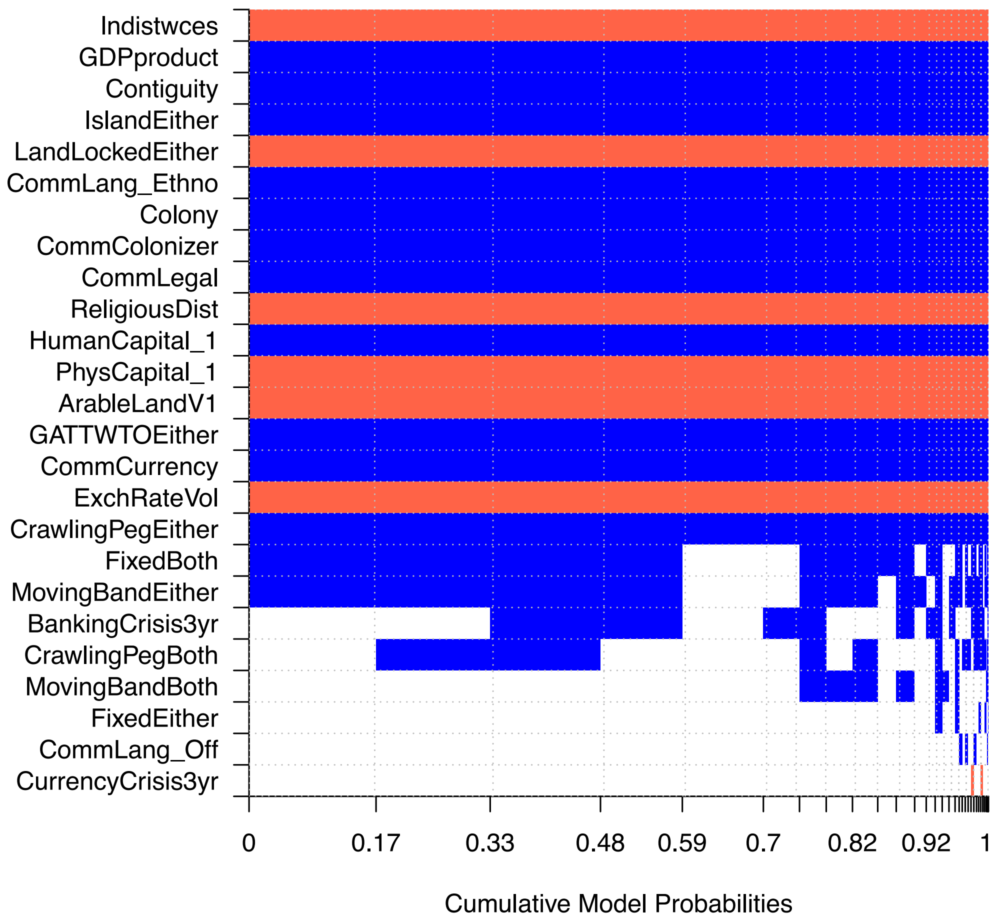


Figure 3- 2: Bayesian Model Averaging, Posterior Model Probabilities

Figure shows posterior model probabilities and the associated variables included in the models. Variables shaded red have negative estimated coefficients; blue shading indicates positive estimated coefficients. Blank shading indicates variable not included in the given model.

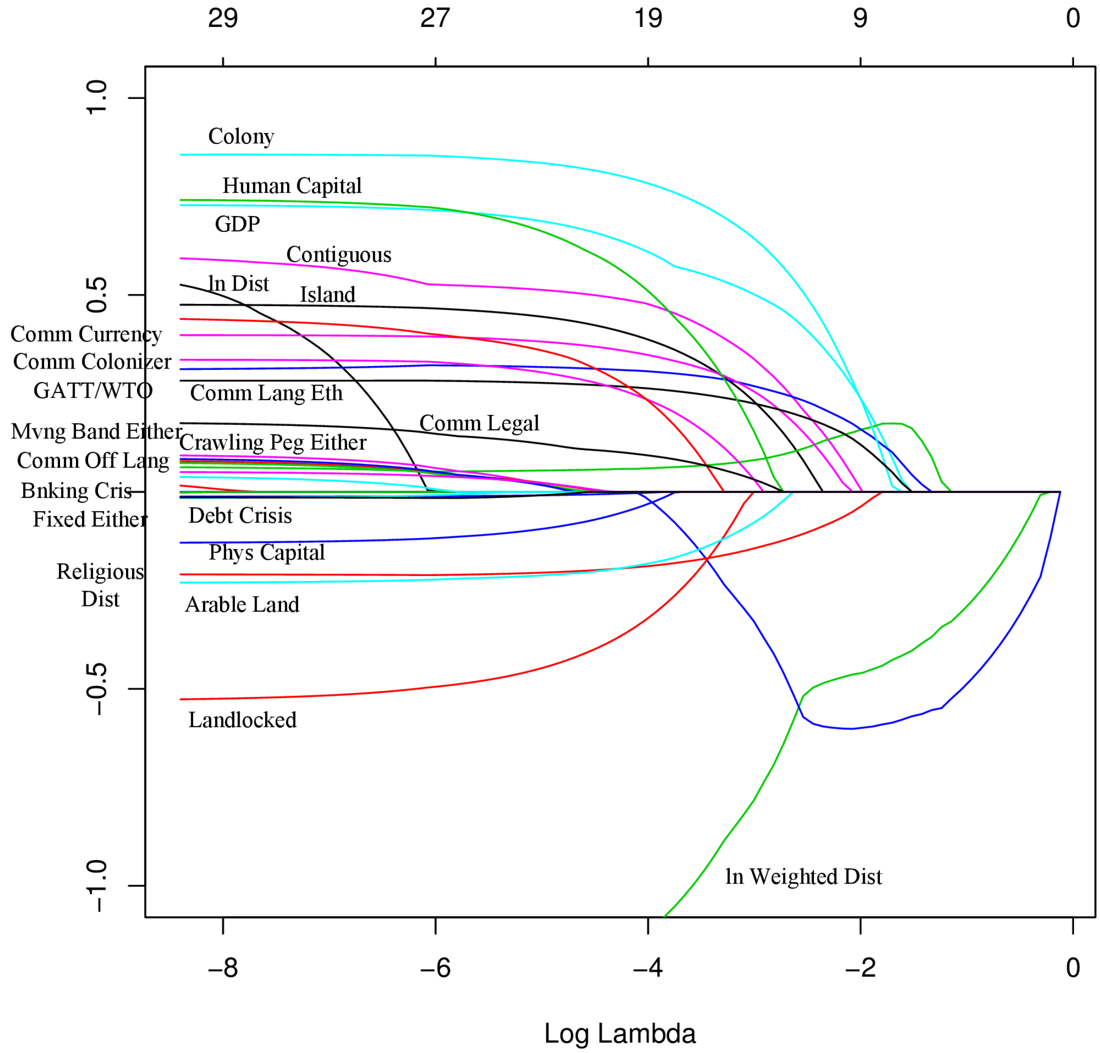


Figure 3- 3: Shrinkage Path for Lasso Estimation

CUMULATIVE BIBLIOGRAPHY

Afzal, M., Hersh, J., and Newhouse, D. (2016). “Building a Better Model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka”. Mimeo, World Bank.

Anderson, J. (1979) “A theoretical foundation to the gravity equation”. *American Economic Review*, 69(1), 106-116

Anderson, J. and van Wincoop, J. (2003) “Gravity with gravitas: A solution to the border puzzle,” *American Economic Review,* 93(1), 170-192.

Anselin, Luc. *Spatial econometrics: Methods and models.* Vol. 4. Springer Science & Business Media, 2013.

Anselin, Luc, et al. (1996) “Simple diagnostic tests for spatial dependence.” *Regional Science and Urban Economics,* 26(1), 77-104.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483-485.

Athey, S., & Imbens, G. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. arXiv preprint arXiv:1504.01132.

Baier, S. and Bergstrand, J, (2007) “Do free trade agreements actually increase members' international trade?” *Journal of International Economics*, 71(1), 72-95.

Baldwin, R. E., and Murray, T. (1977). “MFN tariff reductions and developing country trade benefits under the GSP.” *The Economic Journal*, 30-46.

Baldwin R. and Taglioni, D. “Gravity for Dummies and Dummies for Gravity Equations.” (No. w12516). National Bureau of Economic Research.

Barro, R. J. and Lee, J. W. (2013). “A new data set of educational attainment in the world, 1950-2010.” *Journal of Development Economics*, 104, 184-198.

Bay, H., T. Tuytelaars, and L. V. Gool. (2006) SURF: Speeded Up Robust Features. *Lecture Notes in Computer Science*, 3951, 404–417.

Beim D. and Calomiris, C. (2001). *Emerging Financial Markets*. Irwin Professional Pub.

Belloni, Alexandre and Chernozhukov, V. (2013). “Least squares after model selection in high-dimensional sparse models” *Bernoulli*, 19(2).

Belloni, A. Chernozhukov, V. and Hansen C. (2014). “Inference on Treatment Effects after Selection among High-Dimensional Controls.” *The Review of Economic Studies,* 81(2), 608-650.

Besley, T., & Ghatak, M. (2006). "Public goods and economic development". *Understanding Poverty*. (pp. 285-302). Oxford: Oxford University Press.

Bhagwati, J. (2008). *Termites in the trading system: How preferential agreements undermine free trade.* Chicago: Oxford University Press.

Broda, Ch. and Romalis. J. (2009). "The Welfare Implications of Rising Price Dispersion." University of Chicago mimeo.

Bun, M. and F. Klaassen (2007) "The Euro effect on trade is not as large as commonly thought." *Oxford Bulletin of Economics and Statistics* 69(4), 473-496.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). "Alternative transformations to handle extreme values of the dependent variable." *Journal of the American Statistical Association,* 83(401), 123-127.

Carrere, C. (2006). "Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model." *European Economic Review*, 50(2), 223-247.

Chaney, T. (2013). "The gravity equation in international trade: An explanation." NBER Working Paper No. 19285.

Clausing, K. A. (2001). "Trade creation and trade diversion in the Canada-United States free trade agreement." *Canadian Journal of Economics*, 34(3), 677-696.

Dalal, N., and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition (CVPR),* San Diego, CA, 2005, pp. 886-893.

De Haan, J. and Sturm, J. E. (200)0. "On the relationship between economic freedom and economic growth." *European Journal of Political Economy*, 16(2), 215-241.

De Loecker, Jan. 2013. "Detecting learning by exporting." *American Economic Journal: Microeconomics*, 5(3), 1-21.

De Sousa, J. "The currency union effect on trade is decreasing over time." *Economics Letters* 117(3), 917-920.

Department of Census and Statistics and World Bank, 2015 "The Spatial Distribution of Poverty in Sri Lanka", available at:
http://www.statistics.gov.lk/poverty/SpatialDistributionOfPoverty2012_13.pdf

Donaldson D., and Storeygard A. "Big grids: Applications of remote sensing in economics", *forthcoming, JEP.*

Drukker, David M., Ingmar R. Prucha, and Rafal Raciborski. "Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances." University of Maryland, Department of Economics (2011).

Edwards, L., & Lawrence, R. (2014). “AGOA rules: the intended and unintended consequences of special fabric provisions.” *African Successes: Modernization and Development.* University of Chicago Press.

Egger P. and Nelson, D. (2011) “How bad is anti-dumping? Evidence from panel data”, *Review of Economics and Statistics*, 93(4), 1374-1390.

Egger, Peter H., and Andrea Lassmann. (2012) “The language effect in international trade: A meta-analysis.” *Economics Letters,* 116(2), 221–224.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. *Econometrica*, 71(1), 355-364.

Elbers, Chris, Peter F. Lanjouw, and Phillippe G. Leite. “Brazil within Brazil: Testing the poverty map methodology in Minas Gerais.” World Bank Policy Research Working Paper Series, (2008).

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). “Mapping city lights with nighttime data from the DMSP Operational Linescan System.” *Photogrammetric Engineering and Remote Sensing*, 63(6), 727-734.

Engstrom, R., Ashcroft, E., Jewell, H., & Rain, D. (2011, April). Using remotely sensed data to map variability in health and wealth indicators in Accra, Ghana. In *2011 Joint Urban Remote Sensing Event (JURSE)*, (pp. 145-148). IEEE.

Engstrom, R., Sandborn, A., Yu, Q. Burgdorfer, J., Stow, D., Weeks, J., and Graesser, J. (2015) Mapping Slums Using Spatial Features in Accra, Ghana. *Joint Urban and Remote Sensing Event Proceedings (JURSE), Lausanne, Switzerland*, 10.1109/JURSE.2015.7120494

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.

Feenstra, R. C. Lipsey, R. E., Deng H., Ma, A. C., Mo, H. (2005). World Trade Flows: 1962-2000 (No w11040). National Bureau of Economic Research.

Felbermayr, Gabriel J., and Farid Toubal. (2010) “Cultural Proximity and Trade.” *European Economic Review* 54(2), 279–293. ScienceDirect. Web. 23 Apr. 2015.

Fernandez, C, Ley, E. and Steel, M. (2001) “Model Uncertainty in Cross-Country Growth Regressions”, *Journal of Applied Econometrics*, 16(5), 563-576.

Foster, James; Joel Greer; Erik Thorbecke (1984). “A class of decomposable poverty measures”. *Econometrica,* 3(52), 761–766.

Frankel, J. (2014). “Mauritius: African success story.” *African Successes: Sustainable Growth.* Chicago: University of Chicago Press.

Frankel J. and Rose A. (2002) “An estimate of the effect of common currencies on trade and income.” *The Quarterly Journal of Economics,* 117(2), 437-466.

Frankel. K. Stein, E. and Wei S. J. (1995). “Trading blocks and the Americas: The natural, the unnatural, and the super-natural.” *Journal of Development Economics*, 47(1), 61-95.

Frazer, G., & Van Biesebroeck, J. (2010). “Trade growth under the African growth and opportunity act.” *The Review of Economics and Statistics,* 92(1), 128-144.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer Series in Statistics.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica*, 1(4), 387-401.

Gabor, D. (1946). “Theory of communication.” *Journal of the Optical Society of America-A*, 2(2), 1455-1471

Gentle, J. E., Härdle, W. K., & Mori, Y. (Eds.). (2012). *Handbook of computational statistics: concepts and methods*. Springer Science & Business Media.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2015). “Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life.” (No. w21778). National Bureau of Economic Research.

Glick, R. and Rose, A. (2002). “Does a currency union affect trade? The time series evidence,” *European Economic Review*, 46(6), 1125-1151.

Graesser, J., A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, (2012) “Image based characterization of formal and informal neighborhoods in an urban landscape,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4), 1164-1176.

Granger, C. and H. Uhlig (1990) “Reasonable extreme bounds analysis,” *Journal of Econometrics*, 44, 159-170.

Halvorsen, Robert and Raymond Palmquist *The American Economic Review*, 70(3), 474-475

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

Head, K. and Mayer, T (2004). “The empirics of agglomeration and trade.” *Handbook of Regional and Urban Economics*, 4, 2609-2669.

Head, Keith, Thierry Mayer, and John Ries (2010). “The erosion of colonial trade linkages after independence.” *Journal of International Economics,* 81(1), 1–14.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). “Measuring economic growth from outer space.” *The American Economic Review*, 102(2), 994-1028.

Heston, A., Summers, R. and Aten, B. (2012). “Penn World Table Version 7.1” Philadelphia: University of Pennsylvania.

Hidalgo, C. A., & Hausmann, R. (2009). “The building blocks of economic complexity.” *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10570-10575.

Holden, M., & McMillan, L. (2006). “Do free trade agreements create trade for South Africa?” *Economic Research Southern Africa.*

Huettner, Frank, and Marco Sunder. (2012) “Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values.” *Electronic Journal of Statistics* 6, 1239-1250.

Israeli, Osnat. (2007) “A Shapley-based decomposition of the R-square of a linear regression.” *The Journal of Economic Inequality*, 5(2), 199-212.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). “Combining satellite imagery and machine learning to predict poverty.” *Science*, 353(6301), 790-794.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). “Prediction policy problems.” *The American Economic Review*, 105(5), 491-495.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).

Krstajic, Damjan et al. (2014). “Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models.” *Journal of Cheminformatics* 6(1), 10. PMC. Web. 22 Apr. 2015.

Krueger, A. O. (1999). “Trade creation and trade diversion under NAFTA.” (No. w7429). National Bureau of Economic Research.

Leamer, E. E. (1983). “Let's take the con out of econometrics.” *The American Economic Review*, 73(1), 31-34.

Leamer, E. E. (1985). “Sensitivity analyses would help.” *The American Economic Review*, 75(3), 308-313.

Leamer, E. E. (2008) “Extreme bounds analysis.” The New Palgrave Dictionary of Economics, Second Edition, 2008.

Levine, R. and Renelt D. (1992). “A sensitivity analysis of cross-country growth regressions.” *The American Economic Review*, 82(4), 942-963.

Lewer, J and Van den Berg (2007). “Religion and international trade: Does the sharing of a religious culture facilitate the formation of trade networks?” *American Journal of Economics and Sociology*, 66(4), 765-794.

Limao, N. and Anthony Venables. (2001). “Infrastructure geographical disadvantage, transport costs, and trade.” *The World Bank Economic Review* 15(3), 451-479.

Linders, Gert-Jan et al. Cultural and Institutional Determinants of Bilateral Trade Flows. Rochester, NY: Social Science Research Network, 2005. papers.ssrn.com. Web. 29 Apr. 2015.

Lovell, M. C. (1963). “Seasonal adjustment of economic time series and multiple regression analysis.” *Journal of the American Statistical Association*, 58(304), 993-1010.

MacKinnon, J. G., & Magee, L. (1990). “Transforming the dependent variable in regression models.” *International Economic Review*, 31(2), 315-339.

Martinez-Zarzoso, I. and Marquez-Ramos L. (2008). “The effect of trade facilitation on sectoral trade.” *The B.E. Journal of Economic Analysis and Policy*. 8(1), 1-46.

Marx, B., Stoker, T. M., & Suri, T. (2013). The Political Economy of Ethnicity and Property Rights in Slums: Evidence from Kenya. Accessed 11/9/2017 from http://cega.berkeley.edu/assets/cega_events/53/WGAPE_Sp2013_Suri.pdf

Mayer, T. and Zignago, S. (2011). “Notes on CEPII's Distances Measures: The GeoDist Database.” CEPII Working Paper 2011-25.

McCallum, J. (1995). “National borders matter: Canada-US regional trade patterns. *The American Economic Review*, 85(3), 615-623.

McMillan, Margaret S., and Dani Rodrik. (2011) “Globalization, structural change and productivity growth.” (No. w17143). National Bureau of Economic Research.

Melitz J. and Toubal, F. (2014) "Native language, spoken language, translation and trade." *Journal of International Economics*, 93(2), 351-363.

Michalopoulos, S. (2012). "The origins of ethnolinguistic diversity." *The American Economic Review*, 102(4), 1508.

Mooney, D. F., Larson, J. A., Roberts, R. K., & English, B. C. (2009). Economics of the variable rate technology investment decision for agricultural sprayers. In Southern agricultural economics association annual meeting, Atlanta, Georgia, January.

Mullahy, J. (1998). "Much ado about two: reconsidering retransformation and the two-part model in health econometrics." *Journal of Health Economics*, 17(3), 247-281.

Mullainathan, S. (2014, August). Bugbears or legitimate threats? (social) scientists' criticisms of machine learning? In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 4-4). ACM.

Newhouse, David; Suarez Becerra, Pablo; Doan, Dung. 2016. *Sri Lanka Poverty and Welfare: Recent Progress and Remaining Challenges*. World Bank, Washington, DC. © World Bank. https://www.openknowledge.worldbank.org/handle/10986/23794 License: CC BY 3.0 IGO.

Nouve, Kofi, and John Staatz. "The African Growth and Opportunity Act and the Latent Agricultural Export Response in Sub-Saharan Africa." Annual Meeting of the American Agricultural Economics Association, Montreal, Quebec, Canada, July. N.p., 2003. 27–30.

Nunn, N., & Puga, D. (2012). "Ruggedness: The blessing of bad geography in Africa." *Review of Economics and Statistics*, 94(1), 20-36.

Pesaresi, M., A. Gerhardinger, and F. Kayitakire, (2008) "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3), 180-192.

Rose. Al. K. (2004). "Do WTO members have more liberal trade policy?" *Journal of International Economics*, 63(2), 209-235.

Rotunno, Lorenzo, Pierre-Louis Vézina, and Zheng Wang (2013) "The rise and fall of (Chinese) African apparel exports." *Journal of Development Economics,* 105, 152-163.

Sala-I-Martin. X. X. (1997). "I just ran two million regressions." *The American Economic Review*, 87(2), 178-183.

Sandborn, A. and Engstrom, R (In Press) Determining the Relationship Between Census Data and Spatial Features Derived From High Resolution Imagery in Accra, Ghana. *IEEE*

Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Special Issue on Urban Remote Sensing.

Sandrey, R. (2006) "Trade Creation and Trade Diversion resulting from SACU trading agreements." Working Paper No 11. Stellenbosch:

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). "Data deprivation: another deprivation to end." World Bank Policy Research Working Paper, (7252).

Shelburne, Robert. (2010). "The Global Financial Crisis and Its Impact on Trade: The World and the European Emerging Economies." UNECE Discussion Paper No. 2010.2.

Shorrocks, Anthony F. (2013). "Decomposition procedures for distributional analysis: A unified framework based on the Shapley value." *Journal of Economic Inequality*, 11,99-126.

Silva J. S. and Tenreyro S. (2006). "The log of gravity." *Review of Economics and Statistics*, 88(4), 641-658.

Smith, S. W., *The scientist and engineer's guide to digital signal processing*. San Diego, CA: California Technical Publishing, 1997.

Sturzenegger, F. and Zettelmeyer, J. (2006). *Debt Defaults and Lessons from a Decade of Crises.* MIT Press.

Subramanian, Arvind & Wei, Shang-Jin, 2007. "The WTO promotes trade, strongly but unevenly," *Journal of International Economics,* 72(1), 151-175.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. 73(3), 267-288.

Tucker CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment,* 8, 127-150.

Watmough, G. R., Atkinson, P. M., Saikia, A., & Hutton, C. W. (2016). "Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: An example from Assam, India." *World Development*, 78, 188-203.

Valencia, F and Laeven, L (2008). Systematic Banking Crises: A New Database." IMF

Valencia, F and Laeven, L (2012). Systematic Banking Crises Database: An Update." IMF

Varian, H. R. (2014). "Big data: New tricks for econometrics." *The Journal of Economic Perspectives*, 28(2), 3-27.

Viner, J. (1950). *The custom unions issue*. New York: Carnegie Endowment for World Peace.

Wang, L. and D. He (1990). "Texture classification using texture spectrum," *Pattern Recognition*, 23(8), 905-910.

Wong, T. H., Mansor, S. B., Mispan, M. R., Ahmad, N., & Sulaiman, W. N. A. (2003, May). Feature extraction based on object oriented analysis. In *Proceedings of ATC 2003 Conference* (Vol. 2021).

Yu, W. P., G. W. Chu, and M. J. Chung (1999). "A robust line extraction method by unsupervised line clustering," *Pattern Recognition*, 32(4), 529-546.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]