THE SOCIO-ECONOMIC DETERMINANTS OF CRIME IN SOUTH AFRICA: AN EMPIRICAL ASSESSMENT

HAROON BHORAT
ADAIAH LILENSTEIN
JABULILE MONNAKGOTLA
AMY THORNTON
KIRSTEN VAN DER ZEE

DPRU WORKING PAPER 201704
MARCH 2017

DEVELOPMENT POLICY RESEARCH UNIT

WORLD BANK GROUP
Jobs

A DPRU Working Paper* commissioned for the

WORLD BANK GROUP
Jobs

THE SOCIO-ECONOMIC DETERMINANTS OF CRIME IN SOUTH AFRICA: AN EMPIRICAL ASSESSMENT

HAROON BHORAT

ADAIAH LILENSTEIN

JABULILE MONNAKGOTLA

AMY THORNTON

KIRSTEN VAN DER ZEE

Working Paper 201704

ISBN 978-1-920633-43-1

March 2017

Summary:

There is a dearth of research on crime in South Africa, which is particularly problematic in this country given the extraordinary high crime rates reported here. Common correlates of crime, such as unemployment, poverty, and inequality, are also at extreme levels in South Africa – making the investigation of the determinants of crime even more pertinent in this context. We combine published crime statistics with demographic data from the 2011 South African Census Community Profiles to investigate which socio-economic factors attract crime at a police precinct level. In particular, we investigate whether, and to what extent, precinct-level unemployment and income and intra-precinct inequality are related to reported crime rates within a particular precinct. The expectation was that resource-acquisition driven crimes (i.e. property and robbery crimes) would be attracted by high levels of income and inequality in a precinct, and low levels of unemployment.

Further, we hypothesised that at some high level of both income and inequality, crime levels would decrease due to individuals beginning to take protective measures against crime. Through a combination of nonparametric and parametric analyses, including an IV regression design, we found support for this protection hypothesis in the case of property crime but not in the case of robbery crime. No socio-economic factors were significantly related to robbery crime in our analyses. We also investigated violent crime, which, due to the interpersonal and psychological nature of such crime, we hypothesised to vary positively with inequality and unemployment and vary negatively with income. Although we did find positive relationships between violent crime and income, we found that at high levels of precinct-level income, violent crime decreased. We did not detect any relationship between inequality and violent crime, nor between unemployment and any crime type.

The interpretation of these findings is that, where significance was found, certain socio-economic factors attract certain types of crime. This is not to say that insignificant findings signal no impact of the indicator on crime, since individual-level driving factors cannot be investigated using precinct-level data.

Keywords:

Crime, South Africa, unemployment, poverty, inequality, property crime, robbery crime, violent crime, precinct-level income

Acknowledgements:
* This research was undertaken by the Development Policy Research Unit, for the Network on Jobs and Development (NJD) programme. The NJD is a partnership financed by the World Bank's Development Grant Facility, and sponsored by the World Bank's Jobs Group.

Working Papers can be downloaded in PDF (Adobe Acrobat) format from www.dpru.uct.ac.za. A limited number of printed copies are available from the Communications Manager: DPRU, University of Cape Town, Private Bag X3, Rondebosch, Cape Town, 7700, South Africa. Tel: +27 (0)21 650 5701, email: sarah.marriott@uct.ac.za

Corresponding author
Haroon Bhorat, haroon.bhorat@uct.ac.za

Recommended citation
Bhorat, H., Lilenstein, A., Monnakgotla, J., Thornton, A. and van der Zee, K. (2017). "The Socio-Economic Determinants of Crime in South Africa: an Empirical Assessment". Development Policy Research Unit Working Paper 201704. DPRU, University of Cape Town.

Contents

List of Tables

List of Figures

List of Equations

1. Introduction

South Africa possess some of the highest reported crime statistics in the world, making the study of crime and its determinants of particular interest in this context. Since the seminal paper by Becker (1968), the body of economic literature around crime has been growing widely. However, the initial insights from this work of Becker (1968) still prove fundamental to economic applications of the incidence and determinants of crime. In the traditional economic model of crime, criminal behaviour depends on, amongst other things, the payoff from committing a crime successfully, conditioned by the likelihood of obtaining legitimate sources of income, together with this commensurate level of legal labour market income. Three socio-economic factors can initially be identified therefore, as playing an important role in our understanding of the incidence of crime: Unemployment, income levels, and the prevalent level of income inequality. High levels of unemployment mean that the probability of gaining a legal income would be low, whilst high levels of legal income simultaneously increases the payoff from criminal activity as well as increasing the opportunity cost of criminal activity. Hence, inequality also influences crime levels by providing a state whereby the benefits of crime are high (in high income or relatively high income areas) while the benefits of legitimate activities are simultaneously low.

South Africa, as is now commonly known, has some of the highest unemployment rates in the world. Unemployment by the standard ILO definition was 27 percent in 2016 while the inclusion of discouraged workers raised the unemployment rate to 36 percent (Statistics South Africa, 2016). It is believed that unemployment in South Africa is structural in nature – the legacy of South Africa's apartheid history has contributed to a skill-dependent economic growth path in a country with a dearth of skills (Bhorat, 2004; Bhorat & Hodge, 1999; Bhorat & Mayet, 2012; Burger and Woolard, 2005; Edwards, 2002). While all age groups suffer from high unemployment levels, the youth are most affected. Youth in South Africa face a daunting narrow unemployment rate of 55 percent (Statistics South Africa, 2016).[1] Unemployment rates also differ vastly across racial groups in South Africa, with 39 percent of Africans being unemployed in the second quarter of 2016, as compared with 27 percent of Coloureds, 14 percent of Indians/Asians, and 8 percent of Whites (DPRU, 2016).

Furthermore, incomes are incredibly low for the majority of South Africans. A number of studies have analysed poverty levels and they suggest that South Africa has the highest poverty relative to their counterparts in developing countries which resembles that of a low-income country (Bhorat et al., forthcoming, Budlender et al.; 2015, Altman; 2006). For example, Bhorat et al. (forthcoming) showed evidence that in 2010, nearly half of South African households (47.3 percent) and more than one-third (37.9 percent) lived below the upper (R577) and lower (R416) poverty lines respectively. Moreover, there are great disparities in poverty levels across racial groups, with a vast majority of Africans living below the poverty line, followed by Coloureds (Bhorat et al., forthcoming). There can be no doubt that a significant proportion of people living in the country are struggling to meet their basic survival needs.

Finally, inequality in South Africa is also exceptionally high by global standards. In a study of 108 economies, South Africa yielded the highest Gini coefficient, 0.62, when calculated as an

[1] Youth in South Africa are defined as individuals between the ages of 15 and 35 (Republic of South Africa, 2009), rather than the International Labour Organisation and United Nations definition of 15 and 25 years. Therefore, over half of all those aged between 15 and 35 years are unemployed and actively searching for work in South Africa.

average over the years between 1960 and 1992 – a period of course covering apartheid rule (Deininger & Squire, 1996). Post-apartheid estimates however suggest that inequality has in fact increased: The Gini coefficient has thus risen from 0.66 in 1993, to 0.70 in 2008 according to analysis by Leibbrandt et al. (2012). In 2010, the Gini was estimated to be 0.69 (Bhorat, Ewinyu, & Monnakgotla, forthcoming).

Given these stark socio-economic conditions, it is no surprise that South Africa has high crime rates. Crime has been characterised as one of the most difficult challenges in post-apartheid South Africa (Demombynes & Özler, 2005). This is due not only to the increasing levels of crime, but also to the perpetuation of social divisions in response to the fear that crime creates (Lemanski, 2004). Violent crime is particularly prevalent in South Africa, and in the minds of South Africans (Shaw & Gastrow, 2001). However, as we show in Section 2.2.2, South Africa also has high rates of property crime. Violent crime includes activities such as murder, assault and sexual offences, while property crime involves burglary and vehicle theft. These are also distinct from robbery crime, which is similar to property crime but includes an element of fear or force in the removal of property.

Given these unique and even extreme conditions, research on the relationship between crime and socio-economic indicators in the South African context is essential. Although some research has been done, the majority is criminological or anthropological in nature. One reason for the paucity of economic research in this area is the lack of reliable micro-data, both from the crime side and from the socio-economic side. In the first instance, it is well-known that data from the South African Police Service on crime statistics does not fully represent all crime in South Africa. This is due to the problem of many individuals not reporting crimes, certain crimes being undocumented in these records – such as organised crime and white-collar crime – and to recording issues that may be present. In addition, survey data on self-reported crime can also be problematic. The South African Victims of Crime Survey, for example, has data quality issues that lead to too few observations for any meaningful conclusions to be drawn, and also cannot be disaggregated to a sufficient degree. In the second instance, most South African survey data, which can be used to look at individual unemployment and income as well as local inequality, are only representative at the national level. Hence, when linking these data to crime statistics, the resulting population-to-crime estimates are unreliable at anything less than the national level.

This research uses similar data to that of Demombynes and Özler (2005). Like the authors, we use small area level data from the South African census to derive socio-economic variables and link this to crimes reported by the South African Police Service (SAPS). This is detailed below in Section 2. Although the issues with SAPS data remain, at present this is the most comprehensive source of reported crime data in South Africa. Furthermore, using the census data means that all socio-economic variables are representative at any level, and are therefore representative at the precinct level, which is our unit of analysis. In our analyses, we use the latest available census data (from 2011) as well as the corresponding crime data. This is in contrast to Demombynes and Özler (2005), who use data from 1996.

The next subsection will briefly discuss the literature surrounding crime and socio-economic indicators, both internationally and in South Africa. We then turn to an overview of the other factors that can influence crime rates, before presenting the classic economic theory of crime in more detail. In the final subsection, we discuss our conceptual approach, which draws on the literature and the economic model to determine what we can expect from the results to follow. In Section 2, we discuss the data used, and in Section 3, we discuss our methodology in more detail. Section 4 presents our results, and Section 5 concludes.

1.1 A Brief Literature Overview

While research in the 1990s and early 2000s generally detected significant but small associations between employment and crime (Chalfin & Raphael, 2011; Freeman, 1983; Mustard, 2010; Piel, 1998), more recent research has documented quite substantial relationships. In a recent review of the literature, Chalfin and McCrary (2017) document that studies which use panel data to assess unemployment-crime relationships (of which there are nine) universally, find evidence in favour of a link between unemployment and crime; particularly property crime. The authors argue that this, as well as an instrumental variables approach, is a far more statistically sound way of measuring such relationships. Of the three studies that use an instrumental variables design, all three find strong relationships between unemployment and property crime. One study found mixed results for violent crime.

Overall, positive correlations are generally found in the empirical literature between inequality and crime and income/poverty and crime. In a meta-analysis of 214 studies, Pratt and Cullen (2005) find that poverty is among the strongest predictors of criminal behaviour. In an earlier review, Hsieh and Pugh (1993) looked at 34 studies and found that in 97 percent of cases both poverty and inequality were found to be positively related to violent crime. Similarly, in a review of 17 time-series studies, Rufrancos et al. (2013) find that all studies provide evidence of significant positive associations between income inequality and crime, including property crime, robbery and homicide.

There are some South African studies that investigate the determinants of crime. Demombynes and Özler (2005) investigate unemployment, income, and inequality, among other things. They find no statistical relationship between crime and employment, except in the case of homicide. They find strong positive relationships between average income in a police precinct and both property crime and violent crime. For property crime, they find a linear positive relationship and conclude that this reflects that income is a proxy for the returns to property crime. For violent crime, they find an inverted U-shape curve: Violent crime increases with income to a point and then decreases. The authors also find a significant positive relationship between inequality and property crime, but not violent crime. In a more recent study, Verrinder (2013) finds a relationship between crime and youth unemployment, but not for the population as a whole. Youth unemployment is particularly important for property crime and less so for violent crime, although both are significant. The author finds no relationship between inequality and crime. Both studies are community-level, cross-sectional analyses.

Social cohesion may also be particularly important in the South African context, which is specifically fragmented in the wake of the apartheid era, with the discriminatory regime resulting in barriers between population and culture groups. In Verrinder's (2013) study, the author investigated both racial homogeneity and linguistic homogeneity in district councils. While racial homogeneity was found to have a strong negative correlation with both property crime and violent crime (although it was stronger for property crime), linguistic homogeneity was not significantly correlated with either crime category investigated. This may be due to a degree of multicollinearity between the two indicators. At a more basic level, racial differences in crime rates may be related to the role that financial support networks play in the decision to participate in criminal activities. Furthermore, Verrinder (2013) notes that wealth and race are correlated in the South African context. In some cases, social networks could help some individuals to enter the labour market, or provide general financial support (Adato et al., 2006).

Education is identified in the literature as another determinant of crime. Jonck et al. (2015) investigated the relationship between crime and education in South Africa from the provincial and national levels. Jonck and others note that individuals are more likely to engage in crime if they drop out of school before obtaining formal education. Furthermore, the proportion of the prison population with incomplete secondary education is higher than that of the general population (Jonck et al., 2015). In contrast, Lochner (2004) asserts that education attainment coupled with skills development increases the probability of entering the formal labour market, and encourages individuals to socialize so that they would prefer not to commit crime.

Several studies have found age as one of the determining factors that influence the engagement in criminal activities (Verrinder, 2013; Demombynes & Özler, 2005). Verrinder (2013) found that youth aged between 15 and 35 are more likely to engage in criminal activities than older age cohorts. Moreover, Jonck et al. (2015) found that the age distribution of imprisoned people differed from the general population, and that the share of 15-49 year olds in the prison population was greater than the general population of the same age. This suggests that youth are more likely to end up incarcerated than older people are.

In terms of gender, the likelihood to commit crime is higher for men than for women. The distribution of the imprisoned population is highly skewed towards men and there is less presence of women in prisons (Jonck et al., 2015). Similarly, Verrinder (2013) found a substantial positive relationship between men and crime; particularly for violent crimes.

Increasing police personnel and allocating more budget to precincts is often cited as a primary crime prevention strategy. This stems from the classic economic theory of crime, which asserts that increased probabilities of capture, and the severity of punishment when captured, should decrease the expected utility from engaging in crime and therefore deter potential criminals. However, the empirical work in this area fails to find a relationship between crime and police level, where police level refers to expenditure on policing and total number of police officers (Lim et al., 2010). There is also a dearth of literature providing empirical evidence, specifically in Africa (Lim et al., 2010). One reason that policing may fail to deter crime is if criminals do not observe changes in policing practices. Individuals may therefore hold incorrect perceptions of the risk of capture (Chalfin & McCrary, 2017). Furthermore, studies investigating this issue suffer from a lack of control over the timing of increases in police personnel and resources: Usually increases in policing are a result of increases in crime creating endogeneity problems and thus rendering it difficult to isolate the anti-crime effects of an increase in resources (Chalfin & McCrary, 2017). Although the literature in this area suggests that these variables are unimportant for crime, there is still reason to believe that long-term differences in policing across precincts may influence crime levels.

1.2 Crime: An Economic Perspective

Economic models of crime stem from Becker's seminal 1968 paper, "Crime and Punishment: An Economic Approach." Here Becker characterises the choice to commit crime as a gamble faced by a rational agent, the outcome of which is dependent on the costs and benefits associated with either committing a crime or not. The benefit of committing a crime is either the monetary or the psychological gain (the former we refer to as resource acquisition crime and the latter is in the case that the crime is not financially motivated), while the expected cost of the crime is a function of the probability of apprehension and the severity of punishment if apprehended. The expected benefit of not committing a crime (which is also the opportunity cost of crime) is the probability of employment and the expected wage in employment. There is no cost to abstaining from crime other than the opportunity cost of the crime itself. Hence, the individual chooses to commit a crime if the following condition holds:

$$(1 - \mathrm{p}_1)\mathrm{U}_{\mathrm{c}1} + \mathrm{p}_1\mathrm{U}_{\mathrm{c}2} > (1 - \mathrm{p}_2)\mathrm{U}_{\mathrm{nc}1} + \mathrm{p}_2\mathrm{U}_{\mathrm{nc}2} \tag{1}$$

Where, on the left hand side, p_1 is the probability of apprehension, U_{c1} is the utility associated with committing a crime and not being caught, and U_{c2} is the utility associated with committing a crime conditional on being caught, p_1. On the right hand side, p_2 is the probability of employment, U_{nc1} is the utility associated with not committing a crime and not being employed, and U_{nc2} is the utility associated with not committing a crime and being employed.

From this one can see that the criminal decision is primarily influenced by six factors: (1) the probability of apprehension, which in this simple formulation would be driven by the prevalence of police personnel and the effectiveness of police action; (2) the utility associated with criminal activity, or the payoff when a crime is successful; (3) the disutility of apprehension, or the severity of punishment intermediated through the criminal justice system; (4) the probability of employment, or the employment rate; (5) the utility of employment, or wages, when employed; and (6) the disutility associated with not being employed. An addition to this model includes time preferences – with respect to the immediacy of sentencing if apprehended – to the list of factors which can influence the criminal decision.

As discussed, we distinguish between three types of crime: Property, robbery, and violent crime. The first two cases are resource acquisition crimes, while the latter is psychologically driven. What follows is related to property and robbery crime specifically, while the benefits and costs associated with engaging and not engaging in crime are completely different for violent crime. Here, the costs of engaging in crime still depend on the likelihood of capture and the sentencing if caught, but individuals who engage in these crimes are likely to be extremely 'impatient', which in econometric terms means that they value immediate gain over future punishment to a great extent. While the benefits of engaging in violent crime are the psychological gains, psychological costs may also be borne by the criminal. Not engaging in crime is only costly insofar as the individual must bear the psychological costs of not committing a violent act. There is no benefit to not engaging in violent crime besides those directly related to the costs of committing the crime.

However, where crime is motivated by resource acquisition, this model points to unemployment, income, and inequality as three socio-economic factors that can affect the decision to commit a crime. In the first instance, when the unemployment rate is high, p_2 is low and hence the expected utility from abstaining from crime is low. Income holds an interesting place in this model, since it can influence both the left and right hand sides of the equation: Higher wages for the individual increases U_{nc2} and therefore the expected utility of abstaining from crime. Hence, higher income decreases the likelihood of criminal activity. Conversely, higher wages of others increases the expected payoff from criminal activity, U_{c1} – especially in the case of property and robbery crime – and hence increases the likelihood of criminal activity. Therefore, areas with high inequality (low U_{nc2} and high U_{c1}) face the highest likelihood of crime since the expected payoff from legitimate activities is low and the expected payoff from illegitimate activities can be very high – if the offender migrates to areas of high income to commit the crime.

Thus far, this discussion has centred on the individual choice to commit crime. However, the literature around crime is spread beyond the level of the individual as the unit of analysis. For example, cross-country studies, studies looking at large metropolitan areas in the United States, as well as studies looking at small areas such as police precincts. The expectations for how variables will interact with observed crime vary depending on the unit of analysis. Given our data, we do not know who commits the crime, and therefore we cannot model the choice at the individual-level (as the Becker model implies); we do not observe where the person committing the crime resides, and only where the crime actually took place. More aggregate datasets have been criticised for making assessments across areas that may not be influenced by the same underlying processes that produce crime (Demombynes & Özler, 2005). However, this criticism is levelled at far more aggregate data than what we use here. The unit of analysis here, while certainly small enough not to suffer from these particular concerns, still cannot be used to model the individual's choice to commit crime. Hence, we expect that the relationships in the classic economic model of crime will differ somewhat in our analysis.

Since we do not observe who commits crime and only where crime occurs, we cannot determine any effects occurring from the right-hand-side of Equation 1. Specifically, even though we observe precinct-level unemployment rates, these are not the unemployment rates associated with the area the criminal lives in, and hence it is impossible to investigate how p_2 influences crime rates. Rather, to the extent that unemployment rates can influence the returns to crime (perhaps through signalling the presence of economic activity), U_{c1}, only then will we see an effect of unemployment on crime. Similarly, we do not observe income levels of the precinct in which the criminal resides, and hence we only observe the effects of income on U_{c1} and not on U_{nc2} – that is, we observe how income attracts crime to an area and not how income can deter it through high paying jobs. Finally, we also observe how inequality can increase U_{c1} by signalling the presence of relative wealth, and not how inequality can decrease U_{nc2} (by indicating relatively low income when compared to others in the precinct). In sum, then, this paper investigates whether, and to what extent, these three socioeconomic factors are related to crime through their influence on the returns to crime in different precincts (the left-hand-side of Equation 1). Whether these variables influence the individual decision to commit crime using other pathways, such as through the expected returns to legitimate activity (the right-hand-side of Equation 1), is not and cannot be touched on here.

1.3 Conceptual Approach

Our expectations regarding the relationship between crime and these three socio-economic variables are based on the literature discussed and the standard economic model employed, but they also vary crucially by category of crime – a fact that is often in our view not sufficiently nuanced in the literature.

For the two resource acquisition crime categories (property and robbery crime) we expect that crime decreases with precinct-level unemployment, increases with precinct-level income, and increases with intra-precinct inequality. In the classic model, higher unemployment rates (i.e. a smaller p_2) should decrease the expected utility of not engaging in crime and therefore make the incidence of crime more likely. However, since we do not observe p_2, there is no reason to believe that relatively high unemployment rates in one precinct should necessarily be associated with higher crime in that precinct. It may be that high unemployment rates result in high crime, but that the crime is committed and reported in precincts with relatively low unemployment rates. Indeed, to the extent that relatively low unemployment in a precinct is still correlated with precinct-level wealth after controlling for precinct-level income, we would expect higher crime levels in precincts with lower unemployment rates. Alternatively, low unemployment rates may be correlated with the opportunity to commit a crime, such as in a busy urban area with much to be stolen, which would result in higher crime in those areas. Hence, we expect precinct-level unemployment to decrease the expected utility of engaging in crime (the only employment-crime pathway that we observe) and therefore to be negatively associated with crime rates. Similarly, income in a precinct increases the expected utility of engaging in crime in that precinct and hence one should observe a positive relationship between precinct-level income crime levels. In terms of intra-precinct inequality, higher inequality signals higher earnings from crime within that precinct relative to migrating to commit crime (i.e. higher expected returns from crime in a precinct) and therefore we can expect higher crime within the precinct itself.

Regarding resource acquisition crime, we also explore a second model that aims to understand the role of protection. In order to protect yourself from crime, a certain level of income is required in order to purchase crime prevention instruments such as alarm systems, armed response services, burglar bars, or private security. Therefore, we need to assess the role of income not only in attracting crime but also in deterring it. What we argue here is that the increase of both precinct-level income and intra-precinct inequality in an area lead to first a rise in crime as both variables attract criminal behaviour. However, at a certain high level of inequality and income, crime levels on average should start to decline as wealthier individuals – aware of their vulnerable status due to inequality – start to mobilise their wealth to protect themselves. It is key that both income and inequality are included in this protection hypothesis because individually they will not necessarily lead to a downturn in crime. Intra-precinct inequality in a precinct with a low level of income may make the relatively better-off aware that they are targets, but without the means to protect themselves, they remain vulnerable to high crime rates. Similarly, an area that is generally rich but not unequal, may not feel the need to erect high walls and electric fences. It is the combination of rising income and inequality that feeds into the protection hypothesis.

Although violent crime is less likely to be financially motivated, it is still likely to be driven by similar factors, but for different reasons and in different ways. Unemployment, lack of income, and inequality all induce psychological costs of their own which may impact on an individual's propensity to commit crime. While these can certainly strengthen the relationships between socio-economic factors and property/robbery crime as well, due to the nature of violent crime they would not be the relationships may not play out in the same way. Although most South Africans are highly fearful of violent crimes propagated by strangers, most violent crimes occur within the home by a known offender (Shaw & Gastrow, 2001). Hence, violent crime, we would argue, is not being perpetuated between outliers in the income distribution but rather by individuals fairly close to each other by the chosen socio-economic status variable. We therefore expect unemployment and inequality to vary positively with violent crime and we expect income to vary negatively with violent crime. We also do not extend the protection hypothesis to our analysis of violent crime and therefore we do not expect to see the same quadratic relationship as in the case of the other two crime categories.

2. Data

The dataset used in this analysis combines data from various sources including the SAPS' Crime Statistics (SAPS, 2011a), the 2011 South African Census Community Profiles (Statistics South Africa, 2011), and data from the 2011 SAPS Annual report (SAPS, 2011b).

The primary source of crime data used for this study is the annually released SAPS' Crime Statistics (hereafter referred to as the 'administrative data'). This dataset documents the number of reported crimes at each of the approximately 1140 police stations in South Africa for a given year. The data describe two major types of crime, namely community reported crime and crime detected as a result of police activity. In the first case, community members are either involved in or observers of a crime, and report this to the appropriate police station (for example, robbery at a residential premises). In the second case, intentional police action results in the uncovering of a crime (for example, a road block resulting in the detection of driving under the influence of alcohol or drugs). For the remainder of this paper we focus on community-reported crime only. More detail on the types of community reported crimes that are available in the administrative data is given in Section 2.2.1.

There are various factors that might impact on whether either of these types of crime are reported. Factors including trauma associated with the event – and thus an unwillingness or inability to report a crime – may limit the number of crimes that are reported by the community. Other factors that could result in crime going unreported include poorly stored or handled data at police stations, incomplete data publication, or police officers unwilling or unable to record the crime (De Kock, Kriegler, & Shaw, 2015). There is also preliminary evidence suggesting that poorly resourced areas in South Africa may also be particularly associated with much lower reporting rates (see Redpath & Nagia-Luddy (2015), who show that very under-resourced areas in the Western Cape also have very low reporting rates). In addition, the efficiency and resources available to a police station and its employees will impact on the number of crimes detected as a result of police action. Huber (2016) also finds that chieftaincy and regional identity (areas with lower levels of trust for the state police), as well as medical facilities (which are often ill-equipped to deal with sexual crimes, and the reporting thereof), often deter the reporting of sexual violent crimes to the police. According to De Kock et al. (2015) approximately 88.7 percent, 45.6 percent and 31.2 percent of the actual number of murders, assaults and theft of personal property respectively, were reported in the April 2013 to March 2014 period. Thus overall, the statistics that are available from the administrative data are very likely to be an underestimate of the actual number of crimes that occurred in each police precinct, and given the above, it is very likely that errors will be non-randomly distributed. As this reported crime data is what is used for our estimates of crime in our analysis, any results that we present are only relevant for reported crime in South Africa, as opposed to the underlying population of crimes that occurred.

Furthermore, in the administrative data, crime is separated into roughly 27 categories. These categories are often aggregates of various other types of crime. For example, the administrative data includes “Drug-Related Crime”, while this category includes crimes such as the use of drugs, drugs trafficking, and drug dealing. The individual rates for each separate crime are not also included. If we believe that the motivators for, say drug use and drug dealing, are different, then it would be useful to analyse these separately. This, however, is not possible. Thus, the structure of the administrative data limits our ability to categorise the various crimes into larger or smaller groups that are most appropriate for analysis. In Section 2.2.1 below, we will outline the way in which we categorise crime, given the limits imposed by the structure of the administrative data.

Although the administrative data documents the actual numbers of crimes reported to each police station, for our purposes it is necessary to present the crime statistics data as crime rates relative to the population size, rather than aggregate crime numbers. This ensures comparability across different areas of different sizes. For example, one murder in area X with 100 000 people is relatively more than one murder in area Y with 300 000 people. The international norm for reporting crime in a comparable way is thus to present the number of crimes per 100 000 people. In order to present the administrative statistics in this way, we first need to identify the geographic regions that each police station services, namely the 1140 police precincts, and secondly we require accurate estimates of the population size of each of these police precincts.

In order to acquire the most accurate population estimates for each police precinct, we require data that is representative at the police precinct level. Thus, this paper makes use of the 2011 South African Census Community Profiles data. This dataset is derived from the 2011 South African Census, and thus represents all South Africans from all regions in the country at a small area level and location. This dataset includes every single census observation, but with statistics aggregated to the small area layer (SAL) level. For example, the dataset will include average employment statistics for each SAL. A SAL is created by combining all enumeration areas (EAs) with a population of fewer than 500 with adjacent EAs within the same sub-place. Hence, this data can be analysed at small geographical units without concerns about representivity. In addition to allowing an accurate estimate of the population size in each precinct, the Community Profiles data includes demographic information such as individuals’ age, population group, language, citizenship, health, as well as economic characteristics – all aggregated to the SAL level. This dataset is the primary source of data for all demographic characteristics included in the analyses to follow.

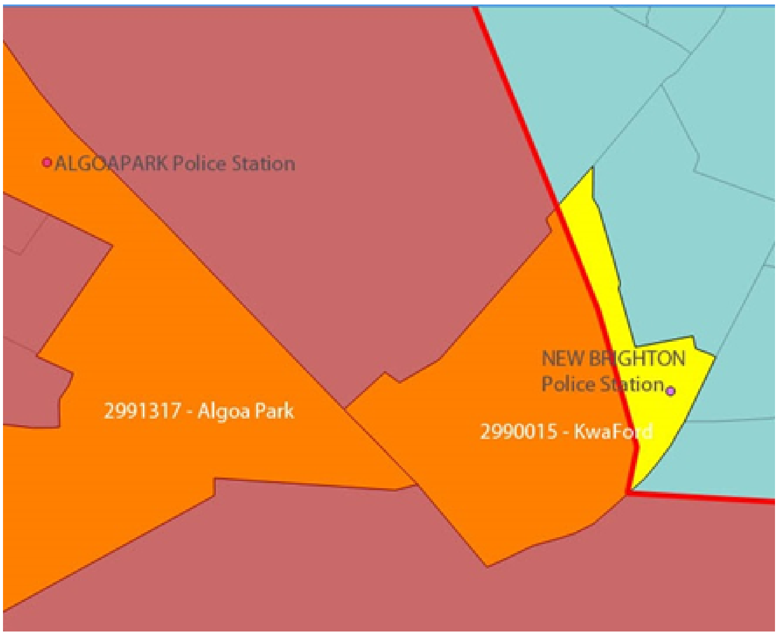
2.1 Combining multiple datasets

Combining the administrative data with the census data involves first allocating each of the approximately 85 000 SALs to one of the police precincts in the administrative data. To do this we utilised, through geospatial mapping, the SAL boundary geographic information system (GIS) data and generated a random point that fell within each SAL's boundary. These generated points were then mapped to the 2015 police station boundary data using a point in the polygon algorithm to determine into which police boundaries these points fell. Using a random point does not guarantee that the majority of the SAL would be in the same precinct as the random point itself. However, a random point was used for this method (as opposed to a central point or centroid), given that the irregular shape of the SALs can often lead to the "central" point falling completely outside of the SALs boundary. Hence, the random point minimises the error potential.

Figure 1 below helps to illustrate this process. The map shows two police precincts and two SALS. The Algoa Park precinct is shown by the red shaded area, while the New Brighton precinct is shown by the blue shaded area. From the map, it can be seen that while the entire Algoa Park SAL falls into the Algoa Park precinct, the KwaFord SAL falls partly into the Algoa Park precinct and partly into the New Brighton precinct. Hence, a random point within the Algoa Park SAL will always fall into the Algoa Park precinct, while a random point in the KwaFord SAL could fall into either precinct. It is not possible to divide the SALs any further, and therefore all information associated with KwaFord must be allocated either to the Algoa Park precinct or to the New Brighton precinct. The random point method means that, statistically, if a larger proportion of KwaFord falls into the Algoa Park precinct than New Brighton, then it is more likely that the randomly generated point will also fall into the Algoa Park precinct, and hence the data associated with KwaFord SAL will be allocated to the Algoa Park precinct.

Furthermore, it can be seen from the map that while the KwaFord SAL has a regular shape, the Algoa Park SAL is irregularly shaped. If all SALs were shaped like KwaFord then taking a point at the centre of the SAL would be the best way to allocate SALs to precincts. However, a central point in an SAL like Algoa Park may fall into a completely different SAL. Hence, we elected to use a random point rather than a central point.

Figure 1. Mapping SALs to Police Precincts



Notes: The red shaded area represents Algoa Park police precinct; the blue shaded area represents New Brighton police precinct.

Ultimately then, every South African in the 2011 Community Profiles data was allocated to a police precinct based on the SAL they were surveyed in, giving us estimates of the population size of each precinct. The already aggregated demographic data in the Community Profiles dataset was then aggregated further to the police precinct level, since the police precinct is our unit of observation in the administrative data. For example, the average age in each SAL was averaged over all the SALs within each precinct, resulting in an average age for each precinct. This was done for all relevant individual characteristic data and economic variables. Therefore, the dataset presented 1140 observations: one for each precinct (described by the 2015 SAPS police boundaries), including the name of the relevant police station, aggregate individual and economic characteristics, as well as the population size for each precinct.

We then combined this dataset with the administrative crime data. In this step, we use the administrative data for the year from 1 April 2011 to 31 March 2012, so that we can directly combine it with the Community Profile data, which was collected in October 2011. By using the 2011 crime data, as opposed to more recent crime statistics, we avoid the bias that would arise from using population growth estimates that would have to be applied to the Community Profiles data and demographic information collected four years prior. Since we used the 2011 crime statistics, we lost 14 police precincts, which, although they existed in the 2015 SAPS boundaries, had police stations that were only built after 2011. We therefore have a total of 1126 police precincts in our dataset.[2]

[2] South Africans living within a precinct that was established after 2015 are lost when the Community Profiles and administrative data are merged: this is approximately 830 000 individuals from 250 000 households, or 1.6% of the 2011 population. Crime rates in these precincts for 2015 were similar to that of the remainder of South Africa in 2015. Exclusion of the new precincts causes an overestimation of the crime rates at the old police stations that would have serviced these 830 000 individuals, due to the fact that the old precincts would have been serving larger populations than what is represented by the

Lastly, we incorporate resource information from the SAPS annual report to control for the level of police "infrastructure" in the various areas. For example, an area with a greater number of police employees might fight crime more effectively than an area with fewer police employees. Unfortunately, information on police infrastructure (such as number of personnel, level of expenditure, etc.) is not available at the police precinct level. However, the SAPS annual report for the 2010/2011 year presents this information at the provincial level.

3. Introducing Crime in South Africa

3.1 Categorising crime

There are 27 types of crime reported by the South African Police Services. Table 1 below displays the relevant crime categories that we consider, and how we have aggregated them further.[3] The table illustrates the various types of crime (reported in the administrative data) that make up our three primary crime categories.

Property crime is made up of crimes that involve the removal (theft) of property, where these crimes do not involve force or threat of harm to the victim. For example, burglary of a property. Robbery also involves the theft of property: However, in this case the victim faces harm or threat of harm, meaning that there is the use of force and fear. For example, common robbery such a mugging would involve threat to and intimidation of the victim. Lastly, violent crime involves the direct use of force and threat upon a victim. For example, assault with the intent to inflict grievous bodily harm.

From the compositions of the crime categories, we can see the major crime types that drive each category. The largest sub-categories of property crime are burglary at residential premises and theft out of a motor vehicle. By far the largest component of robbery is common robbery (which does not involve the use of a weapon, for example grabbing a handbag or cellphone [De Kock et al., 2015]). Common robbery is followed by robbery at residential and non-residential premises. Lastly, the largest component of violent crime is assault with the intent to inflict grievous bodily harm, whilst murder makes up a relatively small proportion of violent crime.

2015 precincts. The 14 new precincts fall within four provinces (five in the Eastern Cape, five in Gauteng, three in Limpopo, and one in the Western Cape), thus we know that these individuals are reasonably spread across South Africa. Given this spread, combined with the fact that this is a very small proportion of the total population, we can assume that the overestimation of the crime rates is relatively small.

[3] The raw SAPS data refers to "Contact Crime" which roughly combines robbery and violent crimes. This is a very broad category with various different types of crime; for example, murder along with common robbery. Such groupings can hamper crime analysis as the determinants of these vastly different types of crime can often vary. We therefore attempt to create categories of crime with similar motivations and relationships with socio-economic indicators, and model these separately.

Table 1: Distribution of Crime, By Category: 2011

Type Of Crime	Sub-Categories	Share
Property crime		**100%**
	Burglary at Non-Residential Premises	13%
	Burglary at Residential Premises	46%
	Theft of Motor Vehicle and Motorcycle	11%
	Theft out of or from Motor Vehicle	24%
	Stock theft	5%
Robbery		**100%**
	Common Robbery	55%
	Robbery at Residential Premises	18%
	Robbery at Non-Residential Premises	17%
	Carjacking	10%
	Truck Hijacking	1%
	Robbery of Cash in Transit	0%
	Bank Robbery	0%
Violent crime		**100%**
	Murder	3%
	Attempted Murder	3%
	Total Sexual Offences	14%
	Assault with Intent to Inflict Grievous Bodily Harm	41%
	Common Assault	39%

As already discussed, we distinguish violent crime from property and robbery crime due to the fact that the latter, and not the former, are primarily financially-motivated. Here we can see an additional distinction between property crime and robbery crime that was not previously obvious. Although property and robbery crime are generally similar in their definitions, and differ almost exclusively with respect to the use of force and fear, it is worth noting the difference in the two with regard to where they mostly take place: While the majority of property crimes take place at a residential premises, the majority of robberies take place outside of a home, and most likely on the street. This suggests that property crime is more premeditated, while robbery crime may be more of momentary decision.

3.2 A snapshot of crime in South Africa

We now present a snapshot of crime in South Africa according to the crime categorization developed in the previous section. Table 2 presents a summary of each category in absolute reported values, as well as the rates per hundred thousand people.

Table 2: Summary Statistics of Categories of Crime, per 100 000 Individuals.

	Mean	Median	Proportion of Total Crime
Property	471	224	25%
Property crime Rate	1195	840	
Robbery	85	34	4.4%
Robbery Rate	154	106	
Violent crime	411	248	22%
Violent crime Rate	1025	870	
Total Crime	1913	929	100%
Total Crime Rate	4769	3562	

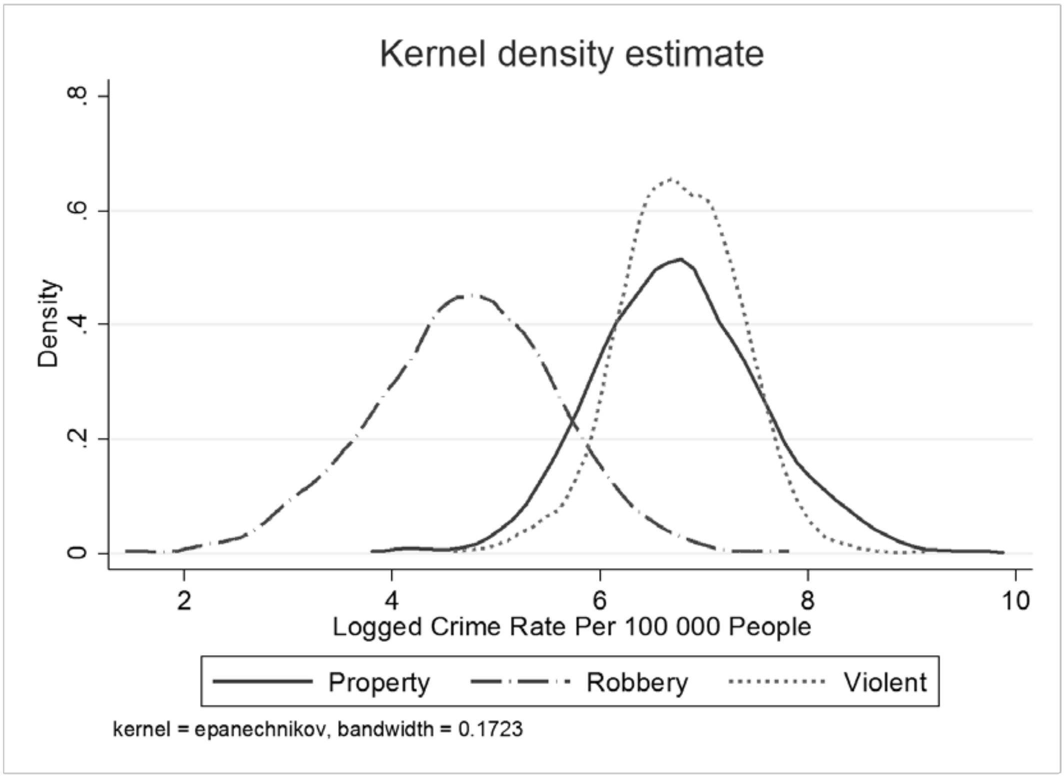
From the table above, we observe that in South Africa in 2011 there were, on average across precincts, 4 769 crimes reported per 100 000 individuals. The actual number of crimes reported across precincts averaged at 1 913 crimes. A quarter of these were property crimes, and almost another quarter were violent crimes. Only about 4 percent were robbery crimes. These three crime categories make up approximately 50 percent of all crime reported in the South African administrative data in 2011. Examples of crime categories that are not included in the three primary crime categories (and therefore make up the remaining 50 percent) are arson, commercial crime, and malicious damage to property. The precinct that reported the highest overall absolute number of crimes was Mitchells Plain in the Western Cape, with almost 29 000 crimes reported in 2011. The precinct with the largest number of crimes per 100 000 people was King Shaka International Airport in Kwazulu-Natal, which had a total of 502 crimes reported, and a population size of only 224.

Apart from constituting the most dominant form of crime, property crime is also the most pervasive, as it is the only category with a non-zero minimum number of crimes reported. For every property crime reported, there were 0.86 violent crimes reported and only 0.13 robberies reported. The highest number of property crimes reported was 4 767, which occurred in Mitchells Plain, in the Western Cape. The highest property crime rate was 16 471 crimes per 100 000 people, which occurred in Committees Drift in the Eastern Cape. This is also the area with the lowest population, which may be driving these high numbers. On average, 1 195 property crimes were reported per 100 000 people across the police precincts.

Violent crime was the second largest contributor to overall crime in South Africa in 2011. On average, 1 025 violent crimes were reported per 100 000 people in each police precinct, which is just short of the property crime average. The precinct with the highest absolute number of violent crimes reported was again Mitchells Plain, and the precinct with the largest violent crime rate was King Shaka International Airport in Kwazulu-Natal, where a total of 21 violent crimes were reported and a population size of 224. Robbery is the smallest crime type of the three categories, with an average rate of 154 robberies per 100 000 people in each police precinct. Both the highest absolute number of robberies, as well as the highest robbery rate per 100 000 people, were reported in the Central Johannesburg police precinct in Gauteng, with 1 301 robberies reported for a population of 52 191.

Figure 2 presents the distributions of each of the three crime categories. This figure illustrates the kernel density plots for the logged crime rates (per 100 000 people) for property, robbery, and violent crime. Property and violent crime are indeed similar in frequency, and both have high average crime rates.

Figure 2: Distribution of Logged Crime Rates



Notes: Own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data.

Violent crime is most concentrated around its mean, while property crime presents slightly more variability. Thus, although property crimes makes the greatest contribution to overall crime, as well as having the highest average crime rate, it is also highly variable, while violent crime is the most consistently high across the precincts. Again, robbery crime was the smallest of the three categories. Although robbery crime is the smallest category in terms of its contribution to total crime, it is the most variable of the three, and robbery crime rates can extend as high as property and violent crime rates at the top end of the distribution.

From this graph, we can also see the structural differences between resource acquisition crime and violent crime. If one were to shift the robbery curve up as it were, it would closely overlap with the property crime curve. The distributions are thus extremely similar in shape, differing only in terms of frequency. On the other hand, violent crime actually has a different shape completely; one that is more consistent. This supports our decision to distinguish empirically and analytically between resource motivated crime, and violence.

4. Methodology

This paper seeks to empirically investigate the socio-economic determinants of crime; that is, the role of unemployment, income, and inequality in predicting the incidence of multiple forms of criminal activity. Despite the extensive international literature finding important relationships between these variables and crime, the literature in South Africa is limited. This may be due to the country data and econometric challenges noted above. Nonetheless, the topic remains under-researched in a country with some of the highest crime rates in world, along with some of the highest levels of unemployment, poverty, and inequality. These characteristics make the study of socio-economic indicators and crime in South Africa of particular pertinence.

In testing our hypotheses, we face a number of econometric and data challenges; namely, endogeneity in the form of both simultaneity and omitted variable bias – specifically with regard to the unemployment measures. As a result, we approach the analyses in different ways in order to take advantage of the strengths and weaknesses of different econometric methods in the presence of endogeneity. We begin by assessing the crude relationship between crime and socio-economic variables with the use of non-parametric local polynomial regressions. This offers preliminary insight into whether the nonlinear relationships posited by the protection hypothesis come out. Secondly, we use Ordinary Least Squares (OLS) regression to control for all the socio-economic variables at once, as well as other important explanatory variables discussed below. We investigate both a linear and nonlinear model using OLS to test the basic model and the protection hypothesis. Finally, we will make use of Two Stage Least Squares (2SLS) regression with the use of an instrumental variable (IV) to correct for the endogeneity problems that occur with analyses of this nature. There are drawbacks to each of these approaches (including the IV approach), which is why we examine each step of this process in order to draw conclusions from all of them holistically. The remainder of this section will expand on this methodology, specifically the basic regression model and our variable selection, as well as an overview of the OLS and 2SLS specifications that will be run.

4.1 Basic Model and Variable Selection

In the econometric analysis, we focus on a number of dependent and independent variables, including measure of crime, socio-economic variables, as well as socio-demographic and infrastructural controls. In this section, we outline the basic econometric model, as well as the choice and construction of the variables used.

The general specification used for the regression analysis is as follows:

$$Ln(crime_{ji}) = B_0 + \boldsymbol{\gamma SO_i} + \boldsymbol{B_1 IND_i} + \boldsymbol{B_2 HH_i} + \boldsymbol{B_3 INF_k} + B_4\, ln(crime_{ji})_{lag} + \varepsilon_i \tag{2}$$

In the equation, $crime_{ji}$ represents the crime rate, where the subscript *i* denotes the precinct and the subscript *j*=1,2, and 3, denotes the crime category under examination: $Crime_{1i}$ is the number of property crimes per 100 000 people in precinct i, $crime_{2i}$ is the number of robberies per 100 000 people in precinct i, and $crime_{3i}$ is the number of violent crimes per 100 000 people in precinct i. We take the natural logarithm of each crime rate as this normalises the distribution of crime, and thus the error term, ε_i. Unfortunately, this transformation converts zero crime rates into missing data. While this causes us to lose a few observations (none for property crime, 68 police precincts for robbery crime, and 3 police precincts for violent crime), the benefits of normality outweigh this cost. We therefore interpret all the regression results as the determinants of crime where crime exists.

The term $\boldsymbol{SO_i}$ is a vector that includes two measures of unemployment (*UE_i* and *$JobDep_i$*), one measure of income (*Inc_i*), and one measure of inequality (*$Ineq_i$*). The term *UE_i* is the precinct level strict unemployment rate, while *$JobDep_i$* measures the precinct's job dependency rate. The job dependency rate represents all individuals in the precinct who are without a job as a proportion of the total precinct population, as opposed to only labour market participants, which is the standard unemployment definition. In other words, job dependency measures all of the people in a precinct who do not have their own wage income source (unemployed and non-labour force participants) as a proportion of the precinct population. By measuring unemployment in this way, we hope to tie resource dependency more closely to unemployment, as job dependency captures dependents who are outside of the labour force (e.g. children). In contrast, the standard unemployment rate is perhaps a better measure of job opportunity in a precinct, since it only includes individuals seeking a job.

Since we are expecting unemployment to be related to crime through the presence of economic activity, it is not immediately clear whether a high concentration of job opportunities or a higher concentration of working individuals would better capture this effect. We include these two measures of unemployment in separate regressions, and therefore test which is the most appropriate determinant of crime. As we mentioned in the literature section, we believe that unemployment will vary negatively with precinct crime rates for property and robbery crime, and thus we hypothesise that the $\boldsymbol{\gamma}$ coefficients will be negative for these two measures in these cases. Conversely, we expect it to be positive for violent crime.

Our income measure, Inc_i, as well as the inequality measure, $Ineq_i$, were created using the annual income reported in the Small Area Layer data. This data was collected at the household level and in 12 different bands. This data structure unfortunately collapses most variation in the income data. We use the midpoint of each bracket and apply this to all households in the same bracket. We then use this to calculate the per capita average income for each precinct, which we use as our Inc_i variable. In order to measure inequality we calculated a household income Gini coefficient for each police precinct, by applying the midpoint of each bracket to all the households in that bracket, weighted according to the proportion of households in each income bracket. As we mentioned earlier in the literature section, we expect that income and inequality will be positively correlated with precinct crime rates for property and robbery crime. We therefore hypothesise that the $\boldsymbol{\gamma}$ coefficients will be positive in these cases. For violent crime, we expect income to vary negatively with crime, while we expect inequality to vary positively with crime. Therefore, in this case we expect that the $\boldsymbol{\gamma}$ coefficient will be positive for income and negative for inequality.

The vector $\boldsymbol{IND_i}$ includes controls for precinct level individual demographic characteristics, namely average age, the proportion of the precinct population who are youth (younger than 30), the proportion of males, proportion of various education levels, proportion of each race, language heterogeneity (the number of languages spoken in the precinct), and the proportion of non-citizens. Based on what was found in the literature, we believe that age and education will vary negatively with crime, and that higher proportions of youth, and lower education levels will be associated with higher levels of criminal activity. As previously discussed, the racial and cultural composition of areas can have an impact on the crime rate. For this reason, we include measures of racial, language and citizenship heterogeneity, in order to control for the social cohesion in an area.

We also include $\boldsymbol{HH_i}$, a vector of precinct level household characteristics. This vector specifically includes the average household size, geographical composition (the proportion of urban, tribal and farm land), and provincial dummies. By including household size, we control for incomes varying due to larger and smaller households in the precincts. The geographical variables control for inherent differences in crime across urban and rural areas, as well as across South Africa's provinces, considering the differing economic activity across these regions.

Lastly, $\boldsymbol{INF_k}$ represents a vector of controls for police infrastructure. These variables intend to control for the effectiveness of policing and police resources across the precincts, and because this information is not available at the precinct level, we include provincial level controls. For this reason the subscript *k*=1,2,3,…,9, for each of South Africa's nine provinces. The vector $\boldsymbol{INF_k}$ includes three variables, namely the number of police staff, the number of victim rooms, and the number of police motor vehicles, all per 100 000 people in the province.

The last term, ($crime_{ji}$)*lag,* represents a lagged crime rate equal to the average crime rate (either property, robbery, or violent, depending on the specific regression) over the five years between 2006 and 2010. This allows us to identify the relationship between crime and employment while controlling for the existence of inherently high or low crime areas, given that previous history of crime in an area is an important predictor of current and future crime (Brantingham & Brantingham, 2013).

4.2 Regression Analysis: OLS

We begin with an OLS regression to achieve a *ceteris paribus* interpretation of our socio-economic variables. We set aside concerns about endogeneity for the time being (see Section 4.3 below), and focus on identifying the specifications of the model to assess the protection hypothesis. We explore whether it is better to include income and inequality as level terms or as quadratic terms. We have two specifications, which are as follows (note the expansion of the $\mathbf{SO_i}$ vector):

Specification 1:

$$Ln(crime_{ji}) = B_0 + \gamma_1 UE_i + \gamma_2 JobDep_i + \gamma_3 Inc_{\mathrm{i}} + y_4 Pov_{\mathrm{i}} + \gamma_5 Ineq_{\mathrm{i}} + \mathbf{B_1 IND_i} + \mathbf{B_2 HH_i} + \mathbf{B_3 INF_k} + \mathrm{B}_4\, ln(crime_{ji})_{lag} + \varepsilon_i \quad (3)$$

Specification 2:

$$Ln(crime_{ji}) = B_0 + \gamma_1 UE_i + \gamma_2 JobDep_i + \gamma_3 Inc_{\mathrm{i}} + \gamma_4 Inc_i^2 + \gamma_5 Pov_{\mathrm{i}} + \gamma_6 Ineq_{\mathrm{i}} + \gamma_7 Ineq_i^2 + \mathbf{B_1 IND_i} + \mathbf{B_2 HH_i} + \mathbf{B_3 INF_k} + \mathrm{B}_4\, ln(crime_{ji})_{lag} + \varepsilon_i \quad (4)$$

The first specification examines all of the variables of interest as level variables. In this case, we are investigating whether our socio-economic indicators have a linear relationship with crime. In the second specification, we examine whether income (Inc_i^2) and inequality ($Ineq_i^2$) are more appropriately specified using quadratic terms. In this case, we are investigating whether these two variables have an inverted U-shape relationship with crime.

4.3 Regression Approach: 2SLS

In this section, we attempt to deal with potential sources of endogeneity in our models above, specifically with respect to our measures of unemployment. There are two main reasons why the relationship between crime and unemployment might be endogenous, and why our estimates of this relationship will be biased. Firstly, there is simultaneity between explaining unemployment and crime: Whilst employment levels may influence crime rates, crime rates may also influence employment levels. For example, high employment areas might attract more crime, while high crime areas might discourage economic activity and associated employment. Causality runs in both directions but with potentially opposite effects: While it is highly likely that crime reduces employment, it is also likely that employment attracts crime – at least at our level of analysis. If this is the case, the simultaneity effect will tend to decrease the employment effect found in an OLS regression analysis.

Income is another variable that may admit reverse causality into the model. Precinct level income explains crime since criminals are attracted to wealthier areas. At the same time, crime can explain precinct level income because higher crime rates may prompt wealthier individuals to migrate to areas with lower crime rates. Such a migratory reasoning, however, is more time-dependent than our cross-sectional analysis allows. Once migration has occurred, we might expect crime rates to adjust down in the original location, and therefore we deem the endogeneity of this variable to be of less concern.

Secondly, given our limited data set, it is highly likely that any regression analysis will suffer from omitted variable bias. For example, we cannot control for police resources and activity at the precinct level because this data is not available. Resources are likely to be related to crime – more resources means less crime (Chalfin & McCrary, 2017) – as well as employment, given the fact that more affluent areas have better resourced police stations in South Africa[4]. As noted above, whilst we attempt to control for resources using the SAPS annual report, this is only possible at the provincial level.

In attempting to minimise these endogeneity concerns with our unemployment variables, we use an IV regression design. The instrumental variable we use for our labour market variables is a proxy for cognitive skills: The precinct-level average of a four-part index for ability to remember and concentrate. This is described more fully in Section 6.2.2. By using an IV to predict employment in the first stage, we generate exogenous variation in the employment rate variable that is then used to predict the relationship between property and robbery crime, and employment, in the second stage.

5. Describing South Africa at the Police Precinct Level

Our sample consists of 1 126 police precincts, and all of the summary statistics presented in Table 3 are at this level. In the Census Community Profiles data, there are about 51 million individuals and 14 million households who have been allocated into these 1 126 police precincts. On average, there are about 45 000 people per police precinct living in an average of around 12 000 households. However, population per police precinct varies sharply. The largest police precinct in terms of population size is Inanda in Kwa-Zulu Natal, with a population of 324 863. In contrast, the smallest is Committees in the Eastern Cape, with only 170 people. The distribution is also exceptionally skewed to the right, as portrayed in Figure 3, by a small number of highly populous precincts. The 22 police precincts with a population above 200 000 people are largely urban and mostly have an average per capita annual income below the sample mean, but around the sample median.

In general, gender appears to be balanced in most precincts, with the mean and median of the proportion of males being 0.49; however, this can reach as high as 0.78 in the police precinct home to the coal-mining town of Viljoensdrif. The proportions of different race groups per police precinct also varies, with the mean proportion of Africans across all precincts being the highest at 0.71. There are some precincts that are close to 100 percent populated by a single race. In general, most precincts are home to people who, collectively, speak almost every official language in South Africa. By contrast, the average proportion of non-citizens is very low at just 0.03, but reaches a maximum of 0.6 in the precinct of Table Bay Harbour. Non-citizens tend to cluster together, with only about 80 precincts having a non-citizen population

[4] This is proven for the Western Cape (Redpath and Nagia-Luddy, 2015), but is likely to be the case in general.

above 10 percent. Non-citizens tend to be located in largely urban precincts with a mean and median income that is above average.

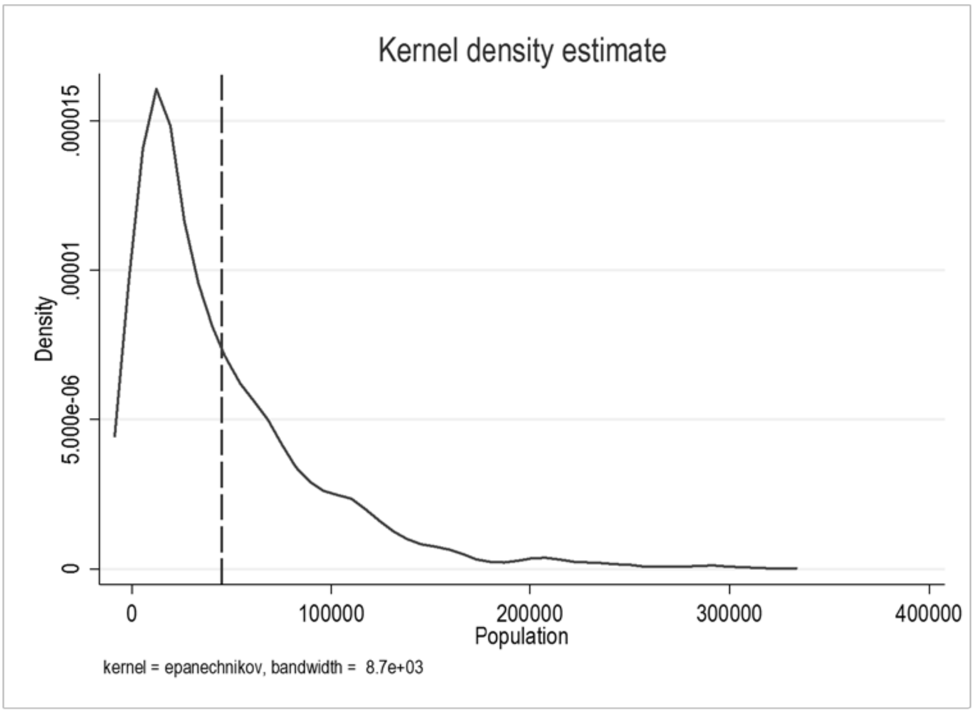
The education variables are interpreted as the average proportion of each education level per police precinct. On average, the proportion of people in the precincts with no schooling is 7 percent. The most common education level is incomplete secondary, with many police precincts having about 27 percent of the population with this education level. This is followed by incomplete primary education at 25 percent. Only 7 percent of the population in the average police precinct have higher education. In general, the education variables appear relatively consistent given the similarity of the mean and median statistics for each education level. The precinct with the highest proportion of highly educated individuals is Claremont in the city of Cape Town in the Western Cape; a precinct close to the University of Cape Town. This is followed by the precinct of Sandton; a prominent business hub in the city of Johannesburg in Gauteng, which is home to over 100 000 people. Piet Plessis and Boshoek, the precincts with the highest proportions of people with no schooling at all, are found in the North West province. These precincts have high proportions of men and are home to very few people. Less than 4 000 people live in Piet Plessis. The primary economic activities in the North West are mining and agriculture.

Table 3. Police Precinct Level Sample Characteristics: Summary Statistics

	Mean	Median	Std. Deviation
Individual Demographics			
Proportion Males	0.49	0.49	0.03
No Schooling	0.07	0.06	0.05
Incomplete Primary Schooling	0.25	0.26	0.08
Complete Primary Schooling	0.05	0.05	0.02
Incomplete Secondary Schooling	0.27	0.27	0.06
Complete Secondary Schooling	0.16	0.14	0.07
Higher Education	0.07	0.04	0.08
Proportion African	0.71	0.88	0.33
Proportion Coloured	0.15	0.01	0.27
Proportion Indian/Asian	0.02	0.00	0.06
Proportion White	0.11	0.05	0.17
Proportion Other Race Group	0.01	0.00	0.01
Number of Languages	11	12	1.46
Proportion Non-Citizens	0.03	0.01	0.05
Proportion Under 30 Years	0.58	0.58	0.08
Household Demographics			
Average Household Size	3.70	3.58	0.88
Proportion Urban	0.59	0.76	0.40
Proportion Tribal Areas	0.25	0.00	0.39
Proportion Farmland	0.16	0.03	0.26
Average Annual Per Capita Income (R)	28 658	17 204	34 167
Labour Market			
Proportion Unemployed	0.25	0.25	0.11
Job Dependency Rate	0.74	0.74	0.13
Inequality			
Gini Coefficient	0.49	0.48	0.08
Population	45 169	29 572	47842.69
Households	12 558	8 060	13718.99

Notes: Own calculations using a dataset derived from SAPS crime statistics and Census Community Profile data.

Figure 3. Population Density per Police Station



Notes: Own calculations using a dataset derived from SAPS crime statistics and Census Community Profile data; Population mean indicated by the dashed line.

The average household size varied from two to 15 people; and the mean household size was just under four people. On average, urban areas made up a mean of 59 percent and median of 76 percent of the space in police precincts. Tribal and farm lands made up much smaller proportions than urban, although there were some precincts that were entirely made up of one of these three types of geographic area. The average proportion of youth under 30 years was high, at almost 60 percent of the precinct population. This highlights the youth bulge in South Africa's population. The mean and median are identical, suggesting a highly consistent distribution.

Precincts with low proportions of youth tend to be urban and situated along the coast (e.g. Table Bay Harbour, Camps Bay, Fish Hoek). The precincts with the lowest proportions of youth are Table Bay Harbour and International Airport King Shaka – also precincts with high proportions of non-citizens and an overrepresentation of men. In general, the Western Cape features highly amongst precincts with low proportions of youth and high average per capita household income. By contrast, the Eastern Cape and KwaZulu-Natal are the provinces with the highest proportions of youth. These precincts tend to be rural and much poorer than the average or median per capita household income; women also tend to be overrepresented in these precincts. For example, Mpisi – the precinct with the poorest average per capita income – has the second highest proportion of youth, at 75 percent. In general, these precincts could possibly be reflecting homestead patterns of migrant workers.

The average per capita annual income was about R28 000 a year. The mean is larger than the median indicating a right-handed skewness to the distribution, which is characteristic of the country's high levels of income inequality. The richest police precinct is Sandton, followed by Rosebank, which are suburban areas in the city of Johannesburg. Sandton was also one of the precincts with the highest proportions of highly educated people. The poorest precinct is Mpisi, which – along with the next five poorest precincts – is in the rural Eastern Cape.

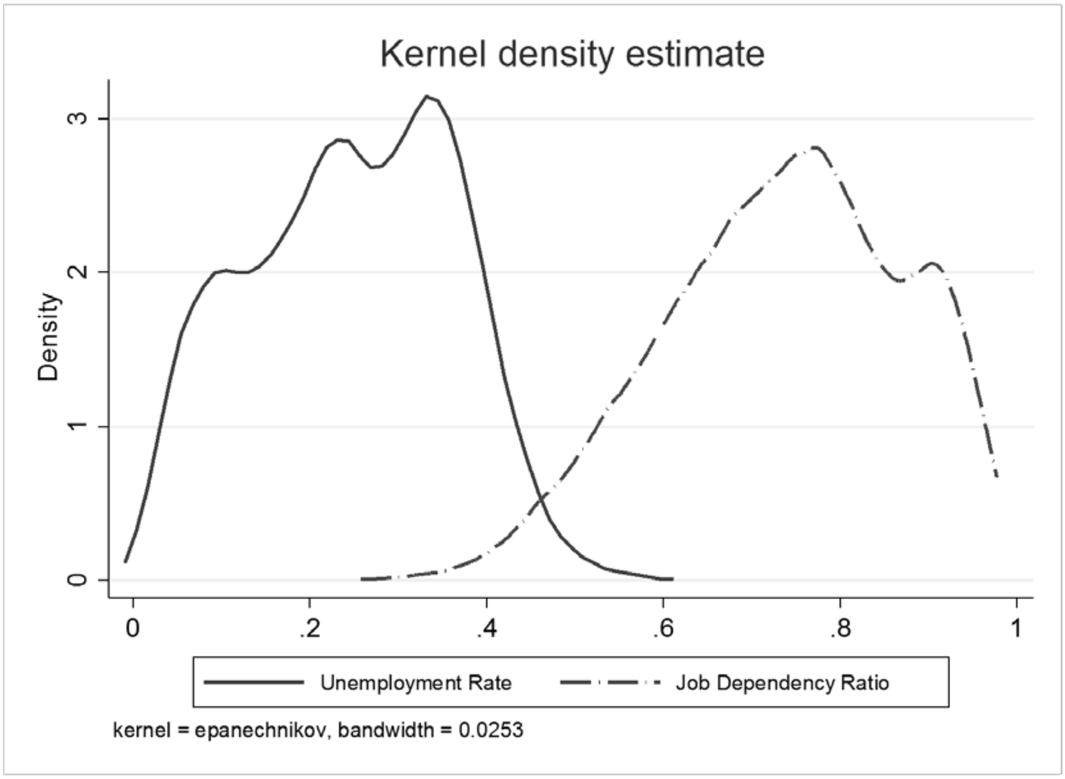
The Gini for the whole population is 0.66, which is consistent with other estimates of the Gini in South Africa (Leibbrandt et al., 2012). The average of the Gini coefficient for police precincts however, was 0.49 and the median was 0.48. This implies a largely consistent distribution around the mean and median, with only a handful of police precincts with Gini coefficients over 0.7. These included International Airport King Shaka (which incidentally, also had a proportion of non-citizens over 10 percent) and Rosebank. The other police precincts with exceptionally high inequality levels were a mixed collection of entirely urban areas and entirely rural areas, as well as a mixture of different provinces.

Our two labour market variables are the unemployment rate and a variation of the unemployment rate, which we call the job dependency rate. In this data, the average precinct unemployment rate was about 25 percent, which mirrors the national unemployment rate (DPRU, 2013). The distribution of the precinct unemployment rate is plotted in Figure 5. The distribution varies widely, with precincts with only 2 percent of the labour force unemployed to those with about 60 percent unemployed. The distribution has several peaks but ultimately is slightly right skewed. The police precincts with unemployment over 50 percent are all located in rural areas in the Eastern Cape, Mpumalanga and Limpopo. As expected, average income in these precincts is below the average income level for the sample. Women are slightly overrepresented in these precincts as well. Precincts with the lowest unemployment rates are not located in any province in particular, but, with the exception of the precinct with the absolute lowest unemployment rate, are typically urban and have incomes above the precinct average. The precinct with the lowest unemployment rate is Augrabies, which is only about 50 percent urban and has a below average income, but an unemployment rate of 1.75 percent.

The average precinct job dependency rate is 74 percent. We expect this rate to be higher than the precinct unemployment rate since the denominator includes the whole precinct population and not just the precinct labour force. The range of the job dependency rate is similar to the unemployment rate, although at a higher magnitude, as also evidenced in Figure 5. The job dependency rate is slightly more normally distributed than the unemployment rate and is skewed to the left. At minimum, a quarter of the precinct population is reliant on the jobs of the other three quarters. At most, 98 percent of individuals in a precinct are reliant on only 2 percent of the population with jobs. The precincts with the lowest job dependency rates are similar to those with the lowest unemployment rates, although not exactly the same. The lowest job dependency rates are found in police precincts in Limpopo and the Northern Cape; are characterised by a high proportion of men; are often highly rural; have high proportions of youth; and generally have a slightly lower than average per capita income level. For example, the police precinct of Tolwe in Limpopo has the third lowest job dependency rate at 37 percent. Tolwe has an average per capita income level of close to average, with 57 percent of the population comprised of men, and 50 percent comprising people under the age of 30.

The precincts with the highest job dependency rate are typically rural with proportions of youth above 70 percent; are mostly located in the Eastern Cape and KwaZulu Natal; have very low average per capita income; and have an overrepresentation of women. The precinct with the highest job dependency rate is Hlababomvu, with a job dependency rate of 98 percent; however, the unemployment rate for this precinct is 25 percent (which is the average unemployment rate for the sample). In other words, high job dependency does not always coincide with high unemployment. Similarly, Mpisi – the poorest precinct in the sample – has the fourth highest job dependency ratio at 96 percent, but also has an average unemployment rate of 22 percent. In fact, the five precincts with the highest job dependency rates all have unemployment rates varying between 21 percent and 36 percent. This emphasises the difference between these two measures of local unemployment.

Figure 4. Densities of Police Precinct Level Unemployment and Job Dependency



Notes: Own calculations using a dataset derived from SAPS crime statistics and Census Community Profile data.

As noted above, the available police infrastructure variables in this data are at the provincial level. The distribution of precincts in each province and the population by province is provided in Table 5, as well as the rate per 100 000 people of police personnel, victim rooms and motor vehicles. The bulk of the precincts are found in the Eastern Cape and KwaZulu-Natal, while the two most populous provinces are Gauteng and KwaZulu-Natal. Gauteng only has 137 precincts for 12 million people, whereas the Eastern Cape has 191 precincts encompassing 6 million people. This is understandable because Gauteng encompasses a relatively small area, albeit being a densely populated province, compared with the Eastern Cape, which has a much larger surface area. The least populous provinces are the Northern Cape and Free State. Although these two provinces make up a large proportion of the surface area of South Africa, a large proportion of land is covered by desert, which – along with distance from economic hubs in Gauteng and the Western Cape – explains the low population density.

The highest police personnel to population rate occurs in the Northern Cape, followed by the Free State, most likely driven by the low population density there. This is followed by the Western Cape, Eastern Cape and Gauteng. It is understandable that many resources are allocated to provinces like the Western Cape and Eastern Cape, since high levels of economic activity in these areas could attract more crime. Resource allocation responds to crime rates, meaning that these higher resource allocations could be reflecting higher crime rates. The same conclusions could be drawn from the number of victim rooms and motor vehicles per provincial population. By contrast, KwaZulu-Natal – which is one of the most populous provinces, and which has the second highest number of police precincts – has some of the lowest resource rates by personnel and motor vehicles. Perhaps this province is neglected, if we consider how highly populated it is. On the other hand, perhaps lower resource allocation to KwaZulu-Natal reflects lower historic crime rates.

Table 4. Precincts, Population, and Police Infrastructure Variables by Province

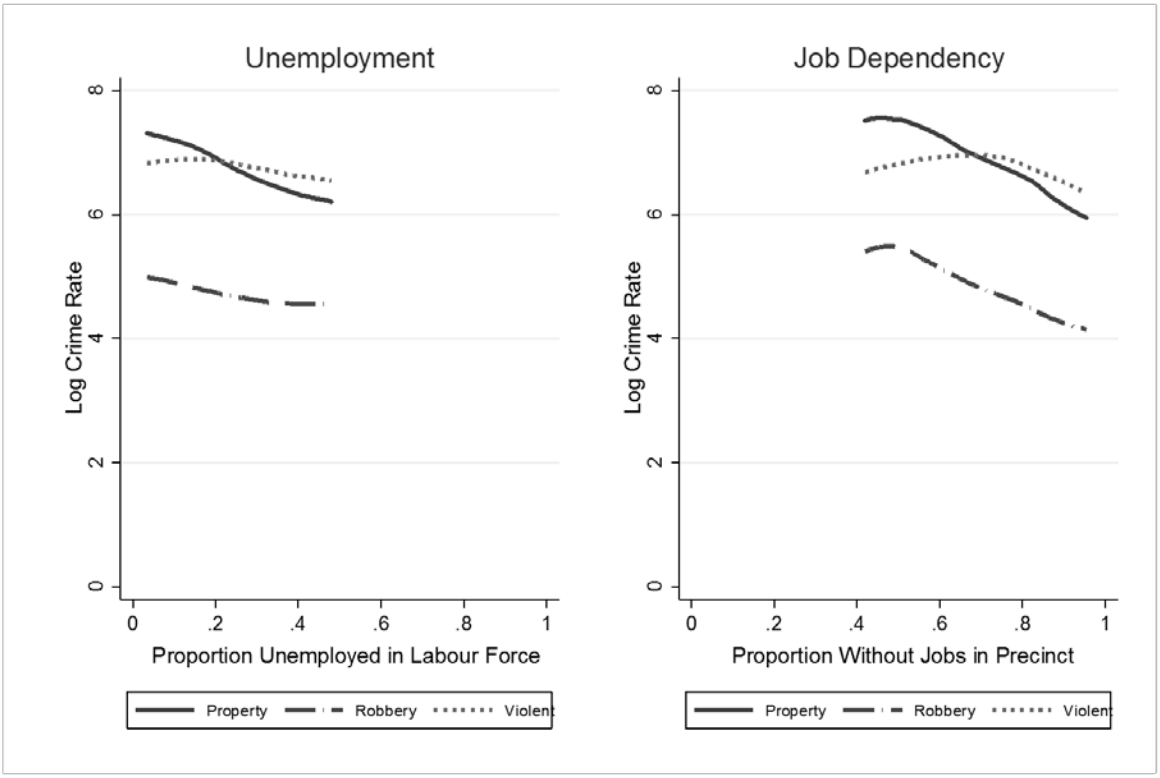
Province	No. of Police Precincts	Proportion of Police Precincts	Population (000s)	Proportion Population	Police Personnel Rate	Victim Room Rate	Motor Vehicle Rate
Eastern Cape	191	0.17	6 373	0.13	315.36	1.52	92.47
Free State	110	0.10	2 745	0.05	436.46	1.89	111.62
Gauteng	137	0.12	12 000	0.24	301	1.22	79.64
KwaZulu-Natal	186	0.17	10 300	0.20	236.73	1.7	67.06
Limpopo	94	0.08	5 139	0.10	239.26	1.38	74.14
Mpumalanga	86	0.08	4 037	0.08	246.52	1.78	64.43
North West	82	0.07	3 408	0.07	283.31	1.17	78.37
Northern Cape	91	0.08	1 158	0.02	604.66	4.92	152.8
Western Cape	149	0.13	5 738	0.11	357.6	3.31	113.23
Total	1 126	1.00	50 898	1.00			

6. Empirical Assessment

6.1 Non-Parametric Analysis

We begin our empirical analysis by running non-parametric local polynomial regressions of each of the socio-economic variables on logged property, robbery, and violent crime rates. These function as an initial step in our understanding of the relationship between crime and the set of socio-economic variables. We begin with precinct-level unemployment and job dependency rates in Figure 7. There is a clear negative relationship between resource acquisition crime (property and robbery crime) and both unemployment and job dependency. Precincts with higher proportions of unemployed people tend to have lower levels of both types of crime. In the same way, precincts with higher levels of job dependency tend to have lower levels of both types of crime. We expect this is because precincts with higher levels of employment are also areas with more resources to attract crime. Surprisingly, there is also a negative relationship between both measures of unemployment and violent crime, although the relationship appears parabolic; particularly in the case of job dependency.

Figure 5. Local Polynomial Regressions of Crime Rates on Unemployment and Job Dependency

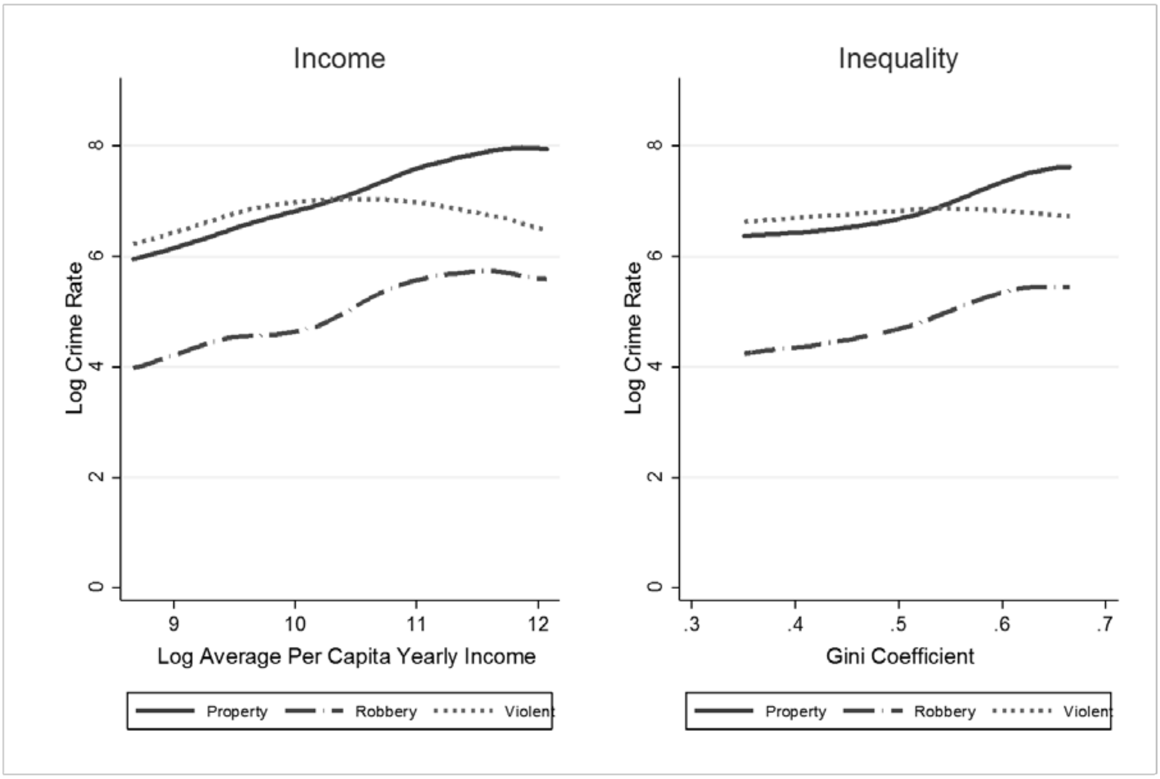


Notes: Own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data. Plotted from the 2nd to the 99th percentile of unemployment and job dependency.

We plot next, non-parametric local polynomial regressions of precinct-level inequality and income on the three crime types in Figure 8. In terms of income, results are partially as expected for property and robbery crime: positive slopes indicate that wealthier areas are more prone to resource acquisition-driven crime. There is some evidence of nonlinearity in the relationship between robbery crime and income; and to a lesser degree between property crime and income, but the evidence is weak. Conversely and surprisingly, violent crime has a very distinct inverted parabolic shape with income; also found by Demombynes and Özler

(2005) for violent crime in 1996. Inequality demonstrates similar relationships as those between income and crime. As the precinct-level Gini coefficient increases, richer households in the area increasingly become targets for crime as relative returns to crime in the area increase. There is a weak suggestion of an inflection point at the top end of the regression plots for robbery and property crime. Demombynes and Özler (2005) found a strictly linear relationship between inequality and property crime. Violent crime appears to be displaying no relationship with inequality in the form of a very horizontal regression plot. This appears to show that the level of violent crime in a precinct is unaffected by the level of inequality. As mentioned above, violent crime is likely to possess motivations other than resource acquisition, which make its relationship to inequality less clear.

Figure 6. Local Polynomial Regressions of Crime Rates on Income and Inequality



Notes: Own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data. Plotted from the 2nd to the 99th percentile of unemployment and job dependency.

In general, the local polynomial regressions presented in this section provide only weak support for the protection hypothesis, which hypothesised that crime would increase and then decrease with income and inequality. Specifically, the theory argued that crime increases with income in an area, as higher income represents higher returns from crime. However, at some point, income is high enough that individuals can start using it to buy protection from crime – which could lead to a downturn in resource acquisition crime. In a similar vein, crime first increases with inequality as increases in intra-precinct inequality represent an increase in the relative gain from crime in the precinct, as opposed to the cost of travelling outside the precinct to commit crime. High inequality can make relatively wealthier individuals in an area aware of their vulnerability to resource acquisition crime, prompting the desire to take protective measures; in which case they mobilise their wealth to do so. The plots can be described as only 'suggesting' turning points towards the top end of the regression plot. There was a negative relationship between resource acquisition crime and the unemployment variables.

Violent crime stood out by consistently having different relationships with socio-economic variables to the resource acquisition crime. Faintly parabolic relationships were found between the labour market variables and income in the case of violent crime; and no relationship was found with inequality.

However, local polynomial regressions are simple regressions that only regress one independent variable on the dependent variable at a time. We are unable to control for crucial explanatory variables, such as historical crime rates, population of youth, and population of men, with these regressions. These results are an important point of departure in the empirical analysis but cannot be considered a sufficiently exhaustive analysis of the relationship between inequality, income and crime; and ultimately, the protection hypothesis. In the next section, we embark on multiple regression analysis in order to achieve a *ceteris paribus* interpretation of the socio-economic determinants of crime.

6.2 Parametric Results

6.2.1 OLS Regression

Results from the OLS regressions are reported in Table 6. We include two specifications, with the difference being that income and inequality enter in level terms in Specification 1, and in quadratic form in Specification 2. A χ^2 statistic for the joint significance of income and inequality for Specification 2 is also reported at the bottom of Table 6. It is interesting to note that the labour market variables are only occasionally significant. Unemployment is significant in the case of violent crime, and job dependency is significant in the case of robbery crime. A 1 percent increase in the precinct unemployment rate is associated with a 27 to 28 percent increase in violent crime, *ceteris paribus*.[5] The sign on the job dependency variable for robbery crime is more in line with expectations: a 1 percent increase in precinct job dependency is associated with an 87 to 95 percent decrease in the precinct robbery crime rate, all else equal. This suggests that crime is low in areas with a high proportion of job-dependent (unemployed and not economically active) people, since this is a proxy for fewer profitable targets and lower returns to crime. The lack of significance for the labour market variables in general could be the result of endogeneity though. As noted above, the coefficients on the labour market variables are subject to reverse causality, since crime in an area could dampen business and employment prospects.

The socio-economic variables that were generally significant – taking joint significance into account – were precinct-level income and inequality. This is broadly consistent with the international literature (see Section 1.1), which generally finds significant and positive linear relationships between both income and crime, and inequality and crime. When included in level terms, income is consistently significant and positive in all crime regressions. This follows through into consistent joint significance of the quadratic terms of income. That is, we find an inverted U-shape between all crime types and income. This is in line with our expectations regarding resource acquisition crime as it fits the protection hypothesis. We find the result with violent crime harder to rationalise, although Demombynes and Özler (2005) also found such a relationship in 1996. Also in support of Demombynes and Özler (2005), we find that including inequality as a level term yields a significant positive relationship with robbery crime, but no significant relationship with violent crime. We go on to also find significant evidence of an inverted U-shape between inequality and both resource-acquisition crimes, and no significant parabolic relationship between inequality and violent crime.

5 This positive sign is hard to rationalise apart from the specification perhaps lacking a square, which would possibly match the inverted U-shape between violent crime and unemployment in Figure 7.

Table 5. OLS Regression Results of the Socio-Economic Determinants of Crime

Depvar:	**Property**		**Robbery**		**Violent**		**Property**		**Robbery**		**Violent**	
Specification:	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Regression Output												
Unemployment Rate	0.09 (0.14)	0.11 (0.14)	0.27 (0.28)	0.28 (0.28)	0.28* (0.14)	0.30* (0.14)						
Job Dependency Rate							-0.09 (0.20)	0.01 (0.20)	-0.95* (0.43)	-0.87* (0.43)	-0.02 (0.19)	0.03 (0.20)
Logged Average Per Capita Income	0.24*** (0.04)	1.77*** (0.45)	0.24** (0.08)	1.81* (0.88)	0.11** (0.04)	1.06* (0.43)	0.22*** (0.04)	1.75*** (0.46)	0.14 (0.08)	1.48 (0.88)	0.08* (0.04)	1.00* (0.44)
Logged Average Per Capita Income^2		-0.08*** (0.02)		-0.08 (0.05)		-0.05* (0.02)		-0.08*** (0.02)		-0.07 (0.05)		-0.05* (0.02)
Gini	-0.00 (0.15)	4.03*** (1.11)	0.79** (0.30)	3.50 (2.63)	0.01 (0.14)	0.88 (1.10)	-0.01 (0.15)	4.03*** (1.11)	0.71* (0.30)	3.60 (2.61)	-0.00 (0.14)	0.89 (1.11)
Gini^2		-3.97*** (1.11)		-2.69 (2.74)		-0.82 (1.12)		-3.98*** (1.11)		-2.89 (2.72)		-0.84 (1.12)
Lagged Crime Rate	0.87*** (0.02)	0.86*** (0.02)	0.71*** (0.02)	0.71*** (0.02)	0.87*** (0.02)	0.86*** (0.02)	0.87*** (0.02)	0.86*** (0.02)	0.72*** (0.02)	0.71*** (0.02)	0.87*** (0.02)	0.86*** (0.02)
N	1110	1110	1037	1037	1107	1107	1110	1110	1037	1037	1107	1107
R^2	0.90	0.91	0.76	0.76	0.84	0.84	0.90	0.91	0.76	0.77	0.84	0.84
Likelihood Ratio Test for Joint Significance $\chi^2 \sim (2)$												
Logged Average Per Capita Income		46.94***		12.61**		13.23**		41.40***		5.69+		9.32**
Gini		13.67**		10.76**		0.87		13.67**		8.91*		0.78

Notes: Standard errors in parenthesis; +p<0.1, * p<0.05, ** p<0.01, *** p<0.001; own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data; the following police precinct-level variables were controlled for: average age; education; average household size; province dummies; proportion of males; number of languages spoken; proportion non-citizens; proportion aged less than 30 years; location; and, the respective average precinct crime rate for the previous five years (2006-2010).

The combined inverted U-shapes of income and inequality for the resource-acquisition crimes lend support to our protection hypothesis. However, we also find partially similar results for violent crime, which is surprising given our understanding of the motivations behind this type of crime. Specifically, we find that whilst precinct-level inequality is not related to violent crime either as a level or as a quadratic term, there is a significant inverted U-shape between violent crime and income. It is hard to rationalise why violent crime would react to income in this way. Higher levels of income could mitigate against some of the poverty-induced triggers for violent crime, which stem from the associated psychological tensions of an unstable livelihood. However, it is harder to understand why violent crime would increase with income[6] for a segment of the distribution.[7]

Finally, in Table 6, we also report the relationship between lagged crime in an area and the 2011 crime rate. Historic crime levels are highly predictive of the 2011 crime rates. A 1 percent increase in the average crime rate for the previous five years leads to a 0.86 to 0.87 percent increase in property and violent crime; and a 0.71 percent increase in robbery crime. This variable is likely to be a proxy for a collection of features that are unobservable in this data; for example, precinct-level resource allocation, hysteria-effects in crime levels, and gang dynamics.

Overall, our findings so far are in support of the protection hypothesis for resource acquisition crime (property and robbery crime), by uncovering statistically significant evidence of an inverted U-shape between both income and inequality. We also find that violent crime has an inverted U-shape with income – consistent with Demombynes & Ozler (2005) – although we do not attribute this to the protection hypothesis and find it hard to explain. Altogether, the labour market variables are mostly insignificant to the prevalence of any type of crime. Within a returns to crime framework, these results can be interpreted as meaning that precinct-level income and inequality signify increased returns to crime, which would undoubtedly just continue to increase if private protection did not enter the equation. It is only when income and inequality reach such heights that they enable the purchase of private protection, that returns to crime in these precincts begin to fall. On the other hand, unemployment – conceptualised either as the standard unemployment rate or as job dependency – does not appear to be a strong signifier of returns to crime. This could be because income from employment can vary widely making some employed people relatively poor targets if their income is in fact very low. Alternatively, the endogeneity of these variables could be undermining their statistical significance; an identification issue dealt with in the next section.

6.2.2 Two Stage Least Squares

In the OLS analysis, we find a general lack of significance of the labour market variables, which could be due to endogeneity in the model. Since labour market conditions can both explain and be explained by the prevalence of crime, the concern is that reverse causality is undermining our ability to accurately estimate the statistical significance of the unemployment and job dependency variables. For example, it is unlikely that businesses and their associated employment levels will flourish in areas rife with crime. In attempting to minimise these endogeneity concerns with our unemployment variables, we use an IV regression design. The instrumental variable we use for our labour market variables is a proxy for cognitive skills: the precinct-level average of a four-part index for ability to remember and concentrate. We link cognitive skills to the ability of an individual to secure, retain, and succeed in employment. However, aside from via the labour market channel, we see no viable link between cognitive

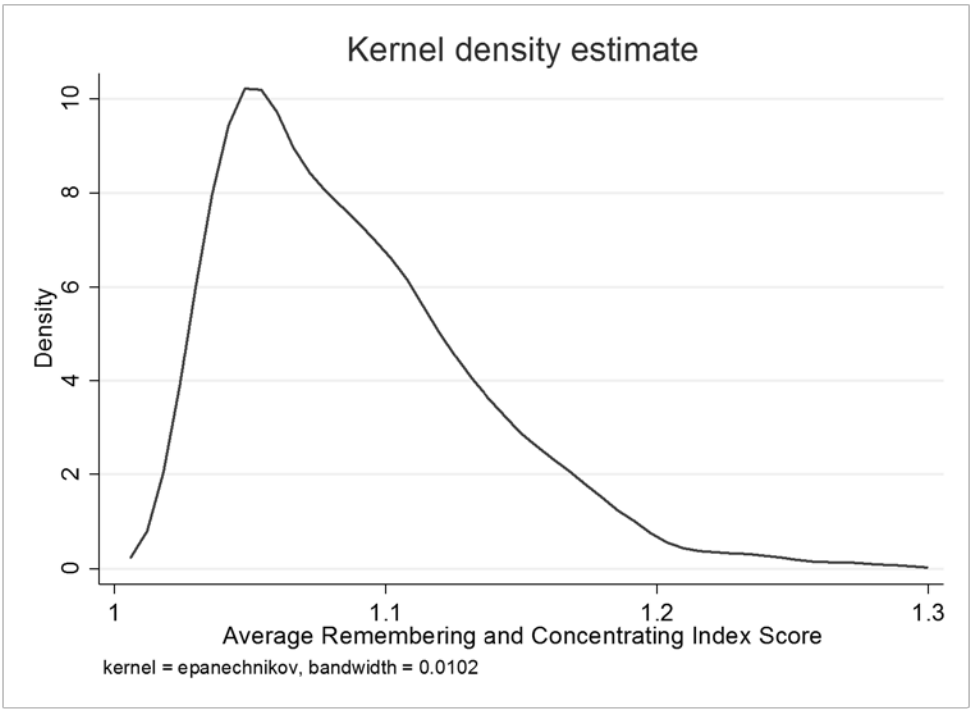
[6] It has been shown that poorer areas have poorer resource allocation in the Western Cape and that to an extent such areas have lower crime reportage rates (Redpath and Nagia-Luddy, 2015).

[7] It is worth noting that Demombynes and Özler (2005) also report the inverted U-shape between income and violent crime but do not attempt to explain it.

skills and resource acquisition crime, particularly when education has been controlled for. We do think there may be a convincing link between cognitive abilities and violent crime, since cognitive ability can be linked to 'impatience' in the classic economic model of crime where psychological costs and benefits are concerned. More specifically, cognitive ability is directly linked to an individual's level of executive functioning (i.e. how highly their frontal lobes perform), which modulates the ability to consider future consequences. Hence, violent crime, which we consider to be driven by psychological motivations, may be more likely to occur when an individual has limited cognitive abilities. We therefore limit the IV regression to property and robbery crime. By using this IV to predict employment in the first stage, we generate exogenous variation in the employment rate variable that is then used to predict the relationship between property and robbery crime and employment in the second stage.

We describe the IV in Figure 8 below. The cognitive ability index ranges from 1 to 4, increasing in difficulty; however, high population numbers (relative to the range of the score) and an extremely skew distribution compress the range to effectively varying between 1.02 and 1.29 with a mean of 1.09. Most people in most precincts have a score of 1 – meaning they have no difficulty remembering or concentrating. The distribution then tapers off quite steeply for the remaining scores that increase in difficulty.

Figure 7. Distribution of the Average Remembering and Concentrating Index



Notes: Own calculations using a dataset derived from SAPS crime statistics and Census Community Profile data.

We begin our analysis by testing whether our IV is statistically related to our endogenous unemployment variables in columns (a) and (b) of Table 9. Since we are interested in testing the protection hypothesis, and we found convincing evidence of this in the OLS results, we only include the quadratic versions of income and inequality in the control variables for these regressions. We find that remembering and concentrating is statistically significant at the 0.1 percent level for job dependency, but fails to yield significance for the unemployment rate. A possible reason for the lack of significance for the unemployment rate is that unemployment in South Africa is structural in nature (Bhorat, 2004; Bhorat & Hodge, 1999; Bhorat & Mayet, 2012; Burger and Woolard, 2005; Edwards, 2002). That is, structural factors such as distance from economic activity and skills mismatch are greater barriers to employment than individual-level effects. Cognitive ability is, however, still relevant for employment, as its significance in the job dependency specification shows. Rather, the effect is more visible at the community-wide, rather than labour force-wide, level. Those who are most (cognitively-) able to work are those who do, when all non-economically active individuals are also considered. Regarding this, we find that a one-unit increase in the average remembering and concentration index for a precinct is associated with a 17 percent increase in the job dependency rate, *ceteris paribus*.

Columns (c) and (d) of Table 9 test the exogeneity condition as far as possible. This is not the exogeneity of the structural model since we cannot test that – rather, we test whether our IV is unrelated to our dependent variables in the absence of the labour market measures. We find that the index fails to raise significance at the 10 percent significance level for any of the dependent variables. Given the evidence in Table 8, we only continue with job dependency, and discontinue with the unemployment rate, in the second stage for the dependent variables of property and robbery crime.

Table 6. Cognitive Index as an Instrumental Variable: Test Results

	Relevance Condition/ First Stage		***Exogeneity Condition***	
Dependent Variable:	**Job Dependency**	**Unemployment**	**Property**	**Robbery**
Column:	**(a)**	**(b)**	**(c)**	**(d)**
Average Remembering and Concentrating Index	0.166***	-0.068	0.223	-0.893
	(0.045)	(0.063)	(0.295)	(0.565)
N	0.929	0.801	0.907	0.765
R^2	1126	1126	1110	1037

Notes: Standard errors in parenthesis; + p<0.1, * p<0.05, ** p<0.01, *** p<0.001; own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data; the following police precinct-level variables were controlled for: income inequality; income inequality squared; average per capita income; average per capita income squared; average age; education; average household size; province dummies; proportion of males; number of languages spoken; proportion non-citizens; race; proportion aged less than 30 years; and, location. For the Exogeneity Condition specifications have an additional control: the average crime rate for the previous five years.

We present our final set of results in Table 10. These are our preferred set of results since they go furthest to deal with endogeneity. The first thing to note is the lack of significance of the job dependency variable. In the previous OLS regression, job dependency was significant in the case of robbery crime; however, this significance disappears when we more carefully identify job dependency. Our analytical conclusion then, is that job dependency is not an important signifier of returns to resource acquisition crime at the precinct level.

Income and inequality represent the returns to crime far more directly, and the relationships identified in the OLS regression are partly reinforced. That is, we find strong statistically significant evidence for the protection hypothesis for property crime. Property crime first increases and then decreases with both income and inequality. The positive relationship between intra-precinct inequality and property crime is one also found by Demombynes and Özler (2005), although our contribution is to uncover the nonlinearities in the relationship. Demombynes and Özler (2005) also explore inequality at a larger local level than just intra-precinct;[8] yet still find a significant relationship between intra-precinct level inequality and crime. They explain that although criminals can travel to richer areas than their own to maximise returns from crime, travel costs and idiosyncratic knowledge of their local area affect the cost-benefit decision to commit crime further afield than the local precinct. Indeed, as local inequality rises, the returns to crime to be made within the precinct increase for those at the bottom end of the precinct income distribution.

The protection hypothesis, however, is not found at all for robbery crime: income and inequality are both jointly insignificant. This could be linked to the motivational differences between robbery and property crime. Property crime, which covers crimes such as house burglary, involves more time planning and deliberation, in turn allowing more time for potential victims to organise counter-moves (i.e. protect themselves). On the other hand, the crime type making up the majority of the robbery crime category is common robbery, which excludes a weapon. We understand this as making robbery crime more opportunistic than property crime and therefore the protection hypothesis is perhaps less relevant, since robbery crime is harder to predict. Since robbery crime is mostly made up of common robbery, this means that it does not occur at a premises, the latter of which is considerably easier to protect. In connection to this, the lagged crime rate has also been consistently lower for robbery crime throughout all the regression analysis, another suggestion that this crime category is more *ad hoc* and contingent.

[8] Specifically, they include a dummy for whether a precinct is the wealthiest in what they term the 'criminal catchment area'; defined as the 'home' precinct and all bordering precincts. In other words, this is more a measure of relative wealth than inequality (since every area, regardless of level of inequality, will have a wealthiest precinct).

Table 7. Second Stage IV Regression Results of the Socio-Economic Determinants of Crime

Depvar: **Specification:**	**Property** **(2)**	**Robbery** **(2)**
Job Dependency Rate	1.56 (2.09)	-9.87 (7.34)
Logged Average Per Capita Income	2.33** (0.90)	-0.94 (2.23)
Logged Average Per Capita Income^2	-0.10* (0.04)	0.02 (0.09)
Gini	4.10*** (1.12)	2.49 (3.20)
Gini2	-3.89*** (1.13)	-2.54 (3.21)
Lagged Crime Rate	0.85*** (0.02)	0.76*** (0.02)
N R^2	1110 0.90	1037 0.66
Likelihood Ratio Test for Joint Significance χ^2 ~ (2)		
Logged Average Per Capita Income	8.67*	0.86
Gini	14.24***	2.34

Notes: Standard errors in parenthesis; + p<0.1, * p<0.05, ** p<0.01, *** p<0.001; own calculations using a dataset derived from SAPS crime statistic data and Census Community Profile data; the following police precinct-level variables were controlled for: income inequality; income inequality squared; average per capita income; average per capita income squared; average age; education; average household size; province dummies; proportion of males; number of languages spoken; proportion non-citizens; race; proportion aged less than 30 years; location; and, the average crime rate for the previous five years.

The lagged crime variables remain large, positive and statistically significant in the IV regression. As previously discussed, this variable is likely controlling for a range of unobservable factors such as resource allocation, hysteria around crime, and gang dynamics. Altogether, these aspects contribute greatly to the prediction of crime in 2011; however, we unfortunately cannot separate out the relative contribution of any of these variables.

The IV regression has reduced the protection hypothesis to only property crime and reinforced that the labour market is not a good signal of returns to crime in a precinct. This leaves us with a cohesive theory about how socio-economic variables signal returns to precinct-level property crime: crime increases and then decreases with income and inequality, and the labour market is insignificant. However, we are left relatively bereft with regards to robbery crime. Not only is the labour market insignificant, but so are income and inequality. In fact, the only variable that is significantly related to robbery crime is its lagged crime rate, which we have already explained is typically lower in magnitude than that of violent or property crime.

7. Conclusion

This paper has attempted to understand how socio-economic factors influence crime in South Africa at the police precinct level. We find that the relationships between crime and these variables are influenced by the type of crime (property, robbery, or violent) and the precinct-level unit of analysis. We find that precinct-level income and inequality are strong signals of returns to property crime. Specifically, we find support for our protection hypothesis, which posits that crime first increases and then decreases with these measures of income and inequality. We argue that crime increases with income, since this represents returns to crime, and crime increases with inequality, since inequality signals a relative increase in returns to crime within that precinct. However, at some level of income and inequality individuals become aware of their vulnerability to acts of crime and take measures to protect themselves, leading to a downturn in property crime.

Protection aside then, there is evidence of a positive relationship between precinct-level income and inequality, and property crime. This finding is supported by Demombynes and Özler (2005), who urge policy-makers to pay attention to precinct-level inequality. Demombynes and Özler (2005) also explain that such a positive relationship with income does not imply that police resource allocation should categorically increase with income, since this would be retrogressive, as it is precisely the wealthy that can afford private protection. Indeed, reports from the Khayelitsha Commission have already laid bare the destructive effect of leaving poorer areas under-resourced in favour of wealthier ones (Redpath and Nagia-Luddy, 2015).

Although robbery crime falls under the bracket of resource acquisition crime, it appears to have relationships with socio-economic variables that are distinct from those of property crime. In our preferred regression, no socio-economic variable was found to be statistically significantly linked to returns to robbery crime. We relate this to structural differences between property and robbery crime; the latter of which is far more opportunistic and contingent. The nature of robbery crime therefore makes it hard to predict. This makes protection harder to accomplish for victims, rendering the protection hypothesis less relevant. Furthermore, it makes resource deployment harder for policy-makers aiming to reduce crime by influencing signals of returns to crime.

Turning to violent crime, we hypothesised that the protection hypothesis would not be relevant, since offenders often know victims and share the same household – rendering protective measures hard to implement. Our results only partly reflect this theory. We find that income is significantly related to violent crime, and also takes the form of an inverted U-shape; yet, there is no significant relationship between intra-precinct inequality and violent crime. This result is consistent with Demombynes and Özler (2005), and stands in contrast to the importance of inequality for other crimes. This implies different foci for policymakers for different types of crime. It is far more important to pay attention to inequality in the case of property crime, than in the case of violent crime. The quadratic result with income remains hard to rationalise. One explanation is that violent crimes are less well-reported in poorer areas as a consequence of a lack of resources. This was partly shown to be the case in the Western Cape by Redpath and Nagia-Luddy (2015).

Overall, unemployment – conceptualised either as the narrow unemployment rate or as a job dependency rate – was not significantly related to crime. Such insignificance is consistent with Demombynes and Özler (2005) for crime in 1996. This result needs to be reconciled with a growing international literature that finds a strong negative relationship between unemployment and crime at the individual level. Indeed, this illustrates the effect of a micro versus a macro level lens, and which structural factors are working on which sides of Becker's model. In individual-level studies, an individual's probability of finding employment, amongst other factors, drives that individual to commit crime. In contrast, when precinct is the unit of analysis, it is inequality, income, and a composite set of factors included in the historical crime rate, that attract crime typically by signalling high returns. The major difference then is the difference between modelling an individual's decision to commit a crime, as opposed to modelling signals of returns to crime in a particular place, the latter of which this paper is concerned with. Our conclusion then is that precinct labour market conditions are not salient signals of returns to crime. Taking studies such as ours together with individual-level literature, we are able to obtain a well-rounded view of the mechanisms behind crime.

The unit of analysis of crime interacts with important econometric and data challenges. Socio-economic variables such as income level and unemployment are endogenous since they explain crime, but crime can also influence them. Our methodological approach in the face of endogeneity was to tackle the analysis from several angles. We ran non-parametric and parametric analyses, and employed an instrumental variable approach for unemployment. The instrumental variable approach was an important advance on the work of Demombynes and Özler (2005) who largely follow a similar methodology to this paper. A preferable option would be to simply have better data. The ideal case would be individual-level longitudinal data sets that include both socio-economic as well as crime data that is representative at a low spatial level (i.e. police precinct or lower). The National Income Dynamics Survey comes close by including many interesting and relevant variables, but lacks representivity. In the absence of such data, a more actionable solution would be questions related to crime protection in the Community Profiles data, and more questions that can be used as instrumental variables. This would allow a more direct testing of the protection hypothesis and improve our identification strategy. In addition, from the administrative data side, administrative data could be provided in a more detailed geographically disaggregated way. For example, the actual coordinates or street address of crimes could be reported.

We find a great deal of variety amongst crime. Property crime has the most foreseeable relationship with socio-economic variables: it increases and decreases with income and inequality, in line with our hypothesis about private protection. We link this to the strong resource acquisition motivation of property crime, with the additional requirement of time spent planning. This contrasts robbery crime which is also motivated by resource acquisition, but which is more opportunistic and contingent in nature making it hard to predict. Resultantly, it has no strong relationship to any socio-economic variable. The motivations behind violent crime are more complex and psychological than those behind resource acquisition crime. Violent crime increases and then decreases with income, but there is no relationship with inequality. The complex motivations behind violent crime make our findings hard to interpret. Overall, crime type and motivation has great bearing on how criminals interpret the relative importance of socio-economic variables as signals of returns to crime.

References

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

Altman, M. (2006). *Low Wage Work in South Africa*. Paper presented to IZA/World Bank Conference on Employment & Development, Berlin, Germany, 25-27 May.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy, 76*(2), 169–217.

Bhorat, H. (2004). Labour market challenges in the post-apartheid South Africa. South African Journal of Economics, 72(5), 940-977.

Bhorat, H., Ewinyu, A., & Monnakgotla, J. (forthcoming). Poverty, inequality and the nature of inclusive growth in post-apartheid South Africa. *Development Policy Research Unit*. Cape Town: DPRU, University of Cape Town.

Bhorat, H., & Hodge, J. (1999). Decomposing shifts in labour demand in South Africa. South African Journal of Economics, 67(3), 155-168.

Bhorat, H. & Mayet, N. (2012). Employment outcomes and returns to earnings in post-apartheid South Africa. *Development Policy Research Unit*. Cape Town: DPRU, University of Cape Town.

Brantingham, P. & Brantingham, P. (2013). Crime Pattern Theory, in *Environmental Criminology and Crime Analysis*. Ed. Wortley, R. & Mazerolle, L. Routledge, 78-91.

Budlender, J., Leibbrandt, M., & Woolard, I. (2015). South African poverty lines: A review and two new money-metric thresholds. *Southern Africa Labour and Development Research Unit*, Working Paper No 151. Cape Town: SALDRU, University of Cape Town.

Burger, R. & Woolard, I. (2005). The state of the labour market in South Africa after the first decade of democracy. *Journal of Vocational Education and Training, 57*(4), 453–476.

Chalfin, A. & McCrary, J. (2017). Criminal Deterrence: A Review of the Literature. *Journal of Economic Literature, 55*(1), 5-48.

Chalfin, A. & Raphael, S. (2011). Work and Crime. In *The Oxford Handbook of Crime and Criminal Justice,* M. Tonry (Ed.). pp. 444-78. Oxford and New York: Oxford University Press.

De Kock, C., Kriegler, A., & Shaw, M. (2015). A citizen's guide to SAPS crime statistics: 1994 to 2015. *Centre of Criminology*. Cape Town: Centre of Criminology, University of Cape Town.

Deininger, K. & Squire, L. (1996). A new data set measuring income inequality. *The World Bank Economic Review, 10*(3), 565-591.

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

DPRU (2013). An Overview of the South African Labour Market for the Year Ending 2013 Quarter 1. *Employment Promotion Programme*. DPRU Factsheet 11. *Development Policy Research Unit*: University of Cape Town.

DPRU (2016). Monitoring the performance of the South African labour market: An overview of the South African labour market for the year ending 2016 quarter 2. *Employment Promotion Programme*. DPRU Factsheet 15. *Development Policy Research Unit*.: University of Cape Town.

Edwards, L. (2002). Globalisation and the skills bias of occupational employment in South Africa. *South African Journal of Economics, 69*(1), 40-71.

Freeman, R. (1983). Crime and Unemployment. In *Crime and Policy,* J. Q. Wilson (Ed.). San Francisco: ICS Press.

Hsieh, C. & Pugh, M. D. (1993). Poverty, income inequality, and violent crime: A meta-analysis of recent aggregate data studies. *Criminal Justice Review, 18*(2), 182-202.

Huber, C. (2016). Understanding the intersectionalities of sexual violence reporting: A comparative analysis of police capacity, sexual violence, and alternative reporting methods in the Eastern and Northern Capes of South Africa. *Politics & Government Undergraduate Theses,* Paper 4.

Jonck, P., Goujon, A., Testa, M. R., & Kandala, J. (2015). Education and crime engagement in South Africa: A national and provincial perspective. *International Journal of Educational Development, 45*, 141-151.

Leibbrandt, M., Finn, A., & Woolard, I. (2012). Describing and decomposing post-apartheid income inequality in South Africa. *Development South Africa, 29*(1), 19-34.

Lemanski, C. (2004). A new apartheid? The spatial implications of fear of crime in Cape Town, South Africa. *Environment and Urbanization, 16*(2), 101-112.

Lim, H., Lee, H., & Cuvelier, S. J. (2010). The impact of police levels on crime rates: A systematic analysis of methods and statistics in existing studies. *Pacific Journal of Police & Criminal Justice, 8*(1), 49-82.

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

Mustard, D. B. (2010). How do labor markets affect crime? New evidence on an old puzzle. *Institute for Study of Labor Discussion Paper 4856.*

Piel, A. M. (1998). Economic Conditions, work, and crime. In *The Handbook of Crime and Punishment,* W. Tonry (Ed.). pp. 302-22. Oxford and New York: Oxford University Press.

Pratt, T. C. & Cullen, F. T. (2005). Assessing Macro-Level Predictors and Theories of Crime: A Meta-Analysis. *Crime and Justice, 32,* 373-450.

Redpath, J. & Nagia-Luddy, F. (2015). 'Unconscionable and irrational': SAPS human resource allocation. *SA Crime Quarterly*, *53*, 15-26.

Republic of South Africa (2009). *National Youth Development Agency Act (Act No. 54 of 2008).*

Rufrancos, H. G., Power, M., Pickett, K. E., & Wilkinson, R. (2013). Income inequality and crime: A Review and Explanation of the time-series evidence. *Social Criminology, 1*(1), 1-9.

SAPS (2011a). South African Police Service Crime Statistics, 2014/2015. Retrieved from https://www.saps.gov.za/services/crimestats.php

SAPS (2011b). South African Police Service Annual Report, 2011/2012. Retrieved from http://www.gov.za/sites/www.gov.za/files/General_infor_1.pdf

Shaw, M. & Gastrow, P. (2001). Stealing the show? Crime and its impact in post-apartheid South Africa. *Daedalus, 130*(1), 235-258.

Statistics South Africa (2011). South African Census Community Profiles 2011. Retrieved from https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/517/study-description

Statistics South Africa (2016). *Quarterly Labour Force Survey, Quarter 1: 2016*. Statistical Release, Pretoria: South Africa.

Verrinder, N. (2013). *Untangling the determinants of crime in South Africa.* Masters Thesis. Cape Town: University of Cape Town

Development Policy Research Unit
University of Cape Town
Private Bag
Rondebosch 7701
Cape Town
South Africa

Tel: +27 21 650 5705
Fax: +27 21 650 5711

www.dpru.uct.ac.za