University of Cape Town

School of Economics

Untangling the Determinants of Crime in South Africa

BY

Noel Verrinder

(VRRNOE001)

SUPERVISED BY

Haroon Bhorat

Minor dissertation presented for the Degree of Masters of Economic Science

in the School of Economics

UNIVERSITY OF CAPE TOWN

February 2013

Declaration

This dissertation is submitted in partial fulfilment of the degree Masters of Economic Science at the University of Cape Town.

I know the meaning of plagiarism and declare that all of the work in this dissertation, save for that which is properly acknowledged, is my own. Each contribution to, and quotation in, this dissertation from the work(s) of other people has been attributed, and has been cited and referenced according to the Harvard system of referencing.

I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as their own.

Noel Verrinder

———————————

February 2013

Acknowledgments

Firslty, I would like to thank my supervisor Haroon Bhorat for his advice given during our meetings and determination in securing financial assistance. Without it the past year would have been far more of a struggle. I would also like to thank Waseema Petersen and Fahieda Akherwaray for organising my meetings with Haroon.

Finally, I thank my family and friends for their encouragement; in particular I would like to thank Robyn Verrinder for proofreading my work.

Abstract

This study investigates the relationships between several key economic, sociodemographic and infrastructural factors and crime in South Africa. Using the five March waves, 2003 to 2007, of the Labour Force Survey and crime statistics for the five-year period from April 2003 to March 2008 from the South African Police Service, this paper examines factors associated with aggregate property and violent crime rates at the district council level. The study uses pooled OLS and province fixed effects to estimate a reduced form equation. The coefficient on unemployment is initially found to be insignificant; however, after dividing the measure into youth and mature unemployment, a positive relationship between youth unemployment and both violent and property crime is revealed. Although, no significant relationship between income inequality and crime, even after decomposing the measure into between- and within-racial group components is found, racial heterogeneity within a district, perhaps a better measure of wealth inequality is found to be significant, with a particularly strong relationship with property crimes. Furthermore, holding all else constant, districts with a larger proportion of males and people aged between 15 and 35, on average experience more violent crimes per capita. Lastly, the number of police stations per capita is found to have a positive relationship with property crime, suggesting that there is endogeneity problem, which to investigate would require more extensive data on policing.

Table of Contents

List of Figures

List of Tables

I. INTRODUCTION

Over the past two decades, crime has become entrenched in the social landscape of South Africa. It has permeated the lives of all and affects many aspects of life. Beyond the direct effect on safety and lowering the utility of civilians, it has broader social and economic implications.

Crime increases the risks and costs of conducting business, and thus detracts investment by either deflecting resources towards security or making it unprofitable to invest entirely (Demombynes and Özler, 2005). This is not only relevant to the formal sector; crime has been found to be one of the largest perceived barriers to starting informal business in urban townships (Cichello *et al.,* 2006; Devey *et al.,* 2005; Kingdon and Knight, 2007), thus blocking a channel for upward social mobility.

The level of human capital is, also, negatively affected; crime is noted as being one of the greatest determinants driving the emigration of skilled professionals and, simultaneously, inhibits the immigration of skilled foreigners (Demombynes and Özler, 2005). Furthermore, rampant street crime can deter children from attending school and increase the associated opportunity costs of education, thus limiting human capital accumulation (Silverman, 2004).

At the public level, resources are diverted from other development imperatives towards the prevention of crime, enforcement of the law and correctional services. In the 2005/06 South African budget, R43,6 billion was allocated to the criminal justice system, representing 10,4% of the national budget (Altbeker, 2005). Moreover, violent crime has placed a burden on the health system, by weighing on the capacity of the trauma wards in public hospitals with inflicted injuries (Petra Brysiewicz, 2001)

Since Becker (1968) first attempted to model criminal behaviour, there has been much research on the factors behind an individual's choice to commit crime (Calvó-Aremengol and Zenou, 2004; Imai and Krishna, 2004; Lochner, 2004; Silverman 2004; Verdier and Zenou 2004) and the drivers of aggregate crime rates (D'Allesso *et al.*, 2002; Demombynes and Özler, 2005; Durlaf *et al.*, 2010; Ehrlich, 1973; Fajnzylber *et al.,* 2002; Imrohoroglu *et al.,* 2004; Levitt, 1997). The literature has attempted to merge social and economic theory to develop models of crime.

The purpose of this paper is to investigate the ability of various sociodemographic, economic and infrastructural characteristics have in explaining the variation of crime across different districts in South Africa.

First, studies, which have attempted to model crime, are reviewed in section II, after which the construction and sources of the data used in this paper are described in section III. Section IV discusses the methodology and presents a motivation for the variable choice. The results of the regressions are presented in section V, followed by a discussion in section VI. Finally, the paper concludes in section VII.

II. Review of Related Studies

This section reviews pertinent studies, which model crime both on the aggregate and from the perspective of the individual. The first focus is on the studies, which have researched social and economic relationships with aggregate crime rates. These are used to establish a framework for the identification strategy employed in this paper. The second part introduces studies, which have modelled crime from the perspective of the individual, which are used to inform on the variable choice in section V.

Demombynes and Özler (2005) use cross sectional data from 1996 to determine the main factors driving crime in South Africa. In particular their aim is to analyse the effect that inequality, within- and between-racial groups, has on violent and property crime rates. They theorise that variations in welfare drive crime through three channels. Firstly, the differential in returns to legitimate and illegitimate activities incentivise or disincentivise criminal behaviour. Secondly, the lack of social mobility of the poor leads to moral disintegration and unrest. Lastly, different levels of investment in private security result in a spread of crime not indicative of possible gain from committing crime.

Households from the 1996 South African census are matched with precincts from the crime statistics for 1996 reported by SAPS, after which economic variables are generated using data from the October Household Survey and Income expenditure survey. The analysis takes place at the precinct level, however, variables generated for a neighbourhood around each precinct are included. This presents a problem, as the economic variables generated may not be a true representation of the conditions of a precinct. Both the IES and OHS have a sample of about 30 000 households, which may suffer from a sampling bias at the precinct level.

Nevertheless, Demombynes and Özler focus on six different crimes, of which they categorise four into two different groups, property crimes and violent crimes, while leaving murder and robbery with aggravating circumstances uncategorised. They estimate the relationships between these crimes and different welfare measures using a negative binomial regression model, as well as OLS. They find that both models produce similar results.

They find that within-racial group inequality is highly correlated with property crimes, while between-racial group inequality has little correlation. Mean expenditure, used as a

proxy for wealth, in a precinct, which is argued to represent returns from crime, is positively correlated with property crime. Furthermore, if a precinct is the wealthiest in its neighbourhood, on average there will be more property crimes, suggesting that inequality between precincts is an important determinant.

They find that violent crime is associated with higher levels of inequality in the criminal catchment area of precincts, rather than inequality at the precinct level. Mean expenditure of a precinct has a non-linear relationship with violent crime, having an inverted U-shape relationship.

Additionally, they find that unemployment is uncorrelated with violent crime, except in the case of murder. Interestingly, despite the findings that between-racial group inequality is uncorrelated with crime, racial heterogeneity in a precinct is highly correlated with all types of crime.

Imrohoroglu *et al.* (2004) attempt to model the movements in aggregate property crime rates in the USA for the period 1980 to 1996. During this time, the USA experienced a large decrease in the property crime rates. They conjecture that the decrease was a result of increased spending on law enforcement, lower overall unemployment, higher earnings and fewer unskilled workers. However, during the same period, youth unemployment and inequality rose considerably.

To try and explain the movements in crime, Imrohoroglu *et al.* (2004) create a model and calibrate it using data from 1980. Similarly to Demombynes and Özler's results, they find that aggregate unemployment does not have a large impact on crime. Rather than unemployment, the model shows that the most important factors driving the decrease in crime were the aging population, the stronger economy and the increased probability of apprehension.

However, the two strongest relationships with property crime that were uncovered in the study were inequality and the probability of apprehension, with the effect of inequality being larger. This suggests that if inequality had stayed constant over the period from 1980 to 1996, the USA would have seen even larger decreases in the crime rate. Two other noteworthy findings of the model and that are supported by the data, are that approximately half of those who commit crime do not have a high school education and that most crime is committed by the youth.

Imrohoroglu *et al.* (2004) suggest that adding a variable that accounts for social networks and an indicator for social stigma will explain the decrease in the crime rate better. This is in accordance with Imai and Krishna's (2004) theory on criminal history hurting future employment opportunities, but contrary to Silverman's (2004) thesis on street crime and reputation, which is discussed in greater detail on **page six**.

Unlike Imrohoroglu *et al.'s* (2004) focus on property crime, Fajnzylber *et al.* (2002) attempt to disentangle the causes of violent crime. They do this by using panel data from 45 countries for the period 1970 to 1994. Using country level rates for intentional homicide and robbery, they estimate the explanatory power of economic, social, demographic and institutional variables.

Their results show that increases in income inequality are associated with higher violent crime rates, crime trends are countercyclical and that there is a great level of criminal inertia. However, it is important to note that cross country analyses may not produce accurate relationships, as such a high level of aggregation disallows an in depth analysis of local determinants. Though, further support for this relationship has been found by an intra-city study by Scorzafave and Soares (2009), who find that income inequality has a very strong relationship with property crimes in the Brazilian city of São Paulo.

Common in the above aggregate studies, are the findings that there is a significant relationship between inequality and crime, while unemployment does not seem to bare much importance. A study by Levitt (1997) on the effects of policing in the USA, also found that while unemployment has a meagre effect on property crime, it did not appear to have any significant relationship with violent crime.

There have been a number of studies modelling the individual's choice to commit crime. Although this study's focus is on aggregate crime levels, the individual behavioural models are important, as they advise the variable choice for the aggregate model. What determines the individual's choice of whether or not to commit a crime will inevitably drive the choice for many individuals, thus determining the crime rate.

The investigated studies, model the individual's choice as a cost-benefit analysis (Imai and Krishna, 2004; Lochner, 2004; Silverman, 2004; Verdier and Zenou, 2004). If the expected benefit of the crime outweighs the expected cost, then the individual will commit the crime. This conjecture forms the basis of the variable choice for this research paper.

There is a general consensus that earning a low wage or no wage at all decreases the opportunity cost of crime and that if other people are wealthier; the expected payoff of crime is higher (Imai and Krishna, 2004; Lochner, 2004; Verdier and Zenou, 2004). Furthermore, education has been accepted to increase the opportunity cost of crime, by inducing higher expected wages and social conditioning (Lochner, 2004).

However, the role of social capital is less certain. Silverman (2004) asserts that greater social capital and connectivity is associated with a greater benefit of committing crime. He suggests a theory of reputation, where people engage in crime to develop a reputation that will protect them from being a victim of crime themselves. Thus, in a community with greater social connectivity, the signal developed by engaging in crime is better communicated than in one where people are less likely to know each other.

To the contrary, Imai and Krishna (2004) state that social capital increases the cost of committing crime through stigmatization. As with Silverman's theory of reputation, Imai and Krishna base their theory on the increased ability of a tight-nit community to identify a criminal. They argue that the cost of committing a crime is higher for a person in a well-connected community, as criminals are viewed to be less productive, thus their future potential employment and wages are negatively affected. Additionally, they assert that criminals are more likely to be alienated or reported to the authorities in communities with a stronger social network.

Verdier and Zenou (2004) maintain that race is an important determinant of one's engagement in criminal activities, even after controlling for other social and economic factors. They assert that this is a result of statistical discrimination and how it becomes a self-fulfilling prophecy. However, D'Allesso *et al.* (2002) find that this does not extend to violent crimes in the USA.

The following section discusses the construction of the data and provides summary statistics for the sample

III. DATA

The data used in this paper are constructed by merging crime statistics, collected from the South African Police Service (SAPS) website, with the Labour Force Survey (LFS) from 2003 to 2007.

South African crime statistics are released annually and are made available to the public on the SAPS website. The data are collected from all 1127 precincts[1] yearly for the period April to March and the first available year of statistics is 2003. Crimes are grouped into 24 main categories with six subcategories for aggravated robbery; however, only four of the subcategories are available at precinct level, with street and bank robberies only being available at provincial level.

During the five-year period April 2003 and March 2008, there have been multiple changes to the names of police stations, as well as the ceding and establishment of new police stations. All stations established after April 2008, of which there are 12, are dropped from the sample, as they record zeros for all crimes before their creation. Twenty-three new stations were established between April 2003 and March 2008, while there was no merging of precincts during the same period.

Aggregate demographic, sociodemographic and economic variables were generated using five March waves of the LFS for the period 2003 to 2007. The LFS is a biannual household survey of 74 591 adults of working age in 30 000 households across South Africa. The main purpose of the survey is to collect data on the South African labour market. It was run from 2000 to 2007 when the Quarterly Labour Force Survey (QLFS) eventually succeeded it.

The LFS is recorded at the district council level, of which there were 51 during the time frame investigated. These included six metropolitan areas. The 1 115 police precincts were matched to their respective district council. This was achieved by downloading a map overlay of the district councils for Google Earth™ from the Municipal Demarcation Board's website and searching for each police station's geographical location.

Once the police station was found, it was recoded as a match to the district council. In the rare case where a precinct straddled district councils, the district in which the police

[1] A precinct is a jurisdiction of a police station. The whole of South Africa is broken down into precincts, so

station was located is preferred. The sample consists of 51 districts and five years, making a pooled sample of 255.

The next section presents the reduced-form model and discusses the rationale behind the variable choice.

IV. METHODOLOGY

This section introduces the reduced-form specification to be estimated in section V. A motivation for the specification is provided, along with a discussion on each regressand and regressor. Furthermore, an explanation of how each variable is constructed is offered.

The foundation of the aggregate crime model presented in this paper is based on Becker's (1968) assertion that rational individuals decided whether to commit crime based on a thought-out cost-benefit[2] analysis. However, the factors influencing one individual's choice, will influence other individuals' choices as well, and thus affect the aggregate crime rate. Furthermore, the decisions of others to engage in crime will have a feedback effect and alter the costs and benefits of crime for the individual. Thus, each variable included in the reduced-form is justified by how it affects the individual's decision to engage in crime.

The purpose of this study is to investigate the relationship between district level conditions and crime and establish the relative importance of such conditions. Thus, a reduced-form specification is used to estimate these relations.

$$C_{ijt} = E_{jt}\beta_1 + S_{jt}\beta_2 + I_{jt}\beta_3 + \lambda_k + \varepsilon_{ijt} \quad (1)$$

The C_{ijt} represents the logged crime rate 'i', where i is either 'p' (property) or 'v' (violent), for the jth district council during time period 't'. The conditions for district 'j' at time 't' are split into matrices E_{jt}, S_{jt} and I_{jt}, which contain economic, sociodemographic and infrastructural variables respectively.

To control for the unobserved heterogeneity between provinces, such as local crime preferences, the level of government involvement and other unobserved institutional factors, a fixed-effect estimator for provinces, λ_k, is included. Finally, ε_{ijt} represents the error term.

[2] This paper does not assume that all costs and benefits are pecuniary and that other non-pecuniary flows, such as reputation and guilt, are considered by the individual.

Dependents

This paper examines both violent and property crime. Composite rates are constructed for both using the SAPS's definitions with a few adaptions.

The property crime index contains: Burglary at non-residential premises, Burglary at residential premises, Theft of motor vehicle and motorcycle, Theft out of or from motor vehicle and Stock-theft.

In this paper violent crime is defined, as crime from which there is no obvious pecuniary gain. Murder, Total Sexual Crimes, Attempted murder, Assault with the intent to inflict grievous bodily harm and Common assault are included in the index. Common Robbery and Robbery with aggravating circumstances are omitted, as these crimes are likely motivated by the prospect of pecuniary gain.

Independents

This paper examines three aggregate groups of conditions: (i) economic, (ii) sociodemographic, and (iii) infrastructural, into which the selected variables are divided. Each group, and the variables it contains, are discussed below.

(i) Economic variables

Economic variables are defined, in this study, as any variable, which is a direct measure of the level of welfare in a district. These variables can affect the expected benefits and opportunity costs associated with engaging in criminal activities.

Income

The average household income per capita is used as a proxy for wealth. The relationship between income and crime is not clear, as income could have opposing effects on crime. On the one hand, a higher level of income increases the possible reward associated with committing crime, while on the other, higher levels of income increase the opportunity cost of crime and decrease the need to generate wealth through criminal activities.

Furthermore, agents at a higher wealth level may choose to dedicate more resources towards protecting themselves, by purchasing private security, thus increasing the

costs of committing crime in an area with more wealthy households (Demombynes and Özler, 2005).

An initial look at the two-way scatter plots shows a positive relationship between mean household income per capita and both crime indices.

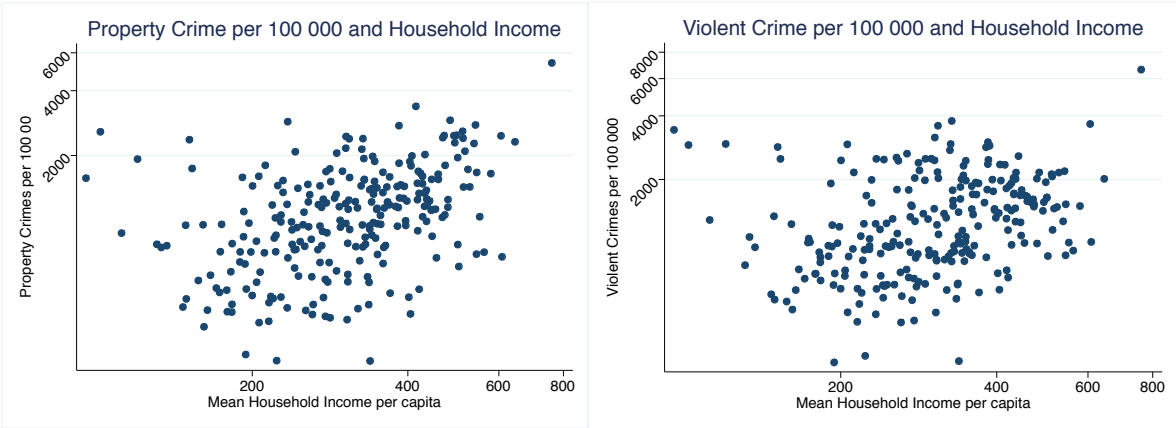

Figure 1. Scatter plots illustrating relationship between mean household income per capita and crime

Unemployment

Despite the literature's findings that unemployment is uncorrelated with crime (Demombynes and Özler, 2005; Imrohoroglu *et al.*, 2004), there is reason to believe that unemployment should be associated with higher property crime rates. If an individual is unemployed, they have lower opportunity costs and a greater need for income than the employed.

It is suggested that youth unemployment may be associated with crime, while aggregate unemployment is not (Demombynes and Özler, 2005; Imrohoroglu *et al.*, 2004). Thus, two different specifications of unemployment will be used in the empirical analysis of this study. Unemployment will be divided into youth and mature unemployment.

Inequality

The main assertion behind the finding that inequality is associated with higher crime rates is that the differential in wealth results in high payoffs for the poor stealing from the rich. However, this does not provide a theory behind inequality's correlation to violent crime.

Persistent inequality may lead to social unrest and resentment, thus increasing the non-pecuniary benefit associated with harming the wealthy (Demombynes and Özler, 2005). The level of social unrest and resentment may be intensified by the sensitivity towards inequality between groups, such as race.

This study uses the Theil measure of inequality, calculated using average household income per capita. The Theil measure is chosen above other measures, such as the gini coefficient for its ability to be decomposed into within- and between-group components.

Education

The impact of education on the decision to commit a crime is multifaceted. Firstly, higher levels of education are associated with higher wages (Lochner, 2004). Thus, assuming that committing a crime takes time away from other productive activities, the opportunity cost of committing such crime will be higher for individuals with a higher level of education.

Furthermore, schools not only provide productivity enhancing education or a signal to employers, but also offer a platform for social and moral moulding (Lochner, 2004). Hence, schooling may increase the guilt cost of committing crime and lower the non-pecuniary gain from engaging in violence. Finally, the youth engage in crime more than any other age group (Demombynes and Özler, 2004; Imrohoroglu *et al.*, 2004), thus keeping teenagers in school will limit the range of possible criminal activities available to them. There is possible reverse causality, as crime, in particular street crime, may inhibit the youth's ability to attend school.

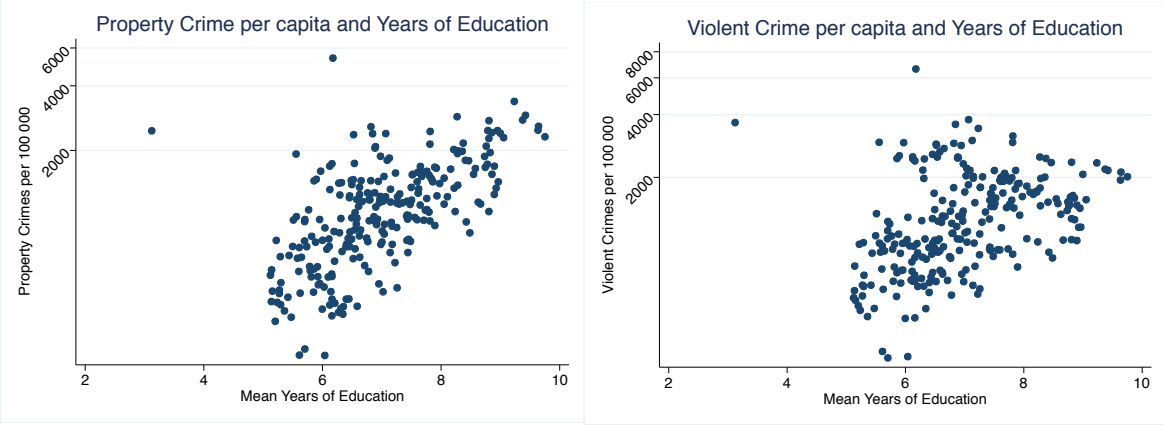

Figure 2 Scatter plot illustrating the relationship between mean years of education and crime

The mean number of completed years of education is used in this analysis. Surprisingly, the scatter plots illustrate a positive relationship. However, without controlling for any other economic factors, it is likely that education is absorbing the role of other pertinent variables, such as inequality.

(ii) Sociodemographic variables

A sociodemographic variable is one, which describes the demographic or social makeup of a district. The sociodemographic make-up of a community determines the social behaviour of the agent and their interactions with others. Furthermore, the sociodemographic characteristics of an individual can indicate or influence their preference for crime.

Age and Gender

Age is widely cited as an important determinant of one's preference to engage in criminal activities (Demombynes and Özler, 2005; Imai and Krishna, 2004; Imrohoroglu *et al.*, 2004; Levitt, 1997; Lochner, 2004). Therefore, a youth variable is included. The variable used is the percentage of people between the ages 15 and 35 within a district.

Gender has also shown to be a powerful indicator of one's probability to commit crime (Imai and Krishna, 2004). Almost all criminals arrested in the USA are male (Imrohoroglu *et al.*, 2004); hence this study includes the percentage of males living in a district. As can be seen in the scatter plot (figure 3), there appears to be some positive relationship between gender composition and crime.

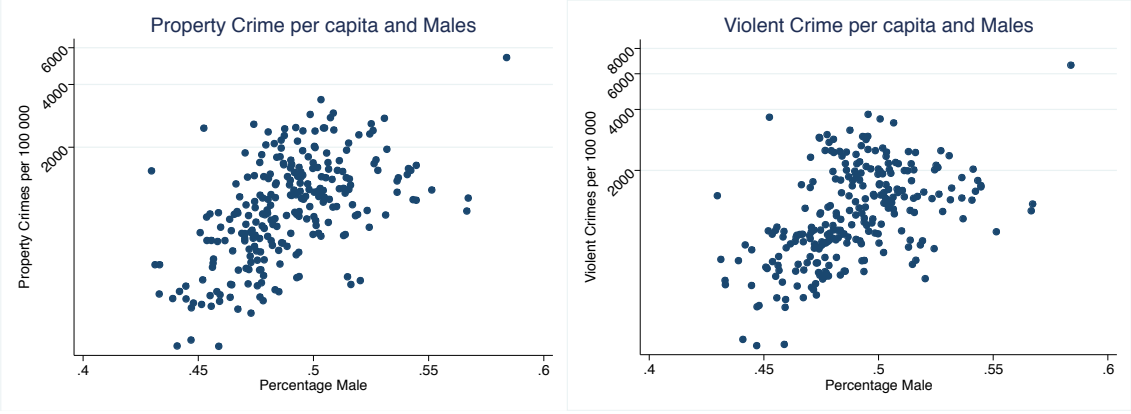

Figure 3 Percentage of Males per district

Racial and Linguistic Make-up

The racial make-up of the districts is taken into account to provide a measure of financial support networks. Wealth is still highly correlated with race in South Africa. Furthermore, social networks in South Africa are cited to still be heavily determined on racial lines (Adato *et al.*, 2006). These social networks are important, as the implications of becoming unemployed or losing income can be cushioned by financial support from others in your social network (Adato *et al.*, 2006). Thus, it is hypothesised that districts with larger populations of weaker financial social groupings will experience higher crime rates; this is investigated in the regression analysis.

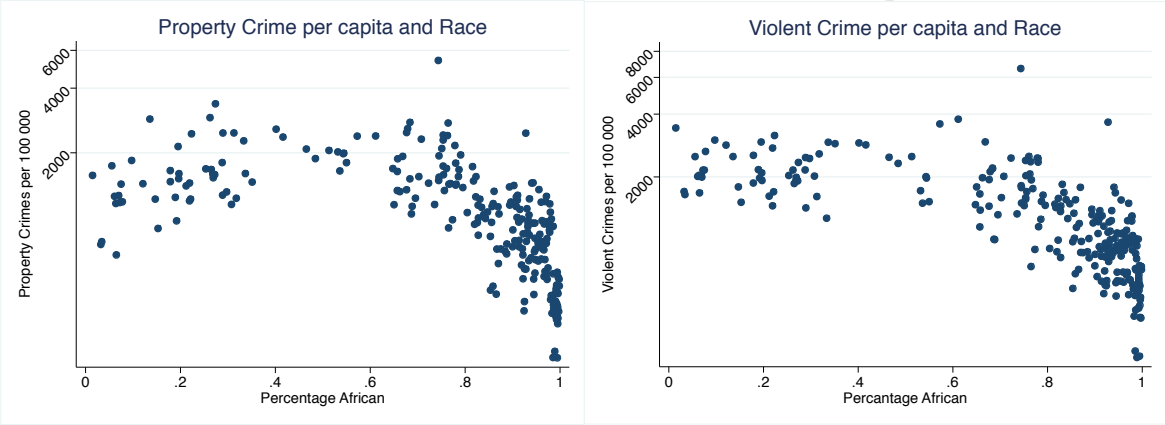

Figure 4 Percentage African in a district

In figure 4, plotting the percentage of Africans in a district against crime, a slight inverted U-pattern can be seen. This could indicate that racial heterogeneity is associated with higher crime rates and is masking the effects of the social support network hypothesized above.

This phenomenon, also found by Demombynes and Özler (2005), may be explained by an individual's preference to commit a crime against someone belonging to a different patent group over someone belonging to their own group. If an individual feels a sense of belonging, they may associate more guilt with committing a crime against someone belonging to there own group. This may be exacerbated by South Africa's racial history and oppression. Another possibility is that race may be a better indicator of wealth than income, and this index is acting as a proxy for inequality.

To control for this an index of racial homogeneity is generated using the sum of the squared percentages of the racial groups. An additional measure is included for linguistic heterogeneity to test whether the preference hypothesis extends to less patent features.

Crime Rate and Population density

Population density is used as a proxy for number of criminal opportunities, which is expected to raise crime rates. A lagged variable for the crime rate is included to test for the persistence of crime in a district.

(iii) Infrastructural variables

Variables relating to government involvement are labelled as infrastructural. The amount of enforcement and investment by government will impact on the costs and benefits of crime.

Police

The SAPS do not release the number of active officers per police station or expenditure by station; therefore, to account for the level of enforcement, the number of police stations per capita is used as a rough measure. There is likely to be an issue of endogeneity, as noted by Levitt (1997), which can be seen in the scatter plot (figure 5), particularly for violent crime.

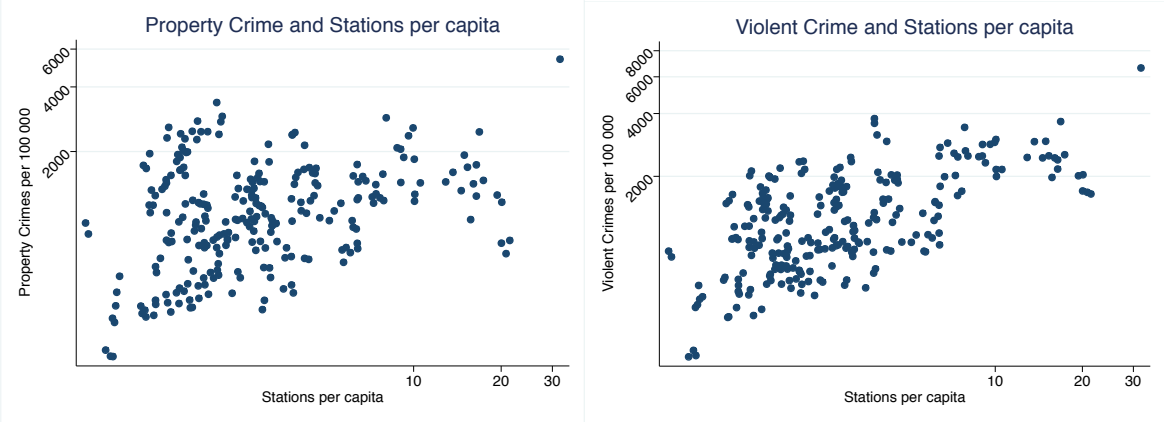

Figure 5 Police Stations per capita

Services

To control for the level of government service delivery, a proxy is used. This proxy is the percentage of households that have needed to collect water in the past week. Low levels of infrastructure can result in lower living standards and thus, may increase the associated benefit of committing crime.

Table 1 Summary Statistics of Sample

	2003	*Std. Dev.*	2004	*Std. Dev.*	2005	*Std. Dev.*	2006	*Std. Dev.*	2007	*Std. Dev.*
CRIME										
Violent crime per 100 000 people	1665.602	*(1076.716)*	1534.979	*(843.7142)*	1310.421	*(617.7342)*	1256.559	*(585.9224)*	1198.443	*(588.5813)*
Property crime per 100 000 people	1451.107	*(937.8034)*	1221.646	*(672.2209)*	1093.839	*(535.7701)*	1055.166	*(508.9711)*	1007.744	*(483.9624)*
ECONOMIC										
Average HH income per capita	275.5394	*(118.9182)*	307.504	*(120.8652)*	311.4364	*(97.92978)*	335.2499	*(105.0601)*	376.8295	*(105.595)*
Unemployment (Broad)	.2643652	*(.0639209)*	.258346	*(.0706688)*	.2534425	*(.0544981)*	.2414537	*(.0565789)*	.2379953	*(.0515699)*
Youth Unemployment	.3151549	*(.0648541)*	.3157577	*(.1036884)*	.2941628	*(.0561197)*	.2804129	*(.0522196)*	.2822528	*(.060809)*
Mature Unemployment	.2461976	*(.0766524)*	.2378456	*(.0809459)*	.2395181	*(.0657491)*	.2287152	*(.0750184)*	.2226362	*(.0625709)*
Theil index of inequality	.9351139	*(.1920228)*	.9121456	*(.1881772)*	.9701698	*(.1835101)*	.9544193	*(.1824012)*	.9130822	*(.1445605)*
Within-race component	.783536	*(.2199574)*	.7790651	*(.2169791)*	.8491531	*(.1926004)*	.8087041	*(.1921511)*	.7857407	*(.1711701)*
Between-race component	.2206637	*(.1425305)*	.2005921	*(.1442545)*	.1882058	*(.1172068)*	.2169387	*(.1226475)*	.1862828	*(.1268405)*
Average years of Education	6.89147	*(1.163586)*	6.923225	*(1.298267)*	6.893144	*(.9991257)*	7.022908	*(.9891925)*	7.138097	*(.9817302)*
SOCIODEMOGRAPHIC										
% of the population between 15 and 35	.3732114	*(.0361594)*	.3714008	*(.0405485)*	.3710958	*(.0305212)*	.3719609	*(.0306863)*	.3696361	*(.032996)*
Average Age	26.36035	*(2.137023)*	26.29477	*(2.215825)*	26.41875	*(1.967778)*	26.6366	*(1.982838)*	26.80041	*(2.014528)*
% Male	.4911982	*(.0296615)*	.4900842	*(.0274414)*	.4885146	*(.0222362)*	.4892484	*(.0217641)*	.4860799	*(.024478)*
% African	.7403542	*(.2895299)*	.7481987	*(.2875483)*	.7487478	*(.2903332)*	.7521495	*(.2874903)*	.751057	*(.2885076)*
% Coloured	.1554799	*(.2574319)*	.1485198	*(.2518874)*	.1536173	*(.2595151)*	.1531912	*(.2565043)*	.1549773	*(.2599423)*
% Indian	.0132086	*(.0345633)*	.0134165	*(.0355533)*	.0137267	*(.0316035)*	.0132115	*(.029459)*	.0128755	*(.028737)*
% White	.0909573	*(.0803038)*	.089865	*(.0822144)*	.0839082	*(.0676793)*	.0814478	*(.0675151)*	.0810902	*(.0647533)*
Population Density	160.5387	*(332.4031)*	162.0624	*(338.2522)*	179.1619	*(403.1506)*	186.4878	*(429.1512)*	189.0632	*(433.8675)*
Index of Racial Homongeneity	.7353952	*(.2037931)*	.7412471	*(.2077242)*	.7455884	*(.1914593)*	.7468588	*(.1904718)*	.7476161	*(.1896808)*
Index of Linguistic Homogeneity	.5922335	*(.2528743)*	.6024526	*(.2563984)*	.5995204	*(.2425191)*	.603777	*(.2455439)*	.5930903	*(.2423528)*
INFRASTRUCTURAL										
Stations per 1000 people	21.45098	*(10.97509)*	21.64706	*(11.26201)*	21.70588	*(11.31953)*	21.78431	*(11.30896)*	21.86275	*(11.35961)*
% of HH who need to fetch water	.1483511	*(.1212434)*	.1032957	*(.0030214)*	.1254726	*(.1154437)*	.1261219	*(.1157329)*	.1174865	*(.1098772)*

Above is a table containing the summary statistics for the variables of interest. The figures are the mean statistics for all districts for each year. The next section presents the results of the pooled OLS and fixed effects regressions, with different specifications.

V. RESULTS

The following section presents the regression output and provides variations of the specification.

Table 2 presents the results from running the pooled OLS regression in column 1 and 2, after which province fixed effects are included in columns 3 to 6, with an additional control for the two-year lagged crime rate in column 5 and 6. All coefficients except those on the measures of inequality, years of education and stations per capita can be interpreted as elasticities. Robust standard errors have been used to control for heteroskedasticity.

Table 2 Pooled OLS and fixed effects regressions

VARIABLES	(1) Property Crime	(2) Violent Crime	(3) Property Crime	(4) Violent Crime	(5) Property Crime	(6) Violent Crime
ECONOMIC						
Average Household Income per capita (log)	-0.218**	-0.184*	-0.210**	-0.158	-0.107	-0.0890
	(0.0890)	(0.108)	(0.0756)	(0.0975)	(0.183)	(0.149)
Unemployment	-0.754	-0.531	0.0326	-0.318	-0.925	-1.026
	(0.766)	(0.607)	(0.953)	(0.765)	(0.900)	(0.752)
Theil index of Inequality	0.249*	0.259*	0.132	0.156	0.0811	0.117
	(0.137)	(0.148)	(0.110)	(0.136)	(0.146)	(0.159)
Average Years of Education	0.111	0.00836	0.173	0.0335	0.0699	-0.0389
	(0.111)	(0.0947)	(0.108)	(0.0942)	(0.0890)	(0.0998)
SOCIODEMOGRAPHIC						
% aged 16-35	2.822**	2.172*	2.672	2.828*	1.588	2.249
	(1.368)	(1.234)	(1.504)	(1.328)	(1.554)	(1.629)
% male	3.226**	3.418***	1.903	2.650*	0.135	1.417
	(1.375)	(0.952)	(1.899)	(1.337)	(1.412)	(1.222)
% African	0.536	0.0580	1.105**	0.253	-0.0409	-0.707*
	(0.351)	(0.298)	(0.364)	(0.387)	(0.262)	(0.312)
Racial homogeneity	-1.736***	-1.278***	-2.060***	-1.355**	-0.540	-0.127
	(0.452)	(0.427)	(0.511)	(0.471)	(0.474)	(0.339)
Linquistic homogeneity	0.203	-0.0365	0.333	0.100	-0.139	-0.114
	(0.302)	(0.229)	(0.280)	(0.125)	(0.241)	(0.173)
Population	-0.168*	-0.124	-0.135**	-0.146**	-0.0333	-0.0631*
	(0.102)	(0.0843)	(0.0548)	(0.0551)	(0.0578)	(0.0277)
Population density	0.0768	-0.0221	0.0410	-0.0184	-0.0198	-0.0299
	(0.0777)	(0.0532)	(0.0629)	(0.0563)	(0.0309)	(0.0272)
Two year Lagged Property Crime rate					0.488***	
					(0.118)	
Two year Lagged Violent Crime rate						0.499***
						(0.105)
INFRASTRUCTURAL						
Stations per capita	0.0324***	0.0236**	0.0464***	0.0249*	0.0132	0.000162
	(0.00968)	(0.00999)	(0.0101)	(0.0111)	(0.00809)	(0.0109)
Proxy for Service delivery	-0.858	-0.971	-0.547	-0.565	-0.411	-0.307
	(0.868)	(0.684)	(1.072)	(0.891)	(0.627)	(0.429)
Constant	7.413***	8.200***	6.880***	8.104***	4.086*	4.533*
	(1.499)	(1.550)	(1.356)	(1.467)	(2.052)	(2.152)
Observations	255	255	255	255	153	153
R-squared	0.722	0.756	0.606	0.540	0.805	0.794
Number of prov			9	9	9	9

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Dependents: Log of crimes per 100 000

Firstly, all coefficients become insignificant when controlling for the respective lagged crime rate. However, it is important to note that the sample size is dramatically reduced from 255 to 153 observations. Although, the sign on income is negative under all specifications, it is only significant at the five percent level for property crimes, with a coefficient of around -0.21. Aggregate unemployment[3] has no significant relationship with either crime index. Nevertheless, to investigate this relationship further, unemployment is split by age in table 5.

The Theil index is used as the measure for inequality, because of its ability to decompose inequality into the within- and between-components. Using this measure, no significant relation between inequality and crime is found; however, the sign of the coefficient is consistently positive. In table 4 inequality is broken into a within- and between-race component, to examine this relationship closer. Despite education having no significant relationship with crime, its sign is consistently positive under the first two specifications.

The percentage of people aged between 16 and 35 and the percentage of males in a district have large positive coefficients, with both being significant for property crime and only the percentage of males being significant for violent crime when looking across all districts. However, when comparing within provinces, these coefficients lose their significance.

Regarding race, looking within provinces, a percentage point increase in the percentage of Africans in a district is associated with a 1.105 percent increase in the number property crimes per capita. Racial homogeneity has a large significant negative correlation on both crime indices, although, the findings show the relationship is stronger with property crime. The index of linguistic homogeneity is found to have no significant correlation.

Interestingly, population density has no significant relation to either crime rate. However, when looking with provinces, holding all else constant, on average the more people in a district, the lower the crime rate.

The number of police stations per capita is associated with higher property crime rates, though the coefficient is only significant for violent crime when looking across all districts.

[3] The broad definition of unemployment is used

Nevertheless, the sign is consistently positive. The coefficient on the proxy for service delivery is consistently negative[4] and found to be insignificant.

To test for a possible non-linear relationship between average household income per capita and crime. A quadratic in income is included in the regression. Table three reports the findings; only the coefficients on the income variables are included, as all the other results remain largely unchanged.

Table 3 Fixed effects regressions with a quadratic for income

VARIABLES	(3) Property Crime	(4) Violent Crime	(7) Property Crime	(8) Violent Crime
ECONOMIC				
Average Household Income per capita (log)	-0.210**	-0.158	-1.239	-1.066
	(0.0756)	(0.0975)	(1.685)	(1.543)
Average Household Income per capita Squared			0.0920	0.0812
			(0.151)	(0.139)
Constant	6.880***	8.104***	9.788	10.67
	(1.356)	(1.467)	(4.933)	(5.000)
Observations	255	255	255	255
R-squared	0.606	0.540	0.608	0.541
Number of prov	9	9	9	9

Robust standard errors in parentheses
*** p<0.01, ** p<0.05
Dependents: Log of crimes per 100 000 people

Column 3 and 4 represent the coefficients of the original specification with province fixed effects, while column 7 and 8 include the quadratic. As can be seen in the table, there is no evidence to suggest a non-linear relationship of this form, in fact, the inclusion of the income variable squared results in much larger standard errors and the loss of significance.

Table 4 presents the decomposition of inequality into within- and between-racial groups. Again column 3 and 4 are the original regressions, while 9 and 10 present the regression with inequality broken down. Inequality remains insignificant, with both measures falling outside of even the 10 percent significance level. Nevertheless, both coefficients are positive for violent and property crime, with between-race inequality having a much larger coefficient with greater significance. Average household income per capita loses its

[4] The proxy is the percentage of people having to walk to collect water for household use, thus the larger the percentage, the lower the level of service delivery.

significance in column 9, along with the coefficients on population in both columns 9 and 10.

Table 4 Fixed effects regressions with inequality decomposed into within- and between-race components

VARIABLES	(3) Property Crime	(4) Violent Crime	(9) Property Crime	(10) Violent Crime
ECONOMIC				
Average Household Income per capita (log)	-0.210**	-0.158	-0.179*	-0.106
	(0.0756)	(0.0975)	(0.0925)	(0.111)
Unemployment	0.0326	-0.318	-0.117	-0.506
	(0.953)	(0.765)	(1.018)	(0.788)
Theil index of Inequality	0.132	0.156		
	(0.110)	(0.136)		
Within Race Inequality			0.0736	0.0750
			(0.108)	(0.132)
Between Race Inequality			0.301	0.315
			(0.242)	(0.190)
Average Years of Education	0.173	0.0335	0.163	0.0168
	(0.108)	(0.0942)	(0.110)	(0.0955)
SOCIODEMOGRAPHIC				
% aged 16-35	2.672	2.828*	2.779	3.023*
	(1.504)	(1.328)	(1.549)	(1.344)
% male	1.903	2.650*	1.886	2.485
	(1.899)	(1.337)	(1.936)	(1.365)
% African	1.105**	0.253	1.122**	0.285
	(0.364)	(0.387)	(0.371)	(0.377)
Racial homogeneity	-2.060***	-1.355**	-2.012***	-1.324**
	(0.511)	(0.471)	(0.542)	(0.468)
Linquistic homogeneity	0.333	0.100	0.312	0.0689
	(0.280)	(0.125)	(0.315)	(0.151)
Population	-0.135**	-0.146**	-0.124*	-0.126*
	(0.0548)	(0.0551)	(0.0601)	(0.0568)
Population density	0.0410	-0.0184	0.0402	-0.0215
	(0.0629)	(0.0563)	(0.0650)	(0.0571)
INFRASTRUCTURAL				
Stations per capita	0.0464***	0.0249*	0.0462***	0.0256*
	(0.0101)	(0.0111)	(0.0104)	(0.0111)
Proxy for Service delivery	-0.547	-0.565	-0.438	-0.411
	(1.072)	(0.891)	(1.038)	(0.850)
Constant	6.880***	8.104***	6.586***	7.688***
	(1.356)	(1.467)	(1.414)	(1.516)
Observations	255	255	253	253
R-squared	0.606	0.540	0.611	0.548
Number of prov	9	9	9	9

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Dependents: Log of Crimes per 100 000

As mentioned above, the relationship between crime and unemployment is investigated in table 5. Unemployment is split into two rates, namely, mature unemployment, which is the

unemployment rate for those above the age of 25, and youth unemployment[5], which is defined as the percentage of people unemployed between the ages of 15 and 25. Inequality remains decomposed.

Table 5 Fixed effects regressions with unemployment split into mature and youth rates

VARIABLES	(9) Property Crime	(10) Violent Crime	(11) Property Crime	(12) Violent Crime
ECONOMIC				
Average Household Income per capita (log)	-0.179*	-0.106	-0.233**	-0.146
	(0.0925)	(0.111)	(0.0953)	(0.119)
Unemployment	-0.117	-0.506		
	(1.018)	(0.788)		
Mature Unemployment			-2.235*	-1.924*
			(0.989)	(0.906)
Youth Unemployment			1.786***	1.195**
			(0.401)	(0.384)
Within Race Inequality	0.0736	0.0750	0.122	0.110
	(0.108)	(0.132)	(0.0954)	(0.121)
Between Race Inequality	0.301	0.315	0.395	0.379*
	(0.242)	(0.190)	(0.226)	(0.172)
Average Years of Education	0.163	0.0168	0.238**	0.0726
	(0.110)	(0.0955)	(0.0776)	(0.0730)
SOCIODEMOGRAPHIC				
% aged 16-35	2.779	3.023*	2.769*	2.988**
	(1.549)	(1.344)	(1.316)	(1.128)
% male	1.886	2.485	0.805	1.694*
	(1.936)	(1.365)	(1.107)	(0.818)
% African	1.122**	0.285	1.492***	0.550
	(0.371)	(0.377)	(0.398)	(0.434)
Racial homogeneity	-2.012***	-1.324**	-2.072***	-1.369**
	(0.542)	(0.468)	(0.535)	(0.472)
Linquistic homogeneity	0.312	0.0689	0.436	0.160
	(0.315)	(0.151)	(0.324)	(0.142)
Population	-0.124*	-0.126*	-0.104	-0.111
	(0.0601)	(0.0568)	(0.0674)	(0.0637)
Population density	0.0402	-0.0215	0.0218	-0.0350
	(0.0650)	(0.0571)	(0.0673)	(0.0580)
INFRASTRUCTURAL				
Stations per capita	0.0462***	0.0256*	0.0416***	0.0223*
	(0.0104)	(0.0111)	(0.0101)	(0.0109)
Proxy for Service delivery	-0.438	-0.411	0.161	0.0371
	(1.038)	(0.850)	(0.782)	(0.673)
Constant	6.586***	7.688***	6.266***	7.450***
	(1.414)	(1.516)	(1.245)	(1.434)
Observations	253	253	253	253
R-squared	0.611	0.548	0.669	0.593
Number of prov	9	9	9	9

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Dependents: Log of crimes per 100 000

[5] As defined by the International Labour Organisation

Youth unemployment has a large significant positive coefficient for property and violent crime, suggesting that there is, in fact, a relationship between crime and unemployment. For a one-percentage point increase in youth unemployment, property and violent crimes will increase by about 1.8 and 1.2 percent respectively. The coefficient on mature unemployment is insignificant at the five percent level, however it has a large negative coefficient for both property and violent crime.

After controlling for youth unemployment and mature unemployment separately, the coefficient on the income variable becomes significant for property crime again at -0.233. While, the coefficient on percentage of young becomes significant for violent crime. For a one-percentage point increase in the percentage of people aged between 15 and 35, there is almost a three percent increase in the number of violent crimes per capita.

The coefficients on racial homogeneity for both indices and the coefficients on percentage of Africans and number of stations per capita for property crimes remain significant in all specifications.

VI. DISCUSSION

This section provides a discussion of the results presented in the section V, as well as highlighting the caveats considered in the research.

Economic

Contrary to Demombynes and Özler's (2005) finding of a strong positive relationship between mean household expenditure and property crime and a significant non-linear relationship with violent crime, this research only finds a negative correlation between income and property crime. This finding is consistent with the theory that aggregate income acts as a proxy for the opportunity costs associated with committing crime.

Surprisingly, the results show no significant relationship between either crime index and inequality. This is unlike the findings of Imrohoroglu *et al.* (2004), who find that inequality is one of the largest determinants of property crime in the USA, and Fajnzylber *et al.* (2002), who find that it is associated with violent crime, globally. Even after decomposing inequality into within- and between-racial groups, there is still no significant relation.

However, unlike Demombynes and Özler (2005), who find that the within component is more important, the author finds that the between component has a more significant relationship with crime, albeit below the five percent level. It is important to note that income may not provide a good measure of wealth; therefore not revealing the true relationship between crime and inequality.

In line with the literature (Demombynes and Özler, 2005; Imrohoroglu *et al.,* 2004), aggregate unemployment is not found to have a significant relationship with violent or property crime. However, after splitting unemployment into youth and mature unemployment, youth unemployment emerges as highly significant, particularly for property crimes. This is consistent with the assertion that there are lower opportunity costs for the unemployed. Surprisingly mature unemployment is found to have a negative relationship with both indices; however, this is finding is only significant at the 10 percent level.

Despite Lochner's (2004) assertion that education is an important determinant of crime, the results do not pick up on any significant relationship. Furthermore, the visible

relationship seen in Figure 2 is likely cofounded by education's correlation to other variables.

Sociodemographic

The relationship between the percentage of young people does not emerge as clearly as it does in Demombynes and Özler's (2005) findings. When looking across the entire country, having more young people living in a district is associated with a higher property crime rate. However, when comparing within provinces, this becomes insignificant, perhaps suggesting that the initial relationship was controlling for the socio-geographic structure of the country, as well as youth unemployment. Only after controlling for youth unemployment, do we see a strong correlation with violent crime.

Looking across all districts in South Africa, the results show a very large positive relationship between males and crime, in particular violent crime. This significance falls away when looking within provinces. However, this is a result of the lack of variation in the gender composition of districts within provinces. Despite this, the initial findings are harmonious with the assertion that males are the main perpetrators of crime.

Despite its categorisation, race may be an economic indicator, as well as a sociodemographic factor. Within provinces, a higher percentage of Africans in a district is associated with a higher level of property crimes, but does not have an effect on violent crime. This is consistent with the hypothesis that a stronger financial social network may cushion economic misfortune.

One of the most robust results is that, holding all else constant; on average a more racially homogenous district will have fewer property and violent crimes per capita. The coefficient is particularly large for property crimes. This is consistent with the hypothesis that people of one group are more willing to commit a crime against someone belonging to another patent group. It is disturbing that this is true for violent crimes, where there is not necessarily a direct pecuniary payoff associated with committing the crime.

That being said, this result should be interpreted with caution, as it is possible that the index is not controlling for racial relations, but instead providing a proxy for wealth inequality, even after controlling for income inequality. Furthermore, there is no evidence in support of the preference theory from the findings on linguistic homogeneity.

Infrastructural

The number of police stations per capita in a district is associated with more property crimes per capita. The final regression indicates that for every additional police station per 100 000 people, holding all else constant, there will be on average a four percent increase in property crimes per capita. This finding is likely the result of an endogeneity problem, similar to that found by Levitt (1997). To test whether there is in fact an endogeneity problem, a better measure of policing would first need to be found, and unfortunately, the SAPS does not release statistics on the number of officers or spending per police station.

The proxy for other service delivery, the percentage of households who needed to collect water for household use, yielded no significant relationship. Providing no evidence that service delivery is directly associated with individual based property and violent crime.

There are two additional caveats to the research worth mentioning. Firstly, a social mobility measure is excluded, which is likely to be an important economic factor. Demombynes and Özler (2005) conjecture that social mobility is likely an important determinant, as the lack of upward social movement results in social unrest and tension. The reason a measure of social mobility was not included is that, with only five years of data, any direct measure looking at a reasonable time to expect social movement would result in a major reduction in this study's sample size.

Secondly, there may be a reporting bias, in that crime is more likely to be reported by some groups. For example, the wealthy may be more likely to report property crimes, as they are required to do so for insurance purposes or as they have a higher level of the understanding of what constitutes a criminal offence. Moreover, people living in high crime areas may become discouraged with the police and may be less likely to report crime.

However, the coefficient on our income variable does offer support for this. If there is a reporting bias, then the coefficient reports a lower bound and underestimates the relationship. Furthermore, Demombynes and Özler test for this and find that it does not alter their results significantly.

VII. CONCLUSIONS

This study set out to investigate the relationship between economic, sociodemographic and infrastructural factors and crime rates. The study uses pooled OLS and a fixed effects model to estimate a reduced form specification for aggregate crime rates in South African districts. The results uncover previously unfound links, reinforce current theory and findings, as well as reject others.

The previously unfound link between unemployment and crime in South Africa emerges after splitting unemployment by age group. Youth unemployment is an important factor associated with crime, particularly property crimes; for a one-percentage point increase in youth unemployment, a district can expect a 1.79 percent increase in property crime. Unlike other studies, no significant relationship is found between inequality and crime when using the Theil index on average household income per capita, even after decomposing into between- and within- racial group components.

However, a significant relationship between racial homogeneity within a district and crime is found, in particular property crime. We suggest caution in interpreting this result as an indication of racial tension and that instead that it is possibly a result of the measure acting as a close proxy for wealth inequality. Furthermore, the percentage of Africans in a district is positively correlated with property crime, supporting the hypothesis on financially supportive social networks. We find that age and gender are important, with districts with more people between the age of 15 and 35 and a higher percentage of males experiencing higher violent crime rates.

Lastly, a positive relationship between our measure of policing and crime suggests that there is an endogeneity problem. To establish the effectiveness of policing on crime, further research is needed, as well as access to policing data.

References

Adato, M., Carter, M. and May, J. (2006). Exploring poverty traps and social exclusion in South Africa using qualitative and quantitative data. *Journal of Development Studies.* 42(2). pp. 226 - 247

Altbeker, A. (2005). Paying for crime: South African spending on criminal justice. Cape Town: Institute for Security Studies.

Becker, S. (1968) Crime and punishment: An economic approach. *Journal of Political Economy.* 76. pp. 169 - 217

Brysiewicz, P. (2001) Trauma in South Africa. *International Journal of Trauma Nursing.* 7. pp. 129 - 132

Calvó-Aremengol, A. and Zenou, Y. (2004). Social netowrks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review.* 45(3). pp. 939 - 958

Cichello, P., Almeleh, C., & Ncube, L. (2006). *Perceived barriers to entry into self-employment in Khayelitsha, South Africa: Crime, Risk, and Start up Capital Dominate Profit Concerns.*

D'Allesso, S., Stolzenberg, L. and Eitle, D. (2002). The effect of racial threat on interracial and intraracial crimes. *Social Science Research.* 31. pp. 392 - 408

Demombynes, G. and Özler, B. (2005) Crime and local inequality in South Africa. *Journal of Development Economics.* 76(2). pp. 265-292

Devey, R., Valodia, I. and Velia, M. (2005). Constraints to growth and employment: Evidence from the Greater Durban Metropolitan Area. University of KwaZulu-Natal

Durlaf, S., Navarro, S. and Rivers, D. (2010). Understanding aggregate crime regressions. *Journal of Econometrics.* 158. pp. 306 - 317

Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation, *Journal of Political Economy.* 81(3). pp. 521 - 565

Fajnzylber, P., Lederman, D. and Loayza, N. (2002) What causes violent crime. *European Economic Review.* 46. pp. 1323 - 1357

Imai, S. and Krishna, K. (2004). Employment, deterrence, and crime in a dynamic model. *International Economic Review.* 45(3). pp. 845 - 872

Imrohoroglu, A., Merlo, A. and Rupert, P. (2004). What accounts for the decline in crime?. *International Economic Review.* 45(3). pp. 707 - 729

Kingdon, G., & Knight, J. (2007). Unemployment in South Africa, 1995 2003:Causes, Problems and Policies. *Journal of African Economies.* 16(5). pp. 813-848.

Levitt, S. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review.* (87)3. pp. 270 - 290

Lochner, L. (2004) Education, work, and crime: A human capital approach. *International Economic Review.* 45(3) pp. 811 - 843

Scorzafave, L. and Soares, M. (2009). Income inequality and pecuniary crimes. *Economics Letters.* 104. pp. 40 - 42

Silverman, D. (2004). Street crime and street culture. *International Economic Review.* 45(3). pp. 761 - 786

Verdier, T. and Zenou, Y. (2004). Racial beliefs, location, and the causes of crime. *International Economic Review.* 45(3). pp. 731 - 760

Appendix

List and description of do-files used to generate data, which are on the accompanying DVD.

Merging files

- Merge LFS 1
    - This do-file merges the 5 waves of the LFS data. The code in this file was adapted from a do-file from the Development Policy Research Unit at the School of Economics at the University of Cape Town
    - The district councils are recoded in this file
- Data cleaning 1 and 2
    - These files combine the crime statistics (already converted from PDF into excel format) downloaded from the SAPS website and reshape the data into a usable format
- Generate DC
    - This do-file recodes the station IDs to match their respective District Councils
- Merging LFS and Stations
    - The station data are merged with the LFS data, one to many

Variable generation

- Precollapse variable generation
    - Economic, sociodemographic and infrastructural variables are generated in this file
- Generating inequality index
    - This file is referenced in the "Precollapse variable generation" do-file. This do-file generates the Theil indices using a while loop. It then generates the Theil indices for Districts, which have insufficient people in certain racial groups.
- Collapsing data
    - The merged data generated is collapsed into district councils. The indices of racial and linguistic homogeneity are generated, as well as population variables. The dependent variables are converted into per capita form and logged.
- Generate Data
    - This do-file just runs all previous do-files.

Analysis

- Descriptive stats
    - This file generates the scatterplots used in section IV
- regressions
    - The regressions output used in section V is generated in this do-file