# Attempting Replication

# *Lessons from - Commodity Price Shocks*

# *and*

# *Civil Conflict: Evidence from Colombia*

By Gabriel German Higinio Hoyos

Student number: 1704465

A Dissertation Submitted to the
Department of Government in Partial Fulfilment of the Requirements
For the Degree of BA International Relations
May 2020

GV-831 Project Politics
Supervisor: Dr Sotirios Karampapas

## Abstract

What insights can be gained from within-country replication? The replication is conducted by verifying the original results and code and repeating the methodology on an extended time series. This paper finds the replication inconclusive through lack of instruction on behalf of the authors, the lack of micro-data to properly repeat the study in an extended time series and the impossibility of generalising such a country-specific study using available microdata under model specifications. However, the dissertation finds that these issues can be offset by increasing funding, maintenance and construction of micro-data in order to properly replicate studies and, in the context of commodity price and civil conflict, attempt to reach a consensus on what mechanisms drive this relationship.

# Acknowledgments

I would like to thank everyone who has supported me throughout these long three years of University to reach this point.

I am deeply indebted to my Quantitative Political Analysis lecturer Dr Dominik Duell who, without his insights and constant help, this dissertation wouldn't have been possible.

I am grateful to Madeleine O'Connor for proof reading and checking this dissertation.

I would also like to thank my supervisor Dr Sotirios Karampampas.

And finally, I would like to thank my Mother and Brother who have helped push me to excel and do my best.

# Table of Contents

# I. Introduction

Successful replication is defined by the faithful recreation of the original study, tracking of differences, checking assumptions in new contexts and the adoption of high-powered replication studies (Brandt et al, 2014; 222). In social science, replications are crucial as they establish the veracity of an effect and estimate its size (Brandt et al,2014; 222). The definition of micro-data with regards to within-country analysis is “Microdata are unit-level data obtained from sample surveys, censuses, and administrative systems. They provide information about characteristics of individual people or entities such as households, business enterprises, facilities, farms or even geographical areas such as villages or towns. They allow in-depth understanding of socio-economic issues by studying relationships and interactions among phenomena.” (World Bank Data Help, 2020). With regards to the established literature on civil conflict and commodity prices, this is sorely needed. There is currently little consensus as to the actual effects and estimates. Many cross-country studies contradict each other, and smaller within-country studies struggle to be generalisable. In order to try and answer these problems, this dissertation attempts to explore if within-country analysis using micro-data can be replicated confidently and what can be learned from such an attempt by replicating *Commodity Price Shocks and Civil Conflict: Evidence from Colombia By Dube & Vargas*. To do this, the replication attempts to verify the original results by replicating the original code with the original dataset on R and attempts to repeat the study on an extended time series of Colombia using macro-data gathered from various sources, using frameworks and ideas drawn from the literature of replication in social sciences. This is done in order to ascertain if the original studies results were correct and if the methodology was robust. This dissertation finds that within-country replication is difficult, and that this attempt proved inconclusive across three distinct parameters of replication. Verification proved for the most part successful, however repeatability proved unsuccessful and generalisability nigh impossible. The lack of available micro-data proves difficult to surmount. However, there are lessons and insights to be gained from such an attempt. This dissertation starts with an in-depth literature review on *Commodity Price Shocks and Civil Conflict: Evidence from Colombia by Dube & Vargas*, the literature on civil conflict and commodity prices and the literature on replication. The paper then moves on to the replication sections, which include verifying the results, repeating them and discussing the lessons learnt. The dissertation concludes that replicating country-specific studies is difficult and results remain inconclusive unless the methodology is generalised and applied to different contexts, through the use of detailed micro-data which is incredibly difficult to find at the country level.

# II. Literature Review - The Replicated Study

Commodity Price Shocks and Civil Conflict: Evidence from Colombia[1]
By Dube & Vargas

## *Historical Context*

The civil war in Colombia started in the 1960s and differs from most civil conflicts as it has been fought on the grounds of political ideology, not on religious, regional or ethnic divisions (Human Rights Watch, 2005). It comprises of three distinct actors; the left-wing guerrillas (FARC and ELN mainly), the Colombian Government and right-wing paramilitaries (the majority of which organised into the AUC in the late 90s).
FARC and ELN fight to overthrow the government and represent the rural poor.
However, in later years their conflict became increasingly economically motivated. In 1996, FARC and ELN had a combined income of $800 million (Human Rights Watch, 2005). The conflict escalated in the 1990s as paramilitary groups emerged and narcotraffickers were defeated. These paramilitary groups were formed by wealthy landowners and organised into the AUC in 1997. Within these paramilitary groups, there have been cases of political involvement including clemency and policy concessions. With this in mind, predation and recruitment are important channels in which conflict responds to economic shocks. Through the AUC predating on government officials and guerrillas recruiting from rural workers. The historical context of this paper is useful as it provides an indication that the causal mechanisms have underlying roots in the causes of the civil conflict.

---

[1] Unless stated otherwise, all information is drawn from the paper: Dube, O. and Vargas, J.F., 2013. Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The review of economic studies*, *80*(4), pp.1384-1421.

## *Causal Mechanisms*

Dube and Vargas adopt a causal framework from Becker (1968) in which workers choose employment either criminal or productive based on whether criminal activity exceeds the wages they would get normally. This can be applied to Colombia's civil war context where conflict is conceptualised as a tool for a violent appropriation of resources, which applies in the case of Colombia as the armed groups all have appropriated resources one way or another. The authors go on to state that increases in commodity prices create contradictory pressures, it may increase appropriation by increasing economic production which is then siphoned. It also may raise wages by increasing commodity production and demand, therefore, increasing labour demand.

To test this, they use a within-country analysis with the assumption that various regions vary in production. They operationalise the opportunity cost mechanism by measuring wages and hours worked, for rapacity effect, they measure the production taxed by these armed groups. For oil production areas, they can measure the effects as there is national royalty rate that specifies the revenue received by each municipality based on the value of production. In summary, a rise in agricultural goods such as coffee should increase work hours in the productive sector and increase wages relative to contestable municipal revenue, which reduces conflict differentially in regions that produce these goods more intensively. Secondly, an increase in the price of natural resources such as oil should increase municipal income but not result in offsetting wage increases, thus increasing conflict differentially in the natural resource region.

## *Data and Methodology*

The data for the Colombian civil war is drawn from the *Conflict Analysis Resource Center,* the data is event-based with more than 21,000 war-related episodes in over 950 Colombian municipalities from 1988 to 2005. It draws from media reports in 25 major newspapers supplemented by reports from a network of Catholic priests who describe incidents of political violence across the country. These reported attacks and clashes form the basis of the dependent variables of the study. With regards to the independent variables, the data is drawn from a variety of sources detailing agricultural commodity production, most notably from the

CEDE (Centre for Study of Economic Development). For oil production, the main measure is the average barrels of crude oil produced per day in each municipality in 1988 (the beginning of sample period). There are 39 oil producing municipalities in the sample. For prices, they gather data from various international and national sources such as the International Coffee Organisation, International Financial Statistics and National Planning Department. Political kidnapping data is drawn from CEDE which originates from Observatory of Human Rights of the Vice-Presidency of Colombia, which is cross referenced with reports from the Department of Security (DAS).

The methodology Dube and Vargas employ follows a difference-in-differences estimator by assessing if changes in commodity prices affect violence disproportionately in municipalities that produce more of these commodities. Time variation stems from movements in annual prices such as international price of coffee and oil, as Colombia accounts for less than 1% of the world oil market and is therefore exogenous. On the other hand, Colombia is among top 10 largest coffee exporters globally and during the sample period ranked second which can lead to bias. Therefore, coffee is instrumented with the internal price of coffee against the export volume of three other leading coffee exporting nations: Brazil, Vietnam and Indonesia. They further instrument coffee intensity with rainfall and temperature to capture the latent coffee production capability of a municipality as the coffee intensity measure in 1997.

The specification is represented in two stages. As shown below (pages: 1397-1398):

Second stage:

$$y_{jrt} = \alpha_j + \beta_t + \delta_r t + \mathrm{Coca}_{jr} t\gamma + (\mathrm{Oil}_{jr} \times \mathrm{OP}_t)\lambda + (\widehat{\mathrm{Cof}_{jr} \times \mathrm{CP}_t})\rho + \mathbf{X}_{jrt}\phi + \varepsilon_{jrt},$$

First Stage:

$$\mathrm{Cof}_{jr} \times \mathrm{CP}_t = \alpha_j + \beta_t + \delta_r t + \mathrm{Coca}_{jr} t\gamma + \sum_{m=0}^{1}\sum_{n=0}^{1}\left(R_{jr}^{m} \times T_{jr}^{n} \times \mathrm{FE}_t\right)\theta_{mn} + \mathbf{X}_{jrt}\rho + \mu_{jrt},$$

The specification is estimated via one-step procedure through 2SLS estimation. Standard errors across all specifications are controlled at the department level to account for serial correlation over time and across municipalities within a department, of which the 978 municipalities are grouped into 32 departments.

## *Results*

The results in Table 2 illustrate that coffee price shocks have a negative relationship with conflict. When the price of coffee decreases, all four measures of violence increase differentially in municipalities that produce coffee more intensively. Table 2 further shows that oil price shocks exert the opposite effect on conflict, Column 2 indicates a rise in oil price increases paramilitary attacks differentially in areas that produce more oil. Comparing both the oil and coffee coefficients show that coffee has a bigger effect; a 10% fall in coffee results in 5% more paramilitary attacks in the average coffee producing region where a 10% rise in oil results in 1% more attacks in the average oil region. The instrumental variable for coffee results is shown to be strong with the Kleibergen-Paap F-statistic being equal to 15.94 which exceeds the Stock-Yogo critical value.

TABLE 2

*The effect of the coffee and oil shocks on violence*

| Dependent variables | (1) Guerrilla attacks | (2) Paramilitary attacks | (3) Clashes | (4) Casualties |
|---|---|---|---|---|
| Coffee int. x log coffee price | −0.611** | −0.160*** | −0.712*** | −1.828* |
| | (0.249) | (0.061) | (0.246) | (0.987) |
| Oil production x log oil price | 0.700 | 0.726*** | 0.304 | 1.526 |
| | (1.356) | (0.156) | (0.663) | (2.127) |
| Observations | 17,604 | 17,604 | 17,604 | 17,604 |

*Notes*: Standard errors clustered at the department level are shown in parentheses. Variables not shown include municipality fixed effects, year fixed effects, log of population, and linear trends by region and municipalities cultivating coca in 1994. The interaction of the internal coffee price with coffee intensity is instrumented by the interaction of the coffee export volume of Brazil, Vietnam, and Indonesia with rainfall, temperature, and the product of rainfall and temperature.
*** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level

Table 3 highlights the second stage of the difference-in-differences. It looks into the opportunity cost and rapacity channels. Oil price shocks significantly increase capital revenue

TABLE 3

*The opportunity cost and rapacity mechanisms*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Opportunity cost mechanism | | Rapacity mechanism | | |
| Dependent variables | Log wage | Log hours | Log capital revenue | Paramilitary political kidnappings | Guerrilla political kidnappings |
| Coffee int. x log coffee price | 0.371* | 0.286** | −0.787 | 0.022 | −0.060 |
| | (0.217) | (0.125) | (0.698) | (0.014) | (0.060) |
| Oil production x log oil price | 1.230 | 0.079 | 0.419** | 0.168*** | −0.066 |
| | (0.894) | (0.314) | (0.203) | (0.009) | (0.206) |
| Observations | 26,050 | 57,743 | 11,559 | 16,626 | 16,626 |
| Sample period | 1998–2005 | 1998–2005 | 1988–2005 | 1988–2004 | 1988–2004 |

*Notes*: Standard errors clustered at the department level are shown in parentheses. In column (1), the dependent variable is the log of hourly wage, defined as the the individuals' earnings in the past month divided by hours of employment in the past month. In column (2), log hours refers to hours of employment during the past month. Variables not shown in all specifications include municipality fixed effects, year fixed effects, and linear trends by region and municipalities cultivating coca in 1994. Columns (1) and (2) also control for education, age, age squared, and indicators of gender and marital status. Columns (3)–(5) additionally control for log population. The interaction of the internal coffee price with coffee intensity is instrumented by the interaction of the coffee export volume of Brazil, Vietnam, and Indonesia with rainfall, temperature, and the product of rainfall and temperature.
*** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level

of the local government. On average the oil price rise from 1998 to 2005 is estimated to have increased capital revenue by 5% in the average municipality. They find that as a result of this, there is a significant rise in paramilitary kidnappings of 7% over 1988-2005 in the average oil municipality, but no significant effect on guerrilla kidnappings. Overall, the results are consistent with the idea that different commodity shocks affect conflict via these channels.

## *Conclusion*

The paper documents that labour-intensive agricultural goods and natural resources affect political violence in opposite directions. These effects permeate across multiple agricultural goods and other natural resources they measured in the stud (the overall effects of price shocks and civil conflict) the authors argue, applies through the causal mechanisms of the opportunity cost and rapacity effect.

# III. Literature Review- Civil Wars and Replication

## *Civil War and Commodity Price Literature*

### *Commodity Price Volatility & Civil Conflict by Richard K. Morgan & Eric Reinhardt*

This paper by Richard K. Morgan and Eric Reinhardt explores the connection between commodity revenue and conflict by expanding upon the bargaining model of war (Fearon 1995). The authors argue that current literature (including the paper this dissertation attempts to replicate) provides many contradictory and null findings (Morgan & Reinhardt, 2014; 0). This is due to the existing literature neglecting the fundamental lesson of rationalist explanations for conflict, which is that observed advantages for one party in a bargaining process will produce more concessions but not more conflict. As such they develop several hypotheses; they are with export commodity price volatility, the degree of unpredictability of future price changes causes a commitment problem that should increase the probability of armed conflict; but neither commodity dependence by itself nor observed commodity price changes have any effect (Morgan & Reinhardt, 2014; 0).
To test these, they use a dataset on monthly data for 71 commodities and 145 non-OECD countries from 1963-2012, which also has indicators of exogenous variability and shocks to world prices along with measures of civil conflict at various intensity levels. They limit their analysis to developing countries as they generally have less diversified economies and thus more vulnerable to commodity price volatility. The dependent variable for their paper is civil conflict. They operationalise this through the use of UCDP/PRIO Armed Conflict Dataset’s listed definition of “internal" or internationalized internal" political or territorial conflicts pitting the government of a state against a formally organized opposition group, producing at least 25 battle-related deaths per year.” (Morgan & Reinhardt, 2014; 16). Since they are only

testing the onset of civil conflict, they exclude country-month measures where conflict is ongoing. For robustness, they include 3 different measures being Civil War, ACLED Conflict, and SCAD conflict.

They have several independent variables, with the foremost being price Volatility (Forecast) and the second most being price Volatility (Observed). Volatility (Observed), is the standard deviation of monthly first differences in the log real price index from t - 13 to t - 1 (t being the measure for 'month'): in other words, commodity price variability over the prior 12 months, lagged by one month (Morgan & Reinhardt, 2014; 18). volatility (Forecast) is measured by the standard deviation (or square root) corresponding to the first-order GARCH variance estimate, to match the scale of volatility (Observed). With these and other control variables, they conduct a multivariate regression to explain the onset of civil disorder using volatility forecasts.

They find volatility and shock to be largely robust and across all 6 models, they find volatility statistically significant to a two-tailed p-value of $<0.05$. Therefore, higher levels of volatility increase the likelihood of conflict onset (Morgan & Reinhardt, 2014; 24). Rather interestingly, they infer that the findings from Dube and Vargas for Colombia do not account for variation across the other OECD countries. Furthermore, in this study that they conduct they sort all 71 commodities into both lootable natural resources) and non-lootable (agricultural) goods and they find that the lootability of commodities has no statistical significance against the onset of civil conflict. This paper contributes to the literature by making the connection between the Bargaining Model of War and Commodity price volatility. The paper further contributes to the literature as it includes a host of countries with a longer time frame with more detail than previous studies by measuring monthly instead of yearly.

However, there are a few limitations, most notably that it only focuses on the onset of conflict. The paper does not cover the full scope of duration, intensity, etc so their findings are still limited. Another significant limitation is that as their study revolves around the uncertainty of income creating the onset of conflict and how price volatility impacts this, they don’t measure whether these commodities provide revenue for the government or rebel groups and how much revenue they gain from these commodities. By excluding this measure, they fail to evaluate how revenue size from these commodities affects the onset of conflict, which certainly is a confounding variable.

## *Concession Stands: How Mining Investments Incite Protest in Africa by Darin Christensen*

This paper by Christensen is more focused in its attempts to explain conflict. By limiting the scope of his paper to mineral resources and conflict in Africa. The paper asks the following questions: do commercial mining investments increase the likelihood of social or armed conflict? If so, when are these disputes most prevalent? And lastly, what mechanisms help explain these conflicts? (Christensen, 2019; 65). To measure this, Christensen employs geospatial, time-series data with a difference-in-differences research design (Christensen, 2019; 71). The mining data is drawn from three separate repositories of data to geo-locate commercial mining projects and determine start years. The paper further employs several conflict data sets to geo-locate protests, riots, and other low-level social conflicts (Christensen, 2019; 71). Using this methodology, Christensen finds that in sub-Saharan Africa, the probability of protests more than doubles with mining and that areas receiving investments do not have differential trends prior to the introduction of mining (Christensen, 2019; 96). Christensen contributes to the literature by reaffirming the link between mineral resources and conflict, through the mechanism of mineral price "supercycles" which lead to heightened and unmet expectations from neighbouring communities and their development dividend from the commodity boom. Which in turn leads to increased protests as price increases (Christensen, 2019; 97).

The paper also augments the literature by showing that, unlike most literature on the resource curse, areas that host commercial mining projects are largely immune from rebel attacks or deadly armed conflict in the vicinity of the mine and surrounding areas. Despite the likelihood of protests and riots increasing in these areas. However, it suffers from some limitations. Despite controlling for year, country, and candidacy in mining regulatory bodies, etc, the analysis struggles to control for unmeasured, individual country-specific reforms with regards to transparency and oversight. Not controlling for these heterogeneous effects potentially leads to omitted variable bias, these confounders could quite significantly affect the results of the paper. On the other hand, there are many strengths to this paper. It combines a qualitatively backed up causal mechanism to explain the quantitative results found. There is a clear causal path described in the study backed up by the robust geospatial data.

## *Economic Shocks & Conflict: Evidence from Commodity Prices by Samuel Bazzi & Christopher Blattman*

This paper seeks to answer the question: “Higher national incomes are correlated with political stability. Is this relationship causal?” (Bazzi and Blattman, 2014; 1). The authors do this by testing three distinct theories that link conflict to export price shocks. The first being that poverty lowers the opportunity cost of insurrection. The second is the theory that states can avert war when they have the revenue to suppress insurgency or buy off opposition. The third theory is that states are prizes that can be seized when power constraining institutions are weak, thus as the ‘value of the prize’ increases, insurrections and coup increase accordingly. To research this, Bazzi and Blattman examine price shocks from 65 globally traded commodities whilst looking at developing countries from 1957 to 2007 (Bazzi and Blattman, 2014; 3). They find that there is little support for the theory of the state as a ‘prize’. Even amongst those developing countries that are fragile, unconstrained, and dominated by extractive commodity revenue. Instead, they find the opposite correlation. They conclude that this reason doesn’t work on a systemic level. With regards to the theory that states can avert war with revenue increases from positive price shocks, they find some evidence that this is the case. This potentially the effects of this theory cancel out the state as a prize theory, however, they do not delve into this explanation. They further find the inverse relationship between prices and war intensity is consistent with opportunity cost accounts, but not exclusively so (Bazzi and Blattman, 2014; 34). To conclude their paper, they suggest that more quantitative country case studies, such as conducted by Dube and Vargas on Colombia, are needed alongside the development of cross-national data. They elaborate that the current literature faces issues in the failure to report insignificant results and as such the current literature runs the high risk of publication bias, especially when papers fail to preregister hypotheses and methods.

The contribution this paper provides to the literature is highlighting the need for future research into country-specific case studies, alongside the development of cross-national studies, in order to test competing theories and not simply confirm evidence of existing models and predictions.

## *Sensitivity Analysis of Empirical Results on Civil War Onset by Hegre and Sambanis*

This paper by Hegre and Sambanis attempts to apply a methodology for organised specification tests to check the robustness of the empirical results for literature on civil war. This is done because they state that the literature of civil war is fraught with empirical results that are not replicable (Hegre and Sambanis, 2006; 508). Therefore, the results fail to be robust. Furthermore, they argue that various studies use different definitions of civil war and analyse varying time periods, which leaves replicators and readers unable to ascertain if differences in empirical results are due to the differing definitions surrounding civil conflict and time periods measured by the researchers, or if the results are simply not robust. To overcome these challenges, they use the same definition of civil war and analyse the same time period.

To evaluate the robustness of the models they evaluate, Hegre and Sambanis implement a 'Sensitivity Analysis' approach in their methodology based on Sala-i-Martin's (1992) test with some variations to account for the intricacies and differences in civil war methodology. This sensitivity analysis method tests a limited number of independent variables on each dependent variable and measures their upper and lower confidence bounds whilst converting some variables to log measures to nullify the effects of extreme observations within these variables. However, the methodology suffers some drawbacks, much of the data on civil war is filled with missing observations and thus the sample sizes vary across models. Furthermore, some important variable effects are not expressed in linear terms, e.g. the hypothesised inverted U-shape association between civil war risk and democracy level and the parabolic association between civil war risk and primary commodity export dependence (Hegre and Sambanis, 2006; 513). Capturing the true effect of these variables is difficult under the sensitivity analysis methodology. Especially when these may directly influence numerous other variables tested in the model.

They conclude that, whilst a range of economic, political and commodity variables pass their tests, many do not. They stress that model specification needs to take into account the fragility of inference without proper robustness checks, such as control variables, large sample sizes, standardised measurements for repeatability and measure across studies and the consideration that different estimation methods produce different results across studies

(Hegre and Sambanis, 2006; 532). Despite this, they stress that this sensitivity analysis does not replace careful theorising. Which is needed because, as the authors state, there is no single model for civil war. Therefore, finding consensus on robust results must be taken into account when theorising and planning what methodology must be undertaken.

## *Overview of the Literature*

The literature on price shocks and civil conflict is an area with a lack of consensus, with regards on how to create a study that provides clear insights on the relationship between the two variables and the results they garner from research into the topic. Whilst they find consensus in the idea that price shocks do have a relationship with civil conflict, consensus ends there, the salient causal mechanisms vary study to study and country to country. Papers vary from broader, cross-national meta-analysis such as the study by Morgan and Reinhardt to more concise, literature by Christensen and Dube & Vargas. Whilst cross-national studies prove valuable as their results are perhaps more generalisable, they fail to take into account country-specific effects such as cultural, religious and demographic factors more easily controlled by country-specific studies. On the other hand, country-specific studies lack the generalisability of these large N studies. Whilst their measured effects may prove salient in their country of choice, extrapolating such results to further cases proves difficult unless replications are conducted using the same methodology on these countries. Overall, the literature lacks one formal 'model' of price shocks and civil conflict which arise from all the differences above.

## *Replication in Social Science Literature*

## *Why Most Published Research Findings Are False by John P.A. Ioannidis*

In this essay by Ioannidis, he illustrates how the vast majority of research findings are false. He further states that the probability of a research claim being true depends on study power, bias, number of studies on the same question and the relationship of true to no relationships among the relationships probed in each scientific field (Ioannidis, 2005; 696) amongst many other factors. Ioannidis argues that with the introduction of bias, the positive predicted value (PPV) calculated in Bayes' theorem depresses the PPV. Even when studies meet the requirements for the type I and II error rate, there is still a 36% probability that a study's results reporting a positive result are incorrect (Ionnadis, 2005). Ioannidis argues that research designs must take into account pre-study odds. To ascertain the chances, they are testing a true rather than non-true relationship (Ioannidis, 2005; 701). This approach, alongside large N studies and low-bias meta-analyses across the literature, can help mitigate the chances that studies fall foul of low statistical power and pitfalls in hypothesis creation.

Ioannidis contributes to the literature surrounding replication and causal inference by highlighting the common issues most studies encounter and illustrating the baseline probability of published findings being wrong. Through this, Ioannidis identifies several factors that can influence the research such as;

- the smaller the studies conducted in a scientific field, the less likely the research findings are to be true
- the smaller the effect sizes in a scientific field, the less likely the research findings are to be true, the greater the number and the lesser the selection of tested relationships in a scientific field
- the less likely the research findings are to be true
- the greater the flexibility in designs, definitions, outcomes, and analytical models in a scientific field, the less likely the research findings are to be true
- the greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true
- the 'hotter' a scientific field (with more scientific teams involved), the less likely the research findings are to be true (Ioannidis, 2005).

This paper has had a huge impact on social science methodology by acknowledging the 'replication crisis' and the need for further robustness checks to ascertain causal inferences.

## *Replication in Social Science by Jeremy Freese & David Peterson*

This paper seeks to address the challenges of the 'replication crisis' within the field of sociology. Whilst this doesn't strictly fall under the bounds of international relations, economics and politics, the underlying message on replication within the realm of social science is still important. The first half of the paper talks about the place of replication within sociological literature and culture whilst the second half (the salient part of this paper) discusses the dimensions of replication within 'quantitative' methodology (Freese and Peterson, 2017; 147). They represent quantitative replication as having four distinct goals: verifiability, robustness, repeatability and generalisability. As shown in the figure below.

| | SIMILAR | DIFFERENT |
|---|---|---|
| OLD DATA | Verifiability | Robustness |
| NEW DATA | Repeatability | Generalization |

To verify results, the replication must follow the methodology of the paper, using the same code and the same data. To ascertain if further data or information is needed to gain the same results. Freese and Peterson argue this requires open data repositories with detailed methodological appendices underlying where and when the data was accessed. For robustness checks, the authors suggest that sensitivity analyses, robustness checks and preregistration of studies can all allay the concerns of 'p-hacking', which corroborates with much of the conclusions drawn from Hegre and Sambanis. With regards to repeatability, the authors discuss the need for taking existing methodological approaches and applying them to new sets of data, with careful consideration that this data properly operationalises the variables in the original methodology. This could be as simple as applying the methodology to another country case study, extending time-series panel data and older data. Generalisability refers to how applicable results are in other settings (which is simply just external validity). The authors question whether explanations that are revised to accommodate new results are salient or if they simply question the validity of the study they are generalising.

This paper contributes to the literature by outlining four key topics that should be covered under a strong and coherent replication. By outlining these and discussing what each contributes in terms of replication, Freese and Peterson provide a basic framework under which quantitative replication can be conducted.

## *The Replication Recipe: What makes for a convincing replication? By Brandt et al.*

Brandt et al. discuss in this paper that there is no real consensus on the components of a convincing close replication study. As such, they have developed a 'Replication Recipe' which outlines standardised criteria for Replication. They discuss further the evaluation of replication results and their limitations. Brandt et al. describe a 'close replication' where the methods and procedures are as close as possible to the original study, mindful of potential differences in coding, syntax, programs etc between the replication and the original. Especially for quantitative methodology. The authors provide a 36-question guide to their 'Replication Recipe' (Brandt et al, 2014; 219). They suggest keeping replications simple, well-constructed and as close as possible whilst keeping track of differences between both studies, check study's assumptions in a new context and evaluate/report results as ethically as possible and pre-registering any replications.

The authors suggest four key ingredients for replication such as; Carefully defining the effects and methods that the researcher intends to replicate, following exactly the methods of the original study, having high statistical power and making complete details about the replication available. This paper by Brandt et al has huge implications for the literature surrounding replication. By expanding upon ideas similar to Freese and Peterson in their paper, Brandt et al. create a guideline for robust replication that works across all social sciences. Through this framework, researchers have a foundation on which to construct any replication going forward based on good methodological practice and science.

## *Overview of the Literature*

The literature on replication is still fairly new, however, it has huge implications on social science. All the papers evaluated reach consensus: more stringent and ordered replication is needed to confirm causal inference, especially with Ioannidis illustrating how even good studies have a large probability of being false. By outlining specific methods and concepts replicators need to take into account outlined by Brandt et al. and Freese & Peterson, a 'gold standard' of social science replication can begin to be drawn up. Whilst this needs further refining and developing, the literature remains in consensus over the solutions to the 'replication crisis'. The replication conducted in this dissertation draws from this literature the framework given by Brandt et al. through answering some of the questions within the 'replication recipe'. It further draws on the literature of Freese and Peterson in attempting to reach the four 'goals' of replication; verifiability, repeatability, robustness and generalisation.

# IV. Verifiability of Original Results

This section of the dissertation focuses on the replication of the original paper, following as faithfully as possible the 'Replication Recipe' by Brandt et al. (2014) As such, Tables 2 and 3 are replicated on R Studio, bearing in mind the original study's syntax was coded on STATA. In order to conduct this replication, the code was created from scratch. The purpose of this section is to verify the methodology employed by Dube and Vargas in their paper. The main models, Tables 2 and 3, are replicated and tested. The effects being scrutinised are the effects of coffee and oil price shocks on civil conflict for Table 2 and for Table 3 the rapacity and opportunity cost mechanisms through which this effect is suggested to take place in the original study (Dube & Vargas, 2013). Replicating this effect is important to analyse if the results are verified and remain significant when the methodology is constructed outside of the original syntax on STATA. It is acknowledged that differences between STATA and R can account for some variation, however much of the methodology remains the same and faithful to the original study. Whilst some findings were found to be verifiable, overall there remains some uncertainty as to the model.
All original details were available from the author including code, data and appendices. These are the tools employed in this replication.

## *Methodology employed*

To begin this replication, Table 1 was replicated. Replicated Table 1 is identical to Table 1 given in the original study down to every observation, mean, median, standard deviation, min and max. This illustrates that the data is complete and should yield the same results based on the correctness of the data.[2]

Next replicated were the models employed in Table 2 of the original study. This is the second stage of the 2SLS model they used. In their STATA code Dube and Vargas use the command 'xtivreg2', this command utilises a generalised method of moments model, or 'GMM' for short. This method is very similar to maximum likelihood estimation. GMM makes assumptions about specific moments in the random variables as opposed to the entire distribution which makes it more robust. They further partial out variables in which they make small-sample adjustments. This model has municipality fixed effects, year fixed effects, log of population, linear trends by region and municipalities cultivating coca in 1994 and the interaction of internal coffee price with coffee intensity instrumented by interaction of coffee export volume in Brazil, Vietnam and Indonesia with rainfall and temperature (to ensure there is no endogeneity problem). The results are shown below. Coffee is significant in all four models with the same direction of effect. It clearly illustrates that as coffee prices increase, guerrilla attacks, paramilitary attacks, clashes and casualties decrease, with the most notable effect on casualties. Oil production is shown to be only significant on paramilitary attacks with little support for it affecting clashes or casualties.

TABLE 2

*The effect of the coffee and oil shocks on violence*

| Dependent variables | (1) Guerrilla attacks | (2) Paramilitary attacks | (3) Clashes | (4) Casualties |
|---|---|---|---|---|
| Coffee int. x log coffee price | −0.611** | −0.160*** | −0.712*** | −1.828* |
| | (0.249) | (0.061) | (0.246) | (0.987) |
| Oil production x log oil price | 0.700 | 0.726*** | 0.304 | 1.526 |
| | (1.356) | (0.156) | (0.663) | (2.127) |
| Observations | 17,604 | 17,604 | 17,604 | 17,604 |

The overall implications of this table are that coffee and oil affect conflict in opposite ways. Through the partialling out and the standard error clustering by department, standard errors remain very low throughout the majority of the models. This is helped by the large number of observations employed within the model, which helps account for sampling bias, especially as

[2] Please see the Appendix for the comparison between the tables

these observations were drawn from the length and breadth of Colombia. However, it is interesting to note the partialling out done by this regression model, which allows for better identification in the GMM. This may lead to some variation by the partialled out region and year effects being ignored and not accounted for within the table, which can lead to bias. particularly if standard errors and correlations are altered as a result. In order to check this, the regression models and tables were replicated.

The Replicated Results of Table 2 models were estimated in R using the PLM command and running a 2SLS like indicated in the original methodology. All other models in the table follow the same methodology. The PLM employed only measures the second stage of the 2SLS as all control variables and the region-year interactions are found on both sides of the 2SLS. The model also accounts for municipality and year effects. By using the within model, the coefficients and intercepts are measured by the departmental clustered, thus giving a better calculation for standard errors.

In order to extract these standard errors, the VcovHC command is used to estimate the heteroskedasticity-consistent covariance estimators. The type of variance-covariance used was 'HC2' as, whilst there are 17604 observations, the sample size is still relatively small and only within-country. By finding the square root of this variance-covariance matrix and cutting the data diagonally, the standard errors for the different models were extracted and implemented. The above code was run for all models, originally, p-values were calculated as well. However, these were not interpretable by Stargazer and therefore the original p-values were used, these are not believed to alter the actual significance of these results as standard errors and coefficient estimates remain the same.

Table 3 was then replicated; this focuses on the opportunity cost and rapacity mechanism. This proved to be far more difficult to replicate and code in R. The original code the authors used in STATA used an 'ivreg2' with interactions between year, origmun and region. When trying to compute these with the plm command as done before, there were multiple syntax errors with regards to the duplication of 'id' and 'time'. Therefore, the model below was employed with the 'REndo' package. The standard errors were extracted the same way as Table 2 for the guerrilla and paramilitary models.

```
### Table 3 Regression: Opportunity Cost and Rapacity Mechanism ###

###Log Wage Regression With Robust SE's###

table3lwreg <- multilevelIV(lwage ~
                            cofintxlinternalp+ oilprod88xlop+gender+age+agesq+married+edyrs+coca94indxyear +
                            as.factor(year)+as.factor(region) + rxltop3cof+txltop3cof +
                            (1|origmun) |   endo(rxltop3cof,txltop3cof),data=wages_commodities)
```

This method was used for Log wage, Log hours and Log Capital Revenue dependent variables, as issues with how the dataset included dummies for origmun, region and time affected the implementation of PLM. This mixed effect model circumvented those issues with the data. However, when attempted on paramilitary and guerrilla Kidnapping variables, this error occurred:

```
Fitting linear mixed-effects model.
Detected multilevel model with 2 levels.
For origmun (Level 2), 978 groups were found.
Error: Matrices must have same dimensions in diag2Tsmart(e1, e2, "d") - e2
```

This error illustrated how the data matrices used to in the paramilitary and guerrilla kidnapping models do not have the same dimensions and therefore the estimates cannot be calculated. Interestingly, when looking at the authors original code, for log wage, log hours and log capital revenue, ‘ivreg2’ is used. For paramilitary and guerrilla kidnappings, the ‘xtivreg2’ command (as used throughout table 2) was implemented. As a result, PLM was used as shown below:

```
###Paramilitary Kidnappings Regression with Robust SE's###

table3parkidreg <- plm(parkidpol ~ cofintxlinternalp+oilprod88xlop +lpop+as.factor(region)*as.factor(year)+coca94indxyear |
                        rxltop3cof*txltop3cof  + lpop +as.factor(region)*as.factor(year),
                       index=c('origmun','year'), data=violence_commodities, model='within')
```

This model, which as shown above can reach very similar results to the original study, should be able to do the same with the paramilitary and guerrilla models. This circumvents the issues ‘multilevelIV’ has with its estimations and data matrix problems and in turn becomes a more accurate estimation for these models.

### *Interpretation of Verified Results*

Using the above method, Table 2 was replicated as shown below. Alongside it is the original Table 2 for comparison.

**Replicated Results of Table 2**

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Guerrilla Attacks | Paramilitary Attacks | Casualties | Clashes |
| | (1) | (2) | (3) | (4) |
| Coffee int. x log coffee price | -0.644*** | -0.183*** | -1.780*** | -0.687*** |
| | (0.111) | (0.031) | (0.665) | (0.137) |
| Oil production x log oil price | 0.526 | 0.703*** | 2.573 | 0.183 |
| | (1.351) | (0.148) | (2.051) | (0.601) |
| Observations | 17,604 | 17,604 | 17,604 | 17,604 |
| $R^2$ | 0.027 | 0.033 | 0.027 | 0.037 |
| Adjusted $R^2$ | -0.035 | -0.028 | -0.034 | -0.024 |
| F Statistic | 614.294*** | 743.386*** | 504.332*** | 749.581*** |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

TABLE 2
*The effect of the coffee and oil shocks on violence*

| Dependent variables | (1) Guerrilla attacks | (2) Paramilitary attacks | (3) Clashes | (4) Casualties |
|---|---|---|---|---|
| Coffee int. x log coffee price | −0.611** | −0.160*** | −0.712*** | −1.828* |
| | (0.249) | (0.061) | (0.246) | (0.987) |
| Oil production x log oil price | 0.700 | 0.726*** | 0.304 | 1.526 |
| | (1.356) | (0.156) | (0.663) | (2.127) |
| Observations | 17,604 | 17,604 | 17,604 | 17,604 |

What the Replication managed were results incredibly similar to that of the original study. Near identical measured effects throughout with (in some cases) higher significance and better standard errors. For example, in coffee price and guerrilla attacks, the coefficient is a stronger negative correlation of -0,644 with a higher significance and 0.100 smaller standard error. The great similarities between results show a methodology that can be verified using the available data. As no partialling out was done in the PLM this may indicate that it had a negligible effect on the results and the methodology remains verifiable and more importantly repeatable for Table 2, even on different statistics programs with slight differences in how they calculate models. Furthermore, the F-statistic on the Replicated Table 2 show that the independent variables add a significant amount of ‘weight’ to the model and act as sufficient explanatory variables.

Table 3 measures the effect of the opportunity cost and rapacity mechanisms through which price shocks affect civil conflict, the original results of the study and the replicated results for paramilitary and guerrilla Kidnappings are shown below:

TABLE 3
*The opportunity cost and rapacity mechanisms*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Opportunity cost mechanism | | Rapacity mechanism | | |
| Dependent variables | Log wage | Log hours | Log capital revenue | Paramilitary political kidnappings | Guerrilla political kidnappings |
| Coffee int. x log coffee price | 0.371* | 0.286** | –0.787 | 0.022 | –0.060 |
| | (0.217) | (0.125) | (0.698) | (0.014) | (0.060) |
| Oil production x log oil price | 1.230 | 0.079 | 0.419** | 0.168*** | –0.066 |
| | (0.894) | (0.314) | (0.203) | (0.009) | (0.206) |
| Observations | 26,050 | 57,743 | 11,559 | 16,626 | 16,626 |
| Sample period | 1998–2005 | 1998–2005 | 1988–2005 | 1988–2004 | 1988–2004 |

*Notes*: Standard errors clustered at the department level are shown in parentheses. In column (1), the dependent variable is the log of hourly wage, defined as the the individuals' earnings in the past month divided by hours of employment in the past month. In column (2), log hours refers to hours of employment during the past month. Variables not shown in all specifications include municipality fixed effects, year fixed effects, and linear trends by region and municipalities cultivating coca in 1994. Columns (1) and (2) also control for education, age, age squared, and indicators of gender and marital status. Columns (3)–(5) additionally control for log population. The interaction of the internal coffee price with coffee intensity is instrumented by the interaction of the coffee export volume of Brazil, Vietnam, and Indonesia with rainfall, temperature, and the product of rainfall and temperature.
*** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level

**Replicated Table 3 Paramilitary and Guerrilla Kidnappings**

| | *Dependent variable:* | |
|---|---|---|
| | Paramilitary Political Kidnappings | Guerrilla Political Kidnappings |
| | (1) | (2) |
| Coffee int. x log coffee price | 0.0005 | 0.004 |
| | (0.005) | (0.008) |
| Oil production x log oil price | 0.027* | 0.027* |
| | (0.015) | (0.014) |
| Constant | 0.150*** | 0.041** |
| | (0.020) | (0.018) |
| Observations | 16,626 | 16,626 |
| $R^2$ | 0.010 | -0.065 |
| Adjusted $R^2$ | 0.010 | -0.065 |
| Residual Std. Error (df = 16621) | 0.261 | 0.297 |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

What the results show is variation between the original and replicated Table, suggesting the opposite effect of price shocks on the rapacity mechanism suggested by the original study. The only significant finding was the effect of oil price on political kidnappings which appears actually more significant than the original results, alongside a stronger measured effect and smaller standard error. However, for the guerrilla political kidnappings and the effect of coffee on paramilitary political kidnappings, standard errors were larger, this possibly affects the coefficient estimate as the intercept tries to adjust for this increased heteroskedasticity. This may be explained by the implementation of the variance-covariance matrix using 'HC2' which may not have been able to be as accurate with its estimations on the clustering of departmental and municipal variables.

Table 3- Log Wage

```
Number of levels: 2
Number of observations: 26050
Number of groups: L2(origmun): 246

Coefficients for model REF:
                  Estimate Std. Error z-score Pr(>|z|)
(Intercept)      6.405e+00  6.807e-02  94.095  < 2e-16 ***
cofintxlinternalp 1.527e-03  9.540e-03   0.160  0.87280
oilprod88xlop    1.965e-01  1.465e-01   1.342  0.17971
```

Table 3- Log Hours

```
Number of levels: 2
Number of observations: 57743
Number of groups: L2(origmun): 246

Coefficients for model REF:
                  Estimate Std. Error z-score Pr(>|z|)
(Intercept)      4.643e+00  3.782e-02 122.754  < 2e-16 ***
cofintxlinternalp 1.664e-02  5.260e-03   3.163 0.001559 **
oilprod88xlop    3.850e-02  8.963e-02   0.430 0.667515
```

Table 3- Log Capital revenue

```
Number of levels: 2
Number of observations: 17964
Number of groups: L2(origmun): 963

Coefficients for model REF:
                   Estimate Std. Error z-score Pr(>|z|)
(Intercept)       7.271e+00  1.837e-01  39.589  < 2e-16 ***
cofintxlinternalp -8.392e-02  2.028e-02  -4.138 3.50e-05 ***
oilprod88xlop     1.580e-01  8.547e-02   1.848  0.06455 .

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The above tables show the Opportunity cost mechanism within Table 3, Log wage, Log Hours and Log capital revenue are all shown. From the results and comparing between the two tables, coffee is only significant in both tables in the Log Hours model where it is significant to a 0.001 level, despite the large standard errors. Oil was found to be significant in the Log Capital revenue model to a 0.05 level, which is identical to the original model. Some variation between the original model and the replication model can be accounted for by the different syntax used and the multi-level GMM employed. In the end, however, differences between the replicated table and the original table cannot be fully explained by the syntax alone, therefore any attempts at assuring reliability .

## *Analysis*

The purpose of this section of replication is to evaluate how verifiable the models are with the original data. One way of doing this is to simply load STATA and run their code exactly, to see if the same results are found. However, doing this simply confirms that their syntax is correct and fails to understand the methodology employed to reach these results. The method undertaken in this section, of recreating the models on R, has had the effect of breaking down and understanding the models and checking verifiability by implementing the same methodology (albeit different syntax) and seeing if the same results are met. This is a far more stringent exercise to evaluate verifiability. However, there are issues with this method.

Starting with coding problems, the different datasets provided, and the way dummy variables were created meant that for Table 3, the opportunity cost variables could not be coded with the PLM command such as in Table 2 and the rest of Table 3. Variations between PLM and multilevelIV commands and how they calculate effects (especially considering multilevelIV uses GMM like the xtivreg2 command used by the authors in STATA) could have heavily influenced the results. This impacts verifiability to a strong degree because any conclusion about if the model was followed correctly and that this is the reason there is variation is cast into doubt. By having different models because the datasets are coded in a way you cannot use one model throughout, could potentially impact the estimators. The process of replication needs to be absolutely thorough and follow exactly the method of estimation shown in the original code.

Another issue that arises when doing this verifiability method is that there is no real instruction on how to conduct it. This is a fault with the replication package itself and the study. While the study illustrates the specification used to create the models and the equations on which the regression is based, there is very little framework on which to build the replication, at least on another statistics system like R. There is no mention of the partialling-out of region and year effects within the paper nor any explanation of why this was done, nor was there an explanation of their code, no comments on why GMM was used and how this affects the results. This made replicating the code on R difficult and has left open the possibility of bias. Whilst the models were constructed faithfully to the original syntax, they went through numerous iterations which varied widely results wise compared to the original paper.

Which begs the question, are the results biased because they were altered until they reproduced as accurately as possible to the original paper? Due to of the lack of comments and lack of clarity as to the exact decisions the authors made on code, there is no way to cross reference and check why certain models varied in the way they did, nor is there an exact way to confirm the models finally settled on for this dissertation are in fact as close as possible to the original models of the paper. It could be equally possible that the results reached by replicating the tables on R are simply the result of another methodology that could be heavily biased by trying to 'prove' results. Overall this paper can only conclude, despite the accuracy of Table 2 and some estimates in Table 3, that it is inconclusive as to whether the models are verifiable to the extent that the models were replicated properly, and results are accurate.

# V. Repeatability by extending the time series

This part of the replication focuses on the repeatability of the methodology by using new data drawn from a variety of sources to as closely resemble the variables in the original dataset. This new data was drawn from 2006-2018 and contains 377 observations of 97 variables, including all the main variables in Table 2 of the original study. For the purpose of this section, only Table 2 of the main models is repeated with the extended time series. There are a number of reasons for this including lack of available data, lack of specification in how some variables were calculated uncertainty on how to measure and code key variables in the other models. However, since Table 2 is the main model to illustrate that commodity price shocks do affect civil conflict, there is still something to be gleaned from conducting this limited repeatability. All tables, code and data can be found in the replication package and appendix. These are the tools employed in this replication.

## *Methodology Employed*

The time series was extended to cover 2006-2018. However, some variables could not be drawn from these dates. Alongside issues with data gathering. The dataset used in the original study by the authors (CERAC dataset) was, firstly, not accessible from their website and secondly has not been updated post 2002 (CERAC, 2020). Therefore, all conflict data and locations had to be drawn from the Uppsala Geocoded Conflict Dataset (Högbladh Stina, 2019). The locations of these had to be specific to the municipality and department level in terms of precision (lower than what was coded as '3' in the Uppsala Dataset). Because of this, 57 entries had to be deleted. The list of these municipalities was then cross referenced with the municipality list found on the CERAC website which had the Origmun code used in the replication and original study alongside the name of the municipality. However, this list proved to be somewhat different from the Uppsala dataset, with some municipalities missing on either end. Because of this, more entries had to be deleted. Each municipality then had its guerrilla attacks, paramilitary attacks, clashes and casualties summed for each municipality in each year of the time series. Unfortunately, the Uppsala dataset did not code for clashes in which the other fighting side (if a military force) was unable to fight back (which in the

CERAC dataset would count as an ‘attack’ rather than a clash), this led to some loss in the accuracy of the counts as they had to be classed as clashes.

Population data for the log population variable also proved tricky to source. Census data for Colombia was taken from 2018 from the department for statistics in Colombia (DANE, 2018), at the end of the time series. This is due to the fact census data for other years proved more difficult to find and varied in terms of accuracy. More data entries were lost because the list of municipalities given in the Census data was lower than what was given by CERAC and Uppsala. This meant a further loss of detail and potentially influential observations.

Departments given by Uppsala, CERAC and DANE also were not coded into region. As a result, this was done manually using the specification given by the original CERAC dataset (Restrepo et al., 2004).

Peso price in comparison to the US dollar was taken from the 2018 average (Netcials, 2020) at the end of the dataset, this was used to calculate international oil price.

Oil production measures and updated coffee census proved a huge issue in finding accurate data. Accurate coffee census data could not be found past the 1997 Census by the NFCG which the original study used. Furthermore, updated oil production data past the 1988 survey from the MME could not be found. Therefore, the data was drawn from the original dataset and cross-referenced with those municipalities already coded for conflict data. This was done under the assumption that oil reserves (whilst they deplete) stay fixed to those municipalities and that coffee production, which requires very specific elevation, temperature and rainfall to grow would not alter significantly or move from those municipalities in which they grow. However, there were a number of curious N/A’s in the dataset where municipalities given in the list by CERAC were not found in the original data. Therefore, these observations had to be removed. All coffee measurements were taken on 125kg bag per peso alongside the top 3 exporters (Federación Nacional de Cafeteros, 2020) (ICO, 2020).

The data was then collated all into one dataset consisting of 377 observations of 97 variables drawn from 2006-2018. This can be found in the dataset ‘colombiadatanew’ supplied in the replication package.

The methodology employed is exactly the same specification from the verified Table 2 in the verification section of this dissertation. It is a Panel model using the PLM command which only measures the 2$^{nd}$ stage of the 2SLS it runs, whilst accounting for clustered standard errors using a variance-covariance matrix created through ‘HC2’ specification, municipality and year fixed effects alongside the log of population and linear trends by region and municipality cultivating Coca in 1994. The instruments for coffee are also included with export volume of

Brazil, Vietnam and Indonesia rainfall, temperature and the product of rainfall and temperature.

## *Interpretation of extended time series results*

Using this methodology, the following table was created.

**Results of Table 2 with extended time series**

| | *Dependent variable:* | | |
|---|---|---|---|
| | Guerrilla Attacks | Casualties | Clashes |
| | (1) | (2) | (3) |
| Coffee int. x log coffee price | 1.824 | -3.451 | 0.156 |
| | (1.309) | (13.108) | (1.552) |
| Oil production x log oil price | -0.0003 | 0.0003 | 0.0002 |
| | (0.0002) | (0.005) | (0.001) |
| Observations | 377 | 377 | 377 |
| $R^2$ | 0.098 | 0.294 | 0.243 |
| Adjusted $R^2$ | -1.387 | -0.870 | -1.005 |
| F Statistic | 31.197 | 54.090 | 45.523 |
| *Note:* | | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |
| | Paramilitary attacks model removed from regression due as there is only one observation in time series | | |

This Table illustrates, with the sample in the dataset, no significance (even at the 10% level) at all across any variables and models. This suggests that according to this sample, we cannot reject the null hypothesis that coffee and oil price shocks have no effect on the dependant variables. We can also see, despite the robust standard errors extracted from the variance-covariance matrix of these models, there are still large standard errors across all measures of coffee and the dependent variables. Oil production, whilst having very small standard errors, had very small coefficients and remained insignificant. However, most importantly of all, the paramilitary Attack model could not be estimated as throughout the entire extended time series there was only one recorded paramilitary attack. The F-statistic of the independent variables also show no significance, which illustrates that these models are not explanatory of the variation in the dependent variables.

## *Analysis*

Taking into consideration the results of the regression, the data gathering, and methodology used, the attempt to repeat the data with an extended time series failed. There are two highly likely reasons for this that link to each other, firstly, that the sample size of the replication was far too small, with only 377 compared to the 17,964 of the original study and secondly there are other confounding variables that influence the conflict outcome independent of coffee and oil production and which cannot or were not controlled for in the methodology taken from the original study. To address the sample size, this is a critical factor in any research. Such a small number for a within-country analysis significantly increases the chance of selection bias. Especially because other datasets such as Uppsala code differently from how the CERAC dataset coded conflict and show systematically lower numbers (Restrepo et al, 2004; 106), this could significantly impact the amount of observations. Furthermore, the amount of observations was also affected by the data gathering process. Because the CERAC dataset was inaccessible and not updated to encompass recent years, outside sources using different labelling for municipalities and departments had to be used in the coding of the new dataset. As mentioned above, this meant many observations had to be removed despite having conflict data on them. By drawing from macro-data in a haphazard fashion because of the lack of instruction, description and availability of the data, (alongside number of observations) detail and accuracy were most certainly lost. Which could have impacted the results significantly. By having this smaller sample size, heteroskedasticity could not be controlled and the impact of extreme observations on the estimation would have been more significant. Thus, increasing standard errors and influencing the slope and intercept of the regression.

The second potential reason for this could have been outside factors that a) affect the number of observations and b) alter the dependant variables independent of the explanatory variables offered in the model. One such suggestion could be the advancement of the Colombian Peace Accords between FARC and the Colombian Government in recent years. Exploratory meetings for the eventual peace deal first took place in March of 2011 (FARC-EP International, 2011). Throughout 2011-2016, the talks broke down, ceasefires were signed, disarmament commenced, and a number of other events took place which affected the Colombian conflict totally independent of commodity price shocks. Which most likely heavily influenced results. Potentially, if some way of coding the progress of the peace talks, breakdowns in ceasefires etc were included in the model, there may have been some significance. However, this is purely speculation and outside the scope of this dissertation.

# VI. Lessons from the replication

## *Generalisation*

To address the external validity (generalisation) of this paper and its methodology, there was to be a section applying the methodology to another country using panel data and the same methodological specifications. However, this proved near impossible. There were several challenges that proved insurmountable. Firstly, finding a country that suffered a significant amount of civil conflict which also happened to have top exports being both agriculturally labour intensive and another commodity being a natural resource like oil proved difficult. An example of this is Mexico. Mexico suffers from a large amount of civil conflict between cartels and the Mexican Government, with total deaths at 25,738 from 1989-2018 (UCDP, 2020). However, Mexico's exports do not fit the specifications that are able to be tested by Dube and Vargas' models.  Mexico's highest percentage export are Cars, which aren't a natural resource or an agricultural good, despite this, 4.7% of its exports are Crude Petroleum, but unfortunately its highest agricultural export is 'Tropical Fruits' which account for 0.83% of exports (OECD, 2020). If perhaps the entire agricultural sector was measured and modelled in the regression, this could perhaps work. Although the model would be greatly altered from it's original to account for the extra measures and collecting the data would be outside the scope and timeframe of this dissertation.

It becomes even more difficult to select a country when Dube & Vargas' specification is followed further and a country with a civil conflict started along politico-ideological lines should be the target of the within-country analysis. In South America only Nicaragua could potentially be considered a candidate economically as its largest agriculture export is coffee and gold is their largest natural resource export at 9.7% and 8% of total exports respectively (OECD, 2020). Despite this, the sample size for Nicaragua is too small to avoid sampling bias and getting a true population representation out of the sample, with only 601 battle related deaths from 1989-2018. Simply put, the specifications Dube & Vargas state to select a candidate for within country-analysis on commodity prices and civil conflict, are far too narrow and mean the model cannot be sufficiently implemented across other countries, without avoiding bias.

Another issue harming the external validity of the model by Dube & Vargas is that there aren't sufficient micro-datasets (such as the CERAC dataset) in order to construct the variables needed to estimate properly. Whilst micro-datasets on individual countries are incredibly useful as they provide municipal and departmental data on a broad range of variables, they are few and far between. Even the CERAC dataset hasn't updated past 2002, finding sufficient data on other countries is incredibly difficult. Whilst this can be somewhat negated by data within macro-datasets, the majority of them lack the detail needed to run country specific regression models. Especially on the municipal, departmental and even time level (if the unit of analysis is monthly or daily). Furthermore, discrepancies in definitions and unit measurements across datasets can affect accuracy. Gathering data from all these sources, cleaning it and coding it takes a considerable amount of time and effort too, which to some extent subtracts from the value of doing the replication.

## *Lessons*

Within the literature of civil conflict and price shocks, there is little consensus. Studies vary in how they approach the challenges of measuring and modelling the mechanisms behind these effects, within-country analysis, such as the type conducted by Dube & Vargas, provide a tremendous amount of detail through the implementation of micro-data and country-specific modelling. These studies provide enormous insight into specific effects and mechanisms within-country. So, with regards to internal validity, they remain verifiable due to the richness of the data in its detail, the large number of observations and strong ability to control endogenous variables and instrument independent variables properly. However, evaluating their external validity proves incredibly difficult. The lack of adequate micro-data even across different sections of time within the same country, impedes the ability of researchers to properly assess the robustness and external validity of such studies.

In short, with this dissertation, the replication worked and failed. The important models with original data and coding were mostly verifiable with only minor differences, yet the extended time series provided little if any evidence of repeatability. Therefore, the results of this replication are inconclusive in terms of conducting a strong replication.

# VII. Conclusion

This dissertation has attempted to replicate the 2014 paper *Commodity Price Shocks and Civil Conflict: Evidence from Colombia By Dube & Vargas*. It attempts to do this by stringently verifying the original code and by extending the time series on the main model and drawing an overall conclusion. This dissertation concludes that the replication was unsuccessful. Whilst the models were verifiable, the repeatability and generalisability of the study could not be confidently addressed. Through the lack of clarification, instructions and framework from the authors, there is an argument that there may be some bias in the verification of the results. In an attempt to 'prove' the original findings. The lack of sufficient data in the extended time series, either through the lack of an updated original dataset or issues arising from drawing country specific variables from macro-data, rendered that section of the replication inconclusive at best. Furthermore, the lack of data significantly impacted the generalisability of the study, by failing to apply the same methodology to another within-country case study, there is no way of confidently saying these findings and mechanisms can sufficiently explain civil conflict worldwide. This is not to say that the original findings by Dube & Vargas are wrong, the analysis within this dissertation simply cannot verify or falsify due to the above difficulties.

Despite these issues, this dissertation contributes by highlighting the need for the construction and maintaining of new micro-datasets within the current literature of civil conflict and commodity prices. There will remain very little consensus in the literature if claims are not verifiable, repeatable, generalisable and robust. The only realistic way for this to occur with regards to within-country studies and analysis, is to create further micro-datasets with the exact variables and instruments needed to conduct stringent replication studies.

# Bibliography

*Baddeley A, Rubak E, Turner R (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London. http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/.*

Bazzi, S. and Blattman, C., 2014. Economic shocks and conflict: Evidence from commodity prices. *American Economic Journal: Macroeconomics*, *6*(4), pp.1-38.

Becker, G.S., 1968. Crime and punishment: An economic approach. In *The economic dimensions of crime* (pp. 13-68). Palgrave Macmillan, London.

Brandt, M.J., IJzerman, H., Dijksterhuis, A., Farach, F.J., Geller, J., Giner-Sorolla, R., Grange, J.A., Perugini, M., Spies, J.R. and Van't Veer, A., 2014. The replication recipe: What makes for a convincing replication?. *Journal of Experimental Social Psychology*, *50*, pp.217-224.

Cerac.org.co. 2020. *CERAC :: Conflict Data*. [online] Available at: <https://www.cerac.org.co/en/resources/conflict-data.html> [Accessed 12 May 2020].

Christensen, D., 2019. Concession Stands: How Mining Investments Incite Protest in Africa. *International organization*, *73*(1), pp.65-101.

*Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The plm Package." Journal of Statistical Software,* ***27****(2), 1–43. doi: 10.18637/jss.v027.i02.*

*DANE 2018. DANE - Censo Nacional De Población Y Vivienda (CNPV) 2018. Estimaciones De Población Ajustada Por Cobertura Censal. [online] Available at: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018> [Accessed 14 May 2020].*

Dube, O. and Vargas, J.F., 2013. Commodity price shocks and civil conflict: Evidence from Colombia. *The review of economic studies*, *80*(4), pp.1384-1421.

FARC-EP International. 2011. *Exploratory Meetings*. [online] Available at: <http://farc-epeace.org/index.php/peace-process/news/item/262-exploratory-meetings> [Accessed 20 May 2020].

*Federación Nacional de Cafeteros. 2020. Coffee Statistics - Federación Nacional De Cafeteros. [online] Available at: <https://federaciondecafeteros.org/coffee-statistics/?lang=en> [Accessed 5 May 2020].*

Freese, J. and Peterson, D., 2017. Replication in social science. *Annual Review of Sociology*, *43*, pp.147-165.
*Gui, Raluca & Meierer, Markus & Algesheimer, René. (2018). REndo: An R Package to Address Endogeneity Without External Instrumental Variables.*

Hegre, H. and Sambanis, N., 2006. Sensitivity analysis of empirical results on civil war onset. *Journal of conflict resolution*, *50*(4), pp.508-535.

Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

*Högbladh Stina, 2019, "UCDP GED Codebook version 19.1", Department of Peace and Conflict Research, Uppsala University*

HUMAN RIGHTS WATCH. (2005), Smoke and Mirrors: Colombia's Demobilization of Paramilitary Groups (NewYork:
Human RightsWatch)

*Ico.org. 2020. International Coffee Organization - Historical Data On The Global Coffee Trade. [online] Available at: <http://www.ico.org/new_historical.asp> [Accessed 8 May 2020].*

Ioannidis, J.P., 2005. Why most published research findings are false. *PLos med*, *2*(8), p.e124.
*Millo G (2017). "Robust Standard Error Estimators for Panel Models: A Unifying Approach." Journal of Statistical Software,* ***82****(3), 1–27. doi: 10.18637/jss.v082.i03.*

Morgan, R.K. and Reinhardt, E., 2015. Commodity price volatility and civil conflict. *Emory University, Atlanta, GA*.
*Restrepo, J., Gutierrez, G., Marín, M. and Ronderos, N., 2014. CERAC :: Database Of The Armed Conflict In Colombia. [online] Cerac.org.co. Available at: <https://www.cerac.org.co/en/resources/bdcc/> [Accessed 12 May 2020].*

Netcials. 2020. *US Dollar (USD) To Colombian Peso (COP) Yearly Average Exchange Rates*. [online] Available at: <https://www.netcials.com/forex-yearly-average-exchange-rate/USD-COP/> [Accessed 16 May 2020].

Oec.world. 2020. *OEC - Mexico (MEX) Exports, Imports, And Trade Partners*. [online] Available at: <https://oec.world/en/profile/country/mex/> [Accessed 18 May 2020].

Oec.world. 2020. *OEC – Nicaragua (NIC) Exports, Imports, And Trade Partners*. [online] Available at: <https://oec.world/en/profile/country/nic/> [Accessed 18 May 2020].

Minenergia.gov.co. 2020. *Office Of The Minister Of Mines And Energy*. [online] Available at: <https://www.minenergia.gov.co/web/ingles/the-minister> [Accessed 16 May 2020].

Restrepo, J.A., Spagat, M. and Vargas, J.F., 2006. Special data feature; the severity of the Colombian conflict: cross-country datasets versus new micro-data. *Journal of Peace research*, *43*(1), pp.99-115.

RESTREPO, J., SPAGAT, M. and VARGAS, J. F. (2004), "The Dynamics of the Colombian Civil Conflict: A New Data Set", *Homo Oeconomicus*, **21**, 396–428.

*RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.*

*StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.*
*Statista. 2020. OPEC Oil Prices 1960-2020 | Statista. [online] Available at: <https://www.statista.com/statistics/262858/change-in-opec-crude-oil-prices-since-1960/> [Accessed 3 May 2020].*

*Sundberg, Ralph, and Erik Melander, 2013, "Introducing the UCDP Georeferenced Event Dataset", Journal of Peace Research, vol.50, no.4, 523-532*
*Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.*

Uppsala Conflict Data Program (Date of retrieval: 17/05/20) *UCDP Conflict Encyclopedia: www.ucdp.uu.se, Uppsala University*

*World Bank Data Help Desk. 2020. What Do We Mean By Microdata? – World Bank Data Help Desk. [online] Available at: <https://datahelpdesk.worldbank.org/knowledgebase/articles/228873-what-do-we-mean-by-microdata> [Accessed 16 May 2020].*

# Appendix

**Table 1 Descriptive Statistics Original Paper**

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| department | 17,964 | 38.156 | 26.917 | 5 | 15 | 68 | 99 |
| origmun | 17,964 | 118,602,657,286,044,432.000 | 2,697,611,692,315,258,368.000 | 5,002 | 15,646 | 68,162 | 70,823,700,099,005,153,280 |
| multsplit | 17,964 | 0.008 | 0.089 | 0 | 0 | 0 | 1 |
| gueratt | 17,964 | 0.490 | 1.502 | 0 | 0 | 0 | 41 |
| paratt | 17,964 | 0.076 | 0.399 | 0 | 0 | 0 | 11 |
| clashes | 17,964 | 0.473 | 1.344 | 0 | 0 | 0 | 25 |
| casualties | 17,964 | 1.894 | 6.884 | 0 | 0 | 1 | 292 |
| govatt | 17,964 | 0.087 | 0.421 | 0 | 0 | 0 | 14 |
| guermass | 17,964 | 0.011 | 0.122 | 0 | 0 | 0 | 4 |
| parmass | 17,964 | 0.047 | 0.291 | 0 | 0 | 0 | 7 |
| guerkidpol | 16,966 | 0.033 | 0.295 | 0.000 | 0.000 | 0.000 | 20.000 |
| parkidpol | 16,966 | 0.032 | 0.262 | 0.000 | 0.000 | 0.000 | 12.000 |
| oilprod88 | 17,964 | 0.003 | 0.053 | 0 | 0 | 0 | 2 |
| coalprod04 | 17,964 | 1.870 | 12.743 | 0 | 0 | 0 | 156 |
| coalres78 | 17,964 | 0.319 | 0.466 | 0 | 0 | 1 | 1 |
| goldprod04 | 17,964 | 0.374 | 2.660 | 0 | 0 | 0 | 34 |
| mining78 | 17,964 | 583.284 | 3,863.974 | 0 | 0 | 0 | 100,349 |
| cofint | 17,604 | 0.832 | 1.542 | 0.000 | 0.000 | 0.987 | 10.585 |
| coca | 7,984 | 0.111 | 0.854 | 0.000 | 0.000 | 0.000 | 24.507 |
| coca94 | 17,964 | 0.071 | 0.581 | 0 | 0 | 0 | 9 |
| coca94ind | 17,964 | 0.050 | 0.218 | 0 | 0 | 0 | 1 |
| evercoca | 17,964 | 0.241 | 0.428 | 0 | 0 | 0 | 1 |
| lpop | 17,964 | -4.350 | 0.963 | -8.832 | -4.984 | -3.735 | -1.357 |
| rainfall | 17,964 | 1,888.327 | 1,003.273 | 160 | 1,159 | 2,360 | 9,200 |
| temperature | 17,964 | 21.392 | 4.965 | 3.900 | 17.700 | 26.400 | 28.900 |
| yrspropara | 11,976 | 0.761 | 1.418 | 0.000 | 0.000 | 0.000 | 9.000 |
| lcaprev | 11,755 | 7.058 | 2.483 | -6.908 | 6.869 | 8.396 | 11.456 |
| linternalp | 17,964 | 0.642 | 0.240 | 0.252 | 0.434 | 0.868 | 0.985 |
| ltop3cof | 17,964 | 3.433 | 0.255 | 3.087 | 3.232 | 3.741 | 3.845 |
| lop | 17,964 | 4.196 | 0.373 | 3.442 | 3.901 | 4.482 | 4.813 |
| lgoldp | 17,964 | 0.096 | 0.288 | -0.347 | -0.194 | 0.235 | 0.619 |
| lcoalp | 17,964 | -2.164 | 0.257 | -2.522 | -2.380 | -1.930 | -1.744 |
| ltop3coal | 17,964 | 12.657 | 0.111 | 12.462 | 12.544 | 12.743 | 12.862 |
| lsilverp | 17,964 | -4.165 | 0.209 | -4.502 | -4.316 | -4.020 | -3.663 |
| lplatp | 17,964 | 0.382 | 0.334 | -0.119 | 0.082 | 0.725 | 0.889 |
| cofintxlinternalp | 17,604 | 0.535 | 1.076 | 0.000 | 0.000 | 0.588 | 10.427 |
| rxltop3cof | 17,964 | 6,482.265 | 3,486.835 | 493.870 | 3,943.623 | 8,085.737 | 35,372.030 |
| txltop3cof | 17,964 | 73.435 | 17.938 | 12.038 | 59.581 | 87.326 | 111.114 |
| rtxltop3cof | 17,964 | 142,888.100 | 93,272.560 | 11,676.940 | 87,684.440 | 175,591.100 | 944,433.300 |
| oilprod88xlop | 17,964 | 0.014 | 0.224 | 0 | 0 | 0 | 8 |
| yrsproparaxoil88xlop | 11,976 | 0.007 | 0.107 | 0.000 | 0.000 | 0.000 | 2.634 |
| yrsproparaxlop | 11,976 | 3.172 | 5.953 | 0.000 | 0.000 | 0.000 | 43.319 |
| pipelenxlop | 17,964 | 0.299 | 1.128 | 0 | 0 | 0 | 18 |
| coalprod04xlcoalp | 17,964 | -4.046 | 27.769 | -392 | 0 | 0 | 0 |
| coalres78xltop3coal | 17,964 | 4.033 | 5.898 | 0.000 | 0.000 | 12.534 | 12.862 |
| goldprod04xlgoldp | 17,964 | 0.036 | 0.815 | -12 | 0 | 0 | 21 |
| mining78xlgoldp | 17,964 | 55.948 | 1,185.470 | -34,819.290 | 0.000 | 0.000 | 62,098.950 |
| mining78xlsilverp | 17,964 | -2,429.656 | 16,115.940 | -451,734.800 | 0.000 | 0.000 | 0.000 |
| mining78xlplatp | 17,964 | 222.529 | 1,969.260 | -11,949.120 | 0.000 | 0.000 | 89,194.590 |

**Table 1 Descriptive Statistics for Extended time Series**

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| year | 377 | 2,010.483 | 3.787 | 2,006 | 2,007 | 2,013 | 2,018 |
| department | 377 | 18.716 | 8.568 | 5 | 12 | 26 | 33 |
| origmun | 377 | 42,282.660 | 27,941.870 | 5,001 | 18,592 | 73,001 | 99,773 |
| gueratt | 377 | 0.159 | 0.420 | 0 | 0 | 0 | 3 |
| paratt | 377 | 0.003 | 0.052 | 0 | 0 | 0 | 1 |
| clashes | 377 | 1.141 | 0.683 | 0 | 1 | 1 | 5 |
| casualties | 377 | 5.231 | 6.016 | 0 | 2 | 6 | 44 |
| lpop | 377 | 4.554 | 0.621 | 3.246 | 4.119 | 4.729 | 6.870 |
| cofint | 377 | 0.945 | 1.750 | 0 | 0 | 1.1 | 11 |
| oilprod88 | 377 | 0.007 | 0.085 | 0 | 0 | 0 | 2 |
| coca94ind | 377 | 0.117 | 0.322 | 0 | 0 | 0 | 1 |
| coca94indxyear | 377 | 234.756 | 646.681 | 0 | 0 | 0 | 2,018 |
| lop | 377 | 5.344 | 0.107 | 5.080 | 5.255 | 5.444 | 5.509 |
| oilprod88xlop | 377 | 1,260.887 | 3,474.162 | 0 | 0 | 0 | 11,085 |
| linternalp | 377 | 5.791 | 0.107 | 5.660 | 5.668 | 5.870 | 5.989 |
| cofintxlinternalp | 377 | 5.464 | 10.112 | 0 | 0 | 6.3 | 62 |
| ltop3cof | 377 | 4.443 | 0.056 | 4.367 | 4.400 | 4.482 | 4.542 |
| rainfall | 377 | 1,935.937 | 1,249.831 | 0 | 1,186 | 2,650 | 5,910 |
| temperature | 377 | 19.046 | 9.538 | 0 | 17.2 | 25.7 | 29 |
| rxltop3cof | 377 | 8,603.877 | 5,563.378 | 0 | 5,236.3 | 11,678.3 | 26,501 |
| txltop3cof | 377 | 84.620 | 42.381 | 0 | 77.1 | 114.6 | 130 |

**Replicated Results of Table 2**

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Guerrilla Attacks | Paramilitary Attacks | Casualties | Clashes |
| | (1) | (2) | (3) | (4) |
| Coffee int. x log coffee price | $-0.644^{***}$ | $-0.183^{***}$ | $-1.780^{***}$ | $-0.687^{***}$ |
| | (0.111) | (0.031) | (0.665) | (0.137) |
| Oil production x log oil price | 0.526 | $0.703^{***}$ | 2.573 | 0.183 |
| | (1.351) | (0.148) | (2.051) | (0.601) |
| Observations | 17,604 | 17,604 | 17,604 | 17,604 |
| $R^2$ | 0.027 | 0.033 | 0.027 | 0.037 |
| Adjusted $R^2$ | -0.035 | -0.028 | -0.034 | -0.024 |
| F Statistic | $614.294^{***}$ | $743.386^{***}$ | $504.332^{***}$ | $749.581^{***}$ |
| *Note:* | | | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |

**Results of Table 2 with extended time series**

| | *Dependent variable:* | | |
|---|---|---|---|
| | Guerrilla Attacks | Casualties | Clashes |
| | (1) | (2) | (3) |
| Coffee int. x log coffee price | 1.824 | -3.451 | 0.156 |
| | (1.309) | (13.108) | (1.552) |
| Oil production x log oil price | -0.0003 | 0.0003 | 0.0002 |
| | (0.0002) | (0.005) | (0.001) |
| Observations | 377 | 377 | 377 |
| $R^2$ | 0.098 | 0.294 | 0.243 |
| Adjusted $R^2$ | -1.387 | -0.870 | -1.005 |
| F Statistic | 31.197 | 54.090 | 45.523 |
| *Note:* | | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |

Paramilitary attacks model removed from regression due as there is only one observation in time series

# Table 3 Verification (PLM and MultilevelIV sections)

**Replicated Table 3 Paramilitary and Guerrilla Kidnappings**

| | *Dependent variable:* | |
|---|---|---|
| | Paramilitary Political Kidnappings | Guerrilla Political Kidnappings |
| | (1) | (2) |
| Coffee int. x log coffee price | 0.013 | -0.093** |
| | (0.012) | (0.038) |
| Oil production x log oil price | -5.458* | -4.125 |
| | (3.314) | (5.194) |
| Observations | 16,626 | 16,626 |
| $R^2$ | 0.029 | 0.003 |
| Adjusted $R^2$ | -0.036 | -0.064 |
| F Statistic | 1,307.851*** | 338.006*** |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

```
Call:
multilevelIV(formula = lwage ~ cofintxlinternalp + oilprod88xlop +
    gender + age + agesq + married + edyrs + coca94indxyear +
    as.factor(year) + as.factor(region) + rxltop3cof + txltop3cof +
    (1 | origmun) | endo(rxltop3cof, txltop3cof), data = wages_commodities)

Number of levels: 2
Number of observations: 26050
Number of groups: L2(origmun): 246

Coefficients for model REF:
                      Estimate Std. Error z-score Pr(>|z|)
(Intercept)          6.405e+00  6.807e-02  94.095  < 2e-16 ***
cofintxlinternalp    1.527e-03  9.540e-03   0.160  0.87280
oilprod88xlop        1.965e-01  1.465e-01   1.342  0.17971
gender              -1.100e-01  1.079e-02 -10.194  < 2e-16 ***
age                  2.966e-02  2.171e-03  13.658  < 2e-16 ***
agesq               -3.393e-04  2.974e-05 -11.408  < 2e-16 ***
married              1.997e-02  9.631e-03   2.073  0.03815 *
edyrs                6.292e-02  1.232e-03  51.083  < 2e-16 ***
coca94indxyear       3.220e-05  2.971e-05   1.084  0.27851
as.factor(year)1999 -1.481e-02  1.343e-02  -1.102  0.27042
as.factor(year)2000 -2.179e-01  1.422e-02 -15.325  < 2e-16 ***
as.factor(year)2001 -2.286e-01  2.193e-02 -10.421  < 2e-16 ***
as.factor(year)2002 -2.601e-01  2.631e-02  -9.884  < 2e-16 ***
as.factor(year)2003 -2.585e-01  2.620e-02  -9.868  < 2e-16 ***
as.factor(year)2004 -3.668e-01  2.254e-02 -16.274  < 2e-16 ***
as.factor(year)2005 -3.376e-01  2.216e-02 -15.230  < 2e-16 ***
as.factor(region)2  -1.643e-02  4.280e-02  -0.384  0.70098
as.factor(region)3   5.939e-02  7.921e-02   0.750  0.45335
as.factor(region)4  -1.776e-01  3.288e-02  -5.401 6.62e-08 ***
rxltop3cof           1.000e-07  3.792e-06   0.026  0.97896
txltop3cof           2.516e-03  8.788e-04   2.863  0.00419 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Omitted variable tests for model REF:
              df  Chisq p-value
FE_L2_vs_REF  21 25.823   0.213
GMM_L2_vs_REF  3 -6.714   1.000
```

```
Call:
multilevelIV(formula = lhours ~ cofintxlinternalp + oilprod88xlop +
    gender + age + agesq + married + edyrs + coca94indxyear +
    as.factor(year) + as.factor(region) + rxltop3cof + txltop3cof +
    (1 | origmun) | endo(rxltop3cof, txltop3cof), data = hours_commodities)

Number of levels: 2
Number of observations: 57743
Number of groups: L2(origmun): 246

Coefficients for model REF:
                      Estimate Std. Error z-score Pr(>|z|)
(Intercept)          4.643e+00  3.782e-02 122.754  < 2e-16 ***
cofintxlinternalp    1.664e-02  5.260e-03   3.163 0.001559 **
oilprod88xlop        3.850e-02  8.963e-02   0.430 0.667515
gender              -3.293e-01  4.038e-03 -81.551  < 2e-16 ***
age                  2.615e-02  8.719e-04  29.987  < 2e-16 ***
agesq               -3.245e-04  1.169e-05 -27.759  < 2e-16 ***
married              4.259e-03  4.077e-03   1.044 0.296268
edyrs               -2.819e-03  5.405e-04  -5.216 1.83e-07 ***
coca94indxyear      -2.000e-06  1.793e-05  -0.112 0.911207
as.factor(year)1999 -2.934e-02  6.173e-03  -4.753 2.01e-06 ***
as.factor(year)2000 -5.128e-02  6.202e-03  -8.268  < 2e-16 ***
as.factor(year)2001 -9.274e-02  9.567e-03  -9.693  < 2e-16 ***
as.factor(year)2002 -8.942e-02  1.175e-02  -7.613 2.67e-14 ***
as.factor(year)2003 -8.487e-02  1.123e-02  -7.556 4.15e-14 ***
as.factor(year)2004 -1.020e-01  1.044e-02  -9.772  < 2e-16 ***
as.factor(year)2005 -9.466e-02  1.017e-02  -9.311  < 2e-16 ***
as.factor(region)2   1.679e-02  2.587e-02   0.649 0.516477
as.factor(region)3   7.381e-02  4.727e-02   1.562 0.118389
as.factor(region)4  -2.152e-02  2.008e-02  -1.072 0.283915
rxltop3cof          -2.000e-07  2.277e-06  -0.088 0.930024
txltop3cof           1.967e-03  5.274e-04   3.729 0.000192 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Omitted variable tests for model REF:
              df  Chisq  p-value
FE_L2_vs_REF  21 48.473 0.000593 ***
GMM_L2_vs_REF  3 -1.793 1.000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Call:
multilevelIV(formula = lcaprev ~ cofintxlinternalp + oilprod88xlop +
    lpop + coca94indxyear + as.factor(year) + as.factor(region) +
    rxltop3cof + txltop3cof + (1 | origmun) | endo(rxltop3cof,
    txltop3cof), data = violence_commodities)

Number of levels: 2
Number of observations: 17964
Number of groups: L2(origmun): 963

Coefficients for model REF:
                      Estimate Std. Error z-score Pr(>|z|)
(Intercept)          7.271e+00  1.837e-01  39.589  < 2e-16 ***
cofintxlinternalp   -8.392e-02  2.028e-02  -4.138 3.50e-05 ***
oilprod88xlop        1.580e-01  8.547e-02   1.848  0.06455 .
lpop                 6.492e-01  2.475e-02  26.231  < 2e-16 ***
coca94indxyear      -2.435e-04  5.439e-05  -4.477 7.56e-06 ***
as.factor(year)1989 -6.273e-01  1.236e-01  -5.077 3.83e-07 ***
as.factor(year)1990 -1.941e+00  1.167e-01 -16.629  < 2e-16 ***
as.factor(year)1991 -1.742e+00  1.186e-01 -14.684  < 2e-16 ***
as.factor(year)1992 -1.621e+00  1.141e-01 -14.207  < 2e-16 ***
as.factor(year)1993 -1.018e+00  1.132e-01  -8.991  < 2e-16 ***
as.factor(year)1994 -5.964e-01  1.118e-01  -5.333 9.66e-08 ***
as.factor(year)1995  5.464e-03  1.085e-01   0.050  0.95982
as.factor(year)1996  3.248e+00  1.026e-01  31.642  < 2e-16 ***
as.factor(year)1997  3.431e+00  1.040e-01  32.981  < 2e-16 ***
as.factor(year)1998  3.328e+00  1.027e-01  32.398  < 2e-16 ***
as.factor(year)1999  3.521e+00  1.037e-01  33.961  < 2e-16 ***
as.factor(year)2000  3.694e+00  1.028e-01  35.933  < 2e-16 ***
as.factor(year)2001  3.594e+00  1.038e-01  34.624  < 2e-16 ***
as.factor(year)2002  3.829e+00  1.043e-01  36.706  < 2e-16 ***
as.factor(year)2003  3.697e+00  1.035e-01  35.706  < 2e-16 ***
as.factor(year)2004  3.686e+00  1.039e-01  35.471  < 2e-16 ***
as.factor(year)2005  3.787e+00  1.033e-01  36.652  < 2e-16 ***
as.factor(region)2  -2.348e-01  7.656e-02  -3.067  0.00216 **
as.factor(region)3  -6.316e-02  9.377e-02  -0.674  0.50060
as.factor(region)4  -5.367e-01  6.417e-02  -8.364  < 2e-16 ***
rxltop3cof           1.550e-05  6.918e-06   2.240  0.02506 *
txltop3cof           1.912e-03  1.506e-03   1.270  0.20422
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Omitted variable tests for model REF:
              df Chisq  p-value
FE_L2_vs_REF  27 31.16 0.264734
GMM_L2_vs_REF  3 20.83 0.000114 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*Code*

# Verification Section Code

```
### Dataset and packages loading ###

library(plm)
library(ggplot2)
library(tidyr)
library(tibble)
library(spatstat)
library(dplyr)
library(haven)
library(readr)
library(stargazer)
library(estimatr)
library(DiagrammeR)
library(texreg)
library(psych)
library(ivpack)
library(texreg)
library(REndo)
library(finalfit)

# Datasets

hours_commodities <- read_dta("~/Dissertation/Commodity Price Shocks
DATA/Data/origmun_hours_commodities.dta")
migrant_commodities <- read_dta("~/Dissertation/Commodity Price Shocks
DATA/Data/origmun_migrant_commodities.dta")
violence_commodities <- read_dta("~/Dissertation/Commodity Price Shocks
DATA/Data/origmun_violence_commodities.dta")
```

```
wages_commodities <- read_dta("~/Dissertation/Commodity Price Shocks
DATA/Data/origmun_wages_commodities.dta")
municipality_violence_commodities_online_appendix <-
read_dta("~/Dissertation/Commodity Price Shocks
DATA/App_Data/no_split_municipality_violence_commodities_for_online_appendix.dta")
violence_commodities_online_appendix <- read_dta("~/Dissertation/Commodity Price Shocks
DATA/App_Data/origmun_violence_commodities_for_online_appendix.dta")

###########################################################################
###########################################################################
###########################################################################
##################

###Table 1, Key Statistics###

table1 <-
stargazer(as.data.frame(violence_commodities,hours_commodities,wages_commodities), type
= "text",
          omit =
c("department_name","origmun_name","year","1989","1990","1991","1992","1993","1994","1995","1
996","1997","1998","1999","2000","2001","2002", "2003","2004","2005","region"),
                  title = "Table 1 Descriptive Statistics Original Paper",
                  out = "C:/Users/Owner/Documents/Dissertation/table1.html")

###########################################################################
###########################################################################
###########################################################################
#################

### Table 2 Regression ###

###Guerilla Attack Regression with Robust SE's###
```

```
table2greg <- plm(gueratt ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
   rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop +
as.factor(year)*as.factor(region),
  index=c('origmun','year'), data=violence_commodities, model='within') ## Includes
interactions between year and region, rainfall and temperature
##Unable to cluster by department because of syntax errors
summary(table2greg)

#SE Clusters

table2gregVcov <- vcovHC(table2greg, type = "HC2")
table2gregRSE<-sqrt(diag(table2gregVcov))
##Since no clustering by department, robust standard errors by HC2 are done instead


###Paramilitary Attack Regression with Robust SE's###

table2preg <- plm(paratt ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
                rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
              index=c('origmun','year'), data=violence_commodities, model='within')
summary(table2preg)
#SE Clusters

table2pregVcov <- vcovHC(table2preg, type = "HC2")
table2pregRSE <- sqrt(diag(table2pregVcov))


###Clashes Regression with Robust SE's###
```

```
table2clreg <- plm(clashes ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region)|
                rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
              index=c('origmun','year'), data=violence_commodities, model='within')
#SE Clusters

table2clregVcov <- vcovHC(table2clreg, type = "HC2")
table2clregRSE <- sqrt(diag(table2clregVcov))



###Casualties Regression with Robust SE's###

table2careg <- plm(casualties ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
                rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
              index=c('origmun','year'), data=violence_commodities, model='within')

#SE Clusters

table2caregVcov <- vcovHC(table2careg, type = "HC2")
table2caregRSE <- sqrt(diag(table2caregVcov))




###Printed Regression###




stargazer(table2greg,table2preg,table2careg,table2clreg,
        se = list(table2gregRSE,table2pregRSE,table2caregRSE,table2clregRSE),
        omit = c("lpop","coca94indxyear","1989","1990","1991","1992",
              "1993","1994","1995","1996","1997",
```

```
        "1998","1999","2000","2001","2002",
        "2003","2004","2005"),
    title = "Replicated Results of Table 2",
    dep.var.labels = c("Guerrilla Attacks","Paramilitary Attacks","Casualties","Clashes"),
    covariate.labels = c("Coffee int. x log coffee price","Oil production x log oil price"),
    out = "C:/Users/Owner/Documents/Dissertation/table2.html",
    type = "text")


##############################################################################
##############################################################################
####################


na.omit(violence_commodities)

### Table 3 Regression: Opportunity Cost and Rapacity Mechanism ###

###Log Wage Regression With Robust SE's###

table3lwreg <- multilevelIV(lwage ~
             cofintxlinternalp+
oilprod88xlop+gender+age+agesq+married+edyrs+coca94indxyear +
             as.factor(year)+as.factor(region) + rxltop3cof+txltop3cof +
             (1|origmun) |   endo(rxltop3cof,txltop3cof),data=wages_commodities)


summary(table3lwreg)

#Robust SE's

table3lwregVcov <- vcovHC(table3lwreg,type = "HC2")
table3lwregRSE <- sqrt(diag(table3lwregVcov))
```

```
#P Values


table3lwregCOEF <- table3lwreg$coefficients
table3lwregtstat <- table3lwregCOEF/table3lwregRSE
table3lwregDOF <- table3lwreg$df.residual
table3lwregpvalues <- (pt(abs(table3lwregtstat),df=table3lwregDOF))
table3lwregpvalues

###Log Hours Regression With Robust SE's###

table3lhreg <- multilevelIV(lhours ~
                    cofintxlinternalp+
oilprod88xlop+gender+age+agesq+married+edyrs+coca94indxyear +
                    as.factor(year)+as.factor(region) + rxltop3cof+txltop3cof +
                    (1|origmun) |   endo(rxltop3cof,txltop3cof),data=hours_commodities)



summary(table3lhreg)

#Robust SE's

table3lhregVcov <- vcovHC(table3lhreg, type = "HC2")
table3lhregRSE <- sqrt(diag(table3lhregVcov))


#P Values

table3lhregCOEF <- table3lhreg$coefficients
table3lhregtstat <- table3lhregCOEF/table3lhregRSE
table3lhregDOF <- table3lhreg$df.residual
table3lhregpvalues <- (pt(abs(table3lhregtstat),df=table3lhregDOF))
table3lhregpvalues
```

```
###Log Capital Revenue Regression with Robust SE's###

table3lcapreg <- multilevelIV(lcaprev ~
                        cofintxlinternalp+ oilprod88xlop+lpop + coca94indxyear+
                        as.factor(year)+as.factor(region) + rxltop3cof+txltop3cof +
                        (1|origmun) |   endo(rxltop3cof,txltop3cof),data=violence_commodities)

summary(table3lcapreg)

#Robust SE's

table3lcapVcov <- vcovHC(table3lcapreg, type = "HC2")
table3lcapregRSE <- sqrt(diag(table3lcapVcov))

#P Values

table3lcapCOEF <- table3lcapreg$coefficients
table3lcapregtstat <- table3lcapCOEF/table3lcapregRSE
table3lcapregDOF <- table3lcapreg$df.residual
table3lcapregpvalues <- (pt(abs(table3lcapregtstat),df=table3lcapregDOF))
table3lcapregpvalues

###Paramilitary Kidnappings Regression with Robust SE's###

table3parkidreg <- plm(parkidpol ~ cofintxlinternalp+oilprod88xlop
+lpop+as.factor(region)*as.factor(year)+coca94indxyear |
                  rxltop3cof*txltop3cof  + lpop +as.factor(region)*as.factor(year),
                 index=c('origmun','year'), data=violence_commodities, model='within')
```

```
#Robust SE's

table3parkidVcov <- vcovHC(table3parkidreg, type = "HC2")
table3parkidregRSE <- sqrt(diag(table3parkidVcov))


###Guerrilla Kidnapping Regression with Robust SE's###

table3guerkidreg <- plm(guerkidpol ~ cofintxlinternalp+oilprod88xlop
+lpop+as.factor(region)*as.factor(year)+coca94indxyear |
               rxltop3cof*txltop3cof  + lpop +as.factor(region)*as.factor(year),
              index=c('origmun','year'), data=violence_commodities, model='within')


#Robust SE's

table3guerkidVcov <- vcovHC(table3guerkidreg, type = "HC2")
table3guerkidregRSE <- sqrt(diag(table3guerkidVcov))


###Printed Regression###


table3alt <- stargazer(table3lwreg,table3lhreg,table3lcapreg,
             dep.var.labels = c("Log Wage","Log Hours","Log Capital Revenue"),
             covariate.labels = c("Coffee int. x log coffee price","Oil production x log oil
price","Gender","Age","Age Squared","Married",
                    "Education (years)","Log Population","Region and Municipalities
cultivating Coca"),
             out = "C:/Users/Owner/Documents/Dissertation/table3.html",
             type = "text",
             title = "Table 3: Opportunity Cost and Rapacity Mechanism")
```

```
table3guerpar <- stargazer(table3parkidreg,table3guerkidreg,
        omit = c("lpop","coca94indxyear","1989","1990","1991","1992",
               "1993","1994","1995","1996","1997",
               "1998","1999","2000","2001","2002",
               "2003","2004","2005"),
        se = list(table3parkidregRSE,table3guerkidregRSE),
        title = "Replicated Table 3 Paramilitary and Guerrilla Kidnappings",
        covariate.labels = c("Coffee int. x log coffee price","Oil production x log oil price"),
        dep.var.labels = c("Paramilitary Political Kidnappings", "Guerrilla Political
Kidnappings"),
        out = "C:/Users/Owner/Documents/Dissertation/table3par&guer.html",
        type = 'text')
```

# Repeatability Section Code

```
### Datasets Used ###

colombiadatanew <- read_csv("colombiadatanew.csv")
colombiadatanew <- na.omit(colombiadatanew)


### Packages ###

library(plm)
library(ggplot2)
library(tidyr)
library(tibble)
library(spatstat)
library(dplyr)
```

```
library(haven)
library(readr)
library(stargazer)
library(estimatr)
library(DiagrammeR)
library(texreg)
library(psych)
library(ivpack)
library(texreg)
library(REndo)
library(fastDummies)


### Creating Region and Year dummy's plus interactions ###



colombiadatanew <- dummy_cols(colombiadatanew,select_columns=c('year','region')) %>%
  mutate_at(vars(starts_with('year_')),list(region_1=~./region_1,region_2=~./region_2,
                                            region_3=~./region_3,region_4=~./region_4))

write.csv(colombiadatanew, file = "C:/Users/Owner/Documents/Dissertation/Dissertation
Code/colombiadatanew.csv")

############################################################################
############################################################################
############################################################################
##################

###Table 1, Key Statistics###


table1extended <- stargazer(as.data.frame(colombiadatanew), type = "text", omit =
c("1989","1990","1991","1992",
```

```
                                        "2006","2007","2008","2009","2010",
                                        "2011","2012","2013","2014","2015",
                                        "2016","2017","2018","region","X1"),
        title = "Table 1 Descriptive Statistics for Extended time Series",
        out = "C:/Users/Owner/Documents/Dissertation/table1extended.html")

##########################################################################
##########################################################################
##########


### Table 2 Regression ###

###Guerilla Attack Regression with Robust SE's###

table2greg <- plm(gueratt ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
         rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop +
as.factor(year)*as.factor(region),
        index=c('origmun','year'), data=colombiadatanew, model='within') ## Includes
interactions between year and region, rainfall and temperature
##Unable to cluster by department because of syntax errors
summary(table2greg)

#SE Clusters

table2gregVcov <- vcovHC(table2greg, type = "HC2")
table2gregRSE<-sqrt(diag(table2gregVcov))
##Since no clustering by department, robust standard errors by HC2 are done instead

##########################################################################
##########################################################################
######
```

```
#############################Only one paramilitary attack across the whole sample
therefore excluded from regression########################################
############################################################################
############################################################################
######

###Paramilitary Attack Regression with Robust SE's###

table2preg <- plm(paratt ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
                 rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
                index=c('origmun','year'), data=colombiadatanew, model='within')

#SE Clusters

table2pregVcov <- vcovHC(table2preg, type = "HC2")
table2pregRSE <- sqrt(diag(table2pregVcov))



############################################################################
############################################################################
#####

###Clashes Regression with Robust SE's###

table2clreg <- plm(clashes ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region)|
                  rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
                 index=c('origmun','year'), data=colombiadatanew, model='within')
#SE Clusters

table2clregVcov <- vcovHC(table2clreg, type = "HC2")
```

```
table2clregRSE <- sqrt(diag(table2clregVcov))

###Casualties Regression with Robust SE's###

table2careg <- plm(casualties ~ cofintxlinternalp + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region) |
                rxltop3cof*txltop3cof  + lpop + coca94indxyear + oilprod88xlop
+as.factor(year)*as.factor(region),
              index=c('origmun','year'), data=colombiadatanew, model='within')

#SE Clusters

table2caregVcov <- vcovHC(table2careg, type = "HC2")
table2caregRSE <- sqrt(diag(table2caregVcov))



###Printed Regression###




stargazer(table2greg,table2careg,table2clreg,
        se = list(table2gregRSE,table2caregRSE,table2clregRSE),
        omit = c("lpop","coca94indxyear","1989","1990","1991","1992",
              "2006","2007","2008","2009","2010",
              "2011","2012","2013","2014","2015",
              "2016","2017","2018","region"),
        notes = "Paramilitary attacks model removed from regression due as there is only one
observation in time series",
        title = "Results of Table 2 with extended time series",
        dep.var.labels = c("Guerrilla Attacks","Casualties","Clashes"),
        covariate.labels = c("Coffee int. x log coffee price","Oil production x log oil price"),
        out = "C:/Users/Owner/Documents/Dissertation/table2new.html",
        type = "text")
```