The Journal of
International
Policy Solutions

THE JOURNAL *of* INTERNATIONAL POLICY SOLUTIONS

TABLE *of* CONTENTS

The goal of the Journal of International Policy Solutions (JIPS) is to facilitate prescient discussion about challenging policy questions. In this edition, authors address issues ranging from the side effects of sustainable tourism certificates to the impact of shrinking negative lists for trade in China.

As the editor, what has excited me the most about this collection of graduate student policy papers is the diverse range of methods my peers have brought to bear in their policy analysis. One paper uses simple data visualization to demonstrate a lack of gender parity in policy course syllabi at our School of Global Strategy and Policy (GPS) at UC San Diego. Another uses advanced Geographic Information Systems (GIS) analysis to document the extent of the Rohingya refugee crisis and make geographically driven policy recommendations.

The econometric analysis in the journal addresses questions such as, does aid follow conflict in Afghanistan? And, does mobile money adoption affect international remittances? As policy students newly empowered with econometric tool kits, the diverse methods which the authors pull on to address these questions is fascinating to compare.

Amidst today's troubling trend towards blind and excessive nationalism in all corners of the world, facts are easily and frequently forgotten. A comfort for policy analysts-in-training in the face of this tendency is that the depth and breadth of our capacity is ever expanding. With the help of more rigorous policy analysis, as highlighted in this journal, real evidence can be uncovered to better fight against the war on facts.

For this reason I am extremely proud of this edition. I would like to extend my deepest thanks to the willing authors, team of contributing editors, supportive GPS staff, our faculty advisor Professor Liz Lyons, and especially to this edition's managing director Savitri Arvey. Without all these contributions, the journal would simply not have been possible.

Enjoy,

Alicia Krueger
Editor-in-Chief

INFASTRACTURE IMPROVEMETNS IN MEXICAN HOUSEHOLDS: THE "PISO FIRME" PROGRAM

Joanna Valle Luna

University of California San Diego, School of Global Policy and Strategy

According to the Mexican national census, in 2000, more than 1.5 million households, many of which were located in rural low-income municipalities, had dirt floors, which were associated with high levels of marginalization and gastrointestinal diseases. The recently elected federal administration targeted these households with a program called Piso Firme (Firm/Cement floor) with the goal of improving living standards and health conditions in the 50 indigenous municipalities with the lowest Human Development Index (HDI) through the installation of cement floors. Using an innovative program design that shifted responsibility to the beneficiaries, households received materials and instructions to build a modest concrete floor in a room of their home with the promise that the project could be completed in less than a day of work. Wide approval of this program led to its expansion to the 125 poorest municipalities during the following presidential term (2006-2012), and today it remains one of the country's main social programs.

This paper uses a fuzzy regression discontinuity (RD) design to evaluate the results of both pilots in infrastructure improvements and its impact on child mortality. The first pilot showed insignificant results on the household. However, the second pilot increased houses with cement floor in the treated municipalities by 9%, which led to a 5.6% reduction in the mortality rate of children between 0 and 9 years old due to diarrhea and other intestinal-parasite infections. Hence, this evidence supports that under quasi-experimental conditions, which are unlikely settings for infrastructure program analysis, an improvement in infrastructure can have a positive impact on human living standards.

INTRODUCTION

According to the 2000 Mexican national census, more than 1.5 million households, many of which were located in rural low-income municipalities, had dirt floors. These were associated with high levels of marginalization and gastrointestinal diseases. The recently elected federal administration targeted these households through a program called *Piso Firme* (Firm/Cement floor) with the goal of improving living standards and health conditions in the 50 indigenous municipalities with the lowest Human Development Index (HDI) through the installation of cement floor. To achieve this, the program beneficiaries received

materials to apply 5-inch thick concrete in an area of up to 540 feet squared or a full room, along with self-installation instructions. The government advertised that through the program beneficiaries would install a firm floor in their households after only one day of labor. The wide approval of this program led to its expansion to the 125 poorest municipalities during the following presidential term (2006-2012). It remains one of the country's main social programs to improve infrastructure and health outcomes today.

Since 2000, there has been a vast amount of literature analyzing factors that affect children's health. This is partially due to the United Nations member states unanimous adoption of the Millennium Declaration,

which outlined the main development challenges to be achieved by 2015 including improvements in child malnutrition, health conditions, and mortality. Therefore, it was not a surprise that many countries subsequently designed and launched programs that aimed to fulfill or at least make progress on these goals.

The literature on child mortality, health, and potential interventions has focused on three main arguments. The first camp, which includes academics such as Marianne Fay (2005), argues that income levels clearly affect health outcomes.[1] For example, Easterly (1999) found a positive correlation between income and child mortality, as well as the child's nutritional status.[2] Other authors have focused on the behavioral aspects of health, through initiatives that emphasize "handwashing and boiling water" interventions. For example, Jennifer Davis et. al. (2011)[3] studied the effects of informational interventions in safe water storage and hand hygiene in Tanzania, finding better results when the messages were targeted and backed up by information that the beneficiaries can relate to.

The third factor that the literature explores, infrastructure, is the most relevant for the purposes of this paper. Jalan and Ravallion (2001)[4] claim that it is necessary to include an infrastructure component in every health intervention in order to make it effective and long-lasting. Even though their work focused on piped water infrastructure in India, they found that among poor households, diarrhea is less prevalent and less severe among young children who live in households with piped water. Hence, this improvement in infrastructure helps poor households to overcome health "shocks" due to poor environmental conditions.

Literature that focuses on improving child health through infrastructure interventions has mainly focused on access to safe water or health care services. There is no specific reference to changing the physical household infrastructure – the house itself – as a sanitary improvement. On the other hand, research in the medical field has used experimental studies to look at the correlation between dirt floors and infectious diseases, although these experiments have used a small sample size. Zeledón[5] examines the negative impact of dirt floors in 50 rural households in Costa Rica, which made the population more vulnerable to Chagas disease. In this case, since mud houses the perfect environment for the proliferation of the insect that carries this infection, the intervention included the installation of cement floors. In another study of 184 households in Panama, Krause et al. found that kids playing on dirt floors is associated with a 2% to 3% higher incidence of *Ascaris* infections.[6]

As seen in the literature review, most of the studies that focus on the impacts of household infrastructure on health outcomes have been done under a strict experimental approach in order to ensure the collection of sufficient data, which can be a challenge for a regional program such as *Piso Firme.* Overall, all these studies show at a micro-household level that there is a correlation between water, sanitation and hygiene conditions – which includes household infrastructure – and health outcomes. In their work on intestinal related diseases, Esrey, Potash, Roberts, and Shiff (1991) reviewed 49 studies show that on average 22% of the reduction in diarrheal morbidity can be explained by infrastructure improvements. [7]

In the case of *Piso Firme*, the program was evaluated in its early stage when it was a state level program. Before the nationwide launch in 2000, *Piso Firme* began as an infrastructure program in the state of Coahuila, Mexico. In 2005, the state government conducted an impact evaluation of the program, and showed that installing cement floors has a significant and robust effect on reducing the number of diarrhea incidences in children between 0 and 5 years. They found a change between 18% and 20%[8], a marginal effect that is consistent with other findings in international studies.[9] However, it

1 Marianne Fay, Danny Leipziger, Quentin Wodon and Tito Yepes, "Achieving Child-Health-Related Millennium Development Goals: The Role of Infrastructure" in *World Development*, Vol. 33, No. 8, 2005, pp. 1267-1284.
2 Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. Quarterly Journal of Economics, 50, 112–1203.
3 Jennifer Davis , The Effects of Informational Interventions on Household Water Management
4 Jalan, J., & Ravallion, M. (2001). Does piped water reduce diarrhea for children in Rural India? Policy Research Working Paper No. 2664, The World Bank, Washington, DC.
5 Rodrigo Zelodón, The Role of Dirt Floors and of firewood in Rural Costa Rica. The American Journal of Tropical Medicine and Hygiene 33(2):232-5 · April 1984
6 Rachel J. Krause et al., Ascaris and hookworm transmission in preschool children from rural Panama: role of yard environment, soil eggs/larvae and hygiene and play behaviors, Cambridge University, 2015.
7 Brenneman, A. (2002). Infrastructure & poverty link- ages: A literature review. Background Report, The World Bank, Washington, DC.
8 Dr. Paul Gertler, Evaluación de Resultados de Impacto del Programa Piso Firme Estado de Coahuila
9 The coefficient between -0.18 and -0.27, differential effect with respect to the median of the population used in the study, could be compared with 22% reduction on average in diarrheal morbidity explained by infrastructure improvements (Esrey, et all, 1991) or be comparable with the results from Jalan and Ravallion showing that disease prevalence (diarrhea) amongst those with piped water would be 21% higher without it.

is important to be aware of the specificities of those studies and how hey differ from this paper. The Mexican evaluation of this state-level phase of the program was mostly carried out in Torreón, Chihuahua's capital city, through a matching approach with households of the neighboring state, Durango, as a counterfactual. For the rest of the findings, as mentioned previously, the results come from a rigorous experimental environment.

Using the standard errors from the Mexican government's model (Coahuila's impact evaluation; s=2.8%), and multiplying that number by two, we can estimate a minimum detectable effect (MDE[10], measured of 5.6% change in the type of floor (from dirt floor to cement) for the treated households in the program. However, for the health impacts, identifying the MDE becomes more complicated because those randomized control trials have the chance to measure the number of diarrhea incidents, as part of the monitoring household follow-up for the impact evaluation, unlike the national level program, which wasn't designed as a randomized control trial. Given that for this study there is no representative data at a national level for child diarrhea incidences, for computing the MDE on this side, we use child mortality rate from the state study as a proxy. The standard errors from the program evaluation in Coahuila (s=9.61%) for the variable of reduction of diarrhea cases in children is used. A way to extrapolate this number into child mortality due to diarrhea is using data from the World Health Organization (WHO) and the World Bank (WB). According to the World Health Organization, every year diarrhea kills some 525,000 children, and the child population for 2016 according is 1.9 billion; the child mortality rate due to diarrhea is roughly 0.02%. Therefore, the expected MDE in terms of a mortality rate s 0.003%, or 0.02% of the MDE (δ=19.23%) of the dependent variable in Coahuila's evaluation.

The objective of the present paper is to evaluate the program using both pilots, the first version that was launched in 2000 and the subsequent expanded-version in 2006. Based on the literature review, this paper poses the research question if due to the change from a dirt to a cement floor, municipalities that were treated in Mexico experienced a reduction in child mortality

10 The MDE is ex-post statistical measurement as a function of the size of the intervention. Given that the standard error is correlated with finding significance in the coefficients, the MDE is a way to answer the question of what size effect could we statistically detect?

related to diarrhea and intestinal infectious illnesses. Therefore, the main hypothesis of the paper stems from the following causal chain: households with poor infrastructure, like dirt floors, are more exposed to an unhealthy environment since mud hosts the bacteria and parasites that produce intestinal infections. Consequently, the populations living with these conditions are more likely to be affected by intestinal diseases, making the occurrences more severe and frequent.

This paper focuses on child health, in specific fatalities related with diarrhea and parasite-related diseases, for vulnerable populations due to their household environment. The main hypothesis is that the change of the floor material improves the health of the household inhabitants and decreases the fatalities due to these types of infections.

Since the Mexican government selected the target population using a cut-off based on the 2000 Municipality Human Development Index, the program is analyzed using a regression discontinuity design. A specific threshold was set to establish which municipalities were treated: the 50 lowest scores for the 2000 pilot and 125 lowest scores in the case of the expanded pilot in 2006, which also included the 50 municipalities from the first pilot. This approach facilitates this analysis, since it tests not only the implementation of the program itself, meaning the installation of the cement floors in households, but also allows for this research to use that variable as an instrument for assessing improvements in health conditions in the treated municipalities.

DATA

The unit of analysis used for this analysis is at the municipality level, which is Mexico's second-level administrative division after the state, and the unit that the government used to target treated entities for the program. To analyze the proximate outcome of the program, two variables were created using data from the National Census for Population and Households carried out by the National Institute of Statistics and Geography in Mexico (*Censo de Población y Vivienda, INEGI*) for the years 2000, 2005 and 2010. The first outcome variable is the percentage of households with respect to the total number of houses in the municipality with cement floors; also, this paper computed the percentage of households with dirt floors. This allows

one to test if the potential increase in the number of houses with cement floors is not only driven by the construction of new houses or urbanization but also improvements in infrastructure. *(See Figure 1, Appendix C)*

From the same census, this paper also considers the percentage of houses without access to piped water and sewer systems as another source of covariance. For all those variables, the data is given as the total number of households and a ratio was generated to obtain the percentage of houses versus the total number per municipality.

It is important to mention some challenges due to the data, particularly for the first pilot. Since the household data used is from 2000, the same year of the implementation of the program, there is no baseline information for the first pilot, and time endogeneity. Unfortunately, data for the previous census (1995) that could have been used as a baseline lacks complete household information. This shortage of information from the 1995 Mexican census is partly explained by the government's struggle to collect data during and after the 1994 economic crisis, commonly known as the "tequila crisis". On the other hand, for the 2006-2010 pilot, the appropriate baseline was identified, since 2005 and 2010 census data were used for this variable.

For the first time in 2000, the United Nations Development Program (UNDP) created the Municipal Human Development Index for Mexico, which synthetizes the performance of the country's municipalities based on three dimensions: health, education and income. This new development indicator not only showed the level of discrepancies within the country[11], but also provided a tool that the Mexican government decided to use to identify the target population for several social programs. This role of targeting through the index is relevant for the upcoming analysis because it facilitated the regression discontinuity approach, but also because it created some challenges. For example, *Figure 2 (Appendix C)* shows discrepancies between the official ranking published by the UN and ranking of the 125 poorest municipalities according to Mexico's Ministry of Social Development (SEDESOL). SEDESOL's

11 To give some context, according to the 2000 ranking, the municipality with the lowest score was Metlatónoc (Guerrero) with a score of 0.3915 similar to Subsaharan African countries (aka Chad or Burkina Faso showed the same level of development in 2015 HDI); while on the other extreme, the highest score (0.9164) corresponds to the Benito Juárez Delegation in Mexico city, which is 2.3 times the one in Metlatónoc and similar to the development levels in Denmark or Germany.

ranking systematically shifted the whole distribution of municipalities to the left, which allowed the government to treat municipalities that were above the official threshold and otherwise would not have been candidates for social programs. The effect on Piso Firme's results will be discussed further on.

For health outcomes, the analysis uses data from the National Health Information System (SINAIS) for the years 2005 and 2010 (to be consistent with the census data), specifically the National Register of Death Certificates in Mexico. In this case, since this is the only dataset not previously aggregated at the municipality level, the analysis first filtered the global database of deaths in Mexico by municipality, cause (intestinal diseases), and then age (children ages 0-9). Since the analysis aims to look at changes in that outcome for every variable used, it will use the first difference.

In terms of the sample selected for the first regression discontinuity model based on the 2000 pilot, the analysis uses 100 municipalities (50 treated and 50 control). For the second phase of the treatment, the analysis uses 250 municipalities (125 treatment and 125 control).

METHODOLOGY

To analyze the *Piso Firme* program and its impact on household infrastructure improvement and the child diarrhea mortality rate, this paper uses a regression discontinuity (RD) Approach. The identifying assumption is that the treatment assignment is based on whether a unit of analysis falls above or below a cut-off point set by the rating variable, generating precisely a discontinuity in the probability of treatment at that point.

The basic regression discontinuity model is presented below:

$$Y_i = \alpha + \beta T_i + f(r_i) + \varepsilon_i$$

Where:

α = the average value of the outcome for municipalities where T=0 after controlling for the running variable;

Y_i = Outcome variable (Percentage change of households with cement floor and child mortality);

T_i = 1 if the municipality is assigned to the treatment group and 0 otherwise;

r_i = Human Development Index (2000), centered at

zero at the cut-point, with a linear form;
ε_i = error term;
β = marginal impact of the program, for the treated units, at the cut-off point.

In the case of Piso Firme, the rating variable is the Human Development Index 2000 for both pilot programs. For the year 2000, the treatment assignment is consistent with the official cut-off of 0.5345 according to the UN Data, as shown in *Figures 3a and 3b in Appendix C.*

The program's rules of operation for the second pilot stipulated that it would expand the treatment to the poorest 125 municipalities according to the Human Development Index, however, there were a few challenges. The government's official list assigned the treated municipalities lower scores than the official HDI score in 2000, systematically shifting the entire distribution to the left. *(Figure 2)* In other words, the government included some municipalities that would not have qualified according to their UN score.

Also, when ranking both lists, there is a 100% compliance rate until the municipality that was ranked 100, and after number 100 there are gaps or inconsistencies in the assignation of the treatment. This means that some municipalities received the treatment that shouldn't have and some municipalities that were eligible were excluded.

Given these characteristics, the methodology selected for this paper is a fuzzy discontinuous design, which provides the best fit for the data and its structure. Even though the data, in terms of the treatment assignment versus the inception score, does not provide the cleanest threshold identification, the context still provides a valid quasi-experimental assumption for the RD. First, *Figures 5a and 5b* shows that both the targeted and excluded municipalities are relatively close to the cut-off point. Hence, it can be assumed that at the threshold, observations assigned to the treatment and control can be considered random. Second, the design still meets the condition that a unit – in this case a municipality – can only be assigned either to control or treatment; hence, even though the design is fuzzy due to the absence of a "clean cut-off" point, a single program is either included or excluded from the program.

Nonetheless, this design could provide interesting results and discussion regarding how political targeting may or may not affect the outcomes. We will review this in more detail in the results section and during the robustness checks, since this running variable is treated as distance from the cutoff, because it is centered at zero with the RD design.

Hence, the fuzzy RD exploit discontinuities in the probability or expected value of treatment conditional on a covariant. In other words, it accounts for the possibility of having no-shows, treatment units below the cut-off line that didn't receive the treatment, and crossovers, which are the municipalities that should have been control units but were treated; the exact situation that we encounter in the data.

The goal of this paper is to show the impact of the program in floor household (HH) improvement, and how that impacts child mortality. Because the actual treatment is the changing from dirt to cement floor in a HH, and a collateral effect that this paper is exploring as an extrapolation of the program is the health outcome, a two-stage least squares (2SLS) design gives us the best methodological approach. Consequently, this design allows to first look if whether or not the program made a significant impact on the percentage of households with cement floor; and after meeting that condition, build an argument of the potential impact for the health outcomes.

Consequently, the 2SLS design estimation equations are as followed:

[1a] *First stage* with the percentage change of households with a cement floor as an outcome variable.

$$cement\ floor_i = \alpha + \beta T_i + f(r_i) + \varepsilon_i$$

This first stage was tested for both pilots, therefore:
For the 1st pilot, T = 1 for 50 municipalities
For the 2nd pilot, T = 1 for 125 municipalities.

[1b] *Also, the reduced form* but now child mortality due to intestinal diseases as the dependent variable was used for the second pilot set of data.

$$child\ mortality_i = \alpha + \beta T_i + f(r_i) + \varepsilon_i$$

[2] The *second stage* is the instrumented version, using (assignation to the program) as an instrumental

variable (IV) for the actual treatment, which is, the improvements in cement floor. This stage is only tested for the second pilot of the program.

$$child\ mortality_i = \alpha + \beta\widehat{Cement\ floor}_i + \varepsilon_i$$

Where, X_i are household infrastructure variables used as control covariates and $\beta\widehat{Cement\ floor}_i$ is the instrumented version of the percentage of houses with a cement floor, with the dummy for treatment as an IV. This works under two assumptions, first that the assignment of a municipality to the treatment led to an improvement in infrastructure with the installation of the cement floor, which is tested with the first stage of this design. Secondly, the assumption that the treatment, at the verge of the cut-off, can still count as random assignment and that status in the municipality is the causal channel for which changes in the percentage of houses with floors made of this material occur.

Finally, some concerns still remain regarding the design of the methodology. On one hand, it is still unclear why the Mexican government used a ranking variable for the second pilot that does not match the UN scores; especially because they only published their scored list for the 125 lowest municipalities and not for the rest of the country.

On the other hand, isolating the effects of this specific program might be complicated, considering that during Felipe Calderón's presidential term (2006-2012) HDI was used for other social programs besides Piso Firme, such as the *Oportunidades/Progresa* programs; all with the same main goal to target the population living in extreme poverty with conditional cash transfer programs, health services and other basic living standard improvements. In this case, the RD design is still valid in terms of distinguishing the impact attached to the discontinuity, although by using a very specific health variable for children, instead a general child mortality rate, the purpose is trying to detach the variation that comes from other social programs. However, it still might be the case that our results are capturing the effects of some other programs, or on the contrary, because the incidences – in this case deaths related to intestinal problems – are very low, the regression results would turn out to be underpowered.

RESULTS

This paper aimed to analyze two different pilots of the *Piso Firme* program. The Mexican federal government first launched the program during the Vicente Fox administration in 2000 as part of the Micro Regions Strategy *(micro-regiones)*. The idea of that strategy was for different Ministries to design programs that generate "opportunities", and schemes for regions to overcome poverty. The main target was the poorest 50 municipalities according to the HDI in 2000, regions where 92% of the total population is indigenous, and located mostly in the following 7 States: Chiapas, Durango, Guerrero, Nayarit, Oaxaca, Puebla, and Veracruz. *(See Figure 6)* The government launched many programs that included agricultural productivity plans, education, health and nutrition, basic infrastructures and household improvement, which included the Piso Firme program.

Table 1 in Appendix B shows the results of the first stage for the former pilot of the program using changes in the percentage of households with cement floor. According to *Column 1* results, there are no positive effects of the program regarding increasing the percentage of households living with cement floor, on the contrary, there is a negative coefficient for the treated units. A similar correlation coefficient (-0.338) holds in magnitude and direction when we reduce the sample size from 100 to 50 municipalities (*Column 3*), since they are both symmetrical to both sides of the thresholds. This shows evidence of our previous hypothesis of a lack of a baseline for the first pilot, since treated municipalities have between 0.27% and 0.33% less changes in cement floor in households. However, this provides information about the targeting logic for the program, which focused on municipalities with large percentage of households with poor infrastructure and limited sanitized services. In summary, given the absence of a baseline in this first pilot, what the coefficient is capturing is not the marginal effect of the program, but rather the targeting of it. Being T=1 for a municipality, shows less changes in cement floor, which means the results are capturing the households with the poorest floor infrastructure.

This same hypothesis is reinforced by the results shown in *Column 2* when we include more observations on the control side. Since these municipalities

are more developed according to the HDI even without treatment, it is consistent with the fact that those subsets of observations show a 0.6% increase installed floors among the treated units. This change in the sign of the coefficient[12] cannot be attributed to the program, because it simply reflects that we are including more units that already had better household conditions. However, in every case, there is no correlation that is statistically significant at a 95% confidence level.

Additionally, when we replicate the same set of regressions but now use percentage changes of houses with a dirt floor *(Table 2)* they remain insignificant, even though the magnitude of the coefficients become higher.

A comparison between *Column 2* with *Column 1 and 3* in every table shows some interesting results. *Column 2* refers to a subsample that includes more municipalities above the cut-off score as a placebo, which means that having higher HDI is more likely to have better household infrastructure overall. That hypothesis is consistent when we see the coefficient which, for example, indicates the effect of the treatment in reducing the percentage of households with dirt floor changes by almost 1 point; this means that the change from a coefficient of minus 1.17 when having a symmetrical sample, versus a coefficient of -2.12 might be driven only by the fact that those municipalities already had less percentage of houses with dirt floors, or because given the higher level of development, new houses are already constructed with cement floors.

In 2006, the Calderón administration complemented the micro-regions strategy with the *Strategy 100 x 100* as Mexico's development and social core policies with the main goal of improving the socioeconomic conditions in the 125 most disadvantaged municipalities. *(Figure 7)*

Table 3 shows the impacts of the second pilot treatment in terms of the changes in the percentage of households with a cement floor in a municipality. *Column 1* corresponds to the total sample of 250 municipalities and considers the treatment variable as assigned by the Mexican government; in this case, a municipality that benefitted from *Piso Firme* program

shows an increase of 9.1% of the houses with cement floor. This coefficient is statistically significant at a 95% confidence level. This holds even when dropping the municipalities that were not supposed to be treated according to the UN HDI score (Column 3), although the magnitude of the coefficient is lower, decreasing to an increase of 7.9% for the average treatment effect.

These results show that even if the program's targeting was used for political purposes, it still had the positive and intended impact. When the model is applied to the 36 treated municipalities that should not have qualified according to the HDI, the standard error of the model increases from 2.6 to 3.0 (*Column 1 versus Column 3 of Table 3)* but the significance of the estimations holds at a 95% confidence level. It is possible that the government may have used "political targeting" to include ineligible municipalities because when testing subsamples in other models with more higher HDI score units, the standard errors become higher and the model less accurate.

Another important result from Table 3 is the statistical insignificance of the correlations in *Column 2 and 4*, since they correspond to a placebo treatment when including more municipalities above the threshold. These results make sense, since we are including more control units, that are not supposed to be included in the program; and back up the validity of the first stage of the RD design. Additionally, for those two models, the HDI coefficients for both columns become statistically significant, which is a reflection of including more control units that naturally will have higher HDI. In short, this helps validate the unique identification for the RD design, since it shows that the coefficient for the first stage in the models that capture the program as it was implemented is significant and stable and allows the second stage to be valid when using the treatment dummy as an instrument for the percentage change in cement floor.

Exploring how the coefficients change under different subsamples – balanced and unbalanced – is a way to do some robustness checks. For example, when including more units on the untreated side, our coefficients basically capture that the increase of percentage of households with cement floor didn't apply to those who were nor treated by the program.

What we want to compare at this point are two scenar-

---

12 From a decrease of 0.33 % of households with cement floor with the sub-samples of 100 municipalities and 50 municipalities from a marginal increase of 0.6% of the same outcome with an unbalanced sample, using more units on the control side.

ios: the first one, the program as it was implemented (including in the sample the 36 units misplaced by the program in the treatment[13]); and the second one, a simulation of the implementation of the program with not targeting deviation with a clean cut-off point, and with no crossover units.

The same exercise as *Table 3*, as part of the RD model in its reduced form, is repeated with the child diarrhea mortality variable, as the total number of due to diarrhea and other intestinal parasite diseases. using the sample assigned by the government, *Column 1 in Table 5* shows that the average treatment effect of the program on the child mortality variable is the 0.511% reduction in the number of total fatalities, however, this coefficient is not statistically significant. In this case, we can assume that given that the MDE for this variable is around 20% of the median of the population, we might be underpowered, since this number only correspond to 3.14% of the average affected population (children between 0 and 9 years) by this type of disease.

So far, the results have been explained in terms of the proximate outcomes. *Table 6* shows the results of the 2SLS reduced form using the treatment dummy variable as an instrument for the percentage changes of HH with a cement floor. *Panel A* of the same table has two different subsets of municipalities but maintains all the units treated as assigned by the government, meanwhile *Panel B* drops out the crossover units. In this case, when using the treated units as assigned by the government without any other control variables, the program has a significant effect on decreasing 5.68% the total number of children fatalities due to diarrhea, at a 90% confidence level. That result holds even when adding additional covariates like the percentage of houses with no access to water or sewer systems.

ROBUSTNESS CHECKS

A. McCrary Test:
Was the running variable manipulated?

When using an RD design, a McCrary test to assess the internal validity is commonly used as a robustness check. In other words, this test verifies if the density of

observations, according to the running variable in this case, the HDI in 2000, is itself continuous at x_0. *Figure 8* shows the density of the ranking variable, with the cut-off line centered at zero used in the pilot 2000, and the figure displays a continuous line for the observations near it.

The same test was carried out for the 2005 data, and using the cut-off for the poorest 125 municipalities according to the UN score, and the results seem to be consistent *(Figure 9)*. However, as mentioned before, the government assigned a different threshold; when doing the density test, *Figure 10* shows a very small discontinuity at x_0. That discontinuity becomes more evident when instead of plugging in the sample of municipalities for the year 2000, we plugged in the full 2005 sample. *(Figure 11)*

The fact the McCrary test shows some inconsistencies for the second pilot, raises the question why the government decided to maintain an outdated measurement for designing not only Piso Firme, but all of the country's social programs. Using the more recent 2005 Municipality Development Index might make more sense *(Figure 12)*, instead of using a ranking that is unclear and includes some municipalities that shouldn't have been treated.

B. Does political targeting bias the estimations? Selecting different subsamples.

The second robustness check changes the subsample used in the model, which means adjusting for different bandwidths for the program's threshold. For example, for the first pilot the magnitude of the coefficients remains very similar when applying a smaller symmetrical sample size, which means the RD identifying assumption holds. On the contrary, by applying a placebo test and including more variables on the control size, the magnitude of the effect of the treatment increases. That only means that, since those variables are far away from the cut-off score, the probability of being treated is not only lower but also by construction those municipalities are worse counterfactuals.

Another way to show consistency of the results is to, for the first stage of the model, not only regress the cement floor variable on the treatment but also do the same exercise with dirt floors. The decrease in the percentage of households with dirt floor in a treated municipality is another way of confirmation that the program truly had some impact on the household

---

[13] These units are 36 municipalities that according to the HDI cut-off condition should not have been treated, but were included in the Piso Firme Program.

infrastructure, and trying to discard that in our model might be only capturing the overall improvement of HH infrastructure (i.e. construction of houses with already cement floor) regardless of a unit being or not in the program.

The most compelling table in terms of the consistency of the results despite the different windows of entrance into the treatment is *Table 6.* In every regression model with symmetrical samples (*Columns 1,2 5 and 6*), and testing the model with the cross-over units and by dropping those units from the sample, the impact of the Piso Firme program has an effect of decreasing the number of child fatalities by -0.05. This result is important because at some level there is evidence that, regarding the differences in the treatment assignment based on potentially government targeting, since those differences occurred near the threshold, the RD design can still be considered as if random assignment, and in consequence the units close to the cut-off are a good counterfactual. In other words, the units included in the treatment of the program that should not have been included on the program, didn't create a significant difference between control and treated units on average, for the purpose of the program and the outcome variables evaluated in this paper.

CONCLUSION

The data analyzed from the *Piso Firme* Program, mostly in its second pilot, showed that there is a robust and significant effect on improving household infrastructure in the treated municipalities, which was not driven by the construction of new houses with cement floors. For 2000, the lack of significance might be caused by time endogeneity, because the program went on the books in 2000, and the analysis uses 2000 and 2005 data; meaning that there is no clear baseline. It was not possible to use the 1995 census data since it was not complete due to political complications for collecting data that year. Regarding the effects of the program on the health conditions in the treated population, the results from the 2SLS show a -0.5 decrease in the child mortality outcome variable, a coefficient that is still far from the MDE of 18% to 20%. To be underpowered in this case to show an improved significant result makes sense because: first, this type of study normally relies on household level data (since the treatment is at that level and it might be the case that it had heterogeneous treatment, cement floor in

one room versus the entire house), [14] and this study only has aggregated data. Looking for other sources of good health data remains an important challenge in this analysis.

Data on the cases of diarrhea, instead of fatalities, would better capture the potential impacts of the program. Also, grappling with questions from a non-linear design could help to conduct further robust research or additionally added heterogeneous effects, for example like interacting with average rainfall conditions. Finally, further research regarding the political targeting of the treatment selection would be interesting, maybe by including a variable per municipality and per state that captures if the entity shares the same political party as the federal government.

Overall, analysis of infrastructure at the household level provides a golden opportunity to explore the impacts since spillovers are not as big as other types of infrastructure.

14 WHO (World Health Organization) (2002). The World Health Report 2002, Reducing risks, promoting healthy life. Geneva. Available from http:// www.who.int/pub/en/

BIBLIOGRAPHY

Brenneman, A. (2002). Infrastructure & linkages: A literature review. Background Report, The World Bank, Washington, DC.

Center for Global Development, A solid Foundation for Child Health: Mexico's Piso Firme Program.

Comisión Nacional para el Desarrollo de los Pueblos Indígenas, *Desarrollo Indígena en 50 municipios,* 2005.

CONEVAL, Evaluación de Impacto de la Estrategia 100 x 100, 2014.

Davis, Jennifer et al., The Effects of Informational Interventions on Household Water Management, Hygiene Behaviors, Stored Drinking Water Quality, and Hand Contamination in Peri-Urban Tanzania, *American Society of Tropical Medicine and Hygiene*, 2011, pp. 184-191.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. Quarterly Journal of Economics, 50, 112–1203.

Gertler, Paul et al, *Housing, Health and Happiness,* American Economic Journal, 2009.

Gertler, Paul; Evaluación de Resultados de Impacto del Programa Piso Firme Estado de Coahuila, 2005.

Jacob, Robin and Zhu, Pei; *A Practical Guide to Regression Discontinuity*, MDRC, 2012.

Jalan, J., & Ravallion, M. (2001). Does piped water reduce diarrhea for children in Rural India? Policy Research Working Paper No. 2664, The World Bank, Washington, DC.

Marianne Fay, Danny Leipziger, Quentin Wodon and Tito Yepes, "Achieving Child-Health-Related Millennium Development Goals: The Role of Infrastructure" in World Development, Vol. 33, No. 8, 2005, pp. 1267-1284.

PNUD, El Índice de Desarrollo Humano en México: cambios metodológicos e información para las entidades federativas, 2012.

Rachel J. Krause et al., Ascaris and hookworm transmission in preschool children from rural Panama: role of yard environment, soil eggs/larvae and hygiene and play behaviors, Cambridge University, 2015.

Sales Heredia, Francisco; Los 125 municipios con menor Índice de Desarrollo Humano en México, *Centro de Estudios Sociales de Opinión Pública,* Cámara de Diputados, 2013.

White, Halbert and Lu, Xun; *Robustness Checks and Robustness Test in Applied Economics,* UC San Diego, 2010.

WHO (World Health Organization) (2002). The World Health Report 2002, Reducing risks, promoting healthy life. Geneva. Available from http:// www. who.int/pub/e

WHO, Diarrhoeal Disease Fact Sheet, 2017.

Zelodón, Rodrigo and Vargas, Luis; The Role of Dirt Floors and of Firewood in Rural dwellings in the epidemiology of changas disease in Costa Rica, *American Society of Tropical Medicine and Hygiene,* 1984, pp. 232-235.

APPENDIX

Appendix A. Data sources and descriptive statistics

Table A. Descriptive Statistics (Average and SD in parenthesis)				
Variable	2000	2005	2010	Source
Percentage of HH with dirt floor	30.49 (24.83)	24.47 (22.03)	12.82 (11.78)	INEGI
Percentage of HH with cement floor	54.79 (20.48)	60.23 (18.63)	69.56 (14.85)	INEGI
Percentage of HH no water	25.38 (21.53)	20.71 (20.80)	14.43 (17.37)	INEGI
Percentage of HH no sewer system	48.16 (29.02)	31.45 (27.18)	24.70 (24.27)	INEGI
Total number of deaths due to diarrhea	NA	2.96 (8.85)	2.02 (6.31)	SINAIS (SS)
Municipality HDI 2000 (UN)	0.707 (0.075)	NA	NA	UNDP
Municipality HDI 2005 (UN)	NA	0.753 (0.067)	NA	UNDP
Municipality HDI 2000 (Mexican Government, only for 125 municipalities)	0.5312 (0.4355)	NA	NA	SEDESOL

Appendix B. Tables

Table 2. Changes in dirt floor for first pilot (2000-2005)

	(1) Dirt Floor (Sample 50/50)	(2) Dirt Floor (Sample 50/100)	(3) Dirt Floor (Sample 25/25)
Treatment	-1.711	-2.120	-1.171
	(3.504)	(3.623)	(6.686)
HDI (2000)	-28.71	-47.83	13.78
	(45.78)	(43.64)	(283.4)
Constant	6.065	15.95	-16.03
	(25.27)	(24.54)	(153.8)
Observations	100	150	50
R-squared	0.004	0.010	0.005

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3. Changes in cement floor for second pilot (2006-2010)

	PANEL A. Government Treatment Assignment		PANEL B. Without crossover units	
	(1) Cement Floor Sample (125/125)	(2) Cement Floor Sample (75/175)	(3) Cement Floor (125/125)	(4) Cement Floor (75/175)
Treatment	9.103***	2.858	7.933***	-2.717
	(2.649)	(3.122)	(3.015)	(3.643)
HDI (2000)	-3.627	-243.8***	-23.72	-339.0***
	(33.08)	(64.62)	(36.75)	(67.62)
Constant	36.55*	177.0***	47.88**	233.3***
	(19.45)	(38.34)	(21.64)	(40.22)
Observations	251	250	250	251
R-squared	0.090	0.144	0.085	0.168

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4. Changes in dirt floor for second pilot (2006-2010)

	PANEL A. Government Treatment Assignment		PANEL B. Without crossover units	
	(1) Dirt Floor Sample (125/125)	(2) Dirt Floor Sample (75/175)	(3) Dirt Floor (125/125)	(4) Dirt Floor (75/175)
Treatment	-7.868***	-2.194	-6.675**	3.382
	(2.614)	(3.066)	(2.963)	(3.552)
HDI (2000)	11.66	233.6***	31.20	332.4***
	(32.65)	(63.46)	(36.11)	(65.94)
Constant	-41.33**	-171.2***	-52.40**	-229.6***
	(19.19)	(37.65)	(21.26)	(39.22)
Observations	251	250	250	251
R-squared	0.079	0.131	0.076	0.161

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5. Impact on Child Mortality of Second Pilot (2006-2010)

Diarrhea and other intestinal fatalities in children from 0 to 9 years old				
	PANEL A. Government Treatment Assignment		PANEL B. Without crossover units	
	(1) Sample (125/125)	(2) Sample (75/175)	(3) Sample (125/125)	(4) Sample (75/175)
Treatment	-0.511	-0.414	-0.562	-2.717
	(0.407)	(0.464)	(0.442)	(3.643)
HDI (2000)	0.308	1.995	-0.109	-339.0***
	(5.089)	(9.611)	(5.394)	(67.62)
Constant	-0.722	-1.743	-0.474	233.3***
	(2.992)	(5.701)	(3.176)	(40.22)
Observations	251	250	250	251
R-squared	0.013	0.009	0.015	0.168

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6. 2SLS Reduced form - Impact on Child Mortality of Second Pilot (2006-2010)
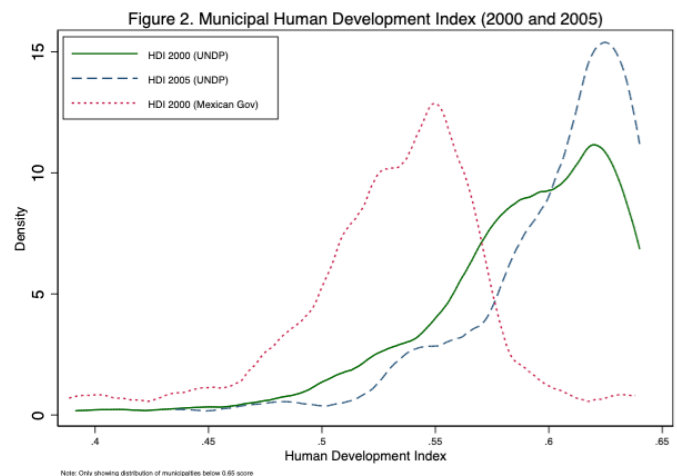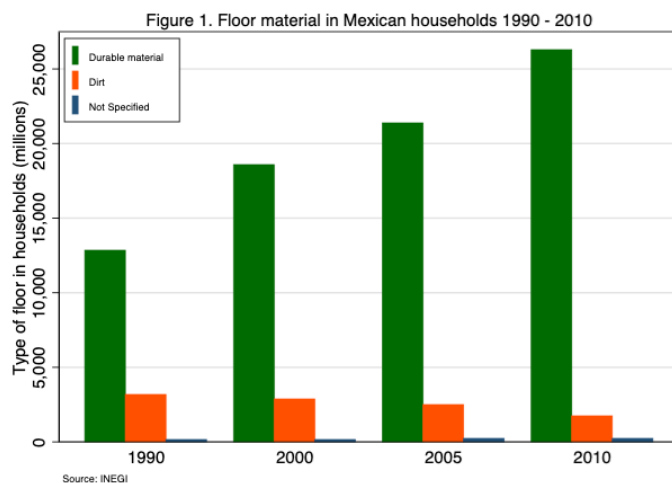
IV = Treatment Piso Firme Program

Diarrhea and other intestinal fatalities in children from 0 to 9 years old							
PANEL A. Government Treatment Assignment				PANEL B. Without crossover units			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample (125/125)		Sample (75/175)		Sample (125/125)		Sample (75/175)	
Cement floor							
-0.0568*	-0.0553*	-0.0426	-0.0427	-0.0591*	-0.0567*	-0.0501	-0.0503
(0.0322)	(0.0324)	(0.0294)	(0.0295)	(0.0327)	(0.0329)	(0.0313)	(0.0314)
No Sewer System							
	0.000149		-0.000276		0.000129		-0.000313
	(0.000307)		(0.000393)		(0.000313)		(0.000392)
No Water							
	0.000276		-1.37e-05		0.000358		4.80e-05
	(0.000454)		(0.000502)		(0.000467)		(0.000513)
Constant							
1.413	1.396	0.822	0.809	1.466	1.423	1.061	1.055
(1.258)	(1.259)	(1.054)	(1.056)	(1.249)	(1.253)	(1.085)	(1.085)
Observations							
251	251	250	250	250	250	251	251

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix C. Figures

Figure 1. Floor material in Mexican households 1990 - 2010

Figure 2. Municipal Human Development Index (2000 and 2005)

Distributions for First Pilot
(2000-2005)

Figure 3a. Municipalties distribution in 2000
(Before First Pilot treatment)

HDI cut-off at 0.5343

Figure 3b. Municipalties distribution in 2005
(After First Pilot treatment)

HDI cut-off at 0.5343

Distributions for Second Pilot
with Official UN Cut Off for 2005
(2000-2005)

Figure 4a. Municipalties distribution in 2005
(Before Second Pilot treatment)

HDI cut-off at 0.6383

Figure 4b. Municipalties distribution in 2010
(After Second Pilot treatment)

HDI cut-off at 0.6383

Distributions for Second Pilot
with Government Assigned Treatment
(2000-2005)

Figure 5a. Municipalties distribution in 2005 (125 municipalities)
(Before Second Pilot treatment)

HDI cut-off for 125 mun at 0.57279998

Figure 5b. Municipalties distribution in 2010 (125 municipalities)
(After Second Pilot treatment)

HDI cut-off for 125 mun at 0.57279998

Figure 6. 50 Mexican municipalities with the lowest HDI in 2000

Figure 7. 125 Mexican municipalities with the lowest HDI

Figure 8. McCrary density test for HDI 2000, with cut-off for pilot in 2000

Figure 11. McCrary density test for HDI 2000, with the highest target score value according to the HDI 2000 published by the government for the second pilot (2005 Municipality sample)

Figure 9. McCrary density test for HDI 2000, with UN cut-off for pilot in 2006

Figure 12. McCrary density test for HDI 2005, hypothetical cut-off for the lowest 125 municipalities.

Figure 10. McCrary density test for HDI 2000, with the highest target score of the 125 municipalities targeted by the government for the second pilot.

Figure 13. RD Plot for cement floor 2010, running variable HDI 2000.

DOES AID RESPOND TO CONFLICT? EVIDENCE FROM AFGHANISTAN

Macklin Scheldrup

Princeton University, Woodrow Wilson School of Public and International Affairs

Development aid is considered a common policy response to conflict and insecurity around the world. In terms of counterinsurgency operations, the ongoing conflict in Afghanistan has been impacted by a large inflow of foreign aid but the production of violence and provision of aid at the subnational level has not been explored. This paper attempts to measure the extent of aid responsiveness to changes in violence levels across Afghanistan. Using lagged fixed effects regression models, it finds that between 2005 and 2013, changes in relative aid commitments at the provincial level were not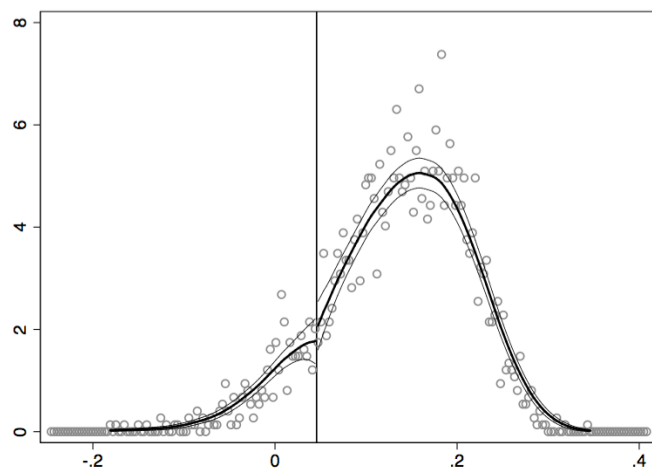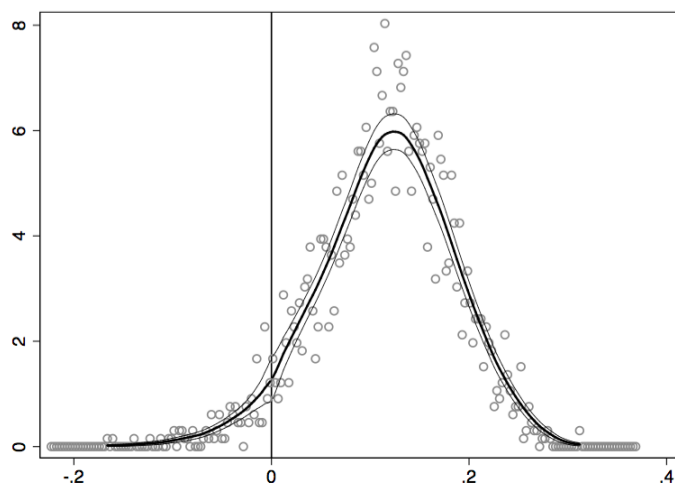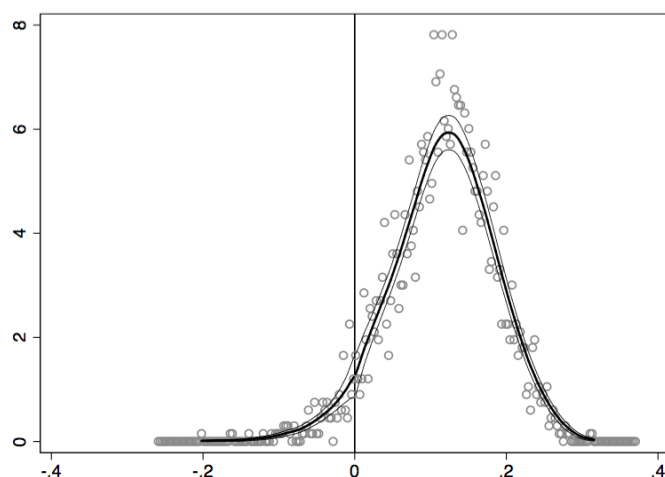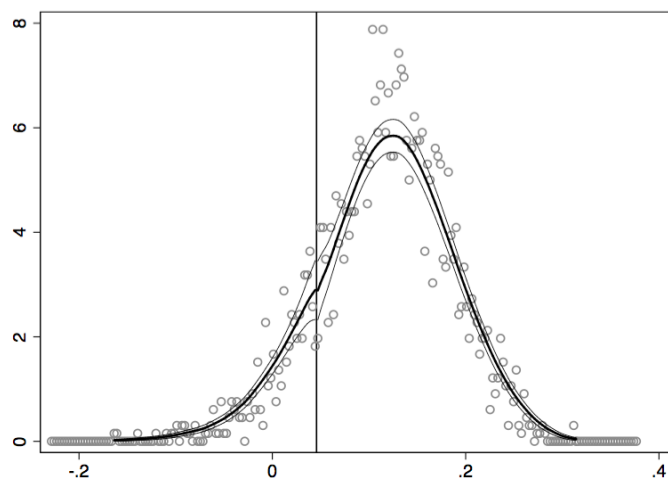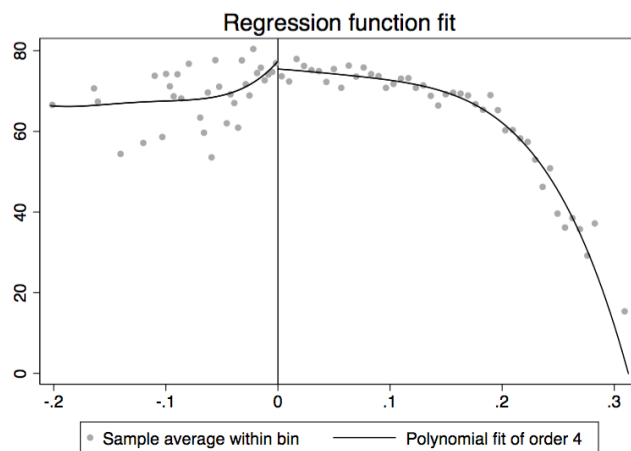 responsive to changes in levels of violence. However, aid was found to be responsive to changes in violence in the largest provinces of the country as well as certain regions. These findings provide evidence that aid and development policy decisions are more reactive in areas of Afghanistan that receive the greatest international attention, regardless of their underlying stability. The paper concludes with policy recommendations to improve aid effectiveness in addressing conflict going forward.

INTRODUCTION

This paper explores the extent to which the commitment of foreign aid to the provinces of Afghanistan was or was not impacted by sub-national variation and changes in violence. It does so through an analysis of the correlation between officially declared aid commitments to provinces and the varying levels of violence experienced by the provinces in the years prior to the aid commitment.

Using a lagged fixed effect ordinary least squares (OLS) regression for province-years between 2005-2014, I find that across the country there was little to no systematic commitment of aid in response to changes in violence. When controlling for province and year, no strong patterns emerge to indicate that greater aid commitments were directly tied to provinces with increased levels of violence. There appears to be a small positive correlation between aid and violence one year after aid commitments which then quickly disappears. These findings are important to understand the patterns of development aid during long-run reconstruction efforts in conflict environments.

I do find, however, that the allocation of aid to the Eastern and Western regions of the country consistently increased after experiencing greater violence. Additionally, the six most populated provinces in the country received aid flows that were highly responsive to changes in violence in the year prior. These results indicate that the political prominence and accessibility of an area within a country increases the extent to which aid allocation responds to variation in violence. Even in a national effort, some areas receive greater attention than others regardless of their stability.

These findings fit within the broader literature on the relationship between aid and the production of violence. There may be a negative relationship between the peacefulness of a province and the aid it receives, but the deterioration or improvement in that peace will only impact later aid provision if the province is of significance (either real or perceived) to decision-makers

in the development community.

CONTEXT

Policymakers have treated aid in conflict and post-conflict settings as a crucial tool for stabilization policy. The provision of official development aid (ODA) within countries is often directed to the areas of greatest concern. In a country without widespread political violence, those areas are usually the poorest places. However, in a counter-insurgency (COIN) effort, aid is often targeted at the places experiencing the greatest levels of insurgent-led violence in an effort to spur government support among a neutral public through the provision of services.[1] Under a "hearts and minds" strategy, aid and its subsequent benefits are intended to garner the support of the local populace by addressing grievances, thus diminishing demand for rebellion.[2] The actual role of aid in addressing the underlying drivers of violence is disputed in the literature, but in the policy realm its utilization for this purpose in conflict and post-conflict environments is often presumed. [3]

Berman et al. posits an information-centric approach to asymmetric conflicts in which both the government and rebels compete for the support of the civilian population. Local communities can share information with the government that will hinder the ability of rebels to operate and therefore should reduce their production of violence.[4] This implies that well-suited development aid to areas "will reduce rebel violence, as it reduces the level of violence that will trigger civilian tips to government, which in turn increases the risk of failure for rebels."[5]

Empirical evidence supporting the link between aid and decreased violence has been found during insurgencies in the Philippines,[6] Iraq,[7] and Vietnam.[8] However, other work has found aid tends to increase violence because insurgents attempt to preemptively derail projects[9] or increase attacks immediately after aid delivery to disrupt implementation.[10] Food aid has also been linked to increasing civil war duration.[11] In Afghanistan, the National Solidarity Program (NSP) was shown to reduce violence in communities with previously low or moderate levels of conflict,[12] yet other aid programs in the country have either had no effect on violence[13] or led to increased levels of conflict.[14]

There is no uniform effect of development aid on violence during insurgencies, the effects are conditional on the size, type, location, and appropriateness of the project being implemented.[15] More modest and locally driven programs in relatively secure areas are the most likely to reduce violence.[16] Aid that responds to local needs increases stability while aid that promotes corruption or is culturally controversial has been shown to provoke conflict.[17] Yet, micro and meso-level impacts of development aid have not translated to macro-level COIN outcomes, likely because of programs' inadequately short time-spans to sustain improvements[18] and overarching concerns with identity and grievances that aid is unable to overcome.[19]

Regardless of its causal effect on reducing violence, development aid was a crucial component of the ISAF's broader COIN strategy in Afghanistan. [20] Alongside the military surge was a massive influx of ODA. According to the Organization for Economic Development (OECD), reported ODA to Afghanistan from all donors increased from $2.8 billion in 2005 to a peak of $6.9 billion in 2011.[21] For the United States government, development aid provision to Afghanistan was explicitly tied to violence reduction, with the United States Senate Committee on Foreign Relations referring to it in 2011 as a "vital tool for promoting stability". [22]

1 Berman, Felter, and Shapiro, Small Wars, Big Data: The Information Revolution in Modern Conflict, 113.
2 Berman, Shapiro, and Felter, "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq."
3 Chou, "Does Development Assistance Reduce Violence? Evidence from Afghanistan."
4 Berman, Felter, and Shapiro, Small Wars, Big Data: The Information Revolution in Modern Conflict, 75.
5 Berman, Felter, and Shapiro, 77.
6 Crost, Felter, and Johnston, "Conditional Cash Transfers, Civil Conflict and Insurgent Influence: Experimental Evidence from the Philippines."
7 Berman, Shapiro, and Felter, "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq."
8 Dell and Querubin, "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies."
9 Crost, Felter, and Johnston, "Aid under Fire: Development Projects and Civil Conflict."
10 Khanna and Zimmermann, "Guns and Butter? Fighting Violence with the Promise of Development."
11 Nunn and Qian, "US Food Aid and Civil Conflict."
12 Beath, Christia, and Enikolopov, "The National Solidarity Programme."
13 Child, "Hearts and Minds Cannot Be Bought: Ineffective Reconstruction in Afghanistan."
14 Iyengar Plumb et al., "Stabilization in Afghanistan: Trends in Violence, Attitudes, Well-Being and Program Activity.," 10.
15 Kapstein, "Aid and Stabilization in Afghanistan. What Do the Data Say?"
16 Berman, Felter, and Shapiro, Small Wars, Big Data: The Information Revolution in Modern Conflict, 148.
17 Child, "Conflict and Counterinsurgency Aid: Drawing Sectoral Distinctions."
18 Kapstein, "Aid and Stabilization in Afghanistan. What Do the Data Say?," 6.
19 Lyall, Blair, and Imai, "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan."
20 "U.S. Government Counterinsurgency Guide," 24–25.
21 "OECD DAC Statistics."
22 U.S. Senate. Committee on Foreign Relations, "Evaluating US Foreign Assistance to Afghanistan."

There has been a constant international military presence in Afghanistan since the fall of the Taliban government in late 2001. During that time, the levels of violence in the country have varied substantially, both at the province level and overall. From a relatively calm start, political violence committed by the Taliban and other insurgent groups has increased.[23] In response, the Afghan government and its foreign partners, the International Security Assistance Force (ISAF), also increased their kinetic military operations.

By the start of the Obama administration in 2009, the Afghan War was at the forefront of policymakers' minds because of growing insurgent territorial control and a continued decline in stability across the country. This led to the "surge" with one hundred thousand international troops deployed to the country to clear areas from the Taliban and then hold onto them while Afghan government forces were built up and trained.

Between 2005 and 2014, the military presence and ODA funding in Afghanistan increased dramatically, and then slowly came down from its 2011 peak. Both before and after the peak, differing levels of violence were experienced across the country.[24] While some areas started with low levels of violence which slowly increased over time, others started with higher levels which declined during the surge (2009-2011). Still others witnessed consistently high or low levels of violence during these years. This variation in violence across provinces and time should have an impact on the location of aid commitments under stabilization strategies as both sides of the conflict have a greater incentive to provide services to areas where tips and information are most valuable.[25] An increase in violence relative to the rest of the country would indicate a comparatively higher value for information from that place. The rest of this paper focuses on testing this proposition.

RESEARCH QUESTION

Is the quantity of aid to an Afghan province responsive to changes in prior levels of violence in that province?

This paper seeks to empirically test the assumption that aid follows violence at the sub-national level during an ongoing insurgency campaign. If aid is being utilized as a policy tool to reduce violence, we should find a predictable pattern of increased aid distribution to more violent locales. In other words, provinces experiencing comparatively higher levels of violence should see increased levels of aid in the years following the increase in violence.

This leads to the following hypothesis:

- H_0: Change in violence levels does not have an effect on changes in aid levels within a province

- H_1: Change in violence levels has a positive relationship to changes in aid levels within a province

These are tested in the subsequent data and methodology sections.

DATA

The AidData dataset was used to measure levels of aid. This data is derived from the Government of Afghanistan's official Development Assistance Database and contains more than fifteen-hundred unique development projects committed to between 2001 and 2014.[26] While it likely does not include ODA from all sources, there is no indication that there is any systematic selection bias in the projects which are included in the database. However, 2014 appears to not be fully reported with total aid commitments in the database for that year less than 10 percent of any other year. For that reason, it is not included in the analysis.

The sums of money designated to every project that had a province assigned was used to produce a figure for each province-year. The dataset does this through the "even splits commitment" category. For a project coded over multiple locations, each one is geocoded separately under the same project ID number. The total commitments for the project are then split evenly across its locations.

Commitments were used in lieu of disbursements because of their wider availability in the dataset and because they better encapsulate the intended placement of aid to locations in certain time periods compared to disbursements which occurred slowly over time. Aid disbursement is often slow and inflexible, as projects with multiple year timespans are implemented. Therefore, the disbursement measure is the better

23 Iyengar Plumb et al., "Stabilization in Afghanistan: Trends in Violence, Attitudes, Well-Being and Program Activity.," 33.
24 Iyengar Plumb et al., 38–39.
25 Berman, Felter, and Shapiro, Small Wars, Big Data: The Information Revolution in Modern Conflict, 78, 127.
26 AidData, "AfghanistanAIMS_GeocodedResearchRelease_Level1_v1.1.1 Geocoded Dataset."

measure to capture the aid allocation policy decisions made in a particular year. However, commitments may never become realized as on the ground disbursements for a variety of reasons. Because the research question focuses on responsiveness of aid decisions and not the broader issue of aid effectiveness in violent reduction, this is less of a concern for this paper.

After separating out aid committed to each province, province-year figures are derived by summing total aid dollars based on the first year of their commitment. This captures the year in which the aid is initially assigned to the province. Most projects are designated to occur over multiple years. Unfortunately, this level of accuracy in the timing of aid provision is not available in the dataset. Still, the first year of commitment is a relevant gauge of how aid is assigned in that time period. The province-year aid dollar amounts provide an estimate of the dedication of aid to a location during a single year.

This is a simple and consistent estimate of aid by location, but its accuracy does suffer for large-scale nationwide initiatives such as the National Solidarity Programme (NSP). Founded in 2003, the NSP was the Afghan government's largest development program at over $2 billion.[27] It was intended to increase the administrative reach of the state through a community-drive development model utilizing village-level block grants. In this dataset, the NSP commitment is coded once in 2003, its founding year, for all provinces, and therefore produces that same amount for each, despite the fact that the disbursement of NSP aid varied across province and took place over more than a decade.[28]

The dataset does not include development assistance which does not fall under the ODA categorization, namely military funded and managed aid such as the Commander's Emergency Response Program (CERP). These programs were meant to make aid provision more flexible and faster by allowing military personnel in the field to decide on its disbursement based on local COIN priorities.[29] CERP and other military sources of aid are important to understanding the overall relationship between aid and violence and based on their purposively responsive and short-time frames, their exclusion may underestimate the impact of violence on overall aid distribution.

The Global Terrorism Database (GTD) is used to measure levels of conflict. GTD codes open-source materials for all events which fall under the *"the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."* [30] Since violence committed by Afghan insurgent groups have explicitly political aims,[31] this database is a strong fit for measuring levels of conflict in this context.

Events in the GTD dataset are coded for their date, location, actor types, and number of casualties. Province-year totals are created in the same manner as aid data. Both event counts and number of casualties are summed by province-year to create two related measures of conflict. Casualties is the conflict variable used throughout the paper because it better captures the impact of violence on the lives of Afghans. Robustness checks using event counts are found in the Annex.

The GTD dataset is better suited for the purposes of this research than alternative event databases because of its political violence distinction as well as the large number of years it includes.

The variables of interest are highly skewed.[32] As a result, they are all logged in the regression analysis. To avoid excluding observations with no violence, one is added to all province-years prior to calculating their natural log. The much smaller number of province-year observations with no aid commitments are excluded from the analysis after the natural log is calculated.

Additionally, the analysis is conducted using the ranks of aid commitments and casualties across provinces in each year. A tie is ranked as the middle of group, so that if the 10th and 11th ranked provinces are equal, they're given the value of 10.5. This method does not take into account changes in the quantities of the variables, but only how each province's figure compares to the rest of the country in a given year. This is a suitable robustness check on the results of the logged regression analysis that avoids the skewness of the independent variable without resorting to the methodologically problematic approach of adding one before a natural log to avoid missing values.

27 SIGAR, "Afghanistan's National Solidarity Program Has Reached Thousands of Afghan Communities, but Faces Challenges That Could Limit Outcomes."
28 Beath, Christia, and Enikolopov, "The National Solidarity Programme."
29 United States Army, "Commander's Guide to Money as a Weapons System."

30 Global Terrorism Database, "National Consortium for the Study of Terrorism and Responses to Terrorism (START)."
31 Osman and Gopal, "Taliban Views on a Future State."

32 Province-year aid commitments – skewness=3.22, kurtosis=15.6
 Province-year casualties – skewness=3.1, kurtosis=14.87

METHODOLOGY

An OLS regression model with lagged variables is used to test for the relationship between conflict and subsequent aid commitments.

This model addresses serial correlation in the time series. Controlling for the previous year's aid level by including it in the regression model ensures that it does not confound the estimates from the regression. Most importantly, it presents the dynamic effects of violence on aid in the following period. This means the model is measuring the impact of changes in violence on changes in aid, not simply underlying correlations between violence and commitments. The model also includes province and time fixed effects to control for non-varying unobserved provincial characteristics and nationwide annual trends in both aid spending and violence.[33] Population estimates for each province-year are also included to take into account any variation in growth rates between provinces over time which may impact aid spending decisions.

The level of aid in the previous year is included to isolate the changes in aid from the previous year. Casualties are lagged by a year, so that casualties in year t-1 is explaining aid in period t. This results in the following model:

$$Log(Aid)_{it} = \beta_0 + \beta_1 Log(Casualties+1)_{it-1} + \beta_2 Log(Aid)_{it-1} + \beta_3 Province_i + \beta_4 Year + \psi Population_{it} + \epsilon_{it}$$

The second model used is also an OLS regression using the same lagged variables model across year ranks of aid commitments and casualties. It too controls for serial correlation across time and province. This produces the following model:

$$Log(Aid)_{it} = \beta_0 + \beta_1 Rank(Casualties)_{it-1} + \beta_2 Log(Aid)_{it-1} + \beta_3 Province_i + \beta_4 Year + \psi Population_{it} + \epsilon_{it}$$

The rank model is an alternative non-parametric measure of the key independent variable to identify trends in changes in aid commitments due to changes in comparative levels of violence. It minimizes the impact of particularly large changes in casualty figures from year to year which might lead to an overestimation of the correlation between aid and violence.

In both cases, the lagging of the casualties' variable is adjusted to run models across different numbers of years prior to the aid commitment. In this way, the models control for different time periods as well as combinations of time periods in investigating the effect of changes in the number of casualties on changes in aid commitments in subsequent years.

RESULTS

More violent provinces are more likely to receive more aid. When averaging casualties and aid commitments across all years by province, there is a statistically significant correlation between casualties and aid.[34] This indicates that aid commitments were directed towards areas of greater conflict, although there may be other causal reasons for this relationship.

The results of the base regression models are displayed in Table 1. The models each include different time lags of casualties. Model 1 displays a weak positive relationship between the previous year's casualties and the subsequent amount of aid committed. The coefficient in Model 1 indicates that when controlling for fixed effects and the previous year's level of aid, a 1% increase in casualties is associated with a 0.04 percent decrease in aid commitments. The magnitude of the positive coefficients of last year's violence stays small and statistically insignificant when controlling for violence in other years.

Models 3-6 shows a positive coefficient for the levels of violence two years prior to the aid commitment in year t. However, this relationship also does not come close to statistical significance. The effect of violence three years prior included in models 5 and 6 shows a statistically insignificant negative relationship. Regressions using the rank casualty variables are shown in Table 2. They too show a small but statistically insignificant relationship between the rank in casualty amounts in a given year t-1 and aid commitments in year t. In this case, moving one rank up in annual casualties is associated with about a 0.007 percent increase in aid commitments the next year. Unlike previous models, rankings two years prior are associated with a negative relationship to aid commitments.

Based on the results in Table 1, the null hypothesis is unable to be rejected. There appears to be no empirical

---

33 Sexton, "Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan."

Table 1. OLS Regression on Log of Province-Year Aid Commitments (2005-13)						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Simple Reg.	Simple Reg.	Simple Reg.	Simple Reg.	Simple Reg.	Simple Reg.
Log of Casualties in t-1	0.037	0.045	0.037			0.038
	(0.045)	(0.043)	(0.045)			(0.044)
Log of Casualties in t		-0.05				
		(0.04)				
Log of Casualties in t-2			0.002	0.059	0.01	0.006
			(0.049)	(0.048)	(0.048)	(0.049)
Log of Casualties in t-3					-0.03	-0.031
					(0.046)	(0.045)
Log of Aid in t-1	0.028	0.026	0.027	0.024	0.025	0.028
	(0.065)	(0.065)	(0.066)	(0.067)	(0.067)	(0.066)
Constant	16.17+	16.28+	16.17+	16.29+	16.29+	16.18+
	(1.16)	(1.18)	(1.18)	(1.19)	(1.19)	(1.17)
Provinces	34	34	34	34	34	34
N	304	304	304	304	304	304
Within R-squared	0.4	0.41	0.4	0.4	0.4	0.4
Robust standard errors clustered at the province level in parentheses.						
+ p<0.01, ** p<0.05, * p<0.1.						
Province and year fixed effects. Population coefficient not displayed.						

evidence for aid commitments following conflict levels in the Afghan context. In fact, there is a consistent, statistically insignificant negative relationship between aid levels and the violence occurring two to three year prior. The previous year's casualty levels does produce consistent positive coefficients to aid, but again the relationship is weak.

Afghanistan is a geographically and ethnically heterogenous country. With the Hindu Kush mountains jutting across the middle of the country, the time to travel between the far northern province of Badakhshan and Kandahar in the south is more immense than their distance would suggest. This kind of separation is part of the reason for the high regional variation in ethnic composition, an important characteristic in the Afghan conflict. Some provinces are upwards of 95 percent of the largest ethnic group, Pashto, while others are below 1 percent.[35]

The diversity of the country is reflected in the NATO strategy of regional commands (RC) with responsibilities for specific parts of the country.[36] Each RC was led by an individual ISAF coalition partner, and acted as a tactical headquarters to provide support to the civilian-military Provincial Reconstruction Teams (PRTs) meant to address the development aspects of stabilization.[37] A combination of historical and ethnic ties to the insurgency and potential divergence in regional-level COIN strategy warrants an exploration of differences in aid responsiveness between these regional commands.

Table 3 below presents the results of Model 3 when only including the provinces of each RC.[38] These results do present some noticeable differences across regions. In RC East, there is a statistically significant positive co-

35 The Asia Foundation, "Afghanistan in 2014: A Survey of the Afghan People."

36 ISW, "International Security Assistance Force (ISAF)."

37 Afghanistan Provincial Reconstruction Team: Observations, Insights, and Lessons.

38 Kabul province falls under its own RC- Capital. Because it is the only province in the RC, there are not enough observations for Stata to run a fixed effects model. When running a basic OLS with only Kabul province we do find estimates with magnitudes above 100, which no other province comes close to. This indicates that aid commitments to Kabul or the coding of aid to Kabul is different than other provinces.

Table 2. Ranked OLS Regression on Log of Province-Year Aid Commitments (2005-13)			
	(7)	(8)	(9)
VARIABLES	Rank Reg.	Rank Reg.	Rank Reg.
Rank of Casualties in *t-1*	0.0068	0.0072	0.0072
	(0.009)	(0.008)	(0.008)
Rank of Casualties in *t-2*		-0.0057	-0.006
		(0.008)	(0.008)
Rank of Casualties in *t-3*			-0.002
			(0.009)
Log of Aid in *t-1*	0.026	0.029	0.028
	(0.065)	(0.066)	(0.066)
Provinces	34	34	34
N	304	304	304
Within R-squared	0.4	0.4	0.4
Robust standard errors clustered at the province level in parentheses. Province and year fixed effects.			
+ p<0.01, ** p<0.05, * p<0.1.			
Estimated constant and population coefficients not displayed			

efficient between aid commitments and the number of casualties two years prior. For RC West, the year prior has a still significant positive association with aid commitments. These general trends persist in the rank regressions shown in Table 4, although the significance of the positive coefficients has disappeared. In its place is a significant negative relationship between aid commitments and violence two years earlier in the German military led RC North. It seems that the RC may be an important determinant of aid responsiveness.

RC East contains the greatest number of provinces among the regional commands. It also is the only RC with the U.S. as the lead nation throughout the conflict. Provinces in RC East were either some of the most consistently peaceful (e.g. Panjshir) or are mountainous Pashto-dominated border provinces which witnessed some of the fiercest fighting of the war (e.g. Kunar). With the largest donor as the lead nation, the U.S. may have prioritized responding to these areas more than others.

RC West had the fewest provinces under its command. Its responsiveness is likely due to the impact of Herat, the country's 2nd largest province. Herat's size and histo-

ry increase the attention paid to it, and the influence of Ismail Khan may also be a factor. The self-styled "Amir of the West," Khan is a prominent anti-Soviet mujahidin and former governor of the province whose patronage network centralized his control in the province.[39] His influence over a crucial region of the country could've made both the central government and foreign donors more receptive to requests for more aid in response to security deterioration. Local political dynamics and their relations to Kabul do appear to matter.

Beyond ethnic composition, the size of Afghan provinces also varies widely. In 2014, the government estimated 4.2 million people lived in Kabul province while only 146 thousand lived in Nuristan, the smallest province. More provinces have below 400 thousand people than have greater than 1 million.[40] Provinces with large urban centers and/or those serving as regional hubs have traditionally received the most attention from international media and policy-makers. They've also hosted the vast majority of international troops and NGOs.

The responsiveness of aid to changes in violence would likely be higher in these environments than they would

39 Leslie, "Political and Economic Dynamics of Herat."
40 Central Statistics Organization, "Settled Population by Civil Division, 2013-15."

Table 3. OLS Regression on Log of Province-Year Aid Commitments: Disaggregated by RC (2005-14)

VARIABLES	(East)	(North)	(South)	(West)
Log of Casualties in *t-1*	0.031	0.063	-0.025	0.35*
	(0.028)	(0.071)	(0.089)	(0.13)
Log of Casualties in *t-2*	0.089**	-0.068	0.141	-0.41
	(0.036)	(0.04)	(0.11)	(0.185)
Log of Aid in *t-1*	-0.093	-0.107	-0.15	-0.014
	(0.101)	(0.071)	(0.127)	(0.277)
Provinces	14	9	6	4
N	124	81	54	40
Within R-squared	0.57	0.48	0.63	0.48

Robust standard errors clustered at the province level in parentheses. Province and year fixed effects

+ $p<0.01$, ** $p<0.05$, * $p<0.1$.

Estimated constant and population coefficient not displayed.

Table 4. Rank OLS Regression on Log of Province-Year Aid Commitments: Disaggregated by RC (2005-14)

VARIABLES	(East)	(North)	(South)	(West)
Rank of Casualties in *t-1*	0.002	0.008	0.009	0.017
	(0.007)	(0.014)	(0.014)	(0.025)
Rank of Casualties in *t-2*	0.011	-0.024**	0.022	-0.028
	(0.007)	(0.007)	(0.022)	(0.035)
Log of Aid in *t-1*	-0.086	-0.11	-0.167	0.038
	(0.1)	(0.067)	(0.113)	(0.3)
Provinces	14	9	6	4
N	124	81	54	40
Within R-squared	0.57	0.49	0.63	0.44

Robust standard errors clustered at the province level in parentheses. Province and year fixed effects

+ $p<0.01$, ** $p<0.05$, * $p<0.1$.

Estimated constant and population coefficients not displayed.

VARIABLES	(10) Log	(11) Log	(12) Log	(13) Rank	(14) Rank	(15) Rank
Casualties in t-1	0.01	0.03	0.03	0.004	0.006	0.006
	(0.045)	(0.044)	(0.043)	(0.009)	(0.009)	(0.009)
Casualties in t-1 * Six largest province	0.361+	0.311+	0.32+	0.028	0.038	0.036
	(0.055)	(0.083)	(0.077)	(0.027)	(0.034)	(0.034)
Casualties in t-2		-0.012	-0.009		-0.009	-0.009
		(0.047)	(0.049)		(0.009)	(0.009)
Casualties in t-2 * Six largest province		0.3+	0.3+		0.037**	0.033**
		(0.078)	(0.095)		(0.016)	(0.015)
Casualties in t-3			-0.017			0.001
			(0.051)			(0.01)
Casualties in t-3 * Six largest province			-0.018			-0.013
			(0.07)			(0.017)
Log of Aid in t-1	-0.001	-0.024	-0.022	0.03	0.036	0.037
	(0.064)	(0.068)	(0.069)	(0.065)	(0.067)	(0.067)
Provinces	34	34	34	34	34	34
N	304	304	304	304	304	304
Within R-squared	0.43	0.45	0.45	0.4	0.41	0.41

Robust standard errors clustered at the province level in parentheses. Province and year fixed effects.

+ p<0.01, ** p<0.05, * p<0.1.

Estimated constant and population coefficients not displayed.

be in smaller, rugged, and more remote province. To test this proposition, the fixed effects model is rerun with the inclusion of an interaction term for being one of the six largest provinces by population.[41] The results are shown in Table 5.

Models 10-12 are using the logged casualties variable and Models 13-16 have a ranked casualties variable. Across all models, the coefficient for the interaction between casualties two years prior and large province status is positive and significant at 95% statistical confidence. The coefficient for t-2 casualties stays negative and insignificant across the models, but the sum of its coefficient and the interaction term coefficient are al-

ways positive. The logged casualties model also has a statistically significant positive interaction term coefficient for the prior year. This indicates that there is a strong positive relationship between aid commitments and prior levels of violence for these large provinces, but not for the smaller provinces receiving less attention.

According to Model 12, holding all else constant, a one percent increase in casualties in the province the previous year is associated with a 0.03 percent increase in aid commitments for the smallest 28 provinces, but a 0.35 percent increase for the six largest provinces. A one percent increase in casualties two years prior is associated with a 0.01 percent decrease in aid commitments for small provinces and a 0.29 percent increase for large provinces. When taking the averages of these six provinces between 2009 and 2011, a one percent increase is

41 In order of size: Kabul, Herat, Nangarhar, Balkh, Ghazni, Kandahar. Cutting-off large province status between the 6th and 7th largest province is not entirely arbitrary. Theses were the largest 6 every year from 2001-14. The difference between Kandahar (1.2 million) and the 7th largest province (Kunduz) is more than 200 thousand. These six also represent the headquarters for each of the 5 RCs.

about 3 more casualties which is associated with $175 thousand more in aid commitments the following year.

This large province effect diminishes over time. The coefficients for the interaction term for casualties three years prior is negative and insignificant. The effect of large province status appears to disappear for casualties three year prior.

These results imply that aid commitments have in fact followed conflict in some particular areas of Afghanistan. This is likely due to the large amount of attention these areas usually receive from policymakers but may also be impacted by the larger urban population in these areas or their place as important regional trading markets.

CONCLUSION & POLICY IMPLICATIONS

The direction of development assistance in conflict environments is an important decision for policymakers. It has traditionally been focused on increasing popular support for the state by targeting the most violent-prone areas of the country which are presumed to have the greatest level of support for the insurgency. Aid commitments should then be adjusted based on the heterogenous and dynamic security situations across the country.

This paper found that while the more violent provinces in Afghanistan did receive more aid, aid commitments were broadly unresponsive to changes in the number of casualties caused by terrorist incidents. Across Afghanistan's 34 provinces over a 9 year-period a one percent increase in casualties was found to be associated with a decrease in aid commitments of between 0.03 and 0.045 percent the following year. The same change in casualties did result in aid commitment increases two and three years later, but the association is never more than a 0.075 percent change and does not reach significance. This holds true across regions of the country.

However, high responsiveness to variations in violence did result in changes to aid commitments in the six largest provinces of the country. A one percent increase in casualties in these provinces was estimated to result in a statistically significant increase of aid commitments between 0.22 and 0.28 percent the following year. On the other hand, changes in violence within the smaller 28 provinces was associated with basically no change

in aid commitments further down the road. These divergent trends grow when limiting the number of years included in the analysis.

Policy recommendations from these findings are presented below:

Aid should be predicated on information. A higher percentage of aid resources should be dedicated to informing the design and location of aid programs with actionable and accurate insight. At a sub-national level, aid is disbursed based on the information available to policymakers and donors. The most reliable and available information comes from more urban, populace, and accessible areas. Theoretically, more aid should be dedicated to areas whose support and information-sharing is most valuable to the COIN campaign. The lack of aid responsiveness to changing conflict conditions in rural areas indicates that this is not occurring. If it were, increasing violence in a province should act as a canary in the coal mine of future instability, thereby raising the value of information from that area and the importance of service delivery.

Resources and time should be dedicated to take into account changing conditions in more remote areas of the country. This could include ensuring that projects have a thorough monitoring & evaluation (M&E) component to respond rapidly to changes in the field, curtail negative unintended consequences, and increase awareness of changes in violence in more remote areas which may be otherwise overlooked.

Projects should be modest, targeted, and locally informed. Aid in conflict zones should be smaller in scope and budget. Numerous smaller projects below $100K allow for a nimbler response to changes in levels of violence across the country compared to large-scale, multi-million dollars reconstruction projects like infrastructure development. Impact evaluations have also shown that smaller projects informed by local preferences and expertise produced better outcomes.[42]

These findings indicate that the broader strategy of preferencing aid towards more violent locations was not flexible in adjusting to conditions on the ground. Additionally, the security and political situation in Afghanistan has steadily deteriorated since 2013, the last year included in this analysis. Aid provisions were neither responsive nor effective. Smaller projects will assist in

42 Kapstein, "Aid and Stabilization in Afghanistan. What Do the Data Say?"

both aims.

This means adapting the approach to aid distribution used in both the NSP and CERP. The NSP follow-on program known as the Citizen's Charter Project (CCP) should continue to be funded into the future in order to ensure aid responsiveness to changes in conflict throughout Afghanistan.[43]

BIBLIOGRAPHY

Afghanistan Provincial Reconstruction Team: Observations, Insights, and Lessons. Handbook 11–16. Ft. Leavenworth, KS: Center for Army Lessons Learned, 2011.

AidData. "AfghanistanAIMS_GeocodedResearchRelease_Level1_v1.1.1 Geocoded Dataset." Williamsburg, VA & Washington, DC, 2016. http://aiddata.org/data/afghanistan-aims-geocoded-research-release-level-1-v1-1-1.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. "The National Solidarity Programme: Assessing the Effects of Community-Driven Development in Afghanistan." *International Peacekeeping* 22, no. 4 (August 8, 2015): 302–20. https://doi.org/10.1080/13533312.2015.1059287.

Berman, Eli, Joseph H. Felter, and Jacob N. Shapiro. *Small Wars, Big Data: The Information Revolution in Modern Conflict*. Princeton, NJ: Princeton University Press, 2018.

Berman, Eli, Jacob N. Shapiro, and Joseph H. Felter. "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq." *Journal of Political Economy* 119, no. 4 (August 2011): 766–819.

Brown, Frances Z. "Local Governance Reform in Afghanistan and the 2018 Election." PeaceBrief. Washington, D.C.: United States Institute of Peace, November 2017. https://www.usip.org/sites/default/files/PB239-Local-Governance-Reform-in-Afghanistan-and-the-2018-Elections.pdf.

Central Statistics Organization. "Settled Population by Civil Division, 2013-15." Demographic Statistics - Population. Government of the Islamic Republic of Afghanistan, 2015. http://cso.gov.af/en/page/demography-and-socile-statistics/demograph-statistics/3897111.

Child, Travers B. "Conflict and Counterinsurgency Aid: Drawing Sectoral Distinctions." *Journal of Development Economics* forthcoming (2018). https://doi.org/10.1016/j.jdeveco.2018.06.003.

———. "Hearts and Minds Cannot Be Bought: Ineffective Reconstruction in Afghanistan." *The Economics of Peace and Security Journal* 9, no. 2 (2014): 43–49. https://doi.org/10.15355/epsj.9.2.43.

---

43 Brown, "Local Governance Reform in Afghanistan and the 2018 Election."

Chou, Tiffany. "Does Development Assistance Reduce Violence? Evidence from Afghanistan." *The Economics of Peace and Security Journal* 7, no. 2 (2012): 5–13.

Crost, Benjamin, Joseph H. Felter, and Patrick B. Johnston. "Aid under Fire: Development Projects and Civil Conflict." *American Economic Review* 104, no. 6 (2014): 1833–56.

———. "Conditional Cash Transfers, Civil Conflict and Insurgent Influence: Experimental Evidence from the Philippines." *Journal of Development Economics* 119, no. 1 (2016): 171–82.

Dell, Melissa, and Pablo Querubin. "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies." *The Quarterly Journal of Economics* 133, no. 2 (2017): 701–64.

Global Terrorism Database. "National Consortium for the Study of Terrorism and Responses to Terrorism (START)." University of Maryland, 2016. https://www.start.umd.edu/gtd.

ISW. "International Security Assistance Force (ISAF)." Institute for the Study of War, 2009. http://www.understandingwar.org/international-security-assistance-force-isaf.

Iyengar Plumb, Radha, Jacob N. Shapiro, Benjamin Crisman, Manu Singh, and James Mao. "Stabilization in Afghanistan: Trends in Violence, Attitudes, Well-Being and Program Activity." Santa Monica, CA: RAND Corporation, 2017. https://www.rand.org/pubs/working_papers/WR1192.html.

Kapstein, Ethan. "Aid and Stabilization in Afghanistan. What Do the Data Say?" United States Institute of Peace Special Report. Washington, D.C.: USIP, June 2017. https://www.usip.org/sites/default/files/2017-06/sr405-aid-and-stabilization-in-afghanistan.pdf.

Khanna, Gaurav, and Laura Zimmermann. "Guns and Butter? Fighting Violence with the Promise of Development." *Journal of Development Economics* 124 (January 2017): 120–41. http://dx.doi.org/10.1016/j.jdeveco.2016.09.006.

Leslie, Jolyon. "Political and Economic Dynamics of Herat." PeaceWorks. Washington, D.C.: United States Institute of Peace, 2015.

Lyall, Jason, Graeme Blair, and Kosuke Imai. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107, no. 4 (November 2013): 697–705.

Nunn, Nathan, and Nancy Qian. "US Food Aid and Civil Conflict." *American Economic Review* 104, no. 6 (2014): 1630–66.

"OECD DAC Statistics." OECD, 2019. http://stats.oecd.org/qwids/about.html.

Osman, Borhan, and Anand Gopal. "Taliban Views on a Future State." New York, NY: New York University Center on International Cooperation, July 2016. https://cic.nyu.edu/sites/default/files/taliban_future_state_final.pdf.

Sexton, Renard. "Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan." *American Political Science Review* 110, no. 4 (November 2016): 731–49. https://doi.org/10.1017/S0003055416000356.

SIGAR. "Afghanistan's National Solidarity Program Has Reached Thousands of Afghan Communities, but Faces Challenges That Could Limit Outcomes." Office of the Special Inspector General for Afghanistan Reconstruction, March 22, 2011.

The Asia Foundation. "Afghanistan in 2014: A Survey of the Afghan People." Survey of the Afghan People. San Franciso, CA: The Asia Foundation, November 2014. https://asiafoundation.org/resources/pdfs/Afghanistanin2014final.pdf.

United States Army. "Commander's Guide to Money as a Weapons System," 2009. https://www.usma.edu/cnrcd/SiteAssets/SitePages/Government%20Publications/CALL%20MAAWS%20Handbook%2009-27%20(APRIL%202009).pdf.

"U.S. Government Counterinsurgency Guide." United States Government Interagency Counterinsurgency Initiative. Washington, D.C.: Department of State, Bureau of Political-Military Affairs, January 2009. https://www.state.gov/documents/organization/119629.pdf.

U.S. Senate. Committee on Foreign Relations. "Evaluating US Foreign Assistance to Afghanistan." U.S. Senate. Committee on Foreign Relations, 2011. https://www.gpo.gov/fdsys/pkg/CPRT-112SPRT66591/pdf/CPRT-112SPRT66591.pdf.

Gender Representation in Graduate Education Materials

Taylor Trummel

University of California San Diego, School of Global Policy and Strategy

This paper examines the gender representation disparity within syllabi in courses from UC San Diego's graduate School of Global Policy and Strategy (GPS). It uses a dataset crafted from the syllabi of the 2017-18 academic year and accounts for the gender composition of assigned readings within the syllabi. It includes variables such as publication year, department, and faculty's gender in the study. It tests the results found in 1,996 assigned readings from 95 International Affairs courses taught by 55 faculty members or lecturers. Overall, it finds women are included in roughly 1/3 of all assigned readings at GPS. This proportion diminishes as I apply stricter, more narrow definitions and constraints on the data. In comparison to Management and Political Science, the study finds Economics to be the department with the most glaring trends in disparity within the disciplines of International Affairs. The study reveals statistically significant correlations between the assignment of female written articles and the faculty member's gender, publication year, and for cases of group authorship. It also uncovers only one faculty member from the sample with gender balanced syllabi. The paper discusses other related research and explains the importance for equal representation in teaching materials. Discussing implicit bias, the paper suggests that rather than intentionally discriminating against female authors, academics may rely on canon and perpetuate the problem by utilizing the publications of disproportionately male expertise when creating their course syllabi. It then suggests several solutions to alleviate the disparity to include the use of the Gender Balance Assessment Tool, including more recent publications, the consultation of sources such as Women Also Know Stuff, and practicing consciousness in assigning more diverse materials.

INTRODUCTION

Since the second wave of feminism, the discrimination women experience in American society is more structural and implicit than obvious and explicit. Today, women are protected by nondiscrimination legislation within institutions and industries in the United States, yet we continue to see disparity in representation across sectors. This paper analyzes the status of women's representation in graduate education materials within the International Affairs discipline. It uses a manually created dataset studying the genders of assigned authors to synthesize a conclusion about the disparity between male-written and female-written teaching tools. The paper describes why representation matters, reports the results, discusses what may be driving these trends and related research, and proposes a solution.

DESCRIPTION OF THE DATA & METHODS

The dataset used for this study contains 1,996 observations from the graduate program of the School of Global Policy and Strategy (GPS) at UC San Diego. Each observation represents a required reading listed in a course syllabus. It does not include optional readings listed in the syllabus nor assigned readings that were

included outside of the syllabus. The principle variable is the gender variable identifying whether the assigned reading contains a female author. It then goes further and documents where there was only one author, more than one author, and if co-authorship resulted in mixed genders in the publication. In addition, the dataset records the assigning faculty, the course assigned, the GPS specialization track, the course disciple, the publication year, the year assigned, and the discipline of the faculty member. The majority of the variables are coded as dummy variables. Others identify what would have otherwise been a string variable as an assigned numeric code. A separate page of the dataset translates the assigned codes for faculty, course name, faculty discipline, and GPS specialization track. There are also two time variables that document the year assigned and the year the assigned reading was initially published. The dataset does not include readings from the GPS language classes or undergraduate international relations courses. It is an almost exhaustive sample of all readings assigned to GPS graduates students during the 2017-18 academic year.

The required readings omitted from the study were few, numbering less than 50, but those that were omitted were done so out of the inability to identify the authors' gender. For example, some works contained authors whose name was not immediately recognizable as male or female. In this instance, I searched the author by name and publication online. There were instances when these searches did not yield information about the author's gender such as through an image or descriptive pronouns.

Finally, I'd like to note the nature of this study and potential limitations. The dataset relies on the researcher's ability to identify an author's gender as male or female. It did not include a variable to measure a gender outside of the binary model of gender. Furthermore, the gender variable is measured solely on the researcher's perception based on pronouns, pictures, and the name of authors. It did not allow those being studied to self-identify their gender. An improvement to the dataset could be to survey the authors included in the study for their self-identified gender, however I find the likelihood of a survey to observe an author's gender unlikely and that any differences that may arise from a group identifying as a gender other than the one observed by the researcher is likely marginal and would not impact the overall trends found in the results.

DEFINITION

For the purpose of this study, a broadest measurement of gender inclusion is represented by assigned readings that contain at least one female author. A more narrow measurement uses the variables that analyze whether there was only one female author as the independent and sole author of a publication. Finally, the most narrow definition of representation looks at assigned readings that contained multiple authors and identifies those that contain a homogenous team of female authors as opposed to mix-gender authorship. These three categories are used throughout the analysis of the results.

RESULTS

1. Overall Findings

(Figure 1.1) Overall, using the broadest measurement of the inclusion of women as authors in assigned materials, I find that only 29% of readings in GPS syllabi contain a female author.

Analyzing the entire dataset, 29% of assigned reading materials contain at least one female author. That means almost 1/3 of assigned readings had at least one female author. More narrowly though, I looked at how many assigned readings were written by an independent female author, as opposed to articles published by a group. To this, I found 19% of assigned readings with one author were written by a woman, and conversely, 81% of assigned readings with only one author were written by men. Even more narrowly defined, if the assigned reading was written by two or more people and consisted of only female authors, this resulted in 4% of assigned readings across all observations containing two or more authors. The opposite, meaning readings written by multiple men and only men, resulted in 59%

	(1) Author's Gender	(2) Author's Gender	(3) Author's Gender
Mixed Gender Co-Authors	0.840***		
	(0.0194)		
Economics Department	-0.0515**	-0.0743**	-0.0877***
	(0.0182)	(0.0282)	(0.0249)
Publication Year		0.00465***	
		(0.000905)	
Multiple Authors			0.227***
			(0.0201)
_cons	0.160***	-9.035***	0.207***
	(0.00883)	(1.817)	(0.0139)
N	1996	1630	1996
adj. R^2	0.484	0.018	0.060
rmse	0.327	0.454	0.441

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(Table 1.1) Regression 1 finds that readings in the syllabi are significantly more likely to occur in instances of mixed gender group authorship as opposed to groups of solely women. Recall that only 4.0% of all assigned readings are written by teams of women, while 59.0% are written by teams of men. Regression 2 finds that women are more likely to be featured as authors in more recent publication years. Regression 3 finds that women are more likely to be found in syllabi as a member of a team as opposed to independent authors. In all 3 tests the department within International Affairs serves as a control variable.

of the materials. The leftover 37% constitutes readings that contain mixed gender authorship. These three definitions of analyzing the representation of women's work reveal that there is a disparity in their representation under three different perspectives of considering the conditions of their inclusion.
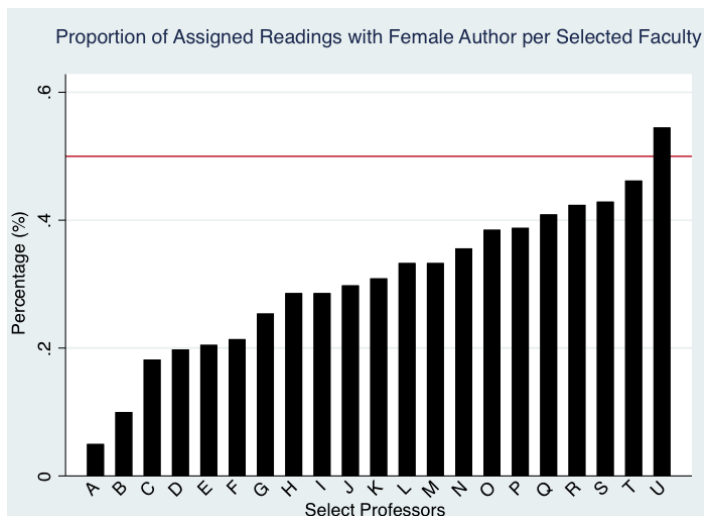
In a regression measuring the correlation between the author's gender and independent variables, I ran three regressions. Regression 1 measures the correlation between the author's gender and the frequency of publishing with at least one other co-author and within a mixed gender team, finding a highly statistically significant positive correlation. This means that female authors are high correlated with the presence of a male author. Regression 2 measures the correlation between the author's gender and the publication year, finding a highly and positively correlated statistically significant result. This means that female authors are strongly correlated with later publication years. Finally, Regression

3 measures the correlation between the author's gender and whether the assigned article was written by more than one person. This too resulted in a strongly positive correlation and statistically significant at the 99.9% confidence interval.

2. Findings about Faculty

How does the gender representation disparity look between faculty? The dataset allows us to look closely at how the gender disparity in reading materials breaks down for each faculty member included in the 2017-18 academic year sample.

Figure 2.1 depicts the results of select faculty in the study using the broadest measurement of gender parity as outlined in the definition, observations that include at least one female author. The names of the professors are omitted for the purpose of this publication. It includes both the professors with the greatest parity of gender and the least parity, as well as a selection of the results

Proportion of Assigned Readings with Female Author per Selected Faculty

(Figure 2.1) Using syllabi available from faculty in the 2017-18 academic year, the study found that only one faculty member, "Professor U" in this anonymized graphic, created syllabi that had a balanced representation of female authors using the broadest measurement of women's representation. Specifications of the results of each faculty member can be made available or requested.

Gender Representation by Faculty Gender

(Figure 3.1) This figure depicts the breakdown in the share of assigned materials by the assigning faculty's gender. The yellow and red quadrants correspond to the share of the materials assigned by male faculty, while the blue and green correspond to those assigned by female faculty. While male faculty assign slightly more of the share of articles containing female authors, they make up a much larger proportion of faculty at GPS. Only 32.7% of the faculty included in this study are women, yet they assign almost half of all articles containing female authors.

in the middle. First, the results reveal that only one GPS Professor assigned readings above parity. Professor U's syllabi contained a proportion of 53% of assigned readings containing at least one female author. Meanwhile, only 5% of Professor A's assigned readings contained at least one female author, marking the greatest source of disparity in assigned readings amongst faculty at GPS.
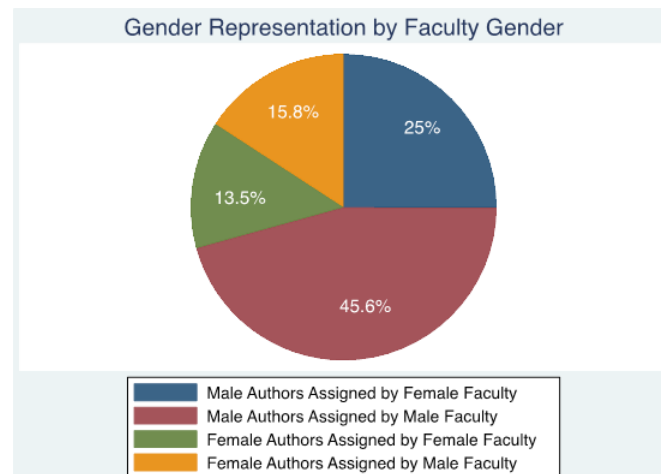
3. Faculty Gender

Do female professors at GPS assign more balanced reading materials than male professors? The dataset allows us to answer this question.

Amongst the 585 assigned articles that have a female author, 46% are assigned by female faculty members, while 54% are assigned by male faculty members in the sample. However, of the 55 UCSD faculty and lecturers included in this sample, 67.2% are male, while 32.7% are female. This signals that while there is almost no difference between the female share of assigned articles dependent uponthe faculty member's gender, female faculty members make up a smaller proportion of those assigning readings. Because of this, a look beyond simple statistics is necessary.

To investigate the strength of the correlation between female authored assigned readings and the faculty member's gender, I ran a regression controlling for the departments. I find a strong statistical significance with a positive correlation, signaling that female faculty members are more likely to assign readings with female authors, as indicated in Table 3.1.

Figure 3.1 illustrates the proportion of female assigned

readings by faculty members' gender. The results depicted in Figure 3.1 are consistent with the overall finding that women were included in roughly only 1/3 of all assigned readings at GPS during the 2017-18 academic school year. What this figure also depicts is the disparity between gender representation among faculty in GPS courses at UCSD during that school year, with male

	(1)
	Author's Gender
Faculty's Gender	0.0920***
	(0.0218)
Economics	0.293
	(0.321)
Political Science	0.311
	(0.321)
Management	0.288
	(0.322)
_cons	-0.0460
	(0.321)
N	1996
adj. R^2	0.009
rmse	0.453

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(Table 3.1) Regression 1 finds that female faculty are more likely to assign readings containing female authors, controlling for department.
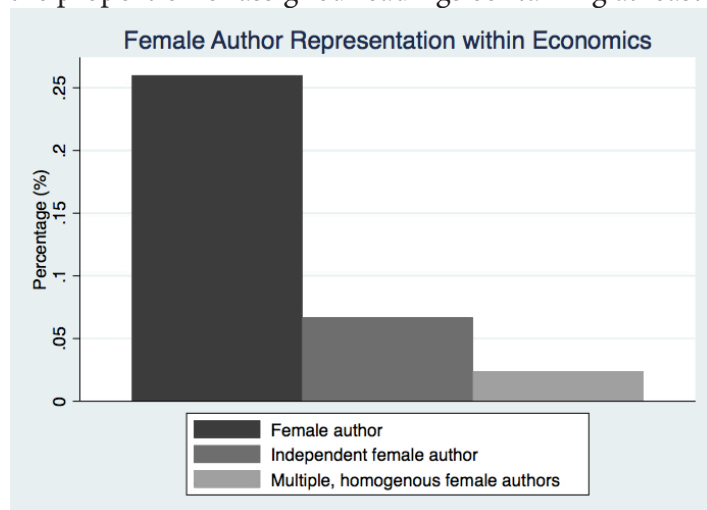
faculty accounting for a much greater share of assigned readings than female faculty. Of the 1,996 assigned readings in the dataset, 38% of them were assigned by female faculty members. While 32.7% of the faculty and lecturers in the sample are female, they constitute 38% of assigned readings, suggesting that female faculty and lecturers may assign slightly more readings than male faculty and lecturers.
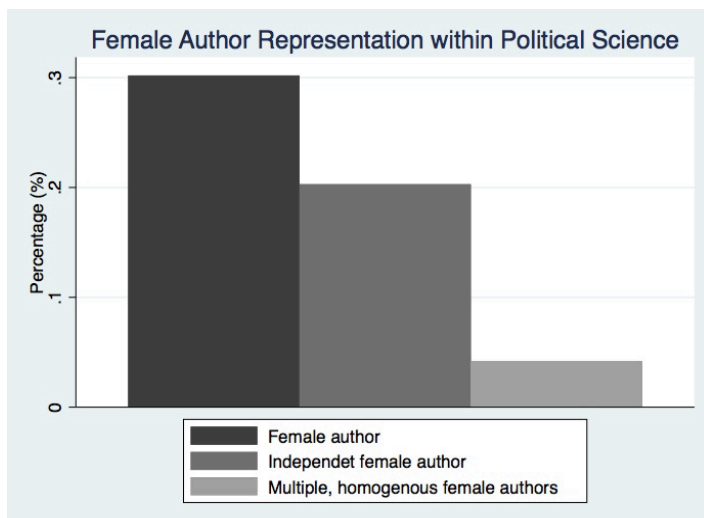
4. Department

How does this trend look between subjects within iternational affairs? The dataset allows us to look closer at the breakdown between departments. To do this, I categorized GPS courses as either a course in Economics, Political Science, or Management.

As depicted in in Figure 4.1, among the courses focused on economics, 26.0% of assigned readings have a female author. This is only 3.3% below the overall average. However, when I narrow the focus, the results are alarming. First, among articles with only one author, only 6.0% of those are written by women. Furthermore, in articles with multiple homogenous authors only 2.0% of assigned readings are written by two or more women. In comparison, articles written by more than one man constitute 68% of the readings, while heterogeneous authorship constitutes the remaining percentage.

Within Political Science, Figure 4.2 depicts the results of the study for GPS. The same overall trend exists as with Economics, however Political Science reveals a greater proportion of independent female authors than Economics. Additionally, in comparison to Figure 4.3 depicting the trends in Management related courses, the proportion of assigned readings containing at least

(Figure 4.2) In addition to the two other measurements as outlined in "Definitions," the study finds papers written by teams of women make up less than 1.0% of all assigned material within Management. The study finds that 30% of assigned readings in Management feature a female author in any capacity.

one female author is very similar to that of Political Science.

Among all assigned readings in Political Science, 30% contain a female author at all. Constraining this further, among assigned materials containing single authors, 20% of readings consist of a single female author. Conversely, the results of this study find 80% of single author articles are written by men in Political Science. Finally, in the narrowest constraints, among assigned materials containing multiple authors, 4.0% represent multiple, homogenous female authors while 54% of assigned readings with multiple authors are written by men homogenously. The remaining 46% of this narrow constraint represent heterogeneous gender authorship. While 4.0% is still strikingly lowly, it is still twice as much as within Economics.
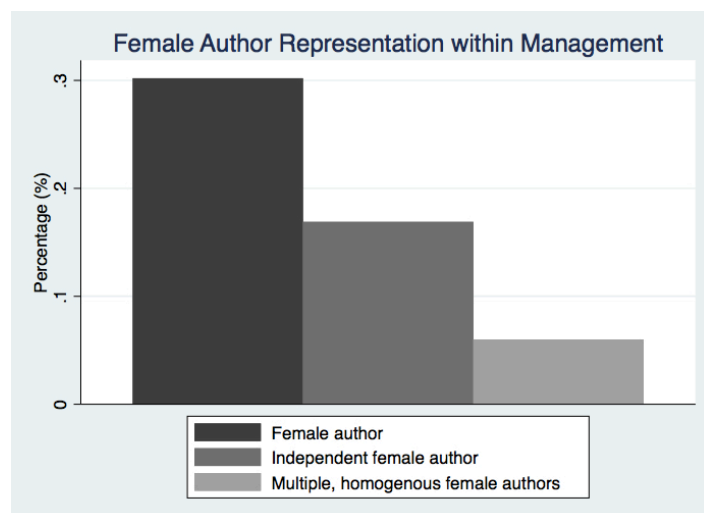

(Figure 4.1) Depicting the three different levels of measurement for women's representation in assigned materials, this figure reveals that within Economics the strictest measurement (multiple, homogenous) finds that papers written by teams of women make up less than 0.05% of all materials.
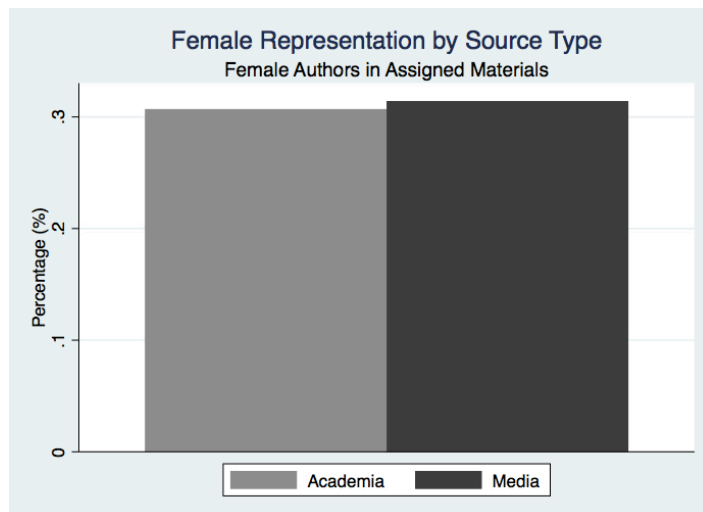

(Figure 4.3) In addition to the two other measurements as outlined in "Definitions," the study finds papers written by teams of women make up less than 1.0% of all assigned material within Management. The study finds that 30% of assigned readings in Management feature a female author in any capacity.

Finally, within Management courses at GPS, there is a similar trend as with the other disciplines. Figure 4.3 depicts an almost identical proportion of female authors as Political Science. Differing from Political Science though, Management constitutes a smaller proportion of independent female authors. Yet, Management also includes a slightly greater proportion of multiple, homogenous female authors in their assigned readings. When compared to Economics, there is a greater proportion of female authors in all categories of the study. The results suggest that depending on the selected definition of representation, either Political Science or Management is performing the best in terms of gender representation.
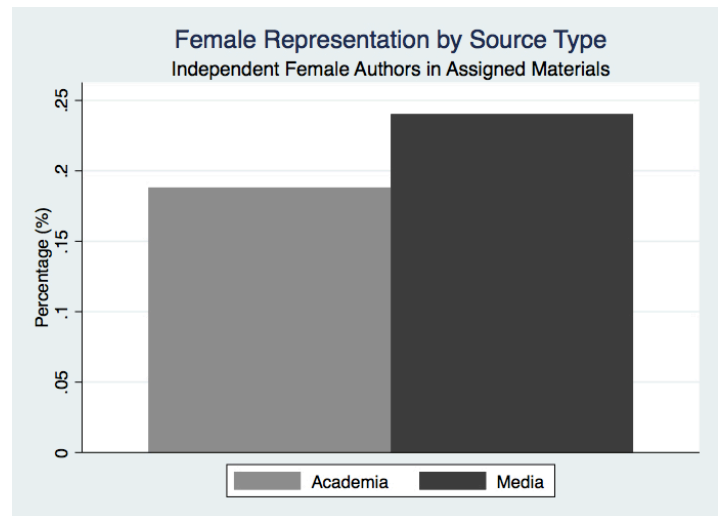
5. Source type

How does the data look between source types? Within this section the differences in gender representation between different types of sources included in syllabi as teaching tools will be discussed. Faculty assign a variety of reading materials throughout the quarter to enable student learning. These sources draw from scholarly published articles to the Economist or the New York Times. A variable to account for the differences between these source types is recorded in order to study if a different trend exists between them. Among the readings selected for graduate teaching materials in international affairs, do women receive greater representation within journalistic or academic sources?

The results depicted in the figures throughout section five illustrate the trend. At first glance, there is an almost equal result between women in journalism and women in academia, with the proportion of women authors

(Figure 5.2) Of readings written by only one author, women in media share a greater proportion than those in academia.

represented in academia only slightly lower (Figure 5.1). Still, women only constitute roughly 1/3 of assigned materials as outlined in the overall trends.

However, when stricter constraints are applied on the sources written by women and focus on those with only one author, the differences between female authors in academia and journalism grow. In Figure 5.2, we can see the increased disparity between women in academia and women in media sources, with single female authors only constituting 19% of works published by only one author. Whereas women in media represent 24% of articles assigned with only one author. This translates to a 1:5 ratio for women in academia and a 1:4 ratio for women in journalism. Conversely, men make up 81% of single authored assigned materials from academia and 76% of single authored assigned materials from journalism.

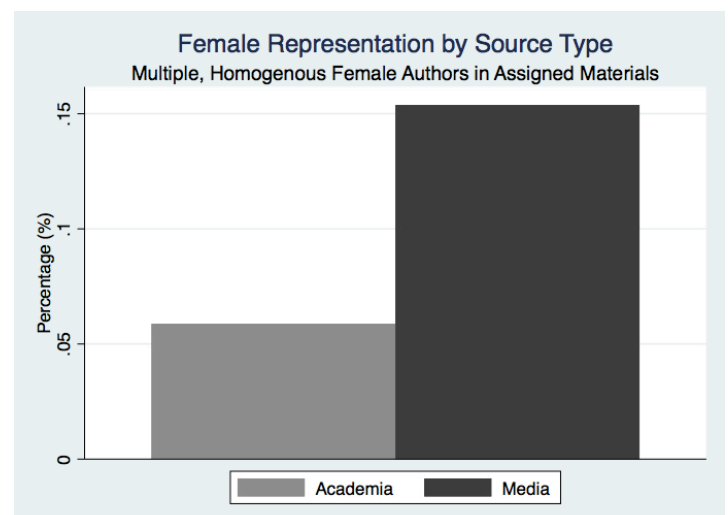**(Figure 5.1)** The study finds almost equal treatment between articles published by women in media and academic sources.
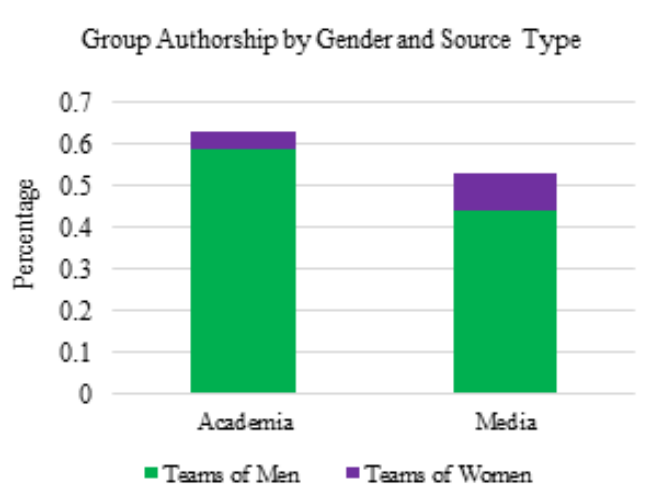

(Figure 5.3) For all readings composed of more than one author, assigned readings in academia written by groups of women made up 6.0% while those in media comprise just over 15.0%.

Group Authorship by Gender and Source Type

(Figure 5.4) Looking more closely at group authorship, men in both the media and academia are included at a much greater proportion than women. The greatest disparity is between men and women in academia, where articles written by groups of men compose just under 60.0% of assigned readings while articles written by groups of women constitute 6.0% of all assigned academic readings with multiple authors.
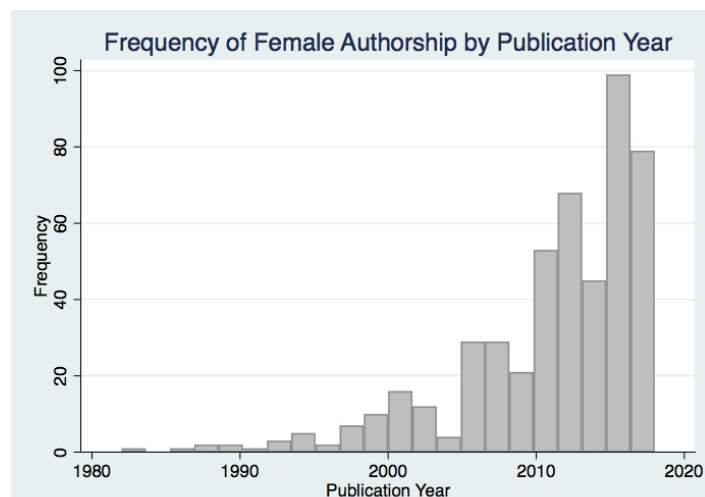
In an even narrower look at women's representation, the proportion of assigned readings with multiple, homogenous authorship by women reveals even greater difference. Figure 5.3 depicts this trend. The results show the disparity between women depending on source types increases once again. Again, women occupy an even smaller share of assigned readings in this category. In academia, homogenous teams of female authors make up roughly 6.0% of all works published with more than one author. In media sources, teams of women make up just over 15% of assigned readings.

Finally, Figure 5.4 visually compares the difference between men and women in this narrowest constraint of this study. The green bars depict the trend for men within the two types of sources use in the study, whereas the purple bars depict the trend previously described for women. The left two bars of Figure 5.4 represent sources in Media and the two bars to the right represent sources in academia. The remaining share not depicted in this bar graph represents works with multiple authors and mix-gendered.

6. Time

How does the publication year impact the author's gender? The study shows that assigned readings written by women that were included in syllabi had been published at far greater frequency 21st century than the 20th, with even greater frequency since 2010. Figure 6.1 depicts this trend with a histogram.

Furthermore, Table 1.1 depicts the results of a regression testing the correlation of publication year on the gender of the author of an assigned reading. It finds a

(Figure 6.1) This histogram counts the number of readings containing a female author by year. The years following 2010 show the greatest frequency. Regression 2 in Table 1.1 also found a positive correlation between publication year and the author's gender.

strong, statistically significant correlation between the two. Figure 6.1 also reveals that a significant number of the articles assigned containing a female author come after 2010, when the frequency of their presence in assigned materials almost doubles.

DISCUSSION

The results of this study beg the question as to why this trend is occurring. Is it that implicit bias is impacting the choices faculty make when selecting their assigned readings? Or is this a proportional representation of the available content published by female writers? Both possibilities can be addressed with research and data from other sources.

First, discussing the role of implicit bias is critical. It impacts everyday actions, choices, and judgement in every part of social life, whether individuals are conscious of their bias or not. "Studies of implicit associations show that gendered attributes are part of almost everyone's preconscious," states Jane Mansbridge of Harvard University. In one study, researchers find that gender affects hiring and promotion patterns. The study finds that male managers respond with a lessened desire to work with an employee who has a woman's name, but not when the employee who has a man's name even if that employee makes a number of irritating negotiating demands (Voeten 2013). This study reveals that implicit assumptions negatively impact the professional careers of women.

Second, to discuss whether this problem is due to the lack of sufficient publications by women, I investigated the data regarding women's enrollment in Political Science departments. Trends show more female Political Science graduate students, faculty, and tenured faculty

decade after decade. In 2010, women made up roughly a third of Political Science faculty and more than half of graduate students (Shames). This data may indicate that results of the GPS focused study are a reflection of available content, with women constituting roughly 1/3 of all Political Science faculty in the United States. This alone is not enough to arrive at this conclusion. Rather, data regarding the output and publication rates between genders is the best indicator of whether this is a reflection of what is available for use in syllabi or due to implicit bias. One such study examines the publication output by graduate student across fields. The results find than female student publish 8.5% fewer articles than men (Pezzoni, et al. 2016). While this may be initially discouraging, section six of the results portion of this paper indicates the rapidly increasing frequency of female authors in articles published after 2010 in syllabi. Coupling this information with the changing demographics of academia, it suggests that this disparity will continue to decrease with time.

Other research in this field has found similar results. For one, current research finds that gender of the instructor for graduate courses matters significantly for what type of material is taught. In the first study of its kind in International Relations, researchers concluded that women tend to cite themselves less than men. Furthermore, they find that men, making up a disproportionate share of scholars in the field, tend to cite men more than women (Maliniak, et al. 2013). Another study similarly finds that on average, female instructors assign significantly more research by female authors than male instructors. This result found by Jeff Colgan of the Watson Institute at Brown University is consistent with the results found in my study of GPS syllabi. Additionally, his study finds that women appear to be considerably more reluctant than men to assign their own research as required readings (Colgan 2017). Related to another variable included in my study, among articles with more than one author, mix gender teams are significantly less likely to cite publications by female authors in contrast to independent female authors (McLaughlin Mitchell, et al 2013).

Examples of related studies do not stop there. In regard to the variables included in my research measuring women's representation in academic sources versus media sources in section five, other scholars too have looked more closely at the media. In analyzing the New York Times' coverage of the U.S. presidential primary, looking at every article from March 2015 through January 2016, researchers find that 80% of the Political Scientists quoted in those articles are men (Taub 2016).

The results of this study focus on the sources of news media, perhaps the equivalent of assigned readings in academic syllabi. In another case, freelance journalist and self-identified feminist Adrienne LaFrance analyzes her own work and discovered she only cites women about a quarter of the time (Leonhardt 2018). While my research study looked at the number of women publishing in media sources, related studies have looked at the sources of these published media authors and discovered a significant disparity.

Issues like this are important to highlight because the trend of associating men with "expertise" contributes to a lifetime of building implicit bias. The absence of women as experts in public sources reinforces stereotypes as to the appropriate role of women. It further stereotypes who is knowledgeable and respected within disciplines. While disparity is diminishing over time, Americans still see women as a significant majority as faculty and students in disciplines such as nursing, English, and gender studies while men are continually a significant majority of faculty and students in disciplines such as economics, statistics, and engineering.

This partially is due to and in turn contributes further to a "role model effect," where young students form ideas of their abilities and aspirations that are reinforced by external stimuli (Klar 2016). The ability to envision oneself in a position such as CEO, engineer, or doctor becomes much easier when an individual is able to see or speak to a role model that represents their experiences, whether it be in race, gender, socioeconomic status, or a plethora of other diverse characteristics. As an example, research has found female students perform significantly better in introductory math and science courses when they are taught by female faculty. In turn, they are more likely to pursue majors in science, technology, engineering or math after positive performance in these fields (Carrell, et al. 2009). In another study related to the role model effect, researchers focused on the visual teaching materials used in the sciences. They find that more men are depicted with a science profession than women and that more women are depicted as teachers than men (Kerhoven, et. al 2016). Subtly, daily interactions and associations contribute to the disparity found in sectors all over the United States, despite the technical potential to have parity. Course syllabi is one such source that serves to influence students and can have long term impacts by contributing to implicit bias.

The importance of equality and representation of diverse perspectives aside, there are also numerous studies that find more tangible benefits for parity and di-

versity. Within business, Gallup finds that companies composed of gender diverse teams had higher revenues than those that were not. Within retail, they find gender diverse teams correlated with 14% higher revenues while those within hospitality had 19% higher average quarterly net profits (Badal 2014). Diversity in representation and ideas has real benefits, whether nuanced or quantifiably testable.

SOLUTIONS

Whether this is a problem of supply or implicit bias, the disparity exists. With increasingly diverse student bodies and gender parity amongst students in many departments, crafting teaching materials that reflect the diverse perspectives of students is important. Diversifying syllabi to include more recent publications is one method to increase the representation of women and reflect current research and thought. While the traditional canon within disciplines such as Political Science and Economics may have an important role, moving away from antiquated syllabi in favor of constantly evolving and dynamic syllabi will inevitably include more modern research and understanding of the world and academia. However, faculty are not left to their own ingenuity to assess their syllabi. Rather, tools exist to help curb this problem.

Dr. Jane Sumner of the University of Minnesota offers a solution to gender disparity in syllabi. The Gender Balance Assessment Tool, which is accessible online, allows faculty to upload their syllabus for an immediate assessment of the balance in their syllabus. This tool evaluates the probabilistic gender of each name on a syllabus and then provides an estimate of what percentage of the authors on a syllabus are women. It even goes a step further and assesses the probabilistic breakdown of race included in uploaded syllabi.

Sources such as Women Also Know Stuff offers a database of materials published by women. It aims to promote the expertise of women by sorting a collection of almost 2,000 female scholars and their research. It seeks to help balance conferences, panels, syllabi, and research. While this platform features women in Political Science, other such databases exist in the fields of Chemistry, History, and many more. Additionally, women in Political Science can register to be a featured scholar in their database. Other online resources exist as well. The website Gender Avenger serves to keep individuals and organizations accountable for their representation of women. It features an online petition for men to sign, pledging that they will not serve on an all-male expert discussion panel and features a page dedicated to rebuttals to common excuses given for not including women in panels.

Another component of incorporating gender into academia comes from the 1991 American Political Science Association recommendation to mainstream gender topics in core required courses in the discipline of Political Science. This mainstreaming helps to normalize the discussion of the topic in the field, rather than saving it for the expertise of one or two faculty members to teach as an elective (Cassese, et al. 2012). It also helps to reinforce the topics of gender as centrally important to the understanding of political affairs. Furthermore, mainstreaming helps to recruit and retain women in the major. Hopefully in turn, the presence of more women in the major leads to more female faculty and publications by women.

Finally, individual and collective consciousness can help to alleviate the problem. Being conscious of one's own implicit bias and assumptions is the first step to better advocating and representing others. This skill is important not only in crafting syllabi, but also as a teacher; from small interactions in daily life to office hours and lectures. Understanding the roots and socialization of one's own perspective opens the mind to understanding the perspectives of others.

With the issues discussed in mind, Amanda Bittner of Memorial University of Newfoundland stated:

> "I assigned course syllabi with very few women on them, and I have certainly written papers that undercite the contributions of my female colleagues. I didn't do these things on purpose. I did them because these issues were not on my radar and I wasn't thinking about them at all."

Daniel Nexon of Georgetown University summarized it well when he said:

> "The 'gender gap' in citations is rooted in syllabi. Our sense of whom we need to cite for a particular argument is based on whom our instructors placed on the syllabus for a particular week. This suggests that with a little bit of effort, and the downstream effects of having a higher percentage of active female scholars, the problem is likely to become less intense over time."

Revealing room for improvement and a self-check,

more progress can be made in the inclusively of expertise. Better gender representation parity in syllabi is one such avenue to use to improve the way faculty and students think about the experts in the world.

CONCLUSION

Sources matter. Diverse experiences shape the way academics think about research questions and write. These thoughts are seen through the published materials writers produce. When syllabi continue to perpetuate the importance of classical publications for study, they are also contributing to the problem of representing significantly more work published by men than women. While the results reveal a more optimistic trend for the 21st century, it is still far from parity. When the students graduate classrooms in International Affairs represent an equal balance of gender diversity, it is worth considering the interests of forward looking students rather than continually stressing the importance of works written several decades ago almost entirely in the perspective of male authors. The results of this account are supported by numerous other research studies focusing on the role of gender in academia. Many variables may be contributing as to why this trend is found today. However, it is safe to argue that implicit bias is affecting the assigned readings used to teach. The evidence found for the School of Global Policy and Strategy are consistent with other studies, but the graduate program uniquely has an opportunity for growth and improvement as a result of this targeted study and suggested solutions.

BIBLIOGRAPHY[1]

Badal, Sangeeta. "Business Benefits of Gender Diveristy." Gallup, January 2014. https://www.gallup.com/workplace/236543/business-benefits-gender-diversity.aspx.

Bittner, Amanda. "Valuing & Supporting the Research of Women in Political Science." PRAXIS, 2017. http://praxispolisci.ca/valuing-supporting-the-research-of-women-in-political-science-by-amanda-bittner/.

Carrell, Scott, et al. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." NBER, 2009. http://www.nber.org/papers/w14959.

Cassese, Erin, et al. "Integrating Gender into the Political Science Curriculum." Political Science and Politics, 45 (2), 2012, pp 238-243. https://pdfs.semanticscholar.org/fc7f/1f2c912b0aff82cab1ff84b5e59c03d13a90.pdf

Colgan, Jeff. "Gender Bias in International Relation Graduate Education? New Evidence from Syllabi." American Political Science Association, 50 (2), 2017, pp 456-460.

"How to beat the Excuses, Updated: Responses to the Top 12 Excuses for a Lack of Women Speakers." Gender Avenger. November 2018. https://www.genderavenger.com/blog/how-to-beat-the-excuses-updated.

Kerkhoven, Anne, et al. "Gender Stereotypes in Science Education Resources A Visual Content Analysis." PLoS One, 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5112807/

Klar, Samara. "Experts Weigh In: Women Also Know Stuff." HuffPost, 2016. https://www.huffingtonpost.com/samara-klar/experts-weigh-in-women-al_b_9404388.html.

Leonhardt, David. "I'm Not Quoting Enough Women." The New York Times, 2018. https://www.nytimes.com/2018/05/13/opinion/women-sexism-journalism-conferences.html.

Maliniak, Daniel, Ryan Powers, and Barbara F. Walter. "The Gender Citation Gap in International Relations." International Organization 67, Fall 2013, pp

---

1 In effort to practice my own proposed solution, I have consciously analyzed who I am citing in my research. Seventy-five percent of my sources come from articles with at least one female author.

889-922.

McLaughlin Mitchell, Sara, et al. "Gendered Citation Patterns in International Relations Journals." International Studies Perspectives, 2013. pp 485-492.

Nexon, Daniel. "The Citation Gap: Results of a Self-Experiment." Duck of Minerva, 2013. http://duckofminerva.com/2013/08/the-citation-gap-results-of-a-self-experiment.html.

Pezzoni, Michele et al. "Gender and the Publication Output of Graduate Students: A Case Study." PLoS One, 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4711938/

Sumner, Jane. "The Gender Balance Assessment Tool (GBAT): A Web-Based Tool for Estimating Gender Balance in Syllabi and Bibliographies." American Political Science Association, April 2018.

Shames, Shauna. "Political Science's Gender Problem." Hippo Reads. http://hipporeads.com/political-sciences-gender-problem/.

Taub, Amanda. "The #ManPanel Problem: Why are Female Experts Still so Widely Ignored? Vox, 2016. https://www.vox.com/2016/3/16/11245454/man-panel-problem-female-experts-ignored.

Voeten, Erik. "Explaining the Gender Gap." Monkey Cage, September 2013. https://www.washingtonpost.com/news/monkey-cage/wp/2013/09/30/explaining-   the-gender-gap/?noredirect=on.

DOES LOOSENING THE "NEGATIVE LIST" STIMULATE FDI INFLOWS AND FOREIGN TRADE?

A Case Study of the Shanghai Pilot Free Trade Zone, 2013-2018

Chengyu Fu
Duke University

The Shanghai Pilot Free Trade Zone (FTZ) is China's first trial to test the FTZ model, and it has achieved growth in both new foreign direct investment (FDI) and foreign trade. Although the Chinese government has claimed that the policies it adopted are responsible for economic development in the Shanghai Pilot FTZ, few empirical studies exist to support these claims. This essay focuses on how the loosening of the "Negative List," one of the key measures taken by Shanghai Pilot FTZ, impacted FDI inflows and foreign trade volumes. A time series analysis provides no evidence that the elimination of restrictions in the Negative List influenced Shanghai's FDI attraction and total international trade. Based on the results, the essay suggests not to rely on loosening the Negative List to stimulate economic growth. Policies including tax relief, currency exchange liberalization, and developing the rule of law are all possible approaches to promote economic globalization in FTZs in the future.

INTRODUCTION

Free Trade Zones (FTZs), defined by the World Bank, refer to "duty-free areas, offering … facilities for trade, transshipment, and re-export operations" (Akinci et al., 2008). With low taxes, tariffs, and institutional barriers, FTZs have been established around the world to stimulate foreign trade and attract foreign direct investment (FDI). Some major FTZs around the world - including those in New York, Amsterdam, Dubai, and Hamburg - have successfully attracted foreign capital and promoted domestic economic growth (Farole and Akinci, 2011).

China has been constructing FTZs since 2013; the first was the Shanghai Pilot FTZ, which has been in development for five years as of 2019. As the first FTZ, the

Shanghai Pilot shoulders the responsibility of attracting foreign investors and improving international commerce. However, the Shanghai Pilot FTZ is more of a "testing ground:" that is, the central government wants to determine whether the FTZ model is suitable for economic development in mainland China. As a result, the FTZ has been widely seen as an exploration of new approaches on how to move forward with reform and opening up – one of the most important political tasks for contemporary China.

The Shanghai Pilot FTZ has achieved development in capital accumulation and foreign trade in its five years of operation: official statistics show that as of June 2018, 8,696 foreign-invested enterprises have been established, collectively bringing in a contractual investment of 110.24 billion dollars (Xinhua News

Table 1: Timeline for the Construction of Pilot FTZs in China	
Date	Location
September 29, 2013	Shanghai
April 21, 2015	Guangdong, Tianjin, Fujian
April 1, 2017	Liaoning, Hubei, Shannxi, Henan, Chongqing, Sichuan, Zhejiang
October 16, 2018	Hainan

Agency, 2018). This rapid growth demonstrates the feasibility of FTZs across China. Table 1 is the timeline of China's FTZ establishment; we can see that after one and a half years' practice, the Shanghai model was utilized by an increasing number of new pilot zones, and now the twelfth FTZ is under construction in Hainan Province. As many Chinese media platforms have concluded, the Shanghai experience is now treated as a great lesson for China's economic reforms. Both the Shanghai Pilot FTZ and the rules, regulations, and institutions it has trialed are expected to be the reference for newly founded pilot FTZs in the future.Of all the features adopted by Shanghai Pilot FTZ, one of the most significant and consistent is the implementation and loosening of the "Negative List" for foreign investors. Traditionally in China, foreign capital's access to the domestic market was limited by the "positive lists" which clarified what sectors were open to non-Chinese market participants. The areas outside these positive lists were wholly closed to foreign investors. On the contrary, the negative list clearly lists all the fields, based on local administrators' executive requirements, that are forbidden or restricted to foreign participants (Hu, 2014). All parties will have equal entry permissions for all the fields outside the negative list (Xinhua, 2015), provided that the participants have filed beforehand.

In the propaganda of official media, the negative lists are said to promote the economic performance of the FTZs because, theoretically, introducing negative lists highly enlarges the areas in which foreign investment can participate (The State Council, 2013), which is exactly the original aim of setting up FTZs. Furthermore, it may encourage entrepreneurs to enter newly developed fields, which are all forbidden under the positive lists, and may therefore promote innovation. Under such logic, the Chinese government has been pushing the elimination of restriction terms in the Negative

List[1], to maximize the effect of the list in attracting FDI inflows and increasing the volume of imports and exports. The reforms and dates can be found in Table 2, and we can find that the number of forbidden fields has been reduced from 190 to only 48, which is a significant decrease in restrictions.

Nevertheless, whether loosening the Negative List has such a positive effect in China is an open question. Some researchers have focused on the list's impact on nation-wide economic performance. For instance, Magiera (2011) and Setiawan (2018) discuss the implementation of negative lists in Indonesia and ASEAN countries respectively, but both of them are studies of the Southeastern Asian region and involve little about the FTZs. Wang (2016) may be the first one to link the negative lists with the economic development of FTZs. He comprehensively analyzes the content, advantages, and the insufficiency of China's negative list policy for increasing FDI and trade. However, he doesn't provide any empirical evidence to prove his arguments. Although some reports and studies in China talk about the effect of loosening the Negative Effect on FTZ's performance (Hu, 2017), they use single cases in which one or two companies chose to enter the Shanghai Pilot FTZ, where the omitted variable bias is easy to observe.

As a result, this essay will empirically investigate whether reducing restriction conditions in the Negative List has an impact on FTZ's economic development, and manage to provide practical policy suggestions for the reforms of the Negative List and the FTZs in China. We focus on the policy changes in Shanghai Pilot FTZ because it is the oldest FTZ in mainland China, with a lifespan of five years. And from Table 2, we can find that the first version of the Negative List was used just for the Shanghai Pilot FTZ, which indicates that this city has experienced all adjustments of the Negative List. Thus, the Pilot FTZ in Shanghai is the ideal subject for our estimation of the effect of loosening the Negative List.

The second part of this essay will discuss data collection and the methodology utilized. The third part will present the main empirical results, and the last part will both draw conclusions and offer suggestions to improve China's Negative List policy and the development of the Shanghai Pilot FTZ.

---

1 To be concise, when I use the "Negative List" later in this essay, I refer to the five versions of the "Negative List for Foreign Investment in China" since 2013.

Table 2: Timeline for the Development of Shanghai Pilot FTZ and the Negative List		
Date	Event	Note
September 29, 2013	The Shanghai Pilot FTZ was formally established; The 2013 version of Negative List was enforced.	The 2013 Version was used just for Shanghai Pilot FTZ, with 190 field restriction terms included.
June 30, 2014	The 2014 version of Negative List was enforced.	The 2014 Version was used just for Shanghai Pilot FTZ, restricting 139 fields.
December 28, 2014	The total area of Shanghai Pilot FTZ was extended to 120.72 km².	120 km² has become the standard area of FTZs in China, reflecting Shanghai Pilot FTZ's role as a benchmarking.
May 8, 2015	The 2015 version of Negative List was enforced.	The 2015 Version was used for four Pilot FTZs, including 122 restriction terms.
July 10, 2017	The 2017 version of Negative List was enforced.	The 2017 Version was used for eleven Pilot FTZs and restricted on 95 fields.
July 28, 2018	The 2018 version of Negative List was enforced.	The 2018 Version was used for twelve Pilot FTZs, including only 48 restricted fields.

DATA AND METHODOLOGY

The independent variable in this study is whether loosening of the Negative List takes place or not. Our study will utilize Chinese government statistics from 2013 to 2018 to conduct a time series analysis of changes to the negative list. This research design is based on the fact that the availability of information related to FTZs in China is highly limited. First, most Pilot FTZs are less than two years old, so the amount of data is far from adequate. Second, most statistics on Pilot FTZs are not revealed to the public and can only be found occasionally in newspapers and TV programs. What's worse, media reports are generally not neutral; the government's power will force a consistently positive tone on reports about policy changes in FTZs, which may limit negative information from being released and bias our estimations on the effects of loosening the Negative List.

Therefore, it's difficult for researchers to evaluate FTZ policy effects by methods including regressions and causal inference. However, we find that the Shanghai Bureau of Statistics publishes local economic information every month, and some of them (including the international trade amount) come from the Shanghai Customs. Although these statistics do not directly reflect what the Pilot FTZ achieves in the same period,

they are sufficient variables to use in our study given the data availability limits on the topic. First, these economic statistics tend to vary as a result of policy changes. Second, they are evaluating the spill-over effect of Shanghai FTZ on the city-level development, which is a good proxy of the Pilot FTZ's economic level. As a result, they serve as the most accurate available values to measure the FTZ's performance.

Thus, for the dependent variables, we collect two types of statistics from the Bureau of Statistics. One is the sum of the import and export volume of the Shanghai Customs area, and the other is the FDI amount promised by contracts. They are significant indicators of economic development, both vary by time, and both are directly influenced by policy changes in the Shanghai Pilot FTZ.

To examine the relationship between the independent and dependent variables, we first assume that the reduction of restriction terms in the Negative List happens occasionally. Therefore, we can treat whether the list is loosened or not as a quasi-random experiment and emphasize its impact on both the short-term changes and the long-term trends in foreign trade and FDI attraction. By "short-term," we mean a change in trade and FDI observed by the sections not explained by monthly or quarterly data just after the implementation of a newly modified Negative List. And by "long-term," we try to examine whether these versions of the Negative List influenced the global trends of FDI

inflows and total foreign trade.

Admittedly, officials administering the FTZ program could decide to loosen the list based on its economic performance, which would be a form of reverse causality. For instance, some reports argue that the revision of the Negative List is a response to the disappointed investors who expect more economic reforms of the Chinese market (Margulies, 2014). However, from Table 2, we don't observe any consistent rule of how and when these eliminations occur in reality. In addition, in Figure 1 and Figure 2, we find no such pattern that the release of new Negative Lists is respondent to the prominent changes in economic development. Thus, we may assume that our argument is not likely to be threatened by endogeneity.

To estimate both short-term and long-term effects of the List, our strategy is to use a time series analysis. Though this method is generally used for prediction of dependent variables, this is not our focus because the policy changes are common and difficult to anticipate in China. On the other hand, these measurements are profoundly affected by global and domestic economic growth, which adds to the bias of predictions. Therefore, what we focus on is just the direct impact of loosening the Negative List on the changes of economic indicators in Shanghai.

Figure 1: Line Chart of Monthly FDI Inflows in Shanghai, 2013-2018

Figure 2: Line Chart of Monthly Foreign Trade in Shanghai, 2013-2018

We first observe the original data line graphs to see if time series models are applicable. Figure 1 and Figure 2 plot the general trend of monthly FDI inflows and foreign trade (both are logarithmically transformed) in Shanghai, and the dashed lines reflect the dates when Negative List restrictions were reduced. From these two graphs, we see that the dependent variables have a global positive trend, and the seasonal factors are also likely to exist, both of which indicate that our model selection is acceptable.[2] Therefore, it's reasonable to use time series models to combine both global trends and seasonal factors. The time series model for the logarithmically transformed variables can be expressed as:

Log (Dependent Indicator) = Trend + Seasonal + Remainders

Since loosening of the Negative List is rare across the 60 months, we assume that it has little impact on both the seasonal factors and the general trend. As a result, we can expect that the effect of the Negative List will mainly be found in the remainder term, and by seeing if the remainders have significant improvement after reducing restrictions on the Negative List, we can infer whether such policy has the anticipated positive outcome. Furthermore, we can judge whether our independent variables have an impact on overall trends by checking if the global trend terms have some inflection points around the times of policy changes related to the Negative List.

To empirically construct the time series models, we use both monthly and quarterly reported numbers of the two dependent variables. We expect the monthly data to reveal the immediate response of global investors on the policy changes, but we also find it possible that there is a time gap between policy enforcement and market reflection. So using both monthly and quarterly statistics will enable us to evaluate the investors' response comprehensively.

Our dataset begins in October 2013 and ends in September 2018, which covers all versions of the Negative Lists except for the first one. We don't extend our database for more historical statistics, because the period we study covers the whole life of Shanghai Pilot FTZ, and thus the time series trends will not be biased by whether FTZ exists or not. Also, the first version of the Negative List overlaps with the foundation of Shanghai Pilot FTZ, which of course impacts the analysis and

2 Figure 5 and Figure 6 in the Appendix plots the line graph of quarterly statistics for the FDI inflows and foreign trade, and they reveal similar patterns with Figure 1 and Figure 2.

should not be contained.

In sum, we have the occurrence of adjustments in the Negative List as the independent variable, and the economic indicators reported by Shanghai Bureau of Statistics (the FDI and the total foreign trade) as dependent variables. We use time series analysis to decompose the economic indicators' changes and examine how the loosening of the Negative List affect the remainders (short-term effect) and the global trend (long-term impact) of the economic performance.

RESULTS

From the dataset, we get Figure 3 and Figure 4, the decomposed time-series figures of monthly FDI inflows and the foreign trade volume. Both pictures consist of four diagrams: the original data, seasonal factors, global trends, and remainder terms, and we also draw four dashed lines to indicate the point when the reduction of restrictions in the Negative List took place. The trends and the remainders are what we focus on, since global trends reflect the overall development speed, while the remainders are indicators of immediate responses to the economic policies. Therefore, if loosening the Negative List has some impact on economic performance, we would expect an increase of either the short-term remainders or the long-term growth trends. Thus, we care about how the trends and remainders vary after the time points in which changes in the Negative List were implemented.

First, there is no overlap between the inflection points of the global trends and the time points when the Negative List is loosened in both Figure 3 and 4. This situation proves that reducing the number of fields restricted in the Negative List will not alter the general trends of economic development in the Shanghai Pilot FTZ, undermining the idea that reducing restrictions in the Negative List will influence long-term economic

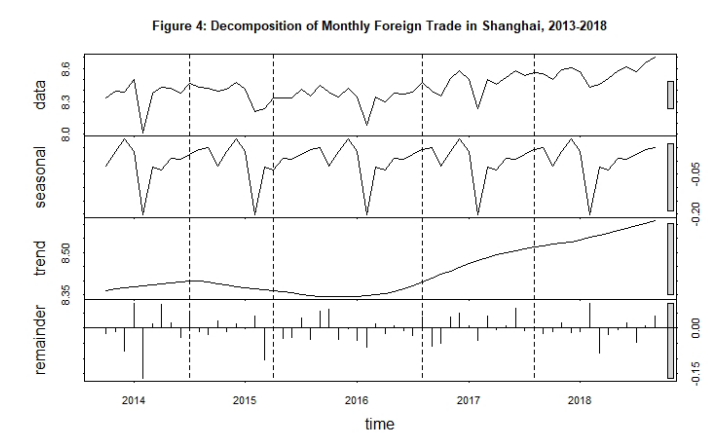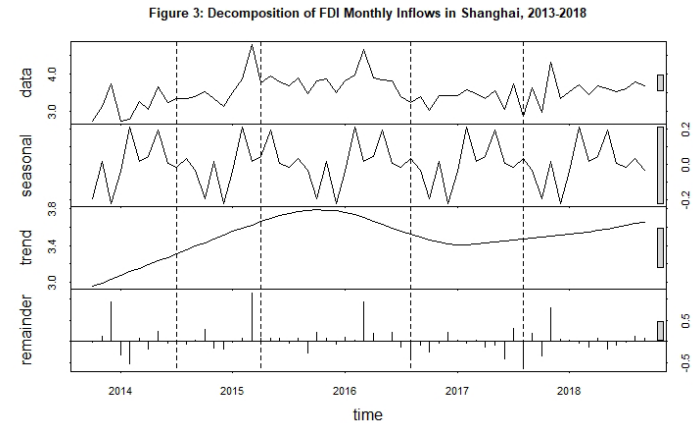

Figure 4: Decomposition of Monthly Foreign Trade in Shanghai, 2013-2018

growth.

Second, if we concentrate on the values of the remainders around the reform points, we cannot find significant patterns, implying that the implementation of the new version Negative Lists (where foreign investors are allowed to enter more industries) may not lead to a growth of remainders, which indicates that we may not observe a short-term growth in economic performance after the reducing the restrictions in the Negative List. On the contrary, we can observe some severe drops of remainder terms after a reduction of the Negative List is enforced, which indicates an inconsistent impact of the policy we study.

To better evaluate this inconsistency, we calculate the values of the remainders and present them in Table 3 and Table 4. From these two charts, we find that the immediate responses of market subjects are difficult to anticipate: in half of the cases, the FDI inflows decrease at the month of policy implementation, and the total trade volume even drops at every point in which the Negative List is revised. Furthermore, for the next two months after policy changes are made, both positive and negative remainders are recorded in our tables. This instability of coefficients means that there is not a consistent positive impact of the policy changes on the short-term economic performance of the FTZ. Therefore, we tend to reject the assumption that loosening

Figure 3: Decomposition of FDI Monthly Inflows in Shanghai, 2013-2018

Table 3: Time Series Remainders for Monthly FDI Inflows Model

Month	Before (t-1)	That Month (t)	Following (t+1)	Following (t+2)
June 2014	0.245530019	-0.03309465	0.039099834	-0.051596862
April 2015	0.066505737	0.074700149	0.064996052	-0.047400262
July 2017	-0.40022195	0.29717424	-0.647001796	0.186926092
July 2018	-0.084635228	-0.006456348	0.12846371	0.066522863

Table 4: Time Series Remainders for Monthly Trade Volume Model

Month	Before (t-1)	That Month (t)	Following (t+1)	Following (t+2)
June 2014	0.0166130845	-0.0295976907	0.0402620453	-0.0103490534
April 2015	0.0029815595	-0.0340476605	-0.0311052533	0.0321247657
July 2017	0.0665391166	-0.0068564648	0.0003279459	-0.0175215492
July 2018	0.0183058309	-0.0460374034	0.0078111351	0.0410529675

the Negative List may promote a short-term increase of FDI and foreign trade.

In addition, we have also drawn the decomposed time-series plots of quarterly FDI inflows and foreign trade with the data we collect, and the results are presented in Figure 7 and Figure 8 in the Appendix. They have revealed similar phenomenon that, both the global trends and the remainder changes are not significantly promoted by the loosening of the Negative List. Table 5 and 6 in the Appendix also proves that no consistent result can be observed regarding the Negative List's effect on short-term economic development. Thus, even if we are calculating the quarterly economic statistics, there is still neither immediate development of FDI attraction and foreign trade nor increase in the overall economic growth after the reduction of restrictions in the Negative List.

CONCLUSIONS AND POLICY SUGGESTIONS

From the results above, first of all, the short-term effects of loosening the Negative List on promoting Shanghai Pilot FTZ's economic performance is in doubt. We examine both monthly and quarterly datasets but find no consistent patterns of remainder changes.

Second, no evidence can be found that reducing restrictions terms in the Negative List will stimulate long-term growth in FDI attraction and total foreign trade volume. The decomposed time series figures tell us that the turning points of global trends fail to overlap with the time points when the elimination of the Negative List is enforced.

The reasons for the lack of incentives are not the focus of this essay. One possible explanation is that the existence of the Negative List itself reflects a limit of the degree to which FTZs can open up to economic globalization. Even if the government is making concessions regarding the content of the List, the List still discourages some investors from entering Shanghai.

 Therefore, loosening the Negative List may have limited positive impact on economic indicators, but the reliability and consistency of such an effect are not proved by our study. In this case, our recommendation

Third, institutional reforms are even more fundamental to FTZ's long-term growth. What makes FTZs, including Hong Kong and Singapore, suc-

is that administrators in both the central government and the Pilot FTZs should not expect stimulation of foreign trade or FDI when reducing restriction terms in the Negative List. Recent news reports indicate that the Shanghai Pilot FTZ has just published a negative list on the service trade, the first one across China. Whether this new negative list will successfully promote foreign investment and business, however, is doubtful, according to our analysis above.

Drawing on experience from FTZs around the world, we would raise three possible approaches through which the Shanghai Pilot FTZ can experiment with other reforms in the future to stimulate FDI. First, tax relief is one of the most common ways to attract foreign investors. Hong Kong, one of the most successful free trade ports around the world, is famous for its low tax rate. The total tax revenue represents 22.9% of Hong Kong enterprises' profits, which is lower than not only the corresponding percentage of the economies in East Asia and the Pacific (34.5%) but also the average 41.3% proportion in the OECD countries (World Bank, 2013). Although it's unlikely that the Chinese government will reduce the tax rate nationally, carrying out such reforms in the Pilot FTZs is possible, and may provide further information on the proper proportions to be used to attract foreign investment and commerce.

Second, financial policies should be modified to increase the convenience of market participants. One of the most significant dimensions is to promote a higher level of liberalization in the exchange of currencies in FTZs. The Central Bank of China has strict control of foreign currencies to avoid the rapid inflow/outflow of capital, reducing the potential risk of financial crisis. This, however, harms the flexibility of international trade. Therefore, with the reality that China will not abandon its currency control, convenience should be created in FTZs: branches of banks from different nations should have some access to the FTZs. Thus, companies can open multiple currency accounts and use a different currency for trade settlement. The Shanghai Pilot FTZ can take this opening up as a short-term financial policy trial, and adjust the details later on according to the effect of a greater currency liberalization. (Ma and McCauley, 2002)

cessful is not just technical issues but also a good environment for international business. It may be difficult for mainland FTZs like Shanghai to thoroughly imitate the political institutions and legal systems of Hong Kong, but there exists a good domestic example, Shenzhen, to follow.

Part of Shenzhen is included in the Guangdong Pilot FTZ, and the city is working hard to promote FDI and foreign trade. Considering Hong Kong has achieved a rigorous rule of law, and that many international investors are initially located in regions accustomed to the common law system, Shenzhen has adopted two ways to improve the quality of its institutions. First, Shenzhen has established an International Committee of Arbitration. To protect the neutrality of the Court, Shenzhen introduced a corporate governance mechanism, with an international council serving as the center, and now one-third of all the council members are from outside mainland China (Zhang, 2016). The arbitration of this Committee has been widely accepted by courts in Hong Kong, indicating Shenzhen's ability to help resolve disputes in international commerce via institutional approaches. Although other FTZs are not close to Hong Kong, they can still learn from Shenzhen's practice and try to establish arbitration institutes that are both neutral and globalized.

Shenzhen's second improvement in institutional quality is that the municipal government allows law offices in Hong Kong and other countries, jointly with local lawyers, to establish representative offices in the FTZ. Consequently, these joint offices are able to provide cross-border legal services to both Chinese and foreign clients, which reduces the investors' expectation of lawsuit risks and draws their attention from Hong Kong to Shenzhen. Therefore, the trial for joint venture law offices has been proven possible in China, and Pilot FTZs, including Shanghai, should take this chance to enhance its nomocracy performance, appealing to foreign investments in the long run.

BIBLIOGRAPHY

Akinci, G., Crittle, J., Akinci, G., & Crittle, J. (2008). Special Economic Zone: Performance, Lessons Learned, and Implication for Zone Development (English). Foreign Investment Advisory Service (FIAS) Occasional Paper. Washington, DC: World Bank.

Farole, T., & Akinci, G. (2011). Special Economic Zones: Progress, Emerging Challenges, and Future Directions. The World Bank.

Gopalan, S., Malik, A. A., & Reinert, K. A. (2013). The Imperfect Substitutes Model in South Asia: Pakistan–India Trade Liberalization in the Negative List. South Asia Economic Journal, 14(2), 211-230.

Hu, J. (2014). A Preliminary Study of Pre-Access National Treatment: Also on the Negative List of Shanghai Free Trade Zone [Translated from Chinese]. Journal of Shanghai Jiaotong University, 48(1), 65-73.

Hu, J. (2017). A Retrospective View on the First Three Years of China (Shanghai) Pilot Free Trade Zone. The Chinese Economy, 50(4), 225-237.

Jiang, L. (2013). Free Trade Zone 'Negative List' Maintains Most Restrictions. Retrieved from www.uschina.org/free-trade-zone-%E2%80%98negative-list%E2%80%99-maintains-most-restrictions

Kim, S. (2009). China and Globalization: Confronting Myriad Challenges and Opportunities. Asian Perspective, 33(3), 41-80.

Margulies, M. (2014). Changes to Shanghai FTZ Negative List Fall Short of Expectations. Retrieved from https://www.uschina.org/changes-shanghai-ftz-negative-list-fall-short-expectations

Krugman, P. (1991). The Move Toward Free Trade Zones. Economic Review, 76(6), 5-25.

Ma, G., & McCauley, R. (2002). Rising foreign currency liquidity of banks in China. BIS quarterly review, September 2002, 67-74.

Magiera, S. (2011). Indonesia's Investment Negative List: an evaluation for selected services sectors. Bulletin of Indonesian Economic Studies, 47(2), 195-219.

Setiawan, S. (2018). Negative List in Services Liberal-

ization for ASEAN Developing Countries. International Journal of Economics and Financial Issues, 8(5), 11-20.

The State Council. (2013). Special Management Measures for Foreign Investment Access in Shanghai Pilot Free Trade Zone (The Negative List, 2013 version) [Translated from Chinese]. Retrieved from http://www.shanghai.gov.cn/nw2/nw2314/nw2319/nw10800/nw11407/nw30984/u26aw37036.html

The State Council. (2014). Special Management Measures for Foreign Investment Access in Shanghai Pilot Free Trade Zone (The Negative List, 2014 version) [Translated from Chinese]. Retrieved from http://www.shanghai.gov.cn/nw2/nw2314/nw2319/nw12344/u26aw39501.html

The State Council. (2015). Special Management Measures for Foreign Investment Access in Pilot Free Trade Zone (The Negative List, 2015 version) [Translated from Chinese]. Retrieved from http://www.gov.cn/zhengce/content/2015-04/20/content_9627.htm

The State Council. (2017). Special Management Measures for Foreign Investment Access in Pilot Free Trade Zones (The Negative List, 2017 version) [Translated from Chinese]. Retrieved from http://www.gov.cn/zhengce/content/2017-06/16/content_5202973.htm

The State Council. (2018). Special Management Measures for Foreign Investment Access in Pilot Free Trade Zones (The Negative List, 2018 version) [Translated from Chinese]. Retrieved from http://www.gov.cn/zhengce/2018-06/30/5302510/files/5616043fde4d432fa8ab542d12db66c4.pdf

Wang, Z. (2016). Negative List in the SHPFTZ and Its Implications for China's future FDI Legal System. Journal of World Trade, 50(1), 117.

World Bank. (2013). Doing Business 2014: Understanding Regulations for Small and Medium-Size Enterprises. Washington, DC: World Bank Group. DOI: 10.1596/978-0-8213-9984-2. License: Creative Commons Attribution CC BY 3.0.

Xinhua: China Introduces Market Access Negative List System. (2015). Chinascope, 77, 39.

Xinhua News Agency. (2018, November 18). Negative List for Five Years: Shanghai Free Trade Zone Attracts Over 100 Billion dollars in Contractual Foreign Investment [Translated from Chinese]. Retrieved from www.gov.cn/xinwen/2018-11/19/content_5341774.htm

Zhang, W., Dai, X. (2016, December 23). Why does Qianhai Attract Foreign Capitals? [Translated from Chinese] Southern Daily, p. AT14.

APPENDIX

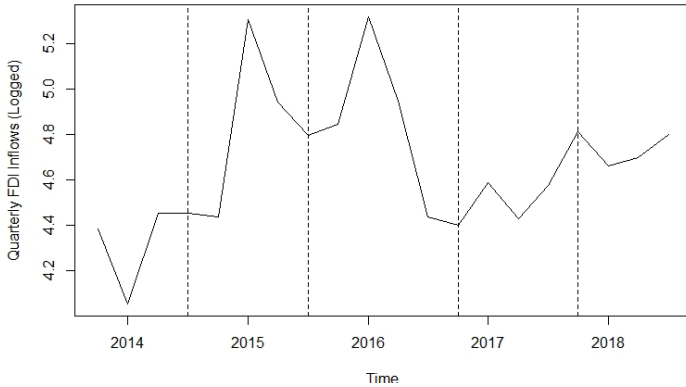**Figure 5: Line Chart of Quarterly FDI Inflows in Shanghai, 2013-2018**

Figure 7: Decomposition of Quarterly FDI Inflows in Shanghai, 2013-2018

Figure 6: Line Chart of Quarterly Foreign Trade in Shanghai, 2013-2018

Figure 8: Decomposition of Quarterly Foreign Trade in Shanghai, 2013-2018

Table 5: Time Series Remainders for Quarterly FDI Inflows Model

Quarter	Before (t-1)	That Quarter (t)	Following (t+1)	Following (t+2)
2014 Q2	-0.681950486	-0.043718291	0.009121203	-0.178620317
2015 Q2	0.230371041	0.043886093	-0.048904206	-0.031146593
2017 Q3	-0.079438266	0.058193711	0.230496558	-0.317062865
2018 Q3	-0.052343653	0.091857911	N/A	N/A

Table 6: Time Series Remainders for Quarterly Trade Volume Model

Quarter	Before (t-1)	That Quarter (t)	Following (t+1)	Following (t+2)
2014 Q2	0.011555012	0.025458928	0.004723102	0.010790723
2015 Q2	-0.007936336	-0.022398923	0.015176421	0.0017066
2017 Q3	0.016894125	-0.009873313	-0.004806066	0.001523271
2018 Q3	-0.005731689	0.004662802	N/A	N/A

Breaking-up Service Homogeneity in Tourism: Certifications for Sustainable Tourism in Costa Rica

Mateo Villamizar Chaparro
Duke University

The effects of sustainable tourism have not been thoroughly researched and few agreements have been reached over its potential benefits for development and environmental efforts. This paper evaluates sustainable certificate policy for tourism operators in Costa Rica. The different studies discussed reveal plausible effects of voluntary sustainable certificates on firms and consumers' behavior from the Ecologic Blue Flag, Code of Conduct Certificate, and Certificate for Sustainable Tourism programs. The fact that the processes are voluntary can leave plenty of room to other actors to take advantage of the holes left by certifying firms. Effects of the programs include increased willingness to pay for green premiums.

INTRODUCTION

Tourism's prominence has risen in the last few years. 2017 was declared the International Year of Sustainable Tourism for Development by the United Nations and its World Tourism Organization (UNWTO). Some figures estimate that the tourism sector accounts for nearly 10% of the World's GDP (World Tourism Organization, 2017). In 2010, tourism services were provided to 940 million people and, according to user growth forecasts, the tourism sector is projected to serve 1.8 billion users by 2030 (Epler Wood, 2017). As a potential tool for poverty alleviation, unemployment reduction, and environmental and patrimonial protection, tourism has become a valuable avenue in achieving the UN Sustainable Development Goals (SDGs) for 2030.

In order to help combat climate change and environmental degradation, the tourism sector has pledged to lower its carbon footprint by decreasing the levels of CO2 emitted by the providers (UNTWO). Sustainable tourism has the potential for growth in upcoming years due to its promising economic benefits and advocacy efforts from the UN (Ayeni, 2013). The sector is also promoting incentives such as skill development programs, loans for tourism-related entrepreneurial activities, and sustainability certificates to encourage companies to adopt sustainable tourism practices.

Despite these optimistic economic forecasts and pro-sustainability intentions, the effects of sustainable tourism have not been thoroughly researched. Few agreements have been reached over the potential benefits sustainable tourism can provide for development and environmental efforts. Some studies have indicated that there is a heterogeneous effect of sustainable tourism on developing and developed countries with developed countries and luxury hotel brands enjoying the positive economic effects of eco-labeling schemes and voluntary sustainable certification programs (Rivera, 2002; Sasidharan, Sirakaya, & Kerstetter, 2002). Moreover, the definition of 'sustainable tourism' is subject to a myriad of interpretations, making it difficult to accurately measure.

This paper aims to evaluate and analyze sustainable certificate policy for tourism operators in Costa Rica. In order to achieve this goal, the first section will try to untangle the concept of sustainable tourism. The second section will discuss the selection of the Costa Rican case. The third section will respond to the enterprises' motivations to undertake the voluntary

Mateo Villamizar Chaparro is a PhD student in Political Science at Duke University. He completed masters degrees at UC San Diego and Universidad de los Andes. He can be reached at sv161@duke.edu

certification process. The fourth section will analyze the factors that drive the demand for certificates taking into account the literature of green premiums. The fifth section will present a succinct discussion of the Costa Rican policy.

THE CONCEPT OF SUSTAINABLE TOURISM

One of the biggest issues within sustainable development is the lack of an agreed upon definition of 'sustainable tourism'. Most authors and international organizations use some variation of the most accepted definitions in attempt to either increase or decrease the number of cases where the definition can apply. The concept consequently loses its connotative precision due to its ambiguity or its extreme narrowness. This results in a conceptual stretching that can leave the concept useless when put to the task of comparing or distinguishing between cases (Collier & Mahon, 1993; Sartori, 2009).

Part of the conceptual conundrum with respect to sustainable tourism comes from a historical lack of shared understanding of the concept of sustainability itself. Sustainable tourism became a prominent topic on the research agenda in the 1970s and shared the same preoccupation about the future of the world, given the economic and growth patterns at that time. However, despite having a clear mission regarding a sustainable future, the actual definition of what sustainability entails was and remains a contested one (Mebratu, 1998). According to Woodhouse (2002), the relationship between the individual and the environment can have three different approaches: utilitarian, romantic, and administrative. A utilitarian relationship represents human dominance over nature, a romantic relationship resembles nature's dominance over humans, and an administrative relationship relates to the reasonable use of nature by humans. In this fashion, the sustainable approach will be dependent on the theoretical consideration of the relationship between humans and nature.

In order to avoid multiple considerations of the same term, this paper will use the definition from the UNW-TO:

"Tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, the environment and host communities" (World Tourism Organization, 2017)

This particular definition seems to imply three different but simultaneous goals that sustainable tourism should achieve (Carter 1993 cited in Liu, 2003). First, sustainable tourism should help the host population meet its economic needs in a way where growth can be sustained for longer periods of time. Second, the activities involved with sustainable tourism must try to meet the demands of the tourists in order to maintain the flow of consumers and the growth within this sector. Finally, sustainable tourism should make sure it is safeguarding the natural environment and minimizing the harmful effects of tourism. For firms or governments that promote this type of tourism, these three goals imply that while making some decisions about tourism operations, it is necessary to make some trade-offs between.

It is important to note the potential weakness of the definition selected for this paper. Firstly, this definition lacks a distinct and unambiguous causal path between sustainable tourism and environmental protection. There might be multiple paths and potential intervening variables mediating the relationship. Some of the most often cited mechanisms where sustainable tourism positively impacts the environment revolve around the reduction of carbon and natural gas emissions; improving the carrying capacity of the environment; control over tourism; and restrictions to mobility.

A second critique is that the definition seems to imply that new tourism operations should be implemented in more environmentally conscious ways, but in turn lacks an explanation of how to alter existing practices or offer explicit recommendations that would make tourism operation more intentional about avoiding negative environmental impacts (Butler, 1999). This particular fact illustrates an idea where sustainable tourism is not a retrospective activity but rather something that will start today and continue for an indefinite period. Related to the previous deficiencies, it is important to mention that the definition is ambiguous about how sustainable tourism and a better environment are measured. This can lead to potentially internal validity concerns and will make the dialogue and possible comparison between different measurement approaches more challenging.

The final weakness is that the definition unravels sustainable tourism as something inherently good without questioning the possible repercussions on the hosting communities (Butler, 1999). There might exist some cases where a sustainable approach to tourism might

bring more harm than good to the environment. For example, Mejía & Brandt (2017) explain that one of the major challenges of tourism in the Galápagos Islands is the possibility that tourists bring alien seeds or bacteria to the protected areas which can have detrimental effects on the island's flora and fauna. On the other hand, Blackman, Naranjo, Robalino, Alpízar & Rivera (2014) explain that good results in certification processes might drive more businesses to the area, increasing the pressure on the environment.

COSTA RICAN CASE:

TOURISM AND VOLUNTARY CERTIFICATES

Costa Rica is a small Central American nation that stands out in the region due to its lack of a military and its progressive set of national environmental policies. In the past decades, the country has experienced a boom in its tourism sector which has made the industry a major engine of growth in the country (Costa Rican Government, 2017). Figure 1 plots the number of foreigner arrivals to Costa Rica since the 1950s and illustrates a clear upward trend (ICT, 2017). From 1950 to 1970, the growth in the number of arrivals was relatively stable, but after 1970 the upward slope becomes more pronounced. According to the 2017 statistics available at the Costa Rican Institute of Tourism (Instituto Costarricense de Turismo, ICT), the majority of tourists are from the United States and Europe and their primary motivation for visiting Costa Rica is vacation. Furthermore, according to a survey conducted by the ICT in 2016, the most popular activities for people staying in the country include "going to the beach" and "ecotourism" (ICT, 2017a).

This has affected the overall economy, as Costa Rica is growing as a result of the dynamism of its ecotourism sector. Figure 2 shows the amount of spending by non-Costa Rican tourists as a percentage of GDP between 2009 and 2015. Overall, tourism has accounted on average for 5.16% of the Costa Rican GDP in the period illustrated in the graph. Likewise, Figure 2 also exhibits the behavior of the total spending by tourists in millions of USDs. In this case, the slope is always positive which means that spending has increased. However, the period between 2009 and 2012 marks a decline in spending as a percentage of GDP, indicating that the country's growth occurred outside this sector. However, after 2012 both lines show a positive slope, displaying a stronger effect of tourism on the country's

growth.

The tourism industry in Costa Rica has been synonymous to the concept of sustainable development as part of the government's bet that this will produce greener and sustainable economic growth. The government has launched three different certification programs for operators of tourism services. The first is the Ecologic Blue Flag (BF) program, adopted in 1998 by permitting communities to certify the sustainability of tourism in their beaches (ICT, 2016a). This voluntary certification evaluates beaches according to the health of the water and the organisms living in it, the sanitation of the water for human enjoyment, and environmental education. Communities must create a series of committees that include individuals from the private sector, tourist operations, the local government, and community leaders. This program is widespread around the country, as can be seen in Figure 3, and it now include 9 additional categories. The evaluation system is well defined, and the certifications are issued annually.

The second program is the Code of Conduct Certificate (CC), which has the primary goal of protecting children and teenagers from risks associated with tourism, such as sexual exploitation and child labor (ICT, 2016b). This program is targeted at lodging centers, tour operators, car rental centers, and theme parks.

The third program is the Certificate for Sustainable Tourism (CST), and will be the central focus of this paper. It was implemented in the early 2000s. According to its webpage, the certifications serve as a mechanism to "make sustainability a practical and necessary reality within the context of the country's competitiveness in tourism, while looking to improve the way that natural and social resources are used, encourage the active

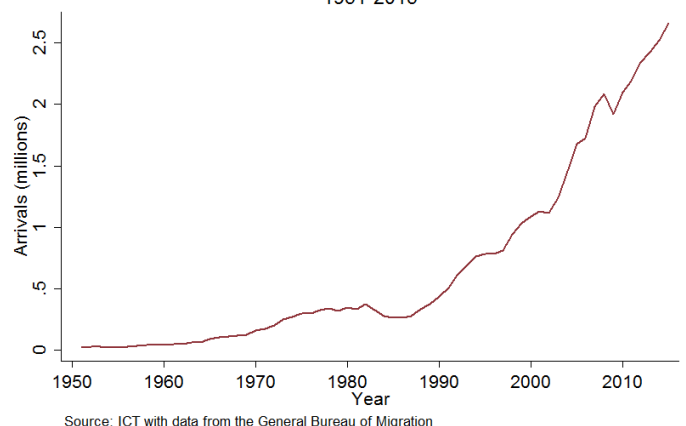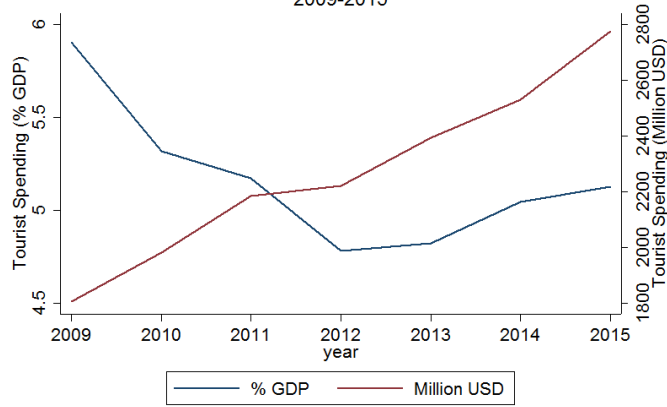Figure 1: Number of Foreigners' Arrivals to Costa Rica
1951-2015

Source: ICT with data from the General Bureau of Migration

Figure 2: Tourist Spending in Costa Rica 2009-2015

Calculations from the authors using ICT (2017) and World Bank data

Figure 3: Map of BF committees in Costa Rica

Source: Instituto Costarricense de Turismo -ICT- (2017a)

participation of local communities and provide a new source of competitiveness within the business sector." (ICT, 2017b). By voluntarily applying for the certificate, the company or tour operator can gain benefits such as the elimination of fees for state sponsored fairs, the promotion of their business with the use of the verified label to attract more consumers, and the chance to be listed on the country's official tourism website. Benefits also increase depending on the rating received by the company. More highly rated firms receive access to additional benefits including access to international and national publicity, training programs for their workers, access to multiple media outlets, and the option of using the country's own tourism brand (Costa Rican Government, 2016).

The CST is given to businesses that comply with a set of objectives related to the physical and biological well-being of the environment; acceptable internal processes linked to waste management; the use of technologies to conserve water and electricity; and the maintenance of good relations with clients and the socioeconomic environment where they work. With these objectives, the Costa Rican National Accreditation Commission conducts a survey to determine the sustainability level of the company, with the rating ranging from 1 (least sustainable) to 5 (most sustainable). The target group for this certificate is the companies that provide tourist lodging, but the certificate was extended to a broader audience by including other non-lodging tourism companies like gastronomical enterprises, tour operators and car rentals. It is important to highlight that some government officials think that the CST is not tied to the community but instead has favored other groups in society (Silva, 2003). This therefore makes it incompatible with our definition of sustainable development, leaving the private enterpris-

es as the full coordinators of the ecofriendly schemes.

The CST program has been evaluated before in a previous series of studies, but the main types of firms being analyzed were hotels (Blackman et al., 2014; Rivera, 2002; Silva, 2003). This is presumably the result of a greater number of lodging firms that apply for the program. However, the impact on other types of tourism service providers is unexplored. Figure 4 shows the total number of accredited firms by subsector in 2016, which totaled 361. The graph displays that the categories with less certificates are Theme Parks (16), Marine and Coastal Agencies (13), Restaurants (6) and Car rentals (9), which might indicate that it is harder for these firms to obtain the certificates, or that they are not interested in doing so. On the other hand, it illustrates more variability between the 236 lodging companies and the 81 tour agencies that are certified. It is important to highlight that the most common category level was 4 out of 5, and that most of the certified agencies were concentrated on the high score spectrum of the CST.

Due to the fact that both tourism and sustainability are key components of the Costa Rican economy, the CST in Costa Rica is a good case study to demonstrate how voluntary sustainable certificates work, specifically, looking at service providers outside lodging in the tourism industry. The next section will explain some of the past studies on green certificates in other contexts.

THE SUPPLY SIDE: WHY ARE VOLUNTARY CERTIFICATES A GOOD IDEA?

As mentioned in the previous section, the CST is a voluntary sustainability certificate (VSC) that firms acquire by satisfying certain conditions established by the certificate and are audited by an independent agency. Voluntary sustainability certificates exist because environmental quality is often something hard to observe when consuming a good. Thus, there is informational asymmetry between consumers and producers about how green a product is, or its supply chain (Podhorsky, 2016). These policy mechanisms occur in contexts where effective regulation is lacking. As a result, they become substitutes of said policies. Good green certificates work when they are context specific to the location, they managed to create an informational agreement between different relevant actors, and audit and assessment procedures are clear and transparent (Buckley, 2002).

Typically, ecolabels or VSCs are the result of an analysis of how tourism operators affect the environment in the place where their services are provided. After this environmental impact evaluation, some assessment criterion are developed which leads to an award if the criterion are met. The process then repeats during yearly assessments. However, this process might be defective as it gives the firms with more resources – namely time and money -- more voice in defining the assessment criteria in a way that is favorable for them (Sasidharan et al., 2002). The definition of the political economy of green certificates should be a topic that receives more focus, especially in qualitative analysis of this subject.
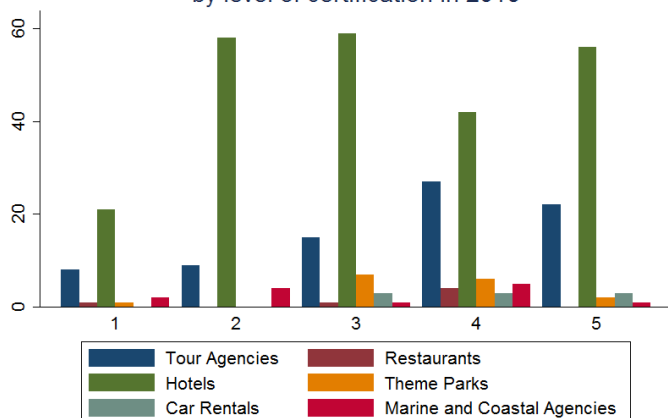
One of the incentives for a firm to obtain a VSC is to send a differential signal to the consumer. This way it is not only able to communicate its concerns for the environment but can also differentiate itself among firms in order to become more competitive. This way, the firms are able to achieve a product heterogeneity that would allow them to charge higher or at least differential prices to the consumer. As a result, shares can increase in existing or new markets (Teisl, Roe, & Levy, 1999) in what Blanco, Rey-Maquieira, & Lozano (2009) define as a demand driven effect. This is important because VSC's are not cost reduction strategies, but rather mechanisms that affect the demand curve of the provided service, which in this case is tourism. Other incentives tourism firms might have, such as lodgings with green certificates, are able to retain highly motivated employees and cultivate guest loyalty (Lee, Hsu, Han, & Kim, 2010).

This signaling by tourism firms that obtain VSCs might not only be targeted to consumers, but also to potential investors. Robinson, Kleffner, and Bertels (2011) find a correlation between the price share of North American companies and the inclusion in the Dow Jones Sustainability Index. Similarly, Foster and Gutierrez (2013) explain that in Mexico, VSCs can give capital holders information about the firm's costs of compliance, which can lead to an increase in future investments. Hence, the effect of the certification processes might not only be through a demand driven mechanism but also through an increase in assets or equity. The examples used do not necessarily correspond to the tourism sector, but it is possible to create an analogy where an increase in demand for a particular service might increase its investment.

As hinted in previous sections, VSCs might not always produce the desired consequences of diminishing environmental damage. For example, if the adoption of a green label increases the consumption of that tourist product, then there will be more demand for it and increased pressure to provide it. If a certified marine and coastal tour guide has perceived a shock in the demand function, she will have an incentive to increase her operations by purchasing a second or third gasoline boat that will enable the tour guide to earn more profit, but in turn could have negative effects on the environment. However, if that tour guide is not the only one perceiving the change, the likelihood of an increase in the number of boats will escalate. Consequently, the damage to the environment could have a multiplicative effect.

Figure 4: Number of CST Firms by level of certification in 2016

Tour Agencies
Hotels
Car Rentals
Restaurants
Theme Parks
Marine and Coastal Agencies

An additional problem with VSCs is that they typically focus on only one environmental issue. In the Costa Rican case, there are three different certificates for sustainability in tourism, each having a different emphasis and diverse requirements. This can potentially create disincentives for firms, as they might choose the least costly certificate instead of the one with the most rigorous assessment. Conversely, the high number of VSCs might generate uncertainty for consumers as they are unable to distinguish between labels (Harbaugh, Maxwell, & Roussillon, 2011). In this case, there will be no demand effect because the consumers are unable to read the signal and the firm incurs greater costs to comply with the VCF. Hence, the firm will be in a worse situation than it was initially without the certificate Additionally, the consumer will not be able to maximize its utility function if part of its preferred consumption bundle comes from sustainable tourism.

Rivera (2002) was one of the first scholars to analyze the CST in Costa Rica. Using a panel estimation for hotels under the CST, he finds that participating in the certification process alone is not enough to yield higher prices or sales. The demand effect is only available for those hotels that hold high environmental standards. However, when analyzing the BF program in Costa Rica, Blackman, Naranjo, Robalino, Alpízar, & Rivera (2014) provide evidence that the program led to higher investments in hotels, and particularly luxury accommodations. In the authors' words "[p]resumably, BF certification gives tourists a credible signal of overall environmental quality of beach communities and therefore increases demand for hotel rooms in certified beach communities. BF's" (p. 50). The interesting finding in this set of results is that the data set used by both studies is the same, or has only gone through minimal changes, but the questions the authors ask and the VSC being analyzed are different. However, given the unavailability of public data or other studies on the subject, this raises questions about the quality of the dataset and their results. Independent of this, both articles use a more statistical technique that enables comparisons across time, which is a step forward from the mostly qualitative approaches to tourism analysis in developing nations.

THE DEMAND SIDE: WILLINGNESS TO PAY AND GREEN PREMIUMS

The coordination for the provision of an environmental good is influenced by the type of alignment present between the public and the private interests. According to Jack, Kousky and Sims (2008), when there is an alignment, optimal provision of the good/service might be easily reached. However, the opposite does not occur. When the interests of the public and the private agents differ, the provision will be suboptimal. This is usually the case with sustainable tourism. That is why the VSC tries to make private entities internalize the negative externality and promote safer conditions for the environment. However, these incentives can be altered by revealing consumers' willingness to pay (WTP) for green premiums. If this is the case, we can pinpoint the demand driven effect explained in the previous section.

Today, individuals realize that their purchasing decisions have direct influence over the environment. This has encouraged people to be willing to pay extra money for products that will be more ecofriendly and less harmful to the environment, giving rise to a socially conscious consumer (Laroche, Bergeron, & Barbaro-Forleo, 2001; Lee et al., 2010). However, it is important to ensure that the signal is well received by potential consumers in order to increase sales. If the signal is received incorrectly or is hard to read, then informational problems will arise between producers and consumers, leading to a suboptimal provision of touristic goods even in the presence of socially conscious agents.

According to some studies, the WTP for environmental premiums is related to two types of values that influence behavior. The first one is collectivism, which "implies cooperation, helpfulness, and consideration of the goals of the group relative to the individual. Being a collectivist means that one may forego individual motivations for that which is good for the group." (Laroche et al., 2001, p. 506). In this regard, individuals who are more collectivist rather than individualist will have, theoretically, higher WTP for green premiums. The second value is that of materialism. "Materialistic tourists therefore are less concerned about overconsuming scarce resources as long as doing so provides other benefits; they evaluate public goods according to personal benefits rather than the value inherent in their existence" (Hultman, Kazeminia, & Ghasemi, 2015, p. 1856). Consumers that are more materialistic will be less willing to pay for the green premiums. On this matter, Meleddu and Pulina (2016) note a study in Italy where less materialistic individuals tend to have bigger WTP for sustainable tourism premiums.

Another factor that might have an influence on the WTP for green services is environmental literacy. This term refers "to knowledge on environmental issues also the ability to synthesize holistically by the personal learning process" (Ramdas & Mohamed, 2014, p. 379). The effects of this particular factor are mixed. Ramdas and Mohamed (2014) find that environmental literacy is essential to understand the WTP for green tourism in small island nations, while Laroche et al. (2001), after conducting a survey in the UK, failed to find statistical differences in environmental literacy between the groups they analyze. More research should be conducted in order to discern not only the meaning of environmental literacy but also the proper way to measure it when realizing WTP techniques.

In academic fields besides tourism, research about the valuation of environmental services seem to find evidence of the existence of green premiums. A study about green electricity conducted in the US, Roe et al. (2001) found that individuals are willing to pay a premium on cleaner energy sources that decrease carbon emissions. Moreover, the work of Greenstone & Jack (2015) explains how willingness to pay can differ in developed and developing countries through the effects in consumption. This particular effect is important to highlight, as most of the contingent valuations of environmental services have been analyzed in developed countries and are assumed to behave the same way in developing ones.

On the references consulted, there is an apparent trend in conducting more qualitative case studies rather than quantitative, specifically large-n analysis. This tends to be the case due to the difficulty of obtaining data. Some tourism providers act in small areas and occasionally in the informal sector, which makes gathering information more difficult. Despite this, some of the research has used representative surveys in developed countries, specifically in the UK, to analyze the attitudes and behaviors towards environmentally friendly touristic attractions or residences. The results of these surveys seem to indicate that families with children, married individuals, and gender were the only variables that generated significant differences in the evaluation of environmental attitudes (Borden, Coles, & Shaw, 2017; Laroche et al., 2001). People who presented these characteristics tend to be more prone to pay an environmental premium. Nevertheless, it prompts us to ask ourselves if these studies are externally valid and replicable, or applicable to other cases.

One of the biggest problems with WTP is that the actual effects are hard to measure, specifically when using a contingent valuation because the stated preferences might be different than the revealed one. There is a chance that under contingency valuation methods the real preferences of the individuals might not be captured, as there is not a commitment to the valuation. For example, in a study by Teisl et al. (1999) the results show that green certificates are more likely to affect the rankings of the products, rather than the choice and subsequent purchase. Accordingly, the calculation of green premiums needs to sometimes be taken with a grain of salt.

FINAL REMARKS

The different studies discussed above reveal the plausible effects of voluntary sustainable certificates on firms and consumers' behavior. Nonetheless, the impacts on the environment are yet to be seen or scrutinized. Most of the literature on green certificates and VSCs tries to understand the logic of accepting the accreditation processes but lack a proper analysis of the impacts of the certificates on the environment. More attention is needed on the results of the certificate assessment and auditing process in order to be able to correctly measure how certified firms might change their environmental priorities after going through the certification process. The fact that the processes are voluntary can leave plenty of room to other actors to take advantage of the holes left by certifying firms. This paper could not find the actual number of hotels or tourist providers in Costa Rica, but certainly Figure 4 is only a subset of the total number of tourism service providers.

As a way of conclusion, it is possible to think that the CST for non-lodging tourism operators might spur a demand driven effect with incentives for small businesses to expand and generate a greater desire to preserve the environment. This will be consequent to the existence of individuals who are willing to pay the green premium of these services. This potential outcome should be considered carefully by the Costa Rican government, as a substantial part of the country's GDP depends on sustainable tourism. It is important to mention that the multiple definitions of sustainable tourism that Costa Rica uses shows the ambiguity in this concept of sustainability. As a result, multiple definitions of the same concepts might generate various types of certificates that could confuse consumers rather than giving them relevant information (Podhorsky,

2016). Consequently, the demand driven effect of VSCs will not be present.

Additionally, is important to collect more data on the tourism sector in order to conduct better analysis. The studies regarding WTP need to go beyond surveys and studies on certification and should utilize more statistical techniques for empirical measures. However, because the certification process sometimes might be prone to corruption schemes, getting this information is costly or very difficult.

It is necessary to create incentives for conducting research on tourism outside of the developed countries. Some of the studies discussed in this paper focus on developing countries and emphasize the qualitative difference between such countries and their developed counterparts. In particular, Greenstone and Jack (2015) explain that "when people are very poor, what little money they have goes toward immediate consumption needs" (p. 12). Thus, the suboptimal provision of certain environmental goods might be socially efficient depending on the level of the country's development.

One final thought is that the theories reviewed seem to agree with the idea that sustainable tourism is inherently good, which suggests that paying a premium to consume services is by extension something good as well. The shortcoming in this argument is that the representative agent is conceived as the green consumer that can be found any place in the world. It is necessary to inquire whether this is the case or if there is heterogeneity between types of consumers, or even within types of green consumers.

BIBLIOGRAPHY

Ayeni, D. (2013). Predicting the Effects of Economic Diversification on Sustainable Tourism Development in Nigeria. *American Journal of Tourism and Management*, *2*(1), 15–21. https://doi.org/1 0.5923

Blackman, A., Naranjo, M. A., Robalino, J., Alpízar, F., & Rivera, J. (2014). Does Tourism Eco-Certification Pay? Costa Rica's Blue Flag Program. *World Development*, *58*, 41–52. https://doi.org/10.1016/j. worlddev.2013.12.002

Blanco, E., Rey-Maquieira, J., & Lozano, J. (2009). Economic incentives for tourism firms to undertake voluntary environmental management. *Tourism Management*, *30*(1), 112–122. https://doi. org/10.1016/j.tourman.2008.04.007

Borden, D. S., Coles, T., & Shaw, G. (2017). Social marketing, sustainable tourism, and small/medium size tourism enterprises: challenges and opportunities for changing guest behaviour. *Journal of Sustainable Tourism*, *25*(7), 903–920. https://doi.org/10.1080/0 9669582.2016.1270952

Buckley, R. (2002). Tourism ecolabels. *Annals of Tourism Research*, *29*(1), 183–208.

Butler, R. W. (1999). Sustainable tourism: A state-of-the-art review. *Tourism Geographies*, *1*(1), 7–25. https://doi.org/10.1080/14616689908721291

Collier, D., & Mahon, J. E. (1993). *Conceptual "Stretching" Revisited: Adapting Categories in Comparative Analysis* (SSRN Scholarly Paper No. ID 1747306). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1747306

Costa Rican Government. (2016, abr). Beneficios del CST. Retrieved June 16, 2017, from http://www.turismo-sostenible.co.cr/index.php?option=com_content&view=article&id=35&Itemid=129

Costa Rican Government. (2017). Costa Rica Tourism Official Website. Retrieved June 16, 2017, from http://www.visitcostarica.com/ict/paginas/informacion.asp

Epler Wood, M. (2017). Introduction. In *Sustainable Tourism on a Finite Planet: Environmental, Business and Policy Solutions:* New York: Routledge. Retrieved from https://www.amazon.

com/Sustainable-Tourism-Finite-Planet-Envi-ronmental/dp/1138217611/ref=sr_1_1?ie=UT-F8&qid=1494539162&sr=8-1&keywords=e-pler+wood+sustainable

Foster, A. D., & Gutierrez, E. (2013). The Informational Role of Voluntary Certification: Evidence from the Mexican Clean Industry Program. *The American Economic Review*, *103*(3), 303–308.

Greenstone, M., & Jack, B. K. (2015). Envirodevonom-ics: A Research Agenda for an Emerging Field. *Journal of Economic Literature*, *53*(1), 5–42. https://doi.org/10.1257/jel.53.1.5

Harbaugh, R., Maxwell, J. W., & Roussillon, B. (2011). Label Confusion: The Groucho Effect of Uncertain Standards. *Management Science*, *57*(9), 1512–1527. https://doi.org/10.1287/mnsc.1110.1412

Hultman, M., Kazeminia, A., & Ghasemi, V. (2015). Intention to visit and willingness to pay premium for ecotourism: The impact of attitude, material-ism, and motivation. *Journal of Business Research*, *68*(9), 1854–1861. https://doi.org/10.1016/j.jbus-res.2015.01.013

Instituto Costarricense de Turismo (ICT). (2016a). Bandera Azul Ecológica. Retrieved June 16, 2017, from http://www.ict.go.cr/es/sostenibilidad/bande-ra-azul.html#criterios

Instituto Costarricense de Turismo (ICT). (2016b). Código de Conducta. Retrieved June 16, 2017, from http://www.ict.go.cr/es/sostenibilidad/bande-ra-azul.html#criterios

Instituto Costarricense de Turismo (ICT). (2017a). In-formes Estadísticos. Retrieved June 16, 2017, from http://www.ict.go.cr/es/estadisticas/informes-esta-disticos.html

Instituto Costarricense de Turismo (ICT). (2017b). Sustainability CST. Retrieved May 12, 2017, from https://www.visitcostarica.com/en/costa-rica/sus-tainability

Jack, B. K., Kousky, C., & Sims, K. R. (2008). Designing payments for ecosystem services: Lessons from previous experience with incentive-based mech-anisms. *Proceedings of the National Academy of Sciences*, *105*(28), 9465–9470.

Laroche, M., Bergeron, J., & Barbaro-Forleo, G. (2001).

Targeting consumers who are willing to pay more for environmentally friendly products. *Journal of Consumer Marketing*, *18*(6), 503–520.

Lee, J.-S., Hsu, L.-T. (Jane), Han, H., & Kim, Y. (2010). Understanding how consumers view green hotels: how a hotel's green image can in-fluence behavioural intentions. *Journal of Sus-tainable Tourism*, *18*(7), 901–914. https://doi.org/10.1080/09669581003777747

Liu, Z. (2003). Sustainable Tourism Devel-opment: A Critique. *Journal of Sustain-able Tourism*, *11*(6), 459–475. https://doi.org/10.1080/09669580308667216

Mebratu, D. (1998). Sustainability and sustainable development: historical and conceptual review. *Environmental Impact Assessment Review*, *18*(6), 493–520.

Mejía, C. V., & Brandt, S. (2017). Utilizing environ-mental information and pricing strategies to reduce externalities of tourism: the case of invasive species in the Galapagos. *Journal of Sustainable Tourism*, *25*(6), 763–778. https://doi.org/10.1080/09669582.2016.1247847

Meleddu, M., & Pulina, M. (2016). Evaluation of individuals? intention to pay a premium price for ecotourism: An exploratory study. *Journal of Behavioral and Experimental Economics*, *65*, 67–78. https://doi.org/10.1016/j.socec.2016.08.006

Podhorsky, A. (2016). Environmental Certifcation Pro-grams: How Does Information Provision Compare with Taxation? Retrieved from http://andreapod-horsky.com/Certification%20Programs_Pod-horsky.pdf

Ramdas, M., & Mohamed, B. (2014). Impacts of Tour-ism on Environmental Attributes, Environmental Literacy and Willingness to Pay: A Conceptual and Theoretical Review. *Procedia - Social and Behavior-al Sciences*, *144*, 378–391. https://doi.org/10.1016/j.sbspro.2014.07.307

Rivera, J. (2002). Assessing a Voluntary Environmen-tal Initiative in the Developing World: The Costa Rican Certification for Sustainable Tourism. *Policy Sciences*, *35*(4), 333–360.

Robinson, M., Kleffner, A., & Bertels, S. (2011). Signal-ing sustainability leadership: Empirical evidence of

the value of DJSI membership. *Journal of Business Ethics*, *101*(3), 493–505.

Roe, B., Teisl, M. F., Levy, A., & Russell, M. (2001). US consumers' willingness to pay for green electricity. *Energy Policy*, *29*(11), 917–925. https://doi.org/10.1016/S0301-4215(01)00006-4

Sartori, G. (2009). Concept Misformation in Comparative Politics. In D. Collier & J. Gerring (Eds.), *Concepts and Methods in Social Science. The Tradition of Giovanni Sartori* (pp. 13–43). New York: Routledge.

Sasidharan, V., Sirakaya, E., & Kerstetter, D. (2002). Developing countries and tourism ecolabels. *Tourism Management*, *23*(2), 161–174.

Silva, E. (2003). Selling Sustainable Development and Shortchanging Social Ecology in Costa Rican Forest Policy. *Latin American Politics and Society*, *45*(3), 93. https://doi.org/10.2307/3177160

Teisl, M. F., Roe, B., & Levy, A. S. (1999). Ecocertification: Why It May Not Be a "Field of Dreams." *American Journal of Agricultural Economics*, *81*(5), 1066. https://doi.org/10.2307/1244085

Williams, C. C., & Millington, A. C. (2004). The diverse and contested meanings of sustainable development. *Geographical Journal*, *170*(2), 99–104. https://doi.org/10.1111/j.0016-7398.2004.00111.x

Woodhouse, P. (2002). Development Policies and Environmental Agendas. In U. Kothari & M. Minogue (Eds.), *Development Theory and Practice: critical perspective* (pp. 136–156). London: Palgrave Macmillan.

World Tourism Organization. (2017). Why Tourism? Retrieved May 11, 2017, from http://www.tourism-4development2017.org/es/por-que-el-turismo/

World Toursim Organization. (2017). Sustainable Development of Tourism. Retrieved May 11, 2017, from http://sdt.unwto.org/content/about-us-5

Technology and Capital Mobility: Estimating the Effects of Mobile Money Adoption on International Remittances

Eddy Chebelyon

University of California San Diego, School of Global Policy and Strategy

This study evaluates the association between introducing mobile money technology and international remittances to home countries using data from the GSMA intelligence mobile money deployment tracker, the World Bank, and the United Nations Population Division. Using a two-way fixed effects model, we find that a country's adoption of mobile money international remittance technology is associated with a $55 increase in remittance per capita in countries with a positive stock of emigrants. This increases to about $73 when using measures above or below median emigrant stock rather than a centered stock of emigrants. On a regional scale, the technology is associated with remittance per capita increases of about $55 in South Asia, $65 in East Asia and the Pacific, and $87 in Europe and Central Asia. In contrast, the study finds no significant relationship between countries adopting the technology and international remittances in Sub-Saharan Africa, Latin America, or the Caribbean; suggesting mobile money firms in these regions capture profits from intra-country capital transfers rather than cross-border remittance transfers. Given the prevalence of South-South migration in these regions where a relationship between the technology and international remittances is not observed, findings may imply that making transfers out of developing countries easier could better facilitate international remittance flows into Sub-Saharan Africa, Latin America, and the Caribbean.

INTRODUCTION

In 2015, some 247 million people lived outside of their countries of birth; up from 175 million in 2000 (Ratha et al. 2016). These emigrants sent an estimated 601 billion USD home in 2015, with approximately 73% of that going to developing countries as international aid. These remittances are about three times as large as the official development assistance given to these nations. Given the likelihood of informal transfers, this amount could be an understatement (Adams Jr and Page 2005).
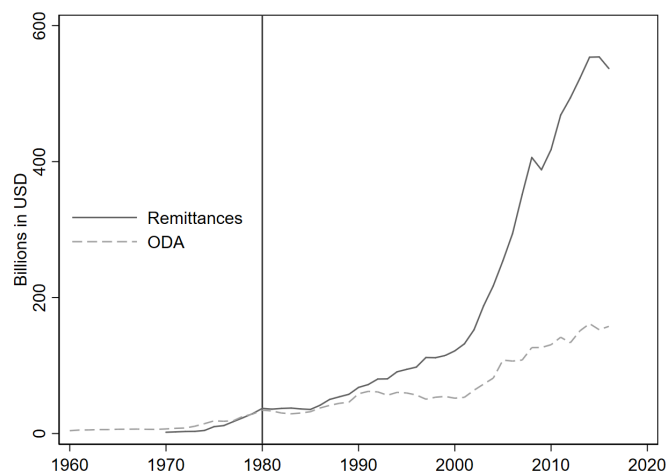
The literature on the role of remittances in economic growth is well established. Mundaca (2009), for instance, finds that remittances have positive and significant long-run effects on growth in Latin America

and the Caribbean by contributing to financial development. Similarly, Giuliano and Ruiz-Arranz (2009) find that remittances boost growth in countries with less developed financial systems by providing them with an alternative way to fund investment and thus overcome liquidity constraints.

On a micro-level, others have argued that remittances induce financial literacy among households, thereby stimulating the demand for financial products and services offered by banks to recipient households (Orozco and Fedewa 2006). Woodruff and Zenteno (2001) study 6,000 small firms in Mexico and estimate that about 20% of capital invested in these microenterprises come from remittances and cumulatively represent about 1.85 billion USD. Anzoategui et al. (2014) finds that Salvadoran households receiving

Figure 1 Remittances received exceeded net Official Development Assistance after 1980, World Bank WDI.

DATA AND METHODS

The measure of mobile money technology adoption used is sourced from the GSMA Intelligence Mobile Money Deployment Tracker. This program works with mobile money firms to create a database of live and planned mobile money services in the developing world. The data contains 276 mobile money firms across 90 countries. Of these firms, 112, or 40% of the firms operating in 56 countries, have the ability to process international remittances. The year in which the firm was launched has been used to construct the treatment variable. This dummy variable equals 1 from the time the country switched on the technology to 2016 (the end of the panel data used in this analysis) and equals 0 otherwise.

Remittance inflow data is sourced from the World Bank's Annual Remittance Database. This measure is the main outcome variable of the analysis. Two control variables from the World Bank's World Development Indicators database are used: GDP per capita, (PPP 2011 international $) and total population. GDP per capita is controlled for because it is likely correlated with both technology adoption and remittance inflows in a way that would bias findings. Population size is used to control for cross-country variation in remittance inflows caused by their size. This is particularly important, first because large remittance recipients tend to be countries with large populations such as India, China, the Philippines, Mexico, Pakistan and Nigeria and secondly, because mobile money firms may be more likely to enter these countries under the assumption that a large population provides a large market for their product.

A country's stock of emigrants is also used as a control variable. To generate this measure, data was taken from the United Nations Population Division on the total number of people living outside of their country of birth and the total number relative to the total population. Given that these data are available every 10 years beginning in 1990, linear interpolation between 1990 and 2000 and each 5-year period afterwards was used to fill the missing data. While this is an imperfect measure, it allows for a sufficiently large sample size necessary to explore variation between year.

remittances are more likely to use deposit accounts but are not more likely to demand and use credit from formal institutions. They hypothesize that remittances relax the credit constraint and encourage households to save.

According to the Global System for Mobile Communications (GSMA), about 5 billion people were connected to mobile services by 2017 (about 3.3 billion were mobile internet users, of which approximately 57% used a smartphone). The ubiquitous nature of the mobile phone has come with seemingly endless possibilities. This ubiquity has led firms to innovate and create technologies that provide a digital transfer of payments through cell phones, better known as mobile money. The technology is fast, secure and has grown in popularity because it gives individuals access to an electronic wallet, which is beneficial for those without bank accounts. It also tends to cost less than traditional transfer methods. These services are typically regulated by national governments and have allowed access to populations locked out of participating in formal financial services, increasing the percent of banked adults from 51% in 2011 to 62% in 2014, according to the World Bank's Universal Financial Access Report. Moreover, according to the European Payments Council (2017), "these new technology solutions provide direct improvement to operations efficiency, ultimately resulting in cost savings". By studying the effects of mobile money international remittance technology on remittance flows, this paper attempts to test one mechanism by which this statement may be true. The following section explains the data and methods used, section III presents and discusses the results while section IV concludes.

Estimation Strategy

The following estimated equation tests the hypothesis:

$$y_t = \alpha_i + \delta_t + \beta_{it} LogGDPperCapita_{it} + \omega Treatment_{it} + vEmigStock_{it} + \tau Treatment_{it} * EmigStock_{it} + \mu_{it} (1)$$

where y_{it} equals the amount of remittance per capita country i receivers in time t. α_i is a country-level fixed effect, δ_t is a time-fixed effect, and μ_{it} is a mean zero error term. τ is the main coefficient of interest and represents the partial marginal effect of a country "switching on" an international remittance mobile money technology conditional on average demanded stock of emigrants. This centered interaction in the model is included because it is expected that the relationship between remittance per capita and the treatment will depend on the magnitude of the stock of emigrants. For instance, we can anticipate that the more emigrants a country has, the more likely it is to receive remittances and respond to a change in the ease of sending remittances.

RESULTS AND DISCUSSION

Descriptive Results

To test whether mobile money adoption had an effect on the flow of international remittances, average annual remittances between countries that received access to international remittance mobile money technology at some point before the end of my sample period, are compared to those that did not. From 1970 to 2000, countries that got access to the technology (treatment) and those that did not (control) appear to have had near parallel co-movement in average annual remittances. Following the turn of the 21st century, treatment countries appear to have had a larger increase in average remittances per year than control countries. This coincides with the timing of the first mobile money firm launched in the Philippines in 2001. After 2001, countries that adopted the technology had higher growth in remittance inflows on average than those that did not.

To test the validity of the assertion on the co-movement in remittances between countries that adopted the technology and those that did not, variable time zero, which is equal to the year in which the country adopted the technology, is generated. Two lag and two lead variables are then constructed around timezero and correspond to 4 years or more than 4 years before and after the technology switched on. Because of the
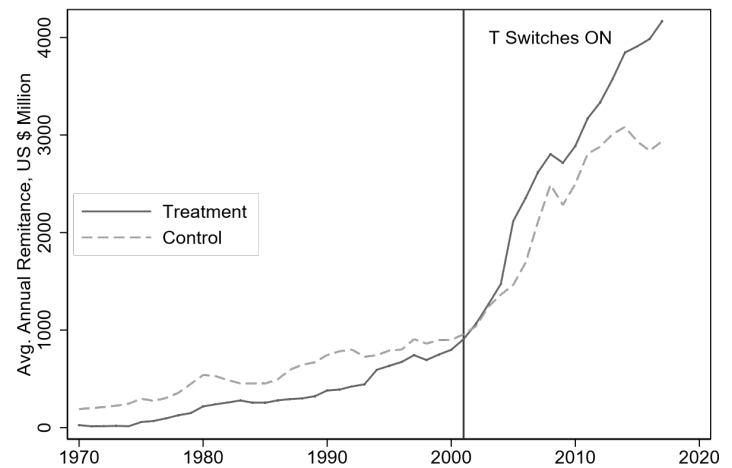

Figure 2 Average annual remittance inflows by mobile technology adopters (treatment) and non-adopters (control), 1970-2016

lag periods, every country that had not adopted the technology before 2011, the median year of adoption, is assigned to the control group.

A t-test is then performed on the difference between the average remittances inflows in treatment relative to control countries by the lead and lags periods around timezero. Figure 3 plots these trends to illustrate whether treatment countries were different than control countries in the average remittances received before and after the treatment countries gained access to the technology.

Figure 3 demonstrates that treatment and control countries were not different before the treatment, but are significantly different after, when more than 4 years had passed since the technology became available.

Empirical Results

A regression analysis was used to examine whether these patterns remain once important confounding factors are controlled for. Table 1 presents estimates of equation 1. Column 1 does not include a control for country wealth, whereas column 2 does. With and without this control, the study finds that a country adopting the mobile money international remittance technology leads to an approximate $54 USD increase in remittance per capita. Moreover, the analysis shows that a 1% increase in emigrants per capita leads to an increase of remittance per capita by between $69 and $75. Not surprisingly, countries that do not have any emigrants living abroad receive no remittance inflows when the technology is introduced, as indicated by the coefficients on *Treatment* which are statistically indistinguishable from zero. Consistent with the

Differences in Remittances Between Treatment and
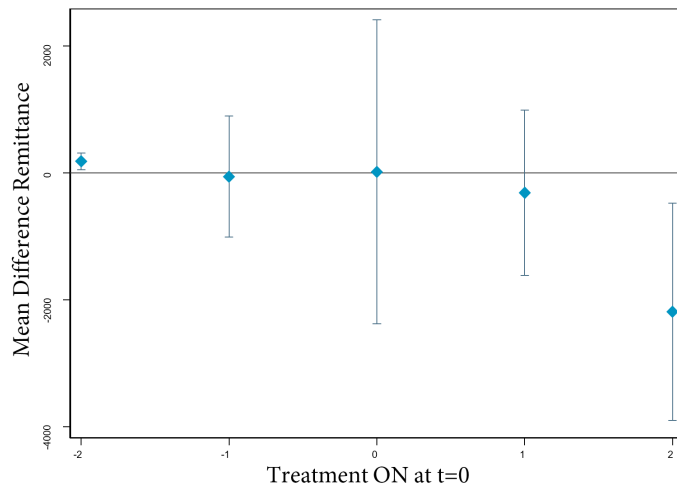Control Before and After Treatment

Figure 3: Average differences in remittance inflows of treatment
and control group, by time period.
Note: -1 and -2 represent 4 and more than 4 years before the
technology was adopted, respectively. Conversely, 1 and 2 rep-
resent 4 and more than 4 years after the technology was adopted.

literature, this study also finds that a GDP per capita
increase of 1% leads to a $55 increase in remittances
per capita (Das et al., 2011).

Table 2 reports results of the estimated effects of mo-
bile money technology on remittance inflows among
countries which have an above-median emigrant pop-
ulation. This coefficient gives the partial effect of the
technology conditional on having above median emi-
grants. The technology increases remittance by about
$73 per person and that countries with above-median
emigrant population and no technology still receive
between $15 and $17 per person. Surprisingly, evi-
dence is found that countries without emigrants lose
about $57 to $60 in remittances per person when the
technology is introduced.[1] Increases in GDP per cap-
ita still lead to increased remittance inflows of about
$51 per person.

Table 3 breaks the main estimation equation by
regions.[2] The technology has positive and significant
effects in three regions, as defined by the World Bank.
In East Asia and the Pacific, the technology increases
remittances by about $70 per person after it has been
adopted, which doubles in South Asia to about $157
and more than triples in Europe and Central Asia to
about $521. In contrast, there are no significant effects
of the technology in Latin America and the Caribbe-

an, the Middle East and North Africa, or in Sub-Sa-
haran Africa. There is also evidence that migrants
from these regions face obstacles moving north or to
regions of higher remittance outflow. Baldwin-Ed-
wards (2006) for instance, finds that Europe welcomes
skilled migrants and rejects asylum seekers who are a
disproportionate representation of these three regions
where the correlation between the technology and
remittances is not significant.

Given that Asia, Europe, and the Pacific traditional-
ly have had larger emigrant populations, represent
significant portions of remittance inflows, and are
rapidly growing or developed, it is assumed that each
country in this region had an equal opportunity to
adopt the technology, and therefore proceed to esti-
mate the reduced-form equation without interacting
the treatment with the stock of emigrants, thereby
estimating a "Quasi-Intention To Treat Effect." Table
4 presents the QITTE estimates for Asia and Europe.
Looking at these regions, treatment increases remit-
tances per capita by between $ 55 in South Asia to
$87 in Central Asia and Europe, though this is weakly
significant.

**Visualizing The Results and Explaining The
Africa Story**

Figure 4 presents the results of the marginal effects
of emigrants on remittance inflows. It shows that
countries that have an emigrant per capita ratio below
10%[3] do not differ, on average, from countries with or
without the technology. However, after a country has
had about 15% of its citizen emigrate, then countries
with the technology have higher remittance inflows
per capita, on average. Of the countries that adopt-
ed the technology and had less than 15% emigrant
stocks, about 52% are in Africa,[4] 11% are in Latin
America and the Caribbean,[5] and about 7% in are
in the Middle East and North Africa[6]. This provides
suggestive evidence as to why the technology did not
have any effect on international remittance inflows
to these reasons. African countries in this sample do
not cross this threshold and so they are isolated and
analyzed on their own.

Figure 5 plots the coefficients for each country in

---

1 Speculatively, this could be because the technology facilitates increased produc-
tion or internal transfers so these countries substitute away from international
remittances.
2 Regional classification used are from the World Bank's Country and Lending
Groups, https://datahelpdesk.worldbank.org/knowledgebase/articles/906519.

3 See Meijering (2002) for details on interpolation methods.
4 Only Burkina Faso and the Democratic Republic of the Congo had more than
15% emigrants and had adopted the mobile money technology.
5 El Salvador and Jamaica.
6 Morocco, Qatar and Tunisia.

Africa that adopted the technology. Estimating the effect of the treatment on the treated African countries, we see that the technology has no significant effect on average. However, by estimating a slope for each country, we see that the technology does have a significant and positive effect on 3 of the 18 countries in the sample: Lesotho, Nigeria and Senegal. While African countries have been at the forefront of mobile technology adoption, this finding provides suggestive evidence that the technology has not been significantly effective in enhancing cross-border capital transfers. This is consistent with much of the evidence in the literature being focused on intra-country mobility of cash and financial inclusion (see Aker and Mbiti, 2010; Asongu, 2013.

CONCLUSION

This study used data the GSMA, the World Bank and the United Nations Population Division to estimate the impact of mobile money international remittance technology on remittance inflows. Using a two-way fixed effects model, the study finds that adopting this technology is associated with an increase in remittances per capita of about $55 conditional on average stock of emigrants and income. This relationship is dominated by European, Asia and Pacific countries which have higher average stocks of emigrants. Though the bulk of mobile money firms are present in Latin America, the Caribbean and South Asia, countries in these regions do not see increased remittances per capita when they adopt the technology. This could suggest that firms are entering these markets with anticipation of maximizing profits from intra-country capital transfers rather than international remittance transfers.

These findings have important implications for firms and public policy. South to North migration appears to be the main driver of remittances, as suggested by results indicating positive effects in regions that form the bulk supply of South-North migration and North-North migration; Asia and Europe respectively. Similarly, mobile money firms have not seen increases from international remittances in countries with higher numbers of south to south emigrants, notably in Sub-Saharan Africa (Shimeles, 2018) and Latin America (International Labour Organization, 2016). This may suggest that making transfers out of developing countries easier (safer, less expensive) may be important for countries in these regions to see increases in international remittances facilitated by technology.

These findings are subject to important limitations that may be addressable as more data becomes available. Specifically, since the stock of emigrants are interpolated, we cannot be sure if the results would hold had the real stock of emigrants been used. To further understand the relationships presented in this paper, a next step would involve studying these relationships on a dyadic structure, with high frequency data and measures for the supply side of mobile money technology. More generally, despite the evidence that the control and treatment countries appear similar before treatment, mobile money technology is not randomly assigned and there may be important unobservable variables related both to whether or not a country gets access to the technology and its international remittances that are biasing my estimates. These findings should therefore be interpreted as robust correlations rather than causal effects.

BIBLIOGRAPHY

Aker, Jenny C, "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, 2010, *2* (3), 46–59.

Christopher Ksoll, and Travis J Lybbert, "Can mobile phones improve learning? Evidence from a field experiment in Niger," *American Economic Journal: Applied Economics*, 2012, *4* (4), 94–120.

Michael A Clemens, and Christopher Ksoll, "Mobiles and mobility: The effect of mobile phones on migration in Niger," 2011.

Isaac M. Mbiti. "Mobile phones and economic development in Africa." *Journal of Economic Perspectives* 24, no. 3 (2010): 207-32.

Anzoategui, Diego, Asli Demirgu¨¸c-Kunt, and Mar´ıa Soledad Mart´ınez Per´ıa,"Remittances and financial inclusion: evidence from El Salvador," *World Development*, 2014, *54*, 338–349.

Asongu, Simplice. "Mobile banking and mobile phone penetration: which is more pro-poor in Africa?." *African Governance and Development Institute WP/13/033* (2013).

Baldwin-Edwards, Martin, "'Between a rock & a hard place': North Africa as a region of emigration, immigration & transit migration," *Review of African Political Economy*, 2006, *33* (108), 311–324.

Das, Anupam, Murshed Chowdhury et al., "Remittances and GDP dynamics in 11 developing countries: evidence from panel cointegration and PMG techniques," *Romanian Economic Journal*, 2011, *14* (42), 3–23.

European Payments Council, "White Paper Mobile Payments," 2017.

Giuliano, Paola and Marta Ruiz-Arranz, "Remittances, financial development, and growth," *Journal of Development Economics*, 2009, *90* (1), 144–152.

Gramer, Robbie, "Watch: The Global Flow of Asylum-Seekers In One Interactive Map," *Foreign Policy*, 2017.

International Labour Organization, "Labour migration in Latin America and the Caribbean," *Regional Office for Latin America and the Caribbean*, 2016.

Jr, Richard H Adams and John Page, "Do international migration and remittances reduce poverty in developing countries?," *World development*, 2005, *33* (10), 1645–1669.

Meijering, Erik, "A chronology of interpolation: From ancient astronomy to modern signal and image processing," *Proceedings of the IEEE*, 2002, *90* (3), 319–342.

Mundaca, B Gabriela, "Remittances, financial market development, and economic growth: the case of Latin America and the Caribbean," *Review of Development Economics*, 2009, *13* (2), 288–303.

Muto, Megumi, "The impacts of mobile phones and personal networks on rural-to-urban migration: Evidence from Uganda," *Journal of African Economies*, 2012, *21* (5), 787–807.

Orozco, Manuel and Rachel Fedewa, *Leveraging efforts on remittances and financial intermediation (Working Paper ITD= Documento de Trabajo ITD; n. 24)*, BID-INTAL, 2006.

Ratha, Dilip, Christian Eigen-Zucchi, and Sonia Plaza, *Migration and Remittances Factbook 2016*, World Bank Publications, 2016.

Shimeles, Abebe, "Foresight Africa viewpoint – Understanding the patterns and causes of African migration: Some facts," *Brookings*, 2018.

Woodruff, Christopher and B Zenteno, "Remittances and Micro Enterprise in Mexico," 2001.

APPENDIX

Table 1: Main Results

	(1)	(2)
	Remittance Per Capita	Remittance Per Capita
Treatment*Emigrant Stock	54.06***	54.60***
	(9.250)	(8.985)
Emigrant Stock	74.70***	69.18***
	(9.740)	(9.624)
Treatment	-8.039	-5.040
	(11.83)	(11.01)
Log GDP Per Capita		54.93***
		(16.21)
Constant	-228.1***	-610.1***
	(21.07)	(114.0)
Observations	3,063	3,007
R-squared	0.988	0.986
Time FE	Yes	Yes
Receiver FE	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Main Results using a dummy for whether a country has above or below median emigrants per capita

	(1)	(2)
	Remittance Per Capita	Remittance Per Capita
Treatment*Median Emigrant Stock	72.77***	72.89***
	(10.75)	(10.64)
Median	14.84**	17.27**
	(6.826)	(7.062)
Treatment	-59.99***	-56.61***
	(6.052)	(5.537)
Log GDP Per Capita		51.70***
		(12.93)
Constant	-145.8***	-510.8***
	(13.89)	(91.77)
Observations	4,135	4,011
R-squared	0.986	0.984
Time FE	Yes	Yes
Receiver FE	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Regional effects of the technology on remittances

	(1)	(2)	(3)	(4)	(5)	(6)
	East Asia & Pacific	LAC & Caribbean	Europe & C Asia	ME & N Africa	S Asia	SSA
Treatment*Emigrant Stock	69.70***	-45.57	520.9***	-59.53	157.7***	-6.442
	(19.32)	(42.51)	(133.2)	(44.36)	(20.34)	(4.874)
Emigrant Stock	17.04	-15.28	247.0***	-73.70*	23.66*	-13.19**
	(34.25)	(15.75)	(38.09)	(43.97)	(12.17)	(5.729)
Treatment	55.33***	11.86	-599.8***	-25.76	115.9***	-12.74
	(17.75)	(18.17)	(176.6)	(22.03)	(12.91)	(8.277)
Log GDP Per Capita	-21.19	-204.2***	77.52***	337.9***	-46.41***	-26.99***
	(22.98)	(52.69)	(28.02)	(69.46)	(13.42)	(9.781)
Constant	264.5	1,809***	-792.6***	-3,123***	-5.203	192.8**
	(260.9)	(487.5)	(229.3)	(641.7)	(87.26)	(78.99)
Observations	391	555	924	260	141	716
R-squared	0.843	0.788	0.898	0.925	0.919	0.829
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Receiver FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Quasi Intention To Treat Estimates for Asia

	(1)	(2)	(3)
	East Asia & Pacific	Europe & C Asia	S Asia
Treatment	65.27***	86.68*	54.98***
	(21.98)	(47.42)	(8.482)
Log Stock of Emigrants	18.91	233.6***	34.26**
	(38.00)	(37.62)	(17.17)
Log GDP Per Capita	-62.22**	84.54***	-17.01
	(24.51)	(27.94)	(11.49)
Constant	730.6**	-193.7	94.45
	(352.3)	(260.6)	(85.66)
Observations	391	924	141
R-squared	0.823	0.897	0.836
Time FE	Yes	Yes	Yes
Receiver FE	Yes	Yes	Yes

Robust standard errors in parentheses
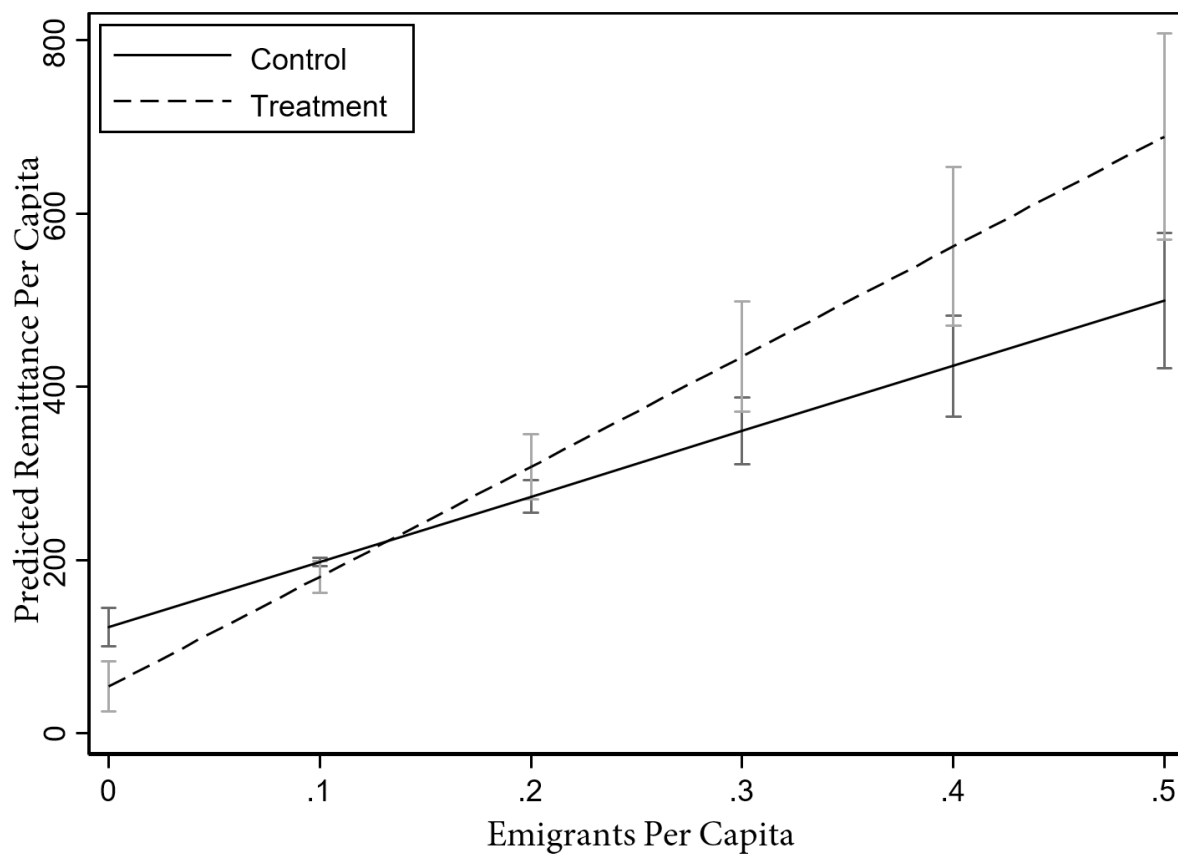
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 4: Adjusted Predictions with 95% CIs

Figure 5: Coefficient plot of predicted margins for remittance inflow for African countries, useing treatment on the treated estimate
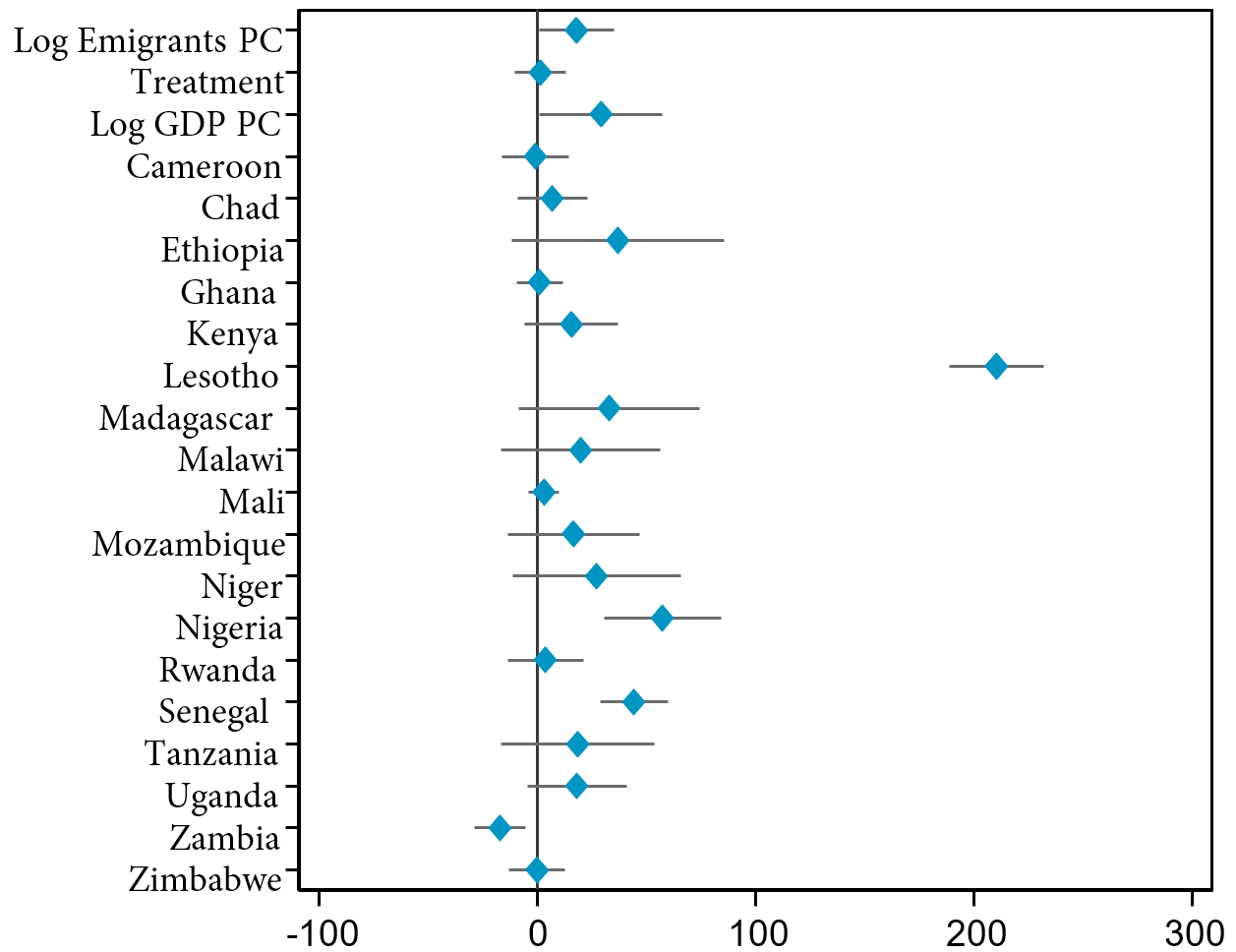
Region	variable	N	mean	sd	min	max	cv
East Asia & Pacific	Remittance Inflow	659	2589.45	8055.94	0.28	63937.65	3.11
	Remittance Inflow Per Capita	629	221.29	510.87	0.01	2899.37	2.31
	Stock of Emigrants	587	944825	1741259	999	9930000	1.84
	Emigrants Per Capita	567	0.19	0.29	0.00	1.11	1.50
	GDP Per Capita	807	16394.79	22239.05	728.03	135319	1.36
Europe & Central Asia	Remittance Inflow	1155	2237.32	3322.50	0.03	25351	1.49
	Remittance Inflow Per Capita	1106	237.85	387.91	0.01	3420	1.63
	Stock of Emigrants	1287	1027820	1220827	9999	5962000	1.19
	Emigrants Per Capita	1242	0.13	0.11	0.02	0.81	0.82
	GDP Per Capita	1284	24007	18025	1043	97864	0.75
Latin America & Caribbean	Remittance Inflow	862	1260	3402	0.01	30534	2.70
	Remittance Inflow Per Capita	829	173.18	222.49	0.00	1982	1.28
	Stock of Emigrants	672	928546	1935923	10000	12500000	2.08
	Emigrants Per Capita	648	0.18	0.21	0.00	0.98	1.17
	GDP Per Capita	886	11745.75	7003.50	1502.03	49903	0.60
Middle East & North Africa	Remittance Inflow	421	1823	2846	2.61	19570	1.56
	Remittance Inflow Per Capita	403	163.63	280.84	0.11	1809	1.72
	Stock of Emigrants	392	542130	705020	9999	2850000	1.30
	Emigrants Per Capita	375	0.07	0.07	0.00	0.31	0.98
	GDP Per Capita	504	26954.56	29515.16	2130.00	12930	1.09
North America	Remittance Inflow	53	2897.62	2049.95	912.16	6613	0.71
	Remittance Inflow Per Capita	50	4452.99	8461.55	4.69	22348	1.90
	Stock of Emigrants	56	593679	576384	10000.00	1314000	0.97
	Emigrants Per Capita	54	0.23	0.29	0.04	1.18	1.23
	GDP Per Capita	78	44280.79	7691.87	30115.28	60675.98	0.17
South Asia	Remittance Inflow	187	7211.16	14466.56	1.30	70389	2.01
	Remittance Inflow Per Capita	179	48	67.82	2.02	342.28	1.42
	Stock of Emigrants	196	3906816	3797317	9999	16500000	0.97
	Emigrants Per Capita	189	0.08	0.09	0.01	0.55	1.18
	GDP Per Capita	193	4123	3101	1062	14232	0.75
Sub-Saharan Africa	Remittance Inflow	922	492.34	2314.42	0.01	22344.54	4.70
	Remittance Inflow Per Capita	886	40.21	70.60	0.00	409.07	1.76
	Stock of Emigrants	980	392513.30	359058.70	9999.00	2220000.00	0.91
	Emigrants Per Capita	940	0.05	0.05	0.00	0.43	0.98
	GDP Per Capita	1189	3916.13	5424.23	247.44	40015.82	1.39

GOVERNMENT INACCESSIBILITY AND COUNTERINSURGENCY: A SPATIAL ANALYSIS OF INSTITUTIONAL PRESENCE IN RAKHINE STATE, BURMA (AUGUST-NOVEMBER 2017)

Ashley Halabi
UC San Diego, School of Global Policy and Strategy

Governments that do not have a sufficient institutional presence in remote areas are more likely to suffer from challenges to their authority from those regions. Insurgents thrive in areas where government reach and law enforcement are weak. They are also advantaged if the region they operate in is characterized with dense forest cover or is close to an international border. Using Geographic Information Systems (GIS) and remote sensing to measure areas and distances and calculate topographic characteristics, this paper will test the hypothesis that as government inaccessibility increases, the likelihood that a village is attacked with a brutal counter-insurgency strategy increases. Analysis is based on the recent attacks instigated by the Burmese in the Rakhine State against Rohingya-dominant villages.

INTRODUCTION

On August 25, 2017, the Arakan Rohingya Salvation Army (ARSA) executed an attack against 30 police stations and outposts in Rakhine State, Burma.[1][2] These attacks continued to escalate through August 31st, triggering a brutal, scorched earth response by the government. Starting on August 26th, the Burmese military initiated a counter-attack on villages in the Maungdaw, Buthidaung, and Rathedaung townships in Rakhine State.[3] At least 232 villages, mostly Muslim-majority populations, were burnt to the ground in the span of two months, which was later characterized by the United Nations High Commissioner for Human Rights as "brutal" and a "textbook example of ethnic cleansing".[4][5]

This paper analyzes the correlation between the subnational Rakhine State government's inaccessibility and its counter-insurgency strategy. The state government may have pursued this strategy because it lacked the capacity to identify and discretely target insurgents who took advantage of their ability to blend in with civilian populations in the Rohingya dominated villages and camouflage provided by the dense forests of the area.

The research question motivating this paper is two-fold: 1. Why did government counterinsurgent forces target villages in the townships closest to the border

---

Figure 1 Damaged Areas in Rakhine State
(August 31 till October 11, 2017)

with Bangladesh and not others that are in the Rohingya-dominated areas? And 2. Is this decision related to the inaccessibility of government in that region, especially since they could have targeted Rohingya-dominated villages closer to the capital and not those surrounded by forests, mountains, or in rural areas?

CONCEPTUAL FRAMEWORK

State capacity and insurgency theories stress the importance of government institutions in preventing rebel groups from mobilizing. Fearon and Laitin (2003) argue that governments that are institutionally and politically weak are more prone to civil conflict, as their weakness fosters the right conditions for insurgencies to thrive.[6] Other theories of insurgency, including that of Tajima (2012), argue that governmental reach allows for better policing in remote regions, thereby preventing outbreaks of communal violence.[7] Tajima further argues that the presence of security outposts in remote regions helps manage communities that might otherwise be violent.[8] Regarding counter-insurgency strategies brought about by inaccessibility, Valentino et al's (2004) theory on mass killing and guerilla warfare explains the use of mass killing as a strategy applied

by governments when the insurgents are supported by the population (i.e. insurgents are indistinguishable from civilians in villages), allowing them to hide from government forces.[9] Building on these theories, this paper tests the proposition that inaccessibility might be correlated with the government's counterinsurgency strategy.

BACKGROUND

After achieving independence from the British in 1948, Burma faced an ethnically and religiously fractionalized country without outside support,[10] and it also had to contend with groups seeking independence and autonomy. The majority-Muslim Rohingya minority ethnic group, for instance, sought greater political representation and more autonomy after they had been deprived of citizenship and identification documents by the newly independent Burmese state.[11] They were confined to Rakhine state and suffered from strict rules and policies that limited their mobility.[12] In that region the insurgent group, Arakan Rohingya Salvation Army (ARSA), quickly gained power and capitalized on the dense forests and mountainous areas in Rakhine State

---

[6]Fearon, James D., and David D. Laitin. "Ethnicity, Insurgency, and Civil War." (2003). P.75
[7]Tajima, Yuhki (2012). "The Institutional Basis of Intercommunal Order"
8 Ibid., p.106

9 Valentino et al (2004) ""Draining the Sea": Mass Killing and Guerrilla Warfare." p.375
10 Fearon and Laitin (2007), "Random Narratives, Burma", p. 2
11 Engy Abdelkader Rutgers University. "The history of the persecution of Myanmar's Rohingya."
12 Ibid.,

to sustain its rebellion for decades.[13] This paper will focus on Rakhine State, and the region within dominated by the Rohingya.

METHODS

This paper tests several hypotheses to assess the theory of whether the Burmese government's institutional presence in Rakhine State is correlated with the likelihood of increased insurgency or local conflict:

Hypothesis 1: Physical proximity to an international border increases the likelihood of insurgency against the government.

Hypothesis 2: Physical inaccessibility of the government to the Rakhine State due to a dense forest cover and mountainous terrain increases the likelihood of insurgency against the government.

Hypothesis 3: Socioeconomic inaccessibility of government institutions increases the likelihood of insurgency against the government.

Table 1: Operationalization of Variables

Dimension	Operationalization
Physical Distance	Distance to International Border (km)
Physical Inaccessibility	Dense Forest Cover (%)
	Elevation Levels
	Road and Railway Access (binary)
Socioeconomic inaccessibility	Urban Areas (Binary)

DATA ASSEMBLY & RESEARCH DESIGN

First, to limit the analysis to the Rohingya-dominated areas in Rakhine State and understand why some villages were targeted for government counterinsurgency actions but not others, the GeoEPR dataset (Wucherpfennig et al. 2011) was used to provide a spatial distribution of ethnic groups in Burma since 1946; though it should be noted that this source only maps the ethnic groups that face discrimination in the State.[14] This was used to create an extent for the study to resemble Rohingya-dominated areas.

The unit of analysis is a point layer. To study the damaged zones burned by the Burmese military between August and November 2017, satellite-imagery from UNITAR was used. This dataset maps the destruction zones after the military attack on Rakhine State on August 25, 2017.[15] The dataset maps the satellite-detected fires from August 25, 2017 through November 20, 2017 in several villages in Rakhine State,[16] as well as the villages destroyed. in 1862. These destroyed areas were converted into points by using the Feature to Point tool in ArcMap 10.5.1. These points are the unit of analysis and the dependent variable. Three kilometer buffers were created around these points to gain a better understanding of the surrounding areas.

To determine why the government attacked only these 1,862 points in the Rohingya-dominated area, 3,000 random points were generated in ArcMap to establish variation for the regression analysis. These 3,000 points were constrained to the Rohingya-dominance shapefile; 1,862 random points and their three-kilometer buffers that did not overlap with the three-kilometer buffers around the damaged points were selected from the 3,000 randomly generated points to serve as the comparison with the points that were attacked. A field was added to each attribute table in ArcMap to create a binary value of whether the point was damaged or not (damaged points were assigned a '1' and random points were assigned a '0'). Finally, the two tables were merged based on the unique identifier and were exported into Google Earth Engine for further remote sensing analysis.

After importing a merged file for the points, three of the four independent variables were calculated. First, the forest cover characteristics of each point was calculated using the Hansen Global Forest Change v1.3 (2000-2015) database in Google Earth Engine Coder.[17] Only the band 'treecover2000', which represents tree cover in the year 2000 defined as canopy closure for all vegetation taller than five meters, was selected as it explains the presence of forest cover in this database. A function was used with the reducer sum to calculate the percentage of forest cover in each point of the feature collection imported from ArcMap. The results were exported as a table to Google Drive and then imported into Stata for regression analysis.

13Fearon and Laitin (2007), Random Narratives p.2
14 Andreas Forø Tollefsen, Halvard Buhaug; Insurgency and Inaccessibility, *p.9*

15 "Destroyed areas in Buthidaung, Maungdaw, and Rathedaung Townships of Rakhine State in Myanmar." UNITAR.
16 "MIMU." Rakhine | MIMU.
17 Hansen, M. C.,et al.. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." Science 342 (15 November): 850–53.
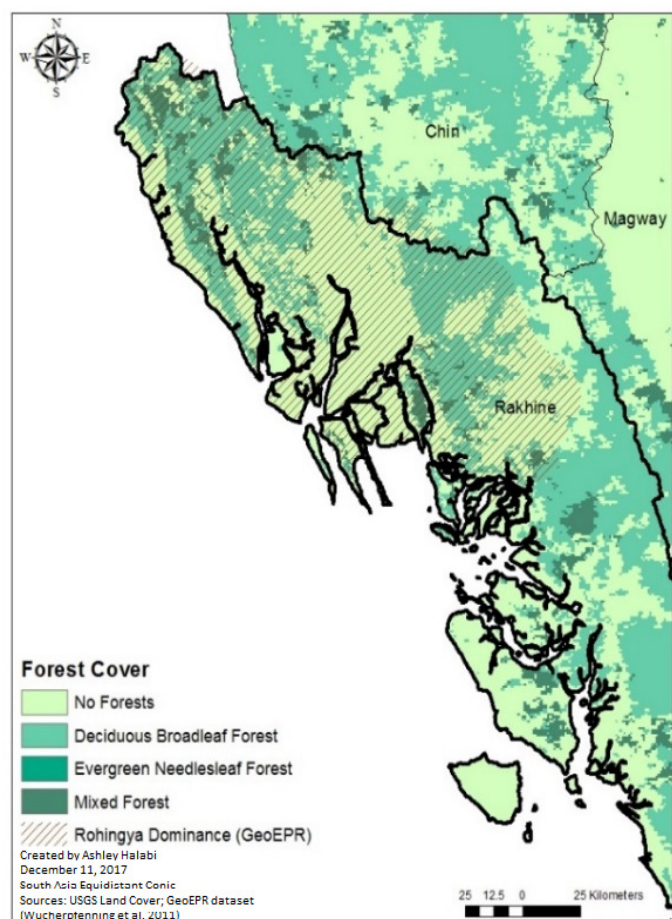
Figure 2: Map of Forest Cover in Rakhine State

Figure 2 shows the dense forest cover that separates Rakhine and specifically the areas that are Rohingya dominated from the rest of the country.

Second, for the elevation, the SRTM Digital Elevation Data 30-m database was used.[18] It has one band with values on elevation that range between -10 to 6,500 meters. To calculate the elevation levels in each of the points, a function was used with the reducer of sum on the feature collection that was imported from ArcMap. This function applied the calculation of elevation levels on every value in the feature collection and the results were then exported as a table to be used in Stata for regression analysis.

Third, is the presence of urban areas in the damaged regions. The Land Cover Type Yearly Global 500-m image collection was imported and only Land Cover Type 2, number 13, which resembles urban areas, was selected. Then, a sum reducer function was used on the feature collection of the merged points that were imported from ArcMap. This function calculated urban area presence for every point in the feature collection. The results were then exported as a table to be used in

18Farr, T.G., et al 2007, The shuttle radar topography mission

Stata for regression analysis.

Finally, the fourth independent variable is the distance between each point, damaged or random, and the closest point in Bangladesh, which was selected in ArcMap using the Polygon to Feature Tool. The Point Distance Tool was also used to generate a table of distances from each point. These were then added to the merged points attribute table to be exported to Stata for further regression analysis.

After the five independent and dependent variables were combined in the merged buffer attribute table, the table was imported into Stata and a regression was conducted to test whether the likelihood of a village being damaged increased when it was located beneath a dense forest cover, at a higher elevation level, in a rural area, and proximate to the Burma-Bangladesh border.

RESULTS

Table 2: Regression Results

	Damaged Points (binary)
Forest Cover (%)	-0.00420***
	(-27.35)
Elevation (meters)	-0.0000890***
	(-4.79)
Urban Area (Binary)	-0.291***
	(-22.61)
Distance to Bangladesh (km)	0.556***
	(56.45)
Constant	0.250***
	(18.45)
Number of Observations	3724
t statistics in parentheses	
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

The table above shows the regression results from Stata. The dependent variable, Damaged Points, was regressed on forest cover, elevation levels, urban area presence, and distance to Bangladesh.

As seen in Table 2, all the coefficients are statistically significant at a 99 percent confidence level, which demonstrates that there is a significant correlation be-

tween these inaccessibility measures and the likelihood that a village was damaged by government counterinsurgency attacks in a Rohingya-dominant area.

As the extent of forest cover increases by 1 percent, the likelihood that a point is damaged decreases by 0.00420 units, keeping in mind that the analysis was constrained to points and does not include the three-kilometer buffer zones. This negative correlation implies that every point with forest cover did not experience government-led counter-insurgency attacks, although this does not refer to the villages that lie behind these forest-covered areas. Moreover, the results show that as the level of elevation increases by 10,000 meters, the likelihood that a village is damaged decreases by 0.89 units. This also indicates that the villages that are located behind these elevated areas were likely attacked and not the mountainous areas themselves. This is further confirmed by the next two variables which show positive correlations with brutal counter-insurgency tactics.

For instance, the urban area variable showed that the likelihood that a village was damaged decreased when the area was urban. This is useful because it illuminates the fact that the Rohingya-dominated area, which is mostly rural, had a greater chance of experiencing a brutal counter-insurgency campaign. Finally, as proximity to the Bangladesh border increased, the likelihood that a village was damaged also increased by 0.556 units, which is statistically significant at the 99.9 percent confidence level. When all other variables are not present, the likelihood that a village was damaged increases by 0.25 units.

As the results of the regression demonstrate, not only is there a correlation between the accessibility of the government and the counter-insurgency attacks, but that there is also a positive correlation between distance to an international border and urban areas and the likelihood that a village was damaged. These variables also help predict where the counter-insurgency efforts were likely to occur in inaccessible regions. Since the analysis was constrained to points and not buffer zones, the results indirectly demonstrate that counter-insurgency tactics occur in areas under dense forest cover and high elevation terrain. This is supported by the urban and the international border results which showed that the areas that were attacked by government forces were not urban, were located in forests, mountains, and were proximate to the Bangladesh border.
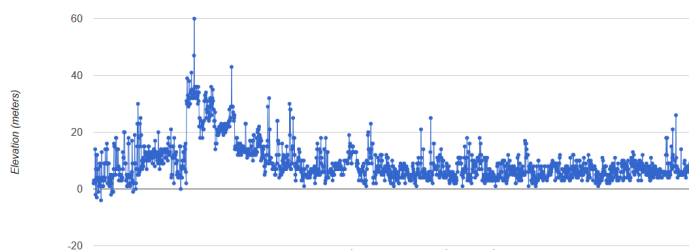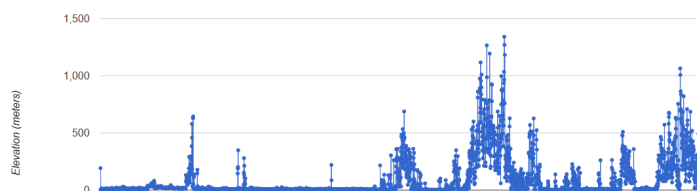
Figure 3: Elevation of Damaged Points

Figure 4: Elevation of Random and Damaged Points

To better explain the elevation results, the charts above show that the immediate points of damaged areas do not have high elevation levels. However, when all the random points are taken into consideration, the elevation rises significantly. This means that the damaged areas are located in mountainous terrains, a factor that influenced government inaccessibility, but was not accurately captured in the regression.

Moreover, these results show that there are many unobserved variables that come into play. There are various explanations for the scorched earth tactic that the government chose to take, and these variables may be merely a small factor in influencing the military's counter-insurgency tactics.

INFASTRUCTURE & ISOLATION

This section analyzes additional information and spatial data on the Rakhine state and its infrastructure.

According to a UNDP report from 2015, the Rakhine State has severe infrastructure shortages including lack of access to roads, railways, electricity, healthcare, education, and policing.[19] This is due to several factors, including the different ethnicities, presence of insurgent groups, and sporadic clashes with the military. [20] To illustrate the region's lack of access to the rest of the country, data from the Myanmar Information Management Unit was used to visualize the roads and railway networks to Rakhine State.[21] The roads and railways were converted into rasters using the Polyline to Raster Tool.

Figure 5 illustrates the physical inaccessibility to the

19 "The State of Local Governance: Trends in Rakhine." UNDP in Myanmar. 2015.
20 Ibid.,
21 "MIMU" Rakhine | MIMU.

Figure 5: Map of Road and Railways in Rakhine State

Figure 6: Map of Euclidean Distance from Rakhine to the Capital

government through the presence of roads or railways. As can be seen from the map, there are no railways that connect the northern Rakhine State which experienced the majority of the violence in August-November 2017, to the rest of the country. Concurrently, only three major (primary) roads connect the whole Rakhine State to the rest of the country; all other roads are secondary or tertiary.

The effect of the lack of roads on provision of other services such as healthcare, education, and policing are exacerbated by the remoteness of the northern region of Rakhine State from the capital. To visualize this, the Euclidean distance raster was used to measure the distance between Rakhine State and the capital Nay Pyi Taw. The Euclidean distance to the capital was calculated with cell size of 5,000 and was masked to Burma.

Figure 6 shows the distance of Rakhine State from the capital. Focusing on the areas dominated by the Rohingya, the map shows that the lack of basic services and transportation is exacerbated by the remoteness of the region from the main institutions of the country.

DISCUSSION & LIMITATIONS

The hypothesis presented in this paper encounters

many challenges that must be overcome in future research. First, the issue of spatial dependence is important to consider in order to avoid inaccurate results. If this analysis were expanded, adding buffers to each point under study would give more accurate results regarding each point's surroundings and how those surroundings influence incidents of violence. To achieve this, a geographically weighted regression can be implemented to account for any auto-correlation. In addition, future research should consider measuring cost distance to the capital through roads and even the time needed to reach the capital or other governmental institutions. Finally, nightlights, proxying for economic activity, can be used as controls.

Many more underlying and unobservable variables must be taken into consideration when analyzing this specific incident. The government's ethnic cleansing campaign is the result of many more political and historical factors than just accessibility. However, as stated earlier, the low accessibility to Rakhine might have been a significant driver of this strategy of counter-insurgency due to informational asymmetries on the government's side.

CONCLUSION & FURTHER RESEARCH

After testing the three hypotheses that impact the likelihood of a counter-insurgency attack in remote regions, this paper can conclude that the presence of physical barriers such as forest cover, mountainous terrain, and proximity to an international border all contribute to government inaccessibility to these regions. This paper determines that some of these measures of inaccessibility increase the likelihood of a counter-insurgency attack against remote places.

Future research can further explore if the insurgent group ARSA benefited from the various safe havens discussed and the government's inaccessibility to Rakhine to mobilize and operate. The ARSA in Rakhine were able to operate freely in the mountains and forests, which allowed them to work on their insurgency tactics and to recruit and train. The government's decision to burn villages in northern Rakhine on August 25, 2017 can be explained by the government's attempt to reach locations that served as safe havens for the ARSA group.

Finally, the paper supports the conclusion that as government inaccessibility increases, the likelihood that a government uses a brutal counter-insurgency strategy increases. This can possibly be due to the information asymmetry of the state in that region and the inaccessibility of the government.

BIBLIOGRAPHY

Andreas Forø Tollefsen, Halvard Buhaug; Insurgency and Inaccessibility1, International Studies Review, Volume 17, Issue 1, 1 March 2015, Pages 6–25, https://doi.org/10.1111/misr.12202

Bhaumik, Subir. "Myanmar has a new insurgency to worry about." South China Morning Post. September 01, 2017. Accessed December 14, 2017. http://www.scmp.com/week-asia/geopolitics/article/2109386/myanmar-has-new-insurgency-worry-about

Center for International Earth Science Information Network (CIESIN), Columbia University http://www.ciesin.org/

United Nations Food and Agriculture Programme (FAO); and Centro Internacional de Agricultura Tropical (CIAT). 2005.

"Destroyed areas in Buthidaung, Maungdaw, and Rathedaung Townships of Rakhine State in Myanmar." UNITAR. http://www.unitar.org/unosat/map/2727.

Farr, T.G., et al, 2007, The shuttle radar topography mission: Reviews of Geophysics, v. 45, no. 2, RG2004, at http://dx.doi.org/10.1029/2005RG000183.

Fearon, James D., and David D. Laitin. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97, no. 01 (2003): 75-90. doi:10.1017/s0003055403000534.

Fearon, James D., and David D. Laitin. "Random Narratives: Burma." (2007) https://web.stanford.edu/group/ethnic/Random%20Narratives/random%20narratives

Hansen, Potapov, Moore, Hancher et al. "High-resolution global maps of 21st-century forest cover change." Science 342.6160 (2013): 850-853.

"The State of Local Governance: Trends in Rakhine." UNDP in Myanmar. 2015. http://www.mm.undp.org/content/myanmar/en/home/library/poverty/TheStateofLocalGovernanceChin/The_state_of_local_governance_trends_in_Rakhine.html

"MIMU" Rakhine | MIMU. http://themimu.info/states_regions/rakhine.

Mizzima. "Think tank releases data on villages burnt in Rakhine State." Mizzima. November 2, 2017. Accessed December 14, 2017. http://www.mizzima.com/news-domestic/think-tank-releases-data-villages-burnt-rakhine-state.

Sundberg, Ralph, and Erik Melander. "Introducing the UCDP georeferenced event dataset." Journal of Peace Research 50, no. 4 (2013): 523-532.

Tajima, Yuhki. "The Institutional Basis of Intercommunal Order: Evidence from Indonesia's Democratic Transition." American Journal of Political Science 57, no. 1 (2012): 104-19. doi:10.1111/j.1540-5907.2012.00631.x.

"The USGS Land Cover Institute (LCI)." The USGS Land Cover Institute (LCI). Accessed December 14, 2017. https://landcover.usgs.gov/.

Valentino, Benjamin, Paul Huth, and Dylan Balch-Lindsay. ""Draining the Sea": Mass Killing and Guerrilla Warfare." International Organization 58, no. 02 (2004). doi:10.1017/s0020818304582061.

Wucherpfennig, Julian, Nils B. Weidmann, Luc Girardin, Lars-Erik Cederman, and Andreas Wimmer. "Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset." Conflict Management and Peace Science 28, no. 5 (2011): 423-437.