Crime and Firm Dynamics
Evidence from South Africa*

Lawrence Edwards
School of Economics, University of Cape Town

Asha Sundaram
School of Economics, University of Cape Town

Draft, Please do not quote
September 2013

Abstract

This paper looks at the impact of crime on firm dynamism in South Africa. Using a unique, new dataset on firms, at the level of the municipal unit over the period 2003 – 2011, this study employs panel-data techniques to look at the impact of crime on firm births, exit and on the growth in the number of firms across regions. Results indicate that after accounting for regional size and time invariant, region specific factors affecting crime and firm dynamics simultaneously, high levels of business-related crime in the region are associated with greater exit and smaller changes in the stock of firms. This suggests that threats to the security of property are associated with lower firm dynamism across regions in South Africa. Our findings highlight the importance for regional economic development of better quality institutions, specifically, a crime-free environment that results in security of property rights.

University of Cape Town, Rondebosch, Cape Town 7701. Lawrence Edwards: Lawrence.edwards@uct.ac.za Phone: +27 21 650 2722, Asha Sundaram: asha.sundaram@uct.ac.za Phone: +27 21 650 2730.

*This paper is part of a World Bank project on private sector growth in less developed regions of South Africa. See Edwards L and Asha Sundaram (2013), 'Private Sector Growth in Marginalized Localities in South Africa: Assessing private sector growth potential', World Bank, Washington D.C. for the original paper. We thank Neil Balchin for excellent research assistance.

1. Introduction

This paper studies the impact of crime on firm activity, which is a key driver of economic growth. A large literature highlights the importance of institutions, especially security of property rights, in ensuring economic dynamism and firm performance. The argument is that while business-related crime can impose direct economic losses upon firms, causing smaller firms to shut down, failure to secure property rights can also discourage firms from investing in technology and physical capital by increasing their uncertainty about returns from such investment. Hence, the cost of dealing with crime inhibits entry of firms, particularly smaller less productive firms, contributing towards to a reduction in the allocative efficiency of domestic industry.

Further, the agglomeration literature also emphasizes absence of crime as a determinant of economic growth in cities. Kahn (2010) points out that regions or cities that see high incidences of crime also find it hard to attract and retain skilled labor, which can result in productivity losses for firms, compromising performance. Evidence from firm surveys also underscores the costs imposed by the presence of crime on firm activity. Figure 1 reveals that crime and access to electricity are reported as the most important constraints to operation by South African firms (World Bank, 2010).

In this study, we use a unique, new dataset on firms, at the level of the municipal unit over the period 2003 – 2011, and employ panel-data techniques to look at the impact of crime on firm births, exit and growth in the number of firms across regions in South Africa. We argue that South Africa is a useful case study to explore this question. The relationship between crime and business activity is well-established in the South African literature. The World Bank Enterprise Surveys of South African firms reveal that South African firms are far more likely to rank crime as a major constraint compared to similar upper-middle income countries. Consequently, the costs of crime as a percentage of revenue are higher in South Africa than in comparator upper-middle income countries (World Bank 2010).

Results indicate that after accounting for regional size and time invariant, region specific factors affecting crime and firm dynamics simultaneously, high levels of business-related crime in the region are associated with greater exit and smaller changes in the stock of firms. We find that qualitatively, our results are robust to measures of various types of crime. Our results suggest that a

one percent increase in the level of business-related crime is associated with a one percent increase in the exit rate of firms and a six percent decrease in the stock of active firms. This suggests that crime is associated with lower firm dynamism across regions in South Africa. Our findings highlight the importance for regional economic development of better quality institutions, specifically, a crime-free environment that results in security of property rights.

The remainder of the paper is structured as follows. Section 2 presents a background overview of trends in the incidence of crime in South Africa. Section 3 present the empirical strategy pursued to identify the relationship between crime and business activity. Section discusses the data used in the analysis. This is followed in Section 5 by the presentation of the results. Section 6 concludes the paper.

2. Background

2.1 Trends in the Incidence of Crime in South Africa

Although the World Bank's 2008 Investment Climate Assessment (ICA) suggests that crime in South Africa has become more problematic as a constraint to doing business, the South African Police Services (SAPS) crimes statistics shown in Table 1 indicates a decline in most categories of reported crimes over the period 2004-2011. Total serious crimes fell by 22 percent (log points) over this period. Crime dependent on police action for detection (Illegal possession of firearms and ammunition, Drug-related crime and Driving under the influence of alcohol or drugs) rose by 80 log points. In contrast, contact (murder, robbery), contact related crime (arson, damage to property) and property related crime all fell by between 22 and 33 percent.

These crimes apply both to businesses and individuals. Isolating the business related crime statistics reveals an increase in the number of business crimes reported from 2004 to 2010 (See last column of Table 1).[1] The share of business-related crimes in total serious crimes rose from 11 percent in 2004 to 15 percent in 2010.

Crime statistics also vary across regions within South Africa. Table 2 presents average crime statistics from 2004 to 2011, normalized by 2001 population levels, for municipal units classified

[1] The measure of business crime includes (a) Robbery at non-residential premises, (2) Truck hijacking, (3) Burglary at non-residential premises, (4) Shoplifting, (5) Theft of motor vehicle and motorcycle, and (6) Commercial crime.

according to four income-population density categories. Municipal units are classified as low-income (high-income) if the average income per capita in 2001 was lower (higher) than the median income per capita across regions. Similarly, municipal units are categorized as low or high population density areas depending on how they compare to the median population density of all regions.

In general, crime rates are higher for high income, high population density areas (average of 74.4 serious crimes per 1000 people) than low income, high population density areas (average of 21.7 serious crimes per 1000 people). Most (90 out of 111) low income, high density municipal units have crimes rates below that of the median municipal unit (see Table A3 in Appendix A).[2] Exceptions include Edenvale, Klaarwater, Ngqungqulu, Cele/Vumengazi and Ntuzuma.

Also presented in Table 2 are average annual changes in reported crimes from 2004 to 2011. Interestingly, crime rates have fallen the least in low income, high density areas. 75 of the 111 low income, high density municipal units experienced above median growth rates in reported crime statistics. Business crime rates have also risen the most in these areas. These trends are anticipated to have hampered private sector growth in these areas. The following section turns towards an analysis of this relationship.

3. Empirical strategy

We are interested in estimating the relationship between business activity and crime rates. Business activity is measured along various dimensions including, the stock firms, changes in stock (entry, exit), firm turnover and firm size. On each of these we are interested in identifying the relationship with crime, after controlling for regional characteristics that are associated with business activity, including market access or the pool of skilled labour, which are indicators of 'potential' for private sector growth, and physical infrastructure. The hypothesis we examine here is that given 'potential' and infrastructure, incidence of crime is significantly associated with firm performance.

One concern with establishing a causal relationship between the incidence of crime and economic performance is the presence of unobservable, region-specific shocks that might potentially affect crime and firm performance simultaneously, which might then bias our estimates. For

[2] As noted earlier, the crime rates per population results in some extreme outlier values. It is not clear to what extent these outliers can be attributed to the mapping procedure followed by Quantec.

instance, it is possible that a positive income shock in a particular region spurs firm activity, while simultaneously increasing resources available for policing, negatively affecting crime. This would bias our estimates of the impact of crime on business activity. A second source of bias is reverse causation, where economic activity in a region might either attract more crime, or reduce crime by improving household incomes.

The ideal way to address these endogeneity concerns would be to adopt an instrumental variables strategy for identification. However, we leave this for future work. In this version of our paper, in order to address the first concern, where data availability allows it, we perform a panel analysis. We account for any time-invariant unobservable shocks that might determine the regional environment and economic performance simultaneously.

To tackle bias introduced due to reverse causation, we relate the incidence of crime in the initial period to economic performance in the region in the current period. The idea here is that in so far as economic performance in the current period is unlikely to influence our regional environmental characteristics of interest in the past, this estimation is free of any bias from reverse causation. Finally, we also utilize different measures of crime to test for robustness of our results to these alternate measures.

We first explore the relationship between the stock of active firms and the provision of utilities across municipal units. Our basic cross-region specification is given by

$$\text{Ln}(Y_i) = \alpha_1 + \eta_{11}\text{Ln}(\text{Market Access})_{i0} + \eta_{12}\text{Crime}_{i0} + \eta_{13}\text{Controls}_{i0} + \varepsilon_{1,i} \qquad (3.1)$$

where the left-hand side variable is the stock of active firms (from the Business Returns database) averaged over the time period 2003-2011, and the right-hand side variables include market access and control variables for access to utilities, area, population and skilled labour. Our hypothesis is that $\eta_{12,} < 0$.

For the panel analysis, the Business Returns database enables the construction of a panel of firm activity by municipal unit over the period 2003 to 2011. The firm activity includes the stock of firms, the change in the stock of firms (net entry), entry rates (firm births/initial stock) and exit rates

(firm exit/initial stock). The Business Register database enables the construction of a panel containing the number of firm births by municipal unit. In addition to the firm data, Quantec provides data on crime levels by municipal unit.

The basic panel data specification is given as:

$$Ln(Y_{it}) = \eta_1 Crime_{i,t0} + \eta_2 Controls_{i,t0} + \psi_i + \psi_t + \varepsilon_{it} \quad (3.2)$$

where ψ_i and ψ_t denote municipal unit and time fixed effects. The dependent variable Y_{it} varies according to whether the focus is on the stock of active firms or changes in the stock of active firms. Quantec also provides estimates of population and gross value added by municipal units. These are included as controls.[3] The panel databases cover the period 2003-2011. Prior to discussing the results, it is important to re-affirm that the estimates do not entirely address endogeneity of the crime statistics.[4]

4. Data

The data used in this paper draw upon a number of unique and relevant databases. Data on firm activity is from the Companies and Intellectual Property Commission (CIPC) database. Crime statistics data by region are constructed by Quantec from station level data reported by the South African Police Services. We use numerous control variables to capture the institutional environment and other region characteristics that might be correlated with firm activity and crime simultaneously. For these variables, we use various data sources, including the 1996 and 2001 Population Censuses[5], and gross value added, capital and employment data obtained from Quantec's Standardized Regional Database. Firm data is at the zipcode level, while data from Quantec is at the municipal-unit level. A mapping was constructed to concord the firm, Quantec and census data to the municipal unit level. The following paragraphs describe each data source and the mapping in detail.

[3] The results are robust to their exclusion.
[4] As discussed earlier, the biases can go in either direction. Businesses may lobby for additional police services resulting in lower crime rates. Businesses also employ private security agents which may reduce crime rates. Alternatively, business activity attracts business related crime.
[5] The Community Survey of 2007 provides information at the level of municipalities. This is problematic for its use in this paper where much of the focus is on sub-municipal level areas, particularly within the metropolitan municipalities.

The firm data come from two sources:

(i) A Business Register database that contains information on the enterprise name, a unique enterprise registration number, company status (e.g. in business, deregistered, dissolved, etc.), physical and postal address including postal code, and registration date. This database was obtained from CIPC during the first quarter of 2012 and reflects the most up-to-date information on the enterprise at the time of download.

(ii) A Business Returns database containing the enterprise registration number and enterprise turnover. By law, all firms are required to lodge their annual returns with the CIPC. The database covers the period 2003 to 2011, but is incomplete (see later).

These databases provide a rich set of information of business entities in South Africa.[6] They are extensive, with the Business Register containing data for over 3 million enterprises and registration dates going as far back as the year 1801. The Business Returns database contains information on 731 922 firms, of which close to 55 percent have provided at least one positive turnover value.[7]

One concern regarding the databases is that the recent (1 May 2011) amalgamation of the Companies and Intellectual Property Registration Office (CIPRO) and the Companies and Intellectual Property Enforcement (CIPE) into the new Companies and Intellectual Property Commission (CIPC) led to significant delays in the registration of companies in 2011 (CIPC, 2012: 10). This was compounded by difficulties associated with digitalisation of the registration process as well as changes in regulations regarding the registration of close corporations.[8]

Nevertheless, by March 2012 the accumulated backlog of registrations had been addressed. Further, in preparation of the implementation of the Companies Act, the CIPC (and previously CIPRO) embarked on an exercise to clean the register by deregistering a large number of perceived

[6] Business entities can be registered as companies, close corporations (new registrations discontinued from 1 May 2011) and cooperatives.

[7] We are unsure why such a large difference exists. Even though there are missing values in the database, as long as these missing values are random across spatial units, we expect our estimates presented later to be consistent.

[8] The contract with the service provider responsible for the implementation of the Enterprise Content Management (ECM) solution was cancelled in June 2010. The registration of close corporations was also discontinued under the new Companies Act (2008). This led to a large number of applications to register close corporations before the closing deadline on 1 May 2011. The composition of registrations has subsequently shifted dramatically from closed corporations to companies.

companies perceived to be inactive.[9] CIPRO also increased enforcement of compliance in the lodging by companies of their annual returns. They issued close to 1, 6 million final deregistration notices through two dedicated deregistration interventions (CIPRO, 2011: 11). CIPC argue that the registers now provide a better reflection of economic activity in South Africa. However, it is acknowledged that the existence of a substantial number of shelf companies on the register continues to distort the database (CIPC, 2012: 10).

We use these databases to construct the stock for firms for the years in our sample, entry and exit of firms and total firm turnover. The Returns database provides information on date of registration. It also provides information on the operational status (active, dissolved, etc.) of the firm. This is useful for determining the total stock of companies in operation at the date at which the data were made available (early 2012). Unfortunately, however, we do not have information on the date when firms closed. We are therefore unable to construct a series of the stock of companies over time using the Register database.

However, we are able to construct a measure of the stock of firms from 2003 to 2011 using the Returns database. This database records turnover of companies in each year. The data are incomplete (See Appendix A), but a measure of the stock of companies in each year can be constructed using information on the date of registration and the final year in which the firm is observed. According to the data, the stock of firms rose from 351,548 in 2003 to 569,258 in 2011. This is substantially less than the close to 1.5 million firms registered as being in operation according to the Returns database.[10]

At this point, we note that the firm data we use has some shortcomings. First, the data available pertains to the enterprise and not the individual subsidiary plant that form part of the business entity. This has a number of implications for the analysis. The address provided generally refers to the location of the head-office. Most head-offices are located in one of the metropolitan municipalities and often within certain main places (e.g. Sandton is the location of many of the head-

[9] Many active firms were also deregistered. These firms have an incentive to request reinstatement as bank accounts are frozen for inactive or non-compliant entities. Nevertheless, the CIPC believes that many active companies remain deregistered (CIPC, 2012: 10).

[10] According to the REGISTER data, there were 1,403,274 whose status is: In Business, Cooperative, Conversion CO/CC or CC/CO, Duplicate Name and CO converted to Cooperative.

office of many of services companies). No information is provided on the number and location of the subsidiary plants. For example, the supermarket chains are mostly registered as being located in the Gauteng area, despite having stores distributed across the country. This exaggerates the regional concentration of companies. However, this is an issue that has plagued studies in the literature that use firm-level, where plant-level data are not available. Second, the database captures information on formal firms and not informal firms. This may lead to a severe under-estimate of the total stock of active firms within marginalized communities, again a common problem with most manufacturing databases across emerging economies.

The primary sources of data for our control variables, including data on employment, population, household income, education, and municipal services including electricity, refuse removal, sanitation and water and access to telecommunication services are the 1996 and 2001 Population Censuses. Data on these variables at the municipal unit are obtained from Quantec Research Ltd.

Finally, this paper draws upon the regional crime statistics provided by Quantec Research Ltd. This data are constructed from the station level data reported by the South African Police Services (SAPS). We use data from Quantec, instead of using data directly from SAPS, to be as consistent as possible with other data used in this analysis even though the exact procedure that underpins this mapping is unclear and various anomalies are evident in the data. For example, the SAPS crime statistics for stations do not necessarily correspond with the Quantec crimes statistics for the spatial units with the same name. [11] The mapping from the geographical areas in the SAPS data to the municipal units and local municipalities is done by Quantec.

Quantec has also constructed a spatially disaggregated database (the Standardized Regional Database) containing information on employment, output, gross value added, wage remuneration, capital formation for 23 industries in 497 municipal units (covering main places in metropolitan municipalities and local municipalities), that we use to generate some of our control variables[12].

[11] For example, 5110 serious crimes were reported to the Khayelitsha police station in 2011. According to the Quantec database, 13 738 serious crimes were committed in the main place Khayelitsha. Another example is Robben island, where according to the Quantec regional crime database around 2000 serious crimes were committed per year from 2004-2011. There were also 101 incidences of driving under the influence of alcohol or drugs in 2012. There is no police station on Robben Island.

[12] The municipal units are made up of 252 local municipalities and 245 main places within 6 metropolitan municipalities.

Quantec also provides data for these 497 municipal units on population, household income and expenditure, and various development indicators (education, access to municipal services, poverty lines, dependency rates, etc.). A severe drawback of this data is that the municipal level unit entries of the industry based data are based on projections of national level aggregate estimates to the lower spatial units. The data are not based on direct statistical inferences from survey results for these spatial units. The data are nevertheless informative in presenting a broad depiction of the spatial composition of employment and output.

A considerable effort was spent on 'cleaning' and preparing the data for subsequent use, and to match the firm data to data from the census and Quantec, to arrive at a database at the municipal unit level. The register data provide details on the enterprise address and postal code. These postal codes could therefore be used to map each enterprise to different spatial units. Unfortunately, the spatial units of the postal codes do not perfectly correspond with those of the Population Censuses.[13] Where possible, postal code areas were mapped to the 497 spatial units using the place names of the postal code database obtained from the South African Postal Office and the main place names of the Census based spatial units. However, many postal codes areas overlap provincial, municipal and main place boundaries. It was therefore not possible to uniquely map each postal code to the 497 spatial units. An aggregated set of 342 spatial units was then constructed. The data Appendix A provides more details.

A preliminary look at the data shows a strong correlation between business activity and crime rates. This is what we expect, since business activity tends to attract business related crime leading to a positive association between these variables.[14] This is revealed in Figure 2 that plots business crime rates against the stock of firms. This highlights the endogeneity of crime and firm activity, as discussed in the empirical section of our paper.

[13] See Lombard, M. (2005) "South African Postcode Geography" Paper presented at the seventh Africa GIS conference, CSIR International Convention Centre, Tshwane (Pretoria), South Africa, 31 October to 4 November 2005.
[14] This association is apparent in the data. A strong positive relationship is found between the share of business crime in total crime and the stock of firms across municipal units. The measure of business crime includes (a) Robbery at non-residential premises, (2) Truck hijacking, (3) Burglary at non-residential premises, (4) Shoplifting, (5) Theft of motor vehicle and motorcycle, and (6) Commercial crime.

5. Results

The regression coefficients for equation 3.1 are presented in Table 3A and B. Table 3B presents additional regressions explaining the stock of active firms in agriculture, mining, manufacturing and services separately. In Table 3A, columns (1) and (2) analyse the relationship between crime and stock of firms, controlling for 'potential', infrastructure quality and service provision, land area and population. Crime is measured by the principal component of the shares of crime arising from contact, property, police detection and other causes. Results suggest a weak negative relationship between the incidence of crime and stock of firms in the region (column 1), but this is not robust to the inclusion of area and population.

Columns (3) through (9) decompose the composite measure of crime into the separate indicators. While most of the coefficients on the crime variables are negative, none of them are statistically significant (at the level of 10 percent or below). Looking at the results by industry presented in Table 3B, we find considerable heterogeneity in results across sectors for the impact of crime on the stock of active firms. Particularly, areas with higher crime rates have lower stocks of mining firms, though the relationship is only significant at the 10 percent level. For all other industries, crime rates are not associated with firm stock.

Overall, the results do not corroborate the negative association between crime rates and business activity suggested by firm surveys, except in the case of mining firms. One reason is that the estimates do not deal with the endogeneity of crime. This will require the construction of an appropriate instrument. Another explanation is that the data allocation process followed by Quantec introduces measurement bias into the estimates. As noted, the station level crime data often deviates substantially from Quantec's municipal unit level data in which that station is located. If the allocation of crime statistics to the municipal units for which no station data are available is based on a quantity measure (e.g. share population, income, etc.), then a positive bias will be introduced. A third reason is that crime statistics may vary substantially even within municipal units. Businesses in high crime municipal units, may locate in areas within these municipal units where crimes rates are relatively low. Unpacking these relationships will require further research.

We now turn to our panel analysis and estimation of equation 3.2. Table 4 reports the estimates based on the Business Returns database, while Table 5 presents estimates based on the

firm birth database. Looking first at (columns 1 and 2 of Table 4), no association is found between the stock of active firms and the principal component for serious crimes. A rises in business crimes, however, is positively correlated with a rise in the stock of firms within municipal units. This result most likely reflects the bias arising from the endogeneity of the business crime variable.[15]

Growth in the stock of active firms (columns 3 and 4), however, is negatively related to the level of business crime. Rises in business crime rates, reduce growth in the stock of active firms within municipal units. The determining factor appears to be firm exit. As shown in columns (7) and (8) of Table 4, exit rates rise as levels of business crime increase. The coefficient on the principal component for crime is also positive, and weakly significant.

The level of crime therefore appears to have a detrimental impact on growth in the stock of firms. This finding is supported by the effect of crime levels on firm births using the Business Register database shown in Table 5. According to these results, a ten percent increase in the total number of severe crimes is associated with a 1.5 percent reduction in the number of firm births.[16] The coefficient is also significant and negative for the principal component for crime, but is not significant for business crime.

The panel data provides additional insights not found in the across-municipal unit based analysis. Whereas crime was found to be uncorrelated with business activity when looking across municipal units, this section finds evidence that rising crime levels within municipal units is associated with declining rates in the growth of the stock of active firms and the number of firm births. The impact of crime levels on firm exit appears to be the explanation.

Finally, we focus on firm turnover across regions as a measure of economic activity. Looking at firm turnover confers several advantages. First, since we now have a measure of the scale of each firm's activity, we are able to look at the relationship between the local economic environment and firm activity differentially across small and large firms. We first consider the total

[15] The coefficient becomes insignificant when estimated using the first-differenced generalized methods of moment (GMM) estimator of Arellano and Bond (1991) and lagged levels of business crime as an instrument for the first difference.

[16] In the estimates based on the individual crime variables, negative and significant coefficients are estimated for contact-related crime and other serious crime.

turnover of firms in the region, calculated as the sum of the firm's average turnover from 2003-2011. We also calculate turnover of the median firm within each municipal unit as the median of the average turnover of firms from 2003-2011. We then calculate mean turnover across firms within each region. The results are presented in Table 6A. The dependent variables are total, median and mean turnover at the regional level in columns (1) through (3), column (4) and (5), respectively.

We find that controlling for 'potential', infrastructure and service provision (principal component for utilities), crime rates are not significantly correlated with the turnover variables. This contrasts the enterprise survey findings (World Bank 2010) that crime impacts more heavily on small firms in South Africa.[17] Finally Table 6B reports estimates separately for agriculture, mining, manufacturing and services sectors. Results are mostly consistent. We do not find evidence that crime is negatively and significantly associated with firm turnover or firm size.

6. Conclusion

In this study, we examine the impact of crime on economic activity in the local area. We find evidence that crime is negatively related to the stock of active firms only in mining. However, there is evidence that a higher incidence of crime is associated with greater exit, and hence a lower growth rate, of firms. We do not find evidence for the impact of crime on firm turnover or size. This indicates that there is role for policy to strengthen institutional quality, by improving policing to ensure security of property rights and a crime free environment to encourage economic dynamism and growth.

[17] We also found no association between crime rates and the 75 percentile to 25 percentile ratio.

References

Arellano, M and Steve Bond (April 1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", The Review of Economic Studies, 58, 277 – 297.

Diaz O L, Plan D and P Pochet (2008), "Household Transport Expenditure in Sub-Saharan African Cities: Measurement and Analysis", Journal of Transport Geography, 16(1), 1-13.

Edwards L and Asha Sundaram (2013), 'Private Sector Growth in Marginalized Localities in South Africa: Assessing private sector growth potential', World Bank, Washington D.C.

Fally, Thibault; Rodrigo Paillacar and Cristina Terra (2010) Economic geography and wages in Brazil: Evidence from micro-data. Journal of Development Economics 91:155–168.

Hanson, G. H. (1998) Market Potential, Increasing Returns, and Geographic Concentration. Journal of International Economics, 67(1): 1-24.

Head, K. and T. Mayer. (2004) Market Potential and the Location of Japanese Investment in the European Union. Review of Economics and Statistics, 86(4): 959-972.

Kahn, M (2010), 'New Evidence on the Trends in the Cost of Urban Agglomeration', in Agglomeration Economics, Edward L Glaeser (Ed.), NBER and University of Chicago Press, February 2010, 339-354.

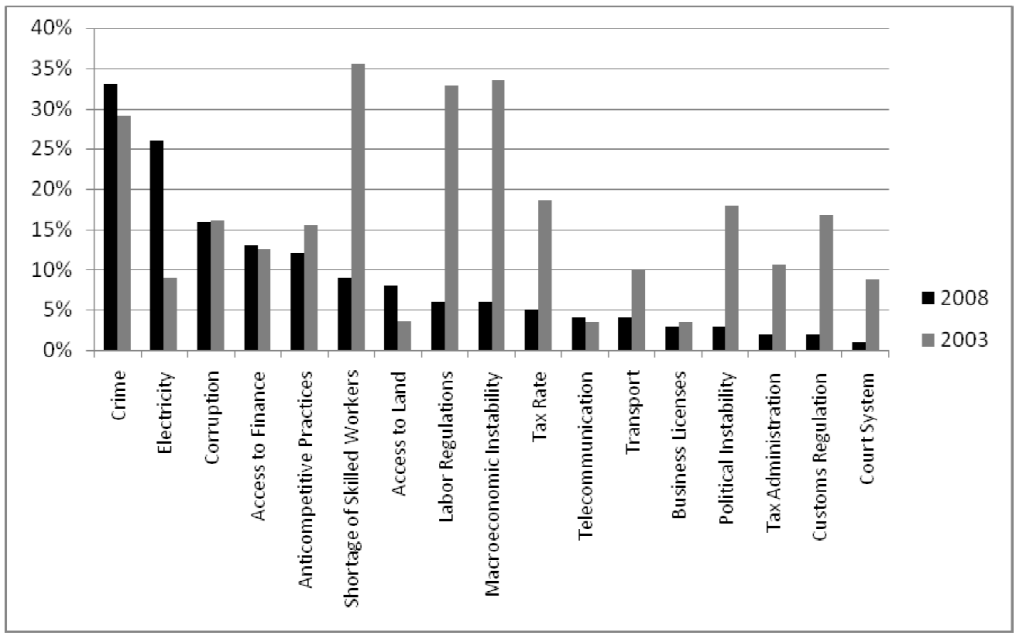
Municipal Demarcation Board (2011) 'MDB Municipal Capacity Assessments 2011 website', http://www.demarcation.org.za/capweb/default.aspx

Nunn N (2007), "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade," Quarterly Journal of Economics, 122(2), 569-600.

Redding S and Anthony Venables (2004), "Economic Geography and International Inequality", Journal of International Economics, 62, 53-82.

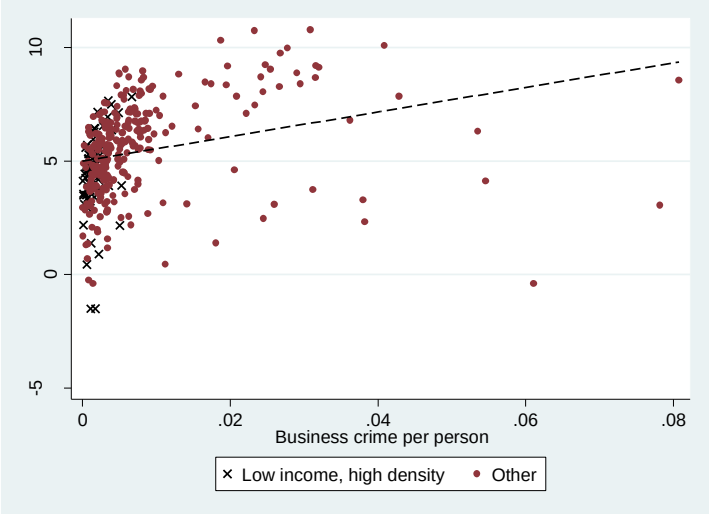
World Bank (2010) South Africa: Second Investment Climate Assessment: Business Environment Issues in Shared Growth, Volume 2. World Bank, Washington DC.

Figure 1: Percent of firms ranking constraints as major obstacle, 2003 vs. 2008.



Source: World Bank, 2010, 'South Africa: Second Investment Climate Assessment, Business Environment Issues in Shared Growth', Volume 2 Full Report (draft), July 2010

Figure 2: Stock of firms (2003-2011) and average business crime as share of 2001 population (2004-2011)



Source: Own calculations using Business Returns database obtained from CIPC and crime statistics from Quantec. The average stock of firms by municipal unit is calculated as the average stock over the period 2003 to 2011. Crime rates are calculated as average number of reported crimes from 2004-2011 as proportion of 2001 population.

Table 1: Reported crime statistics, South Africa

	Total Serious crim	1. CONTACT (CRIME AGAINST THE PERSON)	2. CONTACT-RELATED CRIME	3. PROPERTY-RELATED CRIME	4. CRIME HEAVILY DEPENDENT ON POLICE ACTION FOR DETECTION	5. OTHER SERIOUS CRIME	6. OTHER CRIME CATEGORIES	Business crimes
2004	2 638 702	886 209	167 053	665 318	104 414	734 217	81 491	285 108
2005	2 466 736	847 265	158 969	597 256	129 425	656 737	77 084	264 611
2006	2 237 409	756 112	151 887	570 698	142 259	551 334	65 119	264 252
2007	2 185 699	730 357	151 194	547 258	157 304	542 342	57 244	279 496
2008	2 110 583	692 545	144 364	521 513	171 015	527 574	53 572	286 606
2009	2 148 238	684 199	141 107	532 184	187 382	552 371	50 995	319 581
2010	2 174 741	676 445	138 835	553 368	212 321	540 918	52 854	332 971
2011	2 125 073	638 469	131 860	534 451	231 842	534 866	53 585	316 023
2012	2 137 374	623 487	128 639	536 093	260 209	537 330	51 616	305 808
Log Change 2004-2011	-0.22	-0.33	-0.24	-0.22	0.80	-0.32	-0.42	0.10

Source: Quantec, based on SA Police Service data.

Notes: The measure of business crime includes (a) Robbery at non-residential premises, (2) Truck hijacking, (3) Burglary at non-residential premises, (4) Shoplifting, (5) Theft of motor vehicle and motorcycle, and (6) Commercial crime.

Table 2: Serious crime statistics, by income/population density categories, 2004-2011 average, reported crime per 1000 population

	Total serious crimes	**Contact crimes**	**Contact related crimes**	**Property-related crimes**	**Crime heavily dependent on policy action for detection**	**Other serious crimes**	**Other crime categories**	**Business crime**
Crime rates (per 1000 of the 2001 population) by income/population density categories								
high income, high pop density	74.4	21.0	4.8	19.5	5.4	22.0	1.6	12.1
high income, low pop density	80.7	27.2	5.5	19.3	6.1	19.9	2.7	8.6
low income, high pop density	21.7	9.4	1.5	4.4	1.9	3.9	0.6	1.9
low income, low pop density	25.7	10.0	1.6	6.5	1.7	5.0	0.8	2.4
Total	50.4	16.5	3.3	12.6	3.7	12.9	1.4	6.6
Average change in crime rates, average annual log change, 2004-2011								
high income, high pop density	-2.9%	-5.0%	-3.0%	-2.6%	10.8%	-3.8%	-5.2%	3.4%
high income, low pop density	-3.8%	-4.9%	-4.4%	-3.0%	7.4%	-5.8%	-7.7%	1.3%
low income, high pop density	-2.2%	-4.3%	-2.8%	-1.0%	10.7%	-3.1%	-4.0%	7.1%
low income, low pop density	-3.6%	-4.3%	-3.4%	-3.6%	8.9%	-6.0%	-3.9%	1.5%
Total	-3.2%	-4.6%	-3.3%	-2.7%	9.6%	-4.8%	-5.0%	3.0%

Notes: Average values calculated using 2001 population levels as weights.

Table 3A: Relationship between stock of active firms and crime

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: ln(Average number of active firms, 2003-2011)								
	Base	Principal component	Total	Contact	Contact-related	Property	Police detection	Other	Business
ln (Market access)	0.403*	0.317**	0.315**	0.332**	0.322**	0.311*	0.317**	0.274*	0.236
	[0.206]	[0.158]	[0.160]	[0.158]	[0.163]	[0.167]	[0.147]	[0.159]	[0.156]
Dummy low income, high density	-0.794***	-0.626***	-0.626***	-0.623***	-0.625***	-0.627***	-0.609***	-0.631***	-0.635***
	[0.272]	[0.192]	[0.192]	[0.192]	[0.192]	[0.192]	[0.191]	[0.191]	[0.190]
% no schooling	0.053	-3.606	-3.588	-3.602	-3.640	-3.571	-3.901	-3.366	-3.351
	[3.159]	[3.278]	[3.284]	[3.333]	[3.264]	[3.261]	[3.259]	[3.249]	[3.273]
% primary schooling	-9.324***	-6.485	-6.504	-6.411	-6.481	-6.528	-6.092	-6.716	-6.767
	[2.783]	[4.616]	[4.619]	[4.584]	[4.573]	[4.621]	[4.615]	[4.611]	[4.597]
% higher education	7.179**	9.389***	9.347***	9.453***	9.423***	9.291**	9.844***	8.487**	7.604**
	[3.601]	[3.488]	[3.509]	[3.371]	[3.412]	[3.625]	[3.340]	[3.674]	[3.824]
Principal component utilities	0.257***	0.358***	0.358***	0.358***	0.357***	0.358***	0.355***	0.362***	0.366***
	[0.081]	[0.079]	[0.079]	[0.078]	[0.078]	[0.079]	[0.079]	[0.078]	[0.078]
Principal component crime	-0.129*	-0.009							
	[0.074]	[0.051]							
ln (area)		0.256***	0.255***	0.262***	0.258***	0.254**	0.265***	0.240**	0.230**
		[0.097]	[0.097]	[0.096]	[0.099]	[0.098]	[0.097]	[0.100]	[0.100]
ln (population, 2001)		0.902***	0.904***	0.894***	0.900***	0.906***	0.881***	0.926***	0.937***
		[0.100]	[0.099]	[0.103]	[0.098]	[0.098]	[0.102]	[0.095]	[0.095]
Total serious crime/population			-0.113						
			[1.106]						
Contact crime/population				-1.546					
				[2.422]					
Contact-related crime/population					-4.268				
					[16.702]				
Property crime/population						-0.059			
						[5.678]			
Policy detection crime/population							-10.435		
							[8.523]		
Other crime/population								3.942	
								[5.096]	
Business crime/population									13.898
									[9.877]
Constant	0.788	-9.967***	-9.940***	-10.111***	-10.002***	-9.907***	-9.786***	-9.525***	-9.047***
	[3.124]	[2.435]	[2.431]	[2.381]	[2.478]	[2.506]	[2.268]	[2.453]	[2.429]
Observations	342	342	342	342	342	342	342	342	342
R-squared	0.346	0.725	0.725	0.725	0.725	0.725	0.727	0.726	0.728

Notes: The dependent variable is the average stock of firms by municipal unit over the period 2003 to 2011 and is constructed using the Business Returns database. Police detection refers to the category "crime heavily dependent on policy action for detection". Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table 3B: Relationship between stock of active firms and crime, by industry

	(1)	(2)	(3)	(4)
	Dependent variable: ln(total number of active firms, 2004-2011)			
	Agriculture	Mining	Manufacturing	Services
ln (Market access)	-0.001	0.755**	0.407	0.310**
	[0.237]	[0.298]	[0.253]	[0.153]
Dummy low income, high density	-1.408***	-0.781*	-0.953***	-0.597***
	[0.351]	[0.414]	[0.327]	[0.186]
% no schooling	5.906	6.663	-3.059	-4.180
	[4.375]	[5.450]	[4.445]	[3.193]
% primary schooling	9.458	8.571	-1.360	-6.924
	[6.207]	[7.399]	[6.375]	[4.480]
% higher education	19.971***	22.520***	13.828***	9.108***
	[4.229]	[4.799]	[4.621]	[3.398]
Principal component utilities	0.515***	0.759***	0.509***	0.327***
	[0.113]	[0.155]	[0.114]	[0.076]
Principal component crime	-0.058	-0.172*	0.044	-0.007
	[0.069]	[0.100]	[0.068]	[0.049]
ln (area)	0.528***	0.547***	0.275*	0.233**
	[0.141]	[0.181]	[0.152]	[0.094]
ln (population, 2001)	0.874***	1.061***	1.185***	0.915***
	[0.154]	[0.179]	[0.146]	[0.096]
Constant	-12.431***	-28.012***	-17.347***	-7.671***
	[3.673]	[3.962]	[3.494]	[2.369]
Observations	342	342	342	342
R-squared				0.734

Notes: a value of 1 is added to firm stock of each industry. A tobit estimator is used with lower bound of zero. Robust standard errors in brackets.*** p<0.01, ** p<0.05, * p<0.1

Table 4: Panel analysis of firm dynamics and crime, Business Returns database, 2003-2011

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log stock active firms		Change in log stock active firms		Firm entry rate		Firm exit rate	
ln(population)	-0.241*	-0.443***	0.084	0.129**	0.012	0.065	0.001	-0.004
	[0.126]	[0.124]	[0.066]	[0.065]	[0.075]	[0.073]	[0.020]	[0.021]
ln(gross value added)	0.865***	0.882***	-0.263***	-0.264***	-0.204***	-0.209***	0.012	0.013
	[0.095]	[0.094]	[0.048]	[0.048]	[0.057]	[0.057]	[0.014]	[0.014]
Principal component crime	0.018		-0.014		-0.008		0.008*	
	[0.034]		[0.019]		[0.025]		[0.005]	
ln(Business crime)		0.219***		-0.058**		-0.059		0.010**
		[0.032]		[0.024]		[0.039]		[0.005]
Constant	1.183	2.113*	1.259*	1.070	1.594**	1.355**	-0.102	-0.099
	[1.245]	[1.213]	[0.682]	[0.662]	[0.693]	[0.642]	[0.195]	[0.197]
Observations	2,711	2,711	2,369	2,369	2,704	2,704	2,704	2,704
R-squared	0.988	0.989	0.614	0.617	0.459	0.461	0.462	0.463
Year & municipal unit fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

Notes: the dependent variables are constructed from the Business Returns database as follows: (i) Changes in the stock of measured as log change in stock 2003-2011, (ii) Net entry rates calculated as entry rate - exit rate, (iii) Entry rates calculated as mean entry/stock over period 2004-2011, (iv) Exit rates, calculated as mean exit/stock over period 2004-2011. The principal component for crime is constructed using levels of contact crime, property crime, police detection related crime and other serious crime (all in logs). Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Panel analysis of firm births and crime, Business Register database, 2003-2011

	(1)	(2)	(3)
	Dependent variable: ln(Firms births)		
	Total crimes	Principal component crime	Business crime
Ln(gross value added)	1.048***	1.058***	1.051***
	[0.136]	[0.136]	[0.136]
Ln(population)	-1.727***	-1.725***	-1.715***
	[0.175]	[0.175]	[0.178]
ln (Total serious crime)	-0.152**		
	[0.074]		
Principal component crime		-0.101**	
		[0.048]	
ln(Business crime)			-0.059
			[0.051]
Constant	17.399***	16.103***	16.341***
	[1.901]	[1.853]	[1.821]
Observations	2,714	2,714	2,714
R-squared	0.971	0.971	0.971
Year & municipal unit fixed effects	Y	Y	Y

Notes: Firm births are obtained from the Business Register database. The principal component for crime is constructed using levels of contact crime, property crime, police detection related crime and other serious crime (all in logs). Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 6A: Relationships explaining total firm turnover and firm size across municipal units

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Ln(Total turnover)			Ln(median firm turnover)	Ln(mean firm turnover)
ln (Market access)	0.600*	0.498	0.641**	0.145	0.388**
	[0.332]	[0.329]	[0.311]	[0.121]	[0.162]
% no schooling	-2.625	-3.144	-10.873*	-3.938*	-4.531*
	[4.421]	[4.387]	[5.717]	[2.035]	[2.669]
% primary schooling	-3.957	-4.587	-4.724	-1.663	-2.014
	[4.241]	[4.070]	[7.631]	[1.892]	[2.893]
% higher education	13.077**	14.938***	15.934***	0.417	3.150
	[5.476]	[5.403]	[5.817]	[1.274]	[2.294]
Principal component utilities	0.244**	0.245**	0.384***	-0.023	0.023
	[0.116]	[0.116]	[0.133]	[0.046]	[0.058]
Principal component crime		-0.059	0.003	-0.005	-0.024
		[0.112]	[0.104]	[0.027]	[0.047]
ln (area)			0.479***	0.076	0.148*
			[0.178]	[0.061]	[0.082]
ln (population, 2001)			0.938***	-0.037	0.162*
			[0.181]	[0.063]	[0.089]
Dummy low income, high density	-1.996***	-2.006***	-1.650***	-0.401**	-0.524**
	[0.467]	[0.473]	[0.412]	[0.195]	[0.236]
Constant	13.956***	15.453***	0.406	11.822***	7.570***
	[4.917]	[4.888]	[4.424]	[1.544]	[2.152]
Observations	339	338	338	338	338
R-squared	0.333	0.350	0.597	0.109	0.247

Source: Own calculations using Business Returns database obtained from CIPC. Total turnover by municipal unit is calculated as the sum of firm turnover over the period 2003 to 2011. The median turnover of each firm over the period 2003-2011 is used as the measure of firm turnover. Crime levels are normalized by 2001 population levels. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 6B: Relationships explaining total firm turnover and firm size across municipal units, by industrial sector

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: ln(Total firm turnover)				
VARIABLES	All sectors	Agriculture	Mining	Manufacturing	Services
ln (Market access)	0.641**	0.290	1.235***	0.890**	0.628**
	[0.311]	[0.362]	[0.434]	[0.345]	[0.302]
% no schooling	-10.873*	6.326	-7.451	-13.970**	-11.208**
	[5.717]	[6.017]	[8.327]	[6.871]	[5.515]
% primary schooling	-4.724	12.585*	-10.743	-0.521	-5.981
	[7.631]	[6.422]	[6.809]	[8.027]	[7.355]
% higher education	15.934***	17.622***	4.199	21.385***	15.316***
	[5.817]	[4.844]	[6.371]	[4.790]	[5.638]
Principal component utilities	0.384***	0.433***	0.650**	0.205	0.372***
	[0.133]	[0.137]	[0.312]	[0.204]	[0.130]
Principal component crime	0.003	-0.185**	0.145	0.048	0.010
	[0.104]	[0.093]	[0.362]	[0.138]	[0.104]
ln (area)	0.479***	0.269	0.657***	0.439*	0.457***
	[0.178]	[0.171]	[0.216]	[0.235]	[0.169]
ln (population, 2001)	0.938***	1.014***	1.050***	0.596**	0.942***
	[0.181]	[0.194]	[0.228]	[0.240]	[0.175]
Dummy low income, high density	-1.650***	-0.627	-0.524	-0.427	-1.622***
	[0.412]	[0.656]	[0.865]	[0.738]	[0.384]
Constant	0.406	-1.887	-14.329**	-2.390	0.628
	[4.424]	[5.140]	[6.710]	[4.341]	[4.318]
Observations	338	249	148	254	337
R-squared	0.597	0.341	0.486	0.466	0.601

Source: Own calculations using Business Returns database obtained from CIPC. Total turnover by municipal unit is calculated as the sum of firm turnover over the period 2003 to 2011. The median turnover of each firm over the period 2003-2011 is used as the measure of firm turnover. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Appendix A: Data

Postal code areas

Various difficulties were encountered in mapping the postal code areas in the South African Postal Code System to the Census based spatial units.

- Postal codes often overlap different provincial and municipal boundaries as defined by the Municipal Demarcation Board.
- Postal codes can cover multiple areas (or main places) that are not necessarily contiguous. The main places in the Census spatial units may also cover geographical areas that are not contiguous.
- There is no clear hierarchy for postal codes. Different postal codes can cover combinations of main places that partly overlap with each other. For example, postal code 3610 covers the main places of Pinetown, Hillcrest, Kloof and New Germany in the eThekwini Metropolitan municipality. Postal code 3624 overlaps and covers Kloof, Bothas hill and Assegay. In these cases, it was necessary to aggregate these main places into a single spatial unit.
- Some informal settlements do not have postal codes. Not all farm, tribal and small holding areas get mail delivered to their door and therefore no street Postal codes exist in these areas, unlike in urban areas.
- Each Metropolitan municipality in the 497 Census based classification contains a residual category (e.g. P1D01M01C07: City of Cape Town [Part of P1D01M01]). These are not contiguous areas and it is not possible to create a consistent map between these areas and the postal codes.
- Each District area contains a District management Area, which is often not possible to map to a postal code (can be made up of non-contiguous areas)

In total 342 spatial units were constructed and mapped to postal code areas. It was only possible to map 454 of the 497 Census based spatial units to postal areas. The remaining spatial units are predominantly (a) district municipalities which are often made up of rural non-contiguous areas, (b) Informal settlements, (c) rural municipal units, (d) the residual municipal units in the metropolitan areas (e.g. eThekwini Main Place). The unallocated municipal units are not included in the estimates based on the Business Register and Returns data.

Returns database

The returns database contains data on turnover by company for the period 2003 to 2011. The database however is incomplete. Observations for many companies are only available over intermittent years (e.g. 2007, 2009, 2011, but not in 2008 and 2010). The stock of firms in the database also rises substantially in 2011 in response to the increased enforcement by CIPRO.

An additional problem is that many firms do not report turnover data for every year. Of the 731 922 companies in the database, only 55 percent have provided at least one turnover value. There are many instances where the turnover values are entered incorrectly. This appears to arise from the incorrect conversion of the the turnover data provided by the company to a common unit of measurement, e.g. values in million Rands are not multiplied by 1000 000 000. The largest recorded turnover in the database, for example, is a clearly incorrect 1.09×10^{16} Rands. The following rule was followed in cleaning the data: (a) drop observations with single data value if value > R1bill, (b) if 2

data values, drop both if ratio max/min > 100, (c) If 3 or more data values, drop max and min if ratio of maximum value to minimum value > 100. (d) For remaining, take median turnover of entire period. This leaves turnover values for 377 974 firms. Clearly a more systematic firm by firm cleaning procedure is recommended.

To construct a measure of the stock of firms from 2003 to 2011 it was necessary to fill in the missing firm data in earlier years. We assume that firms exist in every year between their date of registration and the final year in which the firm is observed in Returns database. This gives the following table, where the modified data are in column Updated and the original data are in column Unadjusted.

Table A1: Comparison of stock of firms using Business Returns database

	Updated	Percent	Unadjusted	Percent
2003	351,548	7.33	1,958	0.11
2004	395,881	8.25	4,246	0.23
2005	451,000	9.40	123,571	6.83
2006	495,595	10.33	175,398	9.69
2007	539,258	11.24	165,151	9.12
2008	576,558	12.02	219,945	12.15
2009	574,987	11.98	223,351	12.34
2010	582,164	12.13	168,399	9.30
2011	569,258	11.86	466,292	25.76
Total	4,798,021	100.00		

Note: Own calculations using Returns database.