Local inequality and crime: New evidence from South Africa

The Journal of Economic Inequality

Online Appendix – Not intended for print publication

Author: Nicolas Büttner (single and corresponding author).

Affiliations: ETH Zürich, Development Economics Group, Switzerland (primary); University of Passau, Chair of Development Economics, Germany (secondary).

Contact: nicolas.buettner@nadel.ethz.ch.

ORCID: 0000-0002-4486-3217

Table of content

A. Additional information on SAPS crime statistics

This section provides further details on the SAPS crime statistics, which were used to generate a panel of annual crime rates at the police precinct level from 2000/2001 to 2014/2015 for 1,089 precincts.

Table S1 on the next page presents the categorization of crimes as provided in the official annual SAPS crime statistics, and Table S2 provides an overview of the availability of crime statistics for each crime per SAPS year.

Table S3 and Figure S1 illustrate the creation of new police precincts and the concomitant changes of police precinct boundaries between 2000/2001 and 2014/2015. In total, 51 new police stations were established during this period, affecting the policing areas and precinct boundaries of 45 existing precincts. Table S3 provides for each newly established precinct the existing precinct that previously policed the affected area and thus recorded all crime in this area. It also contains information on the SAPS year in which official crime statistics for the new station were first published and (whenever available) the official opening date of the new precinct. Lastly, it provides the source of information, i.e., whether the information was received from the SAPS Deputy Information Officer, via internet research (archived speeches, press releases, parliamentary monitoring group, twitter), or from the affected stations via e-mail or personal phone call. The information summarized in this table was used together with a shapefile of all police precincts as of September 2014 to create a shapefile of all police precincts as of SAPS financial year 2000/2001. Figure S1 shows the original shapefile from 2014, in which all newly created precincts are colored red, and all affected existing stations are colored green. Merging all green precincts with their corresponding red precinct(s) yielded the police precinct boundaries as of 2000/2001.

Table S1: Official crime categorization of annual SAPS crime statistics

Community reported rimes				
Contact crimes	Contact related crimes	Property related crimes	Other serious crimes	Crimes detected as result of police action
Murder Attempted murder Assault with the intent to inflict grievous bodily harm Common assault Robbery with aggravating circumstances[1] - Carjacking (TRIO) - Robbery – residential (TRIO) - Robbery – nonresidential (TRIO) - Truck hijacking - Robbery of cash-in-transit - Bank robbery Common robbery Sexual offences[2] - Rape - Sexual assault - Contact sexual assault - Attempted sexual offences	Arson Malicious damage to property	Burglary at residential premises Burglary at nonresidential premises Theft of motor vehicle and motorcycle Theft out of or from motor vehicle Stock theft	Other theft Commercial crime Shoplifting	Illegal possession of firearm and ammunition Drug related crimes Driving under the influence of alcohol and drugs Sexual offences detected as a result of police action

Source: Author.
Notes: [1] Sub-categories of 'robbery with aggravating circumstances' are non-exhaustive. The first three subcategories are together labelled as so-called TRIO-crimes.
[2] Sub-categories of 'sexual offences' are exhaustive. Until 2007, there were only two sub-categories, 'rape' and 'indecent assault'.

Table S2: Availability of SAPS crime statistics, by SAPS year and crime

Crime (category)	2000/01	01/02	02/03	03/04	04/05	05/06	06/07	07/08	08/09	09/10	10/11	11/12	12/13	13/14	14/15
Comm. rep. crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Contact crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Murder	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Attempted murder	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Assault GBH	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Common assault	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Robbery aggravated	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Carjacking	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Robbery res.				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Robbery nonres.				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
➢ TRIO				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Truck hijacking	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Robbery of CIT	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Bank robbery	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Common robbery	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sexual offences		✓		✓	✓				✓	✓	✓	✓	✓	✓	✓
- Rape									✓	✓	✓	✓	✓	✓	✓
- Sexual assault									✓	✓	✓	✓	✓	✓	✓
- Cont. sex. assault									✓	✓	✓	✓	✓	✓	✓
- Att. sex. offences									✓	✓	✓	✓	✓	✓	✓
Contact rel. crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Arson	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mal. dam. to prop.	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Property rel. crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Burglary – res.	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Burglary – nonres.	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Theft of vehicle	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Theft out of vehicle	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Stock theft	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other ser. crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other theft	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Commercial crime	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Shoplifting	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Crimes det. police	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Illegal poss. firearm	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Drug related crimes	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Driving alc./drugs	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sex. off. det. police												✓	✓	✓	✓

Source: Author.
Notes: ✓ indicates availability of crime statistics.

Table S3: Establishment of police precincts between 2000/2001 and 2014/2015

New station	Prior station of crime records	SAPS year of first statistics	Month/year of official opening	Source
Emzinoni	Bethal	2001/2002	10/2001	Phone
Zonkizizwe	Heidelberg (Gp)		09/2000	E-Mail
Boipatong	Vanderbijlpark	2003/2004	08/2003	Archived speech[1]
Rosedale	Upington		04/2004	Phone
Tshidilamolomo	Makgobistad		08/2004	Phone
Wembezi	Estcourt		08/2003	E-Mail
Da Gamaskop	Mossel Bay	2004/2005	?	Phone
Harare	Khayelitsha		-/2004	Phone
Lingelethu-West	Khayelitsha		08/2004	Phone
Kleinvlei	Kuilsriver		07/2004	E-Mail
Kopanong	Batho		08/2004	E-Mail
Kuyasa	Colesberg		04/2004	E-Mail
Mangaung	Batho		06/2004	Archived speech[2]
Mbewkweni	Paarl		?	Phone
Philippi East	Nyanga		08/2004	E-Mail
Rabie Ridge	Ivory Park		?	Phone
Kwanokuthula	Plettenberg Bay	2005/2006	05/2004	Phone
Mfuleni	Kuilsriver		?	Phone
Huhudi	Vryburg	2006/2007	12/2006	Phone
Ikamvelihle	Motherwell		12/2006	Phone
Mamelodi East	Mamelodi		-/2007	Parl. monit. group[3]
Ndumo	Ingwavuma		-/2006	Phone
Belhar	Delft	2007/2008	12/2007	E-Mail
Hekpoort	Magaliesburg		10/2006	Phone
Nemato	Port Alfred		10/2007	Phone
Ratanda	Heidelberg (Gp)		06/2007	Phone
Kagisho	Galeshewe	2008/2009	-/2009	Phone
Lwandle	Strand		?	Phone
Augrabies	Kakamas	2009/2010	04/2009	Phone
Tarlton	Krugersdorp		09/2009	E-Mail
Diepsloot	Erasmia	2010/2011	-/2010	Dep. inform. officer
Kwamashu E	Ntuzuma		-/2010	Dep. inform. officer
St Francis Bay	Humansdorp		-/2011	Dep. inform. officer
Hebron	Ga-Rankuwa	2011/2012	-/2011	Press release[4]
Masemola	Nebo		11/2011	Tweet by station[5]
Olievenhoutbosch	Wierdabrug		-/2011	Dep. inform. officer

Pienaar	Kanyamazane		01/2011	Phone
Zamdela	Sasolburg		-/2011	Dep. inform. officer
Bekkersdal	Westonaria		-/2012	Dep. inform. officer
Katlehong North	Katlehong		-/2012	Dep. inform. officer
Mashashane	Seshego		-/2012	Dep. inform. officer
Mlungisi	Queenstown	2012/2013	-/2012	Dep. inform. officer
Moffatview	Booysens		04/2012	Phone
Vaal Marina	Heidelberg (Gp)		-/2012	Dep. inform. officer
Westenburg	Polokwane		-/2012	Dep. inform. officer
Joza	Grahamstown		-/2014	Dep. inform. officer
Kwandengane	Bizana		-/2013	Dep. inform. officer
Lentegeur	Mitchells Plain	2013/2014	-/2013	Dep. inform. officer
Madeira	Mthatha		-/2013	Dep. inform. officer
Sebayeng	Polokwane		-/2013	Dep. inform. officer
Tembisa South	Tembisa	2014/2015	-/2013	Dep. inform. officer

Source: Author

Notes: [1] www.polity.org.za/article/shilowa-opening-of-police-station-in-boipatong-21082003-2003-08-21.

[2] www.gov.za/n-kganyago-hand-over-meloding-police-station.

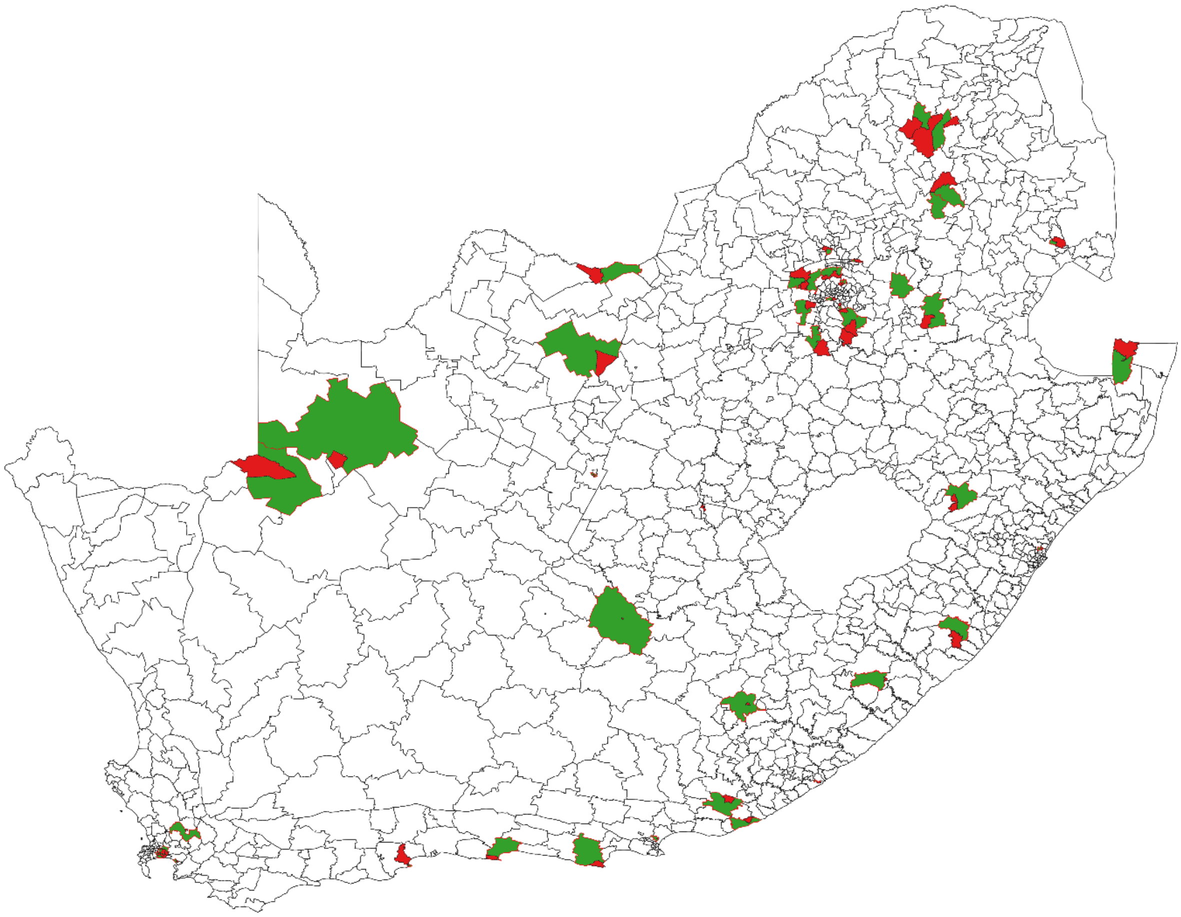
[3] www.pmg.org.za/committee-question/4701/.

[4] www.gov.za/minister-police-nathi-mthethwa-officially-open-klipgat-and-hebron-police-stations-and-launch-safer.

[5] www.twitter.com/1060facts/status/578902503602937856?lang=de.

Small discrepancies between month/year of official station opening and SAPS year of first published crime statistics are possible. In this case the latter was used for the formation of the crime statistics panel.

Figure S1: Creation of shapefile of police precinct boundaries as of 2000/2001



Source: Own calculations using data from ISS and SAPS.
Notes: Newly established precincts are colored red and outlined with a dotted black line. Affected existing precincts are colored green and outlined with a solid red line. Since new precincts lie above the affected old precincts, the area of an old precinct includes both the green and the red area.

B. Additional information on the merging of census and crime data

This section provides additional information on how the census data was allocated to police precincts using 'SAL-to-precinct-weights'. Figure S2 below illustrates the merging procedure exemplarily for the two police precincts (red dashed lines) 'Pinelands' (left) and 'Elsies River' (right) in 2001. SAL boundaries are denoted with solid green lines, and the pixels in varying red tones indicate population distribution, with darker red corresponding to a higher population. Most of the SALs are contained fully in either of the two precincts, hence, all the census households from these SALs can be allocated precisely to their respective precincts ('SAL-to-precinct-weight=1). One SAL though (orange striped, henceforth SAL X) intersects with both precincts. Based on the census data, it is thus impossible to infer whether the households located in SAL X live in 'Pinesland' or 'Elsies River'. Yet, in combination with the population distribution data from WorldPop, it is possible to estimate the probability with which the households in SAL X live in each of the two precincts. For this, I sum up the population data of all pixels in SAL X as well as in the intersecting area of SAL X with 'Pinelands' and in the intersecting area of SAL X with 'Elsies River'. Dividing the estimated population in each intersecting area by the estimated total population of SAL X yields the probability of SAL X inhabitants living in 'Pinelands' and 'Elsies River', respectively. These probabilities serve as the 'SAL-to-precinct-weights'. All households from SAL X are then allocated to both precincts, yet all further calculations of precinct-level variables are executed using these weights.

Figure S2: Creation of 'SAL-to-precinct-weights' for merging of census and crime data

Source: Own calculations using data from SAPS, StatsSA, and WorldPop.
Notes: Red dashed lines indicate police precinct boundaries, green solid lines SAL boundaries, pixels in varying red tones population distribution (darker red: higher population), orange stripes intersection of SAL X with both police precincts. Population data was altered in this example for reasons of clarity and comprehensibility.

C. Small Area Estimation (SAE) of household income

This section provides a detailed description of how the method of Small Area Estimation (SAE) was implemented in order to simulate incomes for the census households. In the following, I summarize how I proceeded in each of the three main steps of SAE, in particular, how the underlying data was created, how variables were selected and models specified, and how simulations were run. For a detailed description of the econometric approach of the method, please refer to Elbers et al. (2003). Alternatively, Demombynes and Özler (2005) provide a brief summary of the econometrics. The user manual of the software PovMap 2.0 (Zhao and Lanjouw, n.d.) contains comprehensive guidelines for each step of the practical implementation. Whenever I describe recommendations, rules, guidelines, or the like, these are taken from Zhao and Lanjouw and are not my original contributions. This section is not to be understood as a comprehensive guideline on how to implement SAE in general, but rather as an effort to be transparent about the implementation in this study.

General set-up

All stages of SAE must be conducted separately for each year, for each province, and for rural and urban households (i.e., by year and stratum). The combination of province and location status (rural/urban) is referred to as 'domain' by Zhao and Lanjouw (n.d.). Importantly, the size of the sub-samples from the surveys should not fall below 300 households, such that in some cases adjacent provinces had to be merged to form one domain. Accordingly, I created the following 15/17 sub-samples in 2001/2011, as shown in Table S4 on the next page.

1) Selection of candidate variables

In principle, all variables that can plausibly serve as explanatory variables for household income, are available in both the census and the survey, have the same definition, and a very similar distribution in both data sources (at the domain level), can serve as candidate variables for the income model. My pool of candidate variables covered four broad classes: demographics, education, occupation (including variables for the household head and for all household members), and housing. Table S5 summarizes all potential candidate variables for 2001 and 2011. Whenever variables had slightly different definitions in the census versus the survey, I harmonized them across datasets (e.g., finer division of highest education level). More generally, it is crucial that all included candidate variables are coded in the exact same manner

Table S4: Overview of survey and census sub-samples based on domains, by year

2001					2011				
Domain	Included provinces	Urban / rural	n(survey)	n(census)	Domain	Included provinces	Urban / rural	n(survey)	n(census)
1	Western Cape	urban	1,920	892,080	1	Western Cape	urban	2,779	1,180,490
2	Eastern Cape	urban	1,572	552,996	2	Eastern Cape	urban	1,657	718,211
3	Northern Cape	urban	856	142,443	3	Northern Cape	urban	1,089	177,443
4	Free State	urban	1,347	471,805	4	Free State	urban	1,831	582,887
5	KwaZulu-Natal	urban	2,131	832,698	5	KwaZulu-Natal	urban	1,828	1,053,949
6	North West	urban	1,271	326,880	6	North West	urban	1,142	481,697
7	Gauteng	urban	3,546	2,002,697	7	Gauteng	urban	3,824	2,820,224
8	Mpumalanga	urban	1,000	299,583	8	Mpumalanga	urban	1,277	408,709
9	Limpopo	urban	848	124,448	9	Limpopo	urban	779	226,712
10	Western Cape	rural	627	83,133	10	Western Cape & Northern Cape & Free State	rural	595	193,136
11	Eastern Cape	rural	1,883	717,614	11	Eastern Cape	rural	1,654	724,939
12	Northern Cape	rural	451	28,293	12	KwaZulu-Natal	rural	1,791	886,349
13	Free State	rural	672	109,391	13	North West & Gauteng	rural	1,403	540,486
14	KwaZulu-Natal	rural	2,256	718,710	14	Mpumalanga	rural	1,021	419,907
15	North West & Gauteng	rural	1,595	455,558	15	Limpopo	rural	2,502	1,051,139
16	Mpumalanga	rural	1,188	308,178					
17	Limpopo	rural	2,218	852,680					

Source: Author.
Notes: Domain 15 in 2001 and domains 10 and 13 in 2011 were created by combining all rural households from adjacent provinces, since the number of rural households sampled in the surveys in these provinces were less than 300.

Table S5: Overview of potential candidate variables for income model, by year

Demographics		Education		Occupation		Dwelling
Head	Members	Head	Members	Head	Members	
			2001			
Sex Age Population group Marital status	Household size Average age Share/number of children Share/number of children Share/number of adults Share/number of elderlies Share/number of males Share/number of male adults	Highest education level Total years of schooling	Years of schooling per capita Total years of schooling Share/number of members with matric	Employment status	Share/number of members with work Share/number of males with work Share/number of members with matric and work	Dwelling type Toilet type Refuse removal type Water access type
			2011			
Sex Age Population group Partner present?	Household size Average age Share/number of children Share/number of children Share/number of adults Share/number of elderlies Share/number of males Share/number of male adults	Highest education level Total years of schooling	Years of schooling per capita Total years of schooling Share/number of members with matric	Employment status	Share/number of members with work Share/number of males with work Share/number of members with matric and work	Dwelling type Toilet type Water access type

Source: Author.
Notes: In case of a continuous variable, the square term was added to the Beta model if this increased the fit of the model. All categorical variables, such as 'highest education level' or 'dwelling type' were created in multiple versions (i.e., with varying grouping of categories) to provide flexibility for the specification of the Beta model. Lastly, the interaction of two candidate variables was also allowed to enter the Beta model, if this increased the fit of the Beta model and if both linear terms and the interaction term were statistically significant.

in both datasets. Households in the survey must be representative of the census households at the domain level for each included candidate variable (household weights are used). For continuous variables, this means that the mean and percentiles should be very similar across datasets. Categorical variables should exhibit a very similar distribution across categories in both datasets. For more flexibility, I applied several alternative groupings to all categorical variables (i.e., if the original variable 'dwelling type' contained eight categories, I included additional variables with six, four, and two categories by grouping adjacent categories). All households with missing data on one of the candidate variables were excluded in both the census and survey datasets. The final products of this stage are household level datasets for each domain in 2001 and 2011, containing all candidate variables that passed the above explained criteria, for the census and for the survey. Importantly, the range of included candidate variables varies across domains, as representativeness is not guaranteed for all variables in all domains.

2) Income model

This stage serves to specify the econometric model of household income, i.e., the model that best explains the variation in per capita household income in the survey, for each domain. It can be divided into the following sub-stages: A.1) Selection of household level explanatory variables (Beta model), A.2) Selection of cluster level variables (Beta model), A.3) Correction for outliers (Beta model), B) Modelling of heteroskedasticity (Alpha model).

A.1) Beta model – selection of household level explanatory variables

In each survey sub-sample, I started with simple (weighted) regressions of log per capita household income on each included candidate variable. The candidate variable with the highest R^2 and $p<0.15$ (for categorical variables all categories must pass $p<0.15$) is selected as the first explanatory variable. From here on, further explanatory variables are added iteratively along the lines of the following criteria: 1) Additional regressors (substantially) increase R^2, 2) additional regressors all have $p<0.15$, 3) if feasible, inclusion of regressors from all four classes.[1] The inclusion of an additional regressor could lead to insignificance of a previously included regressor. In this case, the regressor that contributes more to the R^2 stays in the model and the other one is dropped. Overall, the goal is to maximize the R^2, and it should not be smaller than 0.35. In order to avoid overfitting, the number of included regressors is bounded by the square root of the number of surveyed households. Table S6 on the next page presents, for each subsample, the R^2 after this stage.

[1] Explanatory variables could also enter the model together with their squared term (only for continuous variables) or in the form of interaction terms with other variables (possible for continuous and categorical variables).

A.2) Beta model – selection of cluster level variables

For the simulation of the error term, the error is split into a household component and a cluster component. This stage serves to minimize the cluster error by including cluster level variables.

For this, identifying common geographic clusters in the census and the survey is the first step. Often, household surveys and censuses of the same or nearby years have common enumeration areas (EAs) which can readily serve as clusters. In the case of South Africa, this is not possible, as the primary sampling units (PSUs) of the surveys are formed based on the EA of the preceding census (i.e., the PSUs of the IES 2011 are based on the EAs of the Census 2001) and the census does not contain information on the EA, only on the SAL. Fortunately, I received detailed information from StatsSA on how exactly the survey PSUs were formed from the EAs of preceding censuses. Combining this information with shapefiles of census EAs, I created shapefiles of survey PSUs and overlaid them with the SAL shapefiles. This enabled me to identify (approximately) common geographic clusters between surveys and censuses. Generally, a cluster corresponds to one SAL; sometimes two or three SALs were merged to fit one PSU.

The second step comprises creating cluster level variables from the census data, again from all four classes introduced above. For education, this could be for example the share of adults with at least grade 12 (Matric) education in the cluster. In addition, I calculated mean nighttime light intensity as a proxy for local economic development using data from Li et al. (2021).

The next step serves to select those cluster level variables that most effectively reduce the cluster component of the error term. For this, cluster level regressions are run with the cluster component of the residual on the left-hand side and potential cluster level variables on the right-hand side. Again, selection criteria are maximizing R^2, including only significant regressors and including regressors from several classes. Hereby, the number of cluster level variables should not exceed the square root of the number of clusters in a domain.

After selection of cluster level variables, these were added to the income model. In case one or more previously significant regressors have now turned insignificant, the model needs to be adjusted accordingly. Table S6 presents, for each subsample, the R^2 after adding cluster level variables to the Beta model.

Table S6: Overview of R^2 in Beta model, in different stages, by year

2001				2011			
Domain	R^2 w/o cluster var.	R^2 w/ cluster var.	R^2 after outlier excl.	Domain	R^2 w/o cluster var.	R^2 w/ cluster var.	R^2 after outlier excl.
1	0.3060	0.4132	0.7023	1	0.5111	0.5336	0.6568
2	0.4644	0.4843	0.6667	2	0.4652	0.4934	0.6200
3	0.5138	0.5399	0.6162	3	0.5752	0.5991	0.6403
4	0.4330	0.4410	0.5786	4	0.5381	0.5653	0.6206
5	0.5198	0.5089	0.6400	5	0.4577	0.4647	0.5630
6	0.3881	0.3903	0.5576	6	0.4952	0.5122	0.6083
7	0.4238	0.4430	0.6447	7	0.4203	0.4383	0.5464
8	0.4496	0 5156	0.6645	8	0.4773	0.5020	0.6044
9	0.4528	0.5149	0.5687	9	0.4476	0.5101	0.5429
10	0.3644	0.3732	0.6774	10	0.3144	0.3809	0.5458
11	0.3446	0.3539	0.3934	11	0.3547	0.3932	0.3932
12	0.5424	0.5952	0.7506	12	0.5505	0.5609	0.5725
13	0.4525	0.4655	0.5193	13	0.5641	0.5802	0.5802
14	0.3010	0.3134	0.4102	14	0.3878	0.4394	0.4669
15	0.3896	0.4007	0.4960	15	0.3740	0.3879	0.4096
16	0.4188	0.4334	0.5853				
17	0.4528	0.4318	0.4843				

Source: Author.

A.3) Beta model – correction for outliers

After completion of the Beta model, I identified potential outliers, i.e., households with suspiciously large residuals. Large residuals might be an indication of income misreporting (mostly underreporting) in the household surveys. In most cases, outliers had very low (sometimes zero) incomes. All identified households underwent a plausibility check, i.e., it was assessed whether their low (or zero) income was plausible given their location, demographic composition, educational and occupational structure, and housing situation. If the low incomes seemed implausible and thus misreporting of income was concluded, the household was excluded. After exclusion, the Beta model (including cluster level variables) was run again and, if necessary, adjusted. Table S6 presents, for each subsample, the R^2 after excluding outliers.

B) Alpha model – modelling of heteroskedasticity

Up to this point, for the sake of more flexibility, all steps were conducted in Stata. Then, the data was loaded into the PovMap software, and the Beta model specification was set as described above. The next step, the specification of the Alpha model, serves to model the heteroskedasticity. In this model, the dependent variable is the household component of the residual. Due to the enormous amount of potential regressors for the Alpha model (income, income squared, all regressors of the Beta model, their interactions with income, and their interactions with income squared), I applied the automated selection model of the PovMap software, specifically "Stepwise Selection" and "Forward Selection" of regressors. In the end, all included regressors should be statistically significant and maximize together the R^2 of the household component of the residual.

3) Simulation

The last step of SAE is the simulation of household incomes based on the previously specified Beta and Alpha model. In addition, some choices regarding the simulation set-up have to be made, e.g., regarding the distributional form of the cluster and household error, trimming of simulated household incomes, and the simulation method. I specified normal distributions for both the cluster and the household error. Simulated incomes were trimmed at twice the maximum of the incomes of households in the survey (per domain) in order to avoid extreme outliers. This means that if a simulated income exceeds this trimming threshold, it is set to missing. Across both years and all domains, the share of simulated incomes affected by trimming ranged from 0.002% to 0.51%, with an average of 0.081%. The simulation method was 'Simultaneous drawing', and the number of replications was set to 100.

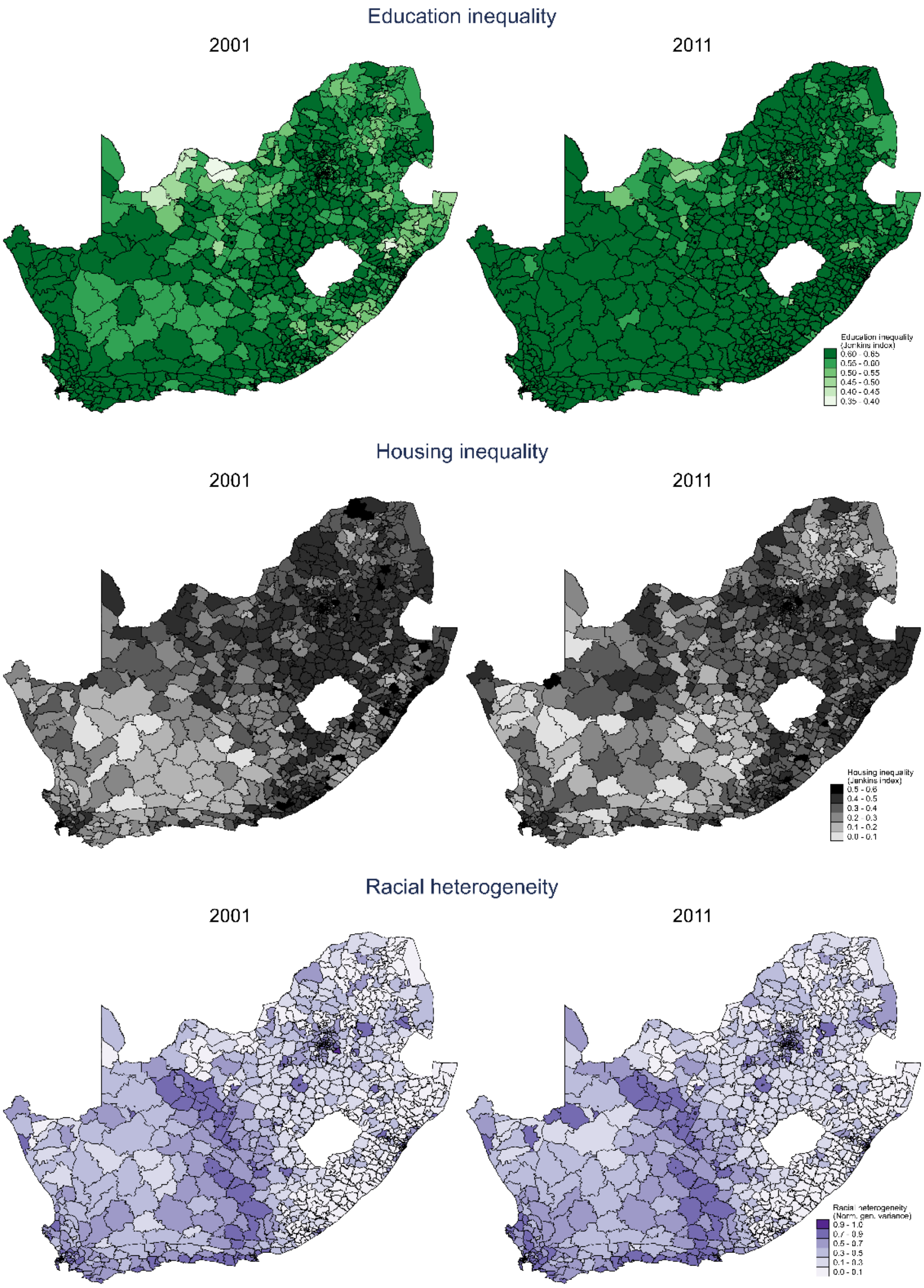
Processing of simulated income variables

The final products of SAE are 100 simulated income distributions for each domain in 2001 and 2011. These datafiles were then appended in order to cover the entire country. Next, variables of interest, i.e., mean per capita household income, the Gini index, mean log deviation and the Theil index were calculated in each police precinct for each of the 100 simulated income distributions. Lastly, for each variable of interest, the mean value was calculated from all 100 'simulated' variables, e.g., the mean of the Gini index for all 100 simulated income distributions. This mean value was then used in the econometric analysis of the study. As a plausibility check, the mean, median, 25th and 75th percentile and Gini index of the simulated household incomes were compared with the weighted equivalents in the survey for each domain. Generally, these statistics were very similar for the simulated census incomes in comparison to reported survey incomes.

D. Additional inequality and crime maps

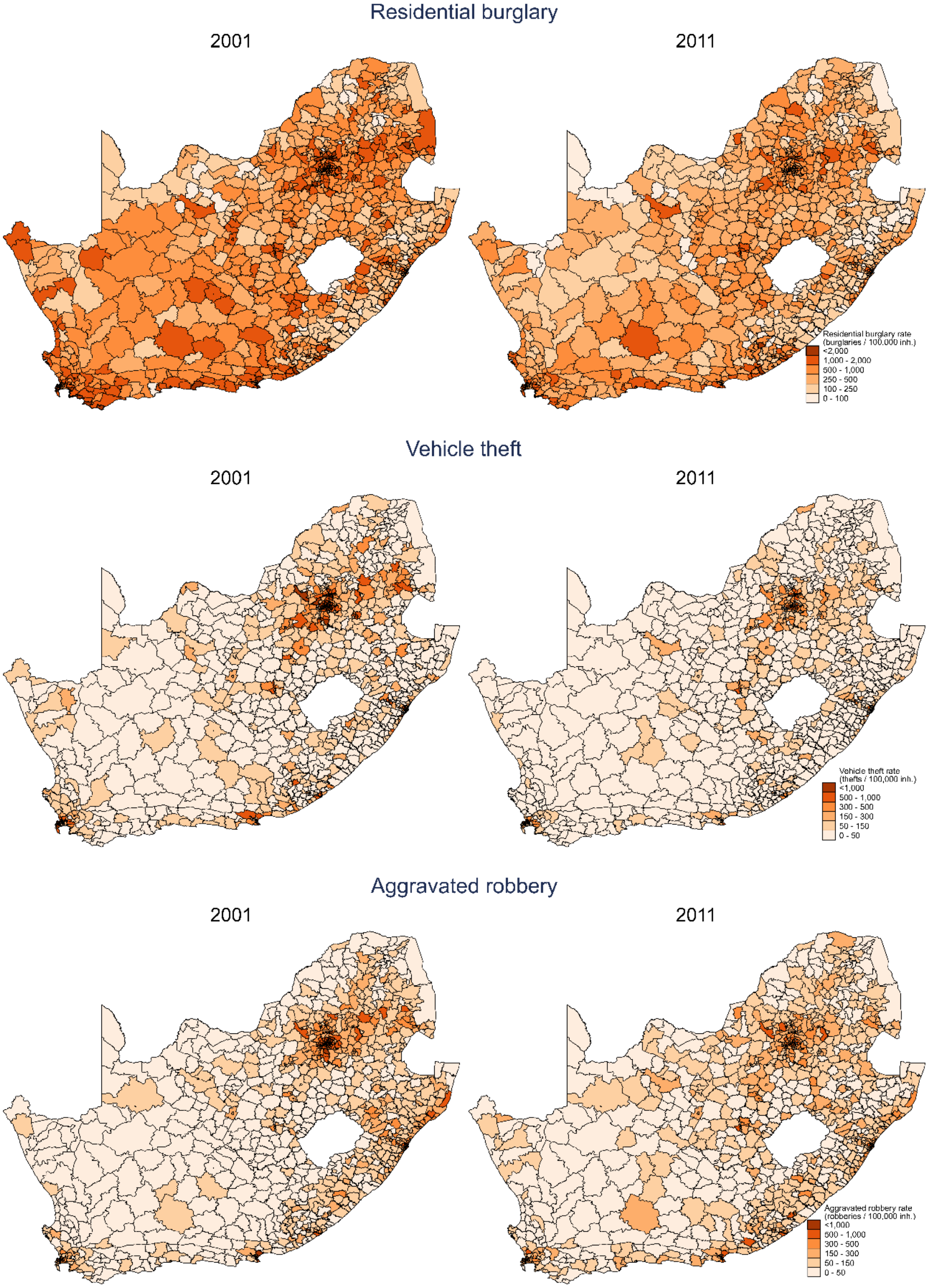
Figure S3 below maps education and housing inequality and racial heterogeneity across South African police precincts in 2001 and 2011. Figures S4 and S5 on the following pages map rates of property crimes and violent crimes, respectively, across South Africa.

Figure S3: Education inequality, housing inequality, and racial heterogeneity across South Africa



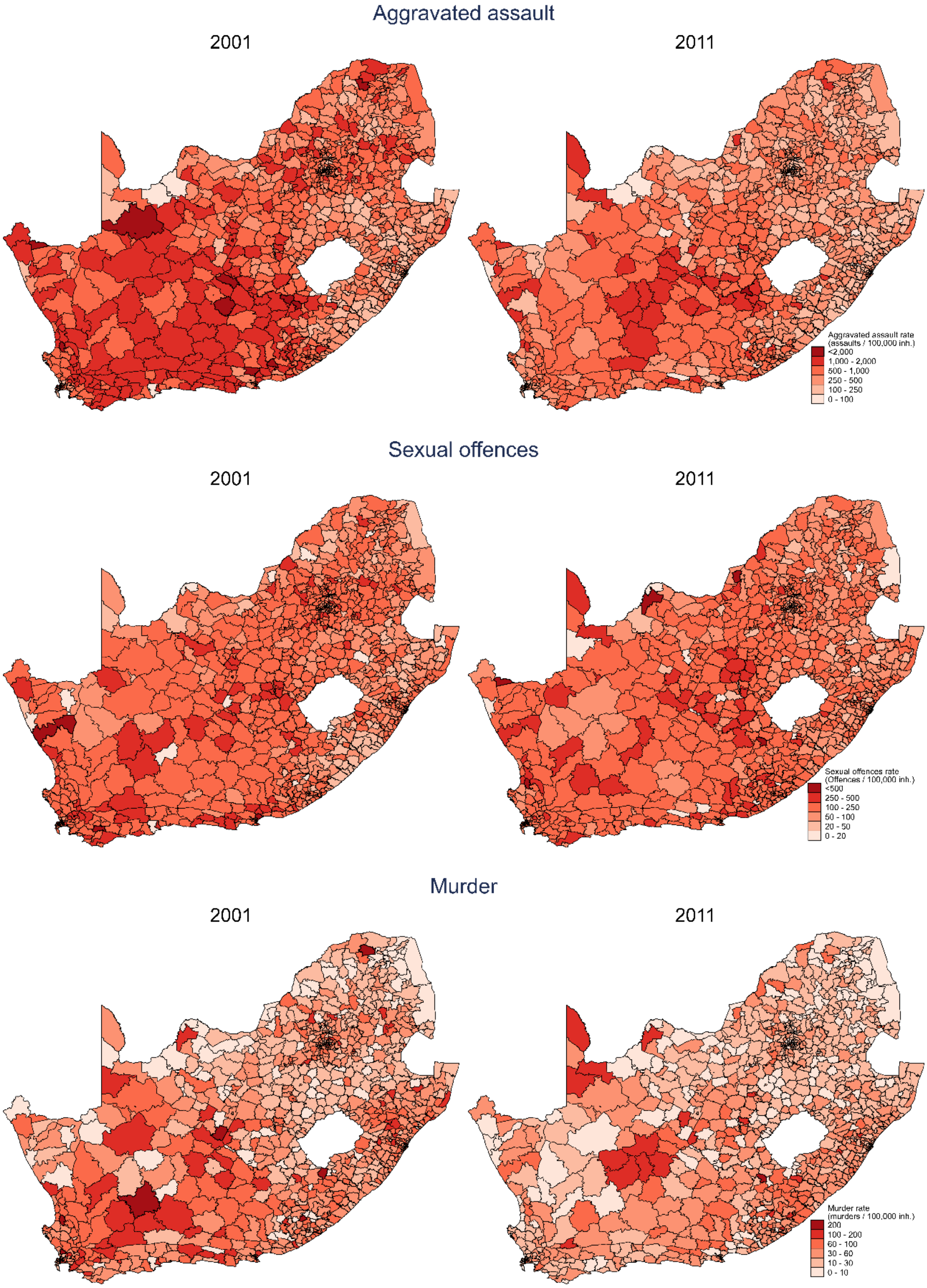
Source: Own calculations using data from SAPS, StatsSA, and WorldPop.
Notes: The darker green/grey/purple, the more education/housing/racial inequality/heterogeneity.

Figure S4: Residential burglary, vehicle theft, and aggravated robbery across South Africa



Source: Own calculations using data from ISS and SAPS.
Notes: The darker orange, the higher the rate of residential burglary / vehicle theft / aggravated robbery.

Figure S5: Aggravated assault, sexual offences, and murder across South Africa



Source: Own calculations using data from ISS and SAPS.
Notes: The darker red, the higher the rate of aggravated assault / sexual offences / murder.

E. Within-estimator approach: regressions with police precinct-fixed effects

This section presents the results from regressions with police precinct- instead of police cluster-fixed effects, i.e., these regressions fully exploit the panel structure of the data set. Yet, due to the small within-variation of the inequality indices, this specification removes most of the variation in inequality and is thus not the preferred specification. As expected, the coefficients on inequality are generally insignificant, independent of the inequality dimension.

Table S7: Regressions of local crime rates on income inequality within police precincts

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(2)	(4)	(6)	(8)	(10)	(12)
Income inequality	0.167	-1.574	-1.260	1.151	1.600	-0.406
(Gini index)	(0.8400)	(1.2495)	(1.3017)	(0.8228)	(1.0623)	(1.6263)
Returns to crime	0.292	0.256	-0.351	-0.497**	-0.184	0.232
(ln mean p.c. hh income)	(0.2159)	(0.2328)	(0.2576)	(0.2382)	(0.2076)	(0.2716)
Costs of crime	-0.424**	0.218	-0.134	-0.412**	-0.455*	-0.072
(ln population density)	(0.1823)	(0.2519)	(0.3001)	(0.1981)	(0.2383)	(0.3171)
Returns from work	-0.638	0.805	0.852	-1.084*	-0.858	-0.271
(unemployment rate)	(0.4822)	(0.9721)	(1.0466)	(0.5712)	(0.7043)	(1.1445)
Crime prone groups	-2.339	0.927	0.272	3.511*	-0.272	1.080
(share of adolescents)	(1.9240)	(2.5356)	(2.8270)	(1.8116)	(2.5978)	(2.8260)
Social cohesion	-0.458	1.211	2.203	-0.002	2.100*	-0.074
(share of recently moved)	(0.6908)	(1.9168)	(1.8099)	(0.7917)	(1.1552)	(1.5918)
Family instability	2.172***	0.116	2.010	-0.642	-1.947**	-0.173
(Share of female-headed hh)	(0.7569)	(1.3081)	(1.4525)	(1.1325)	(0.9301)	(1.6215)
Urbanization	-0.057	0.084	0.380	0.292	0.591*	-1.084***
(share of urban hh)	(0.1741)	(0.3474)	(0.3311)	(0.2418)	(0.3183)	(0.3925)
Gang prevalence	0.836	-3.522*	1.465	1.099	2.497*	-0.489
(share of coloureds)	(0.6944)	(1.9581)	(1.8418)	(1.0246)	(1.3731)	(1.7949)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Precinct-fixed effects	yes	yes	yes	yes	yes	yes
R-Squared	0.268	0.167	0.131	0.328	0.081	0.060
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column refers to one regression. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table S8: Regressions of local crime rates on education inequality within police precincts

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(2)	(4)	(6)	(8)	(10)	(12)
Education inequality	-0.451	0.041	-0.875	0.740	3.036**	2.799
(Jenkins index)	(1.2701)	(1.8316)	(1.7910)	(1.7315)	(1.2930)	(1.9748)
Returns to crime	0.309	0.152	-0.407	-0.439**	-0.120	0.160
(ln mean p.c. hh income)	(0.1926)	(0.2367)	(0.2590)	(0.2142)	(0.1940)	(0.2550)
Costs of crime	-0.425**	0.141	-0.239	-0.345*	-0.345	-0.021
(ln population density)	(0.1805)	(0.2500)	(0.3022)	(0.1894)	(0.2266)	(0.2988)
Returns from work	-0.619	0.672	0.754	-0.949*	-0.610	-0.398
(unemployment rate)	(0.4493)	(0.9527)	(1.0348)	(0.5464)	(0.6796)	(1.0796)
Crime prone groups	-2.342	1.129	0.428	3.437*	-0.573	0.720
(share of adolescents)	(1.9407)	(2.5456)	(2.7887)	(1.8436)	(2.6162)	(2.7200)
Social cohesion	-0.448	1.519	2.521	-0.216	1.749	-0.236
(share of recently moved)	(0.7144)	(1.9468)	(1.8799)	(0.7747)	(1.1138)	(1.5485)
Family instability	2.185***	-0.096	2.011	-0.581	-1.840*	-0.317
(Share of female-headed hh)	(0.7561)	(1.3282)	(1.4645)	(1.1285)	(0.9393)	(1.5749)
Urbanization	-0.069	0.066	0.397	0.269	0.590*	-1.028***
(share of urban hh)	(0.1703)	(0.3560)	(0.3327)	(0.2394)	(0.3152)	(0.3725)
Gang prevalence	0.858	-3.411*	1.517	1.075	2.290*	-0.661
(share of coloureds)	(0.7064)	(1.9668)	(1.8437)	(1.0554)	(1.3537)	(1.7310)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Precinct-fixed effects	yes	yes	yes	yes	yes	yes
R-Squared	0.265	0.175	0.125	0.317	0.092	0.116
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column refers to one regression. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table S9: Regressions of local crime rates on housing inequality within police precincts

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(2)	(4)	(6)	(8)	(10)	(12)
Housing inequality	-0.154	-0.418	-0.928	0.173	0.147	-0.411
(Jenkins index)	(0.2480)	(0.6418)	(0.6494)	(0.2612)	(0.3938)	(0.6656)
Returns to crime	0.308	0.180	-0.399	-0.438**	-0.101	0.225
(ln mean p.c. hh income)	(0.1920)	(0.2393)	(0.2550)	(0.2234)	(0.1956)	(0.2578)
Costs of crime	-0.419**	0.128	-0.207	-0.352*	-0.370*	-0.104
(ln population density)	(0.1758)	(0.2465)	(0.2994)	(0.1805)	(0.2210)	(0.3080)
Returns from work	-0.618	0.632	0.697	-0.952*	-0.670	-0.299
(unemployment rate)	(0.4495)	(0.9513)	(1.0221)	(0.5454)	(0.6804)	(1.1077)
Crime prone groups	-2.308	1.286	0.863	3.369*	-0.476	1.341
(share of adolescents)	(1.9764)	(2.5785)	(2.8728)	(1.7900)	(2.6494)	(2.9105)
Social cohesion	-0.476	1.426	2.371	-0.151	1.921*	-0.010
(share of recently moved)	(0.6730)	(1.9688)	(1.8606)	(0.7598)	(1.1404)	(1.6376)
Family instability	2.183***	-0.061	1.806	-0.550	-1.796*	-0.202
(Share of female-headed hh)	(0.7486)	(1.3175)	(1.4559)	(1.1275)	(0.9454)	(1.6136)
Urbanization	-0.056	0.125	0.448	0.262	0.563*	-1.062***
(share of urban hh)	(0.1739)	(0.3551)	(0.3293)	(0.2468)	(0.3214)	(0.3988)
Gang prevalence	0.850	-3.435*	1.613	1.078	2.457*	-0.430
(share of coloureds)	(0.7001)	(1.9606)	(1.8438)	(1.0141)	(1.3624)	(1.7966)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Precinct-fixed effects	yes	yes	yes	yes	yes	yes
R-Squared	0.266	0.167	0.133	0.323	0.080	0.046
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column refers to one regression. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table S10: Regressions of local crime rates on racial heterogeneity within police precincts

Independent variables	Property crimes: Residential burglary	Property crimes: Vehicle theft	Violent crimes: Aggravated robbery	Violent crimes: Aggravated assault	Violent crimes: Sexual offences	Violent crimes: Murder
	(2)	(4)	(6)	(8)	(10)	(12)
Racial heterogeneity (Generalized variance, norm.)	-0.743 (0.4824)	0.619 (0.7712)	-0.790 (0.8609)	0.343 (0.4278)	-0.995 (0.7442)	-1.263 (0.8429)
Returns to crime (ln mean p.c. hh income)	0.358* (0.1874)	0.112 (0.2540)	-0.361 (0.2706)	-0.458** (0.2316)	-0.025 (0.2026)	0.304 (0.2603)
Costs of crime (ln population density)	-0.452*** (0.1747)	0.168 (0.2444)	-0.245 (0.2984)	-0.339* (0.1852)	-0.427* (0.2401)	-0.156 (0.3109)
Returns from work (unemployment rate)	-0.502 (0.4380)	0.566 (0.9694)	0.847 (1.0650)	-1.005* (0.5520)	-0.498 (0.7148)	-0.107 (1.1227)
Crime prone groups (share of adolescents)	-2.633 (1.9525)	1.322 (2.5159)	0.117 (2.7727)	3.574* (1.8324)	-0.681 (2.5652)	0.726 (2.8023)
Social cohesion (share of recently moved)	-0.220 (0.6927)	1.274 (1.9466)	2.661 (1.8805)	-0.266 (0.7285)	2.311** (1.1608)	0.421 (1.6735)
Family instability (Share of female-headed hh)	2.104*** (0.7391)	-0.012 (1.3066)	1.839 (1.4432)	-0.523 (1.1100)	-1.943** (0.9431)	-0.338 (1.6038)
Urbanization (share of urban hh)	-0.074 (0.1670)	0.080 (0.3583)	0.390 (0.3354)	0.274 (0.2407)	0.562* (0.3104)	-1.096*** (0.3920)
Gang prevalence (share of coloureds)	0.870 (0.7006)	-3.461* (1.9636)	1.520 (1.8510)	1.086 (1.0122)	2.535* (1.3734)	-0.414 (1.8564)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Precinct-fixed effects	yes	yes	yes	yes	yes	yes
R-Squared	0.268	0.174	0.129	0.322	0.086	0.052
Observations	2,170	2,170	2,170	2,170	2,170	2,170

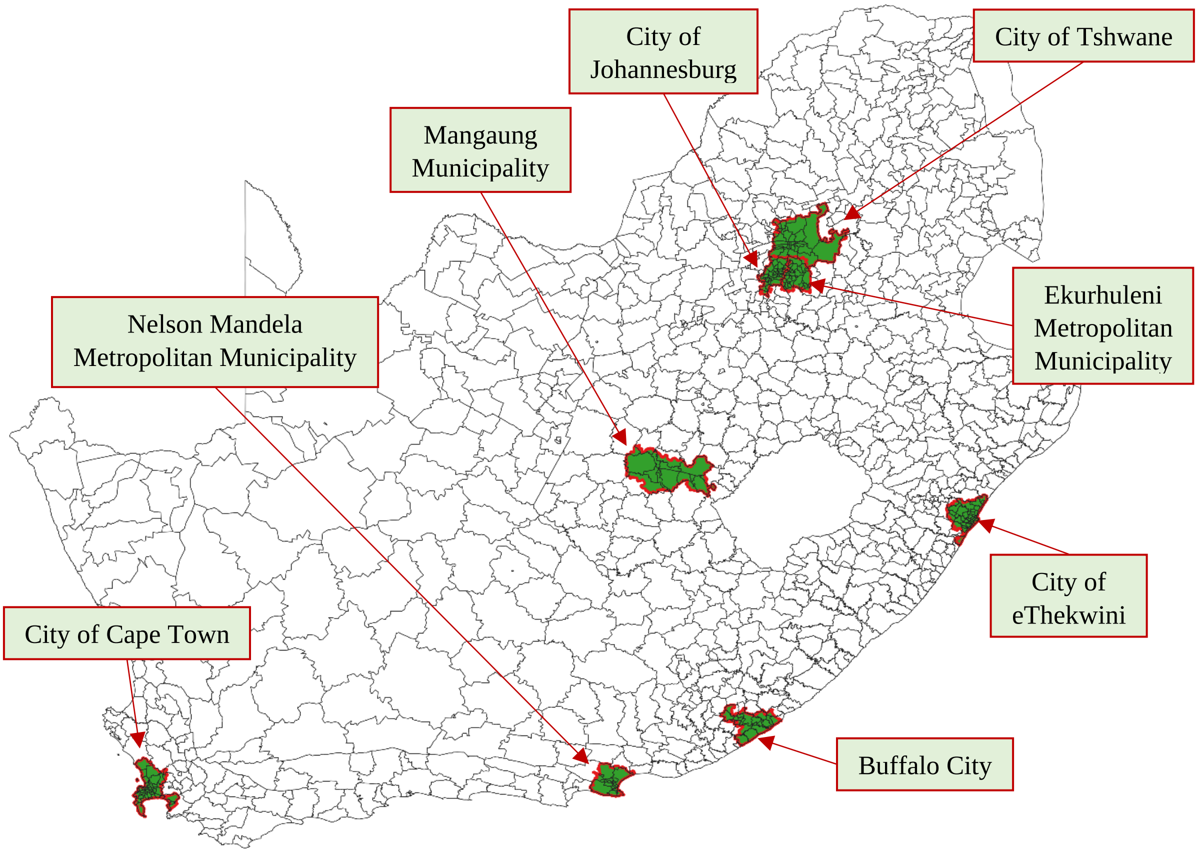
Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column refers to one regression. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

F. Heterogeneity analysis: geographic information on metropolitan municipalities, former colonies, and former Bantustans

This section provides information on the location of South Africa's metropolitan municipalities, the geographic pattern of British and Dutch colonization, and the boundaries of the former Bantustans (or Homelands) under Apartheid used in the heterogeneity analysis.

Figure S6 below maps the locations of all eight metropolitan municipalities in relation to the country's 1,089 police precincts. Three are located in the province of Gauteng (City of Johannesburg, City of Tshwane, Ekurhuleni Metropolitan Municipality), two in the Eastern Cape (Buffalo City, Nelson Mandela Metropolitan Municipality), one in the Western Cape (City of Cape Town), one in the Free State (Mangaung Municipality), and one in KwaZulu-Natal (City of eThekwini). As visible in the map, the boundaries of the metropolitan municipalities frequently intersect with the boundaries of police precincts, i.e., some police precincts lie only partly inside a metropolitan municipality. For the heterogeneity analysis, I consider a police precinct as part of a metropolitan municipality if at least half of its area lies within the boundaries of a metropolitan municipality.

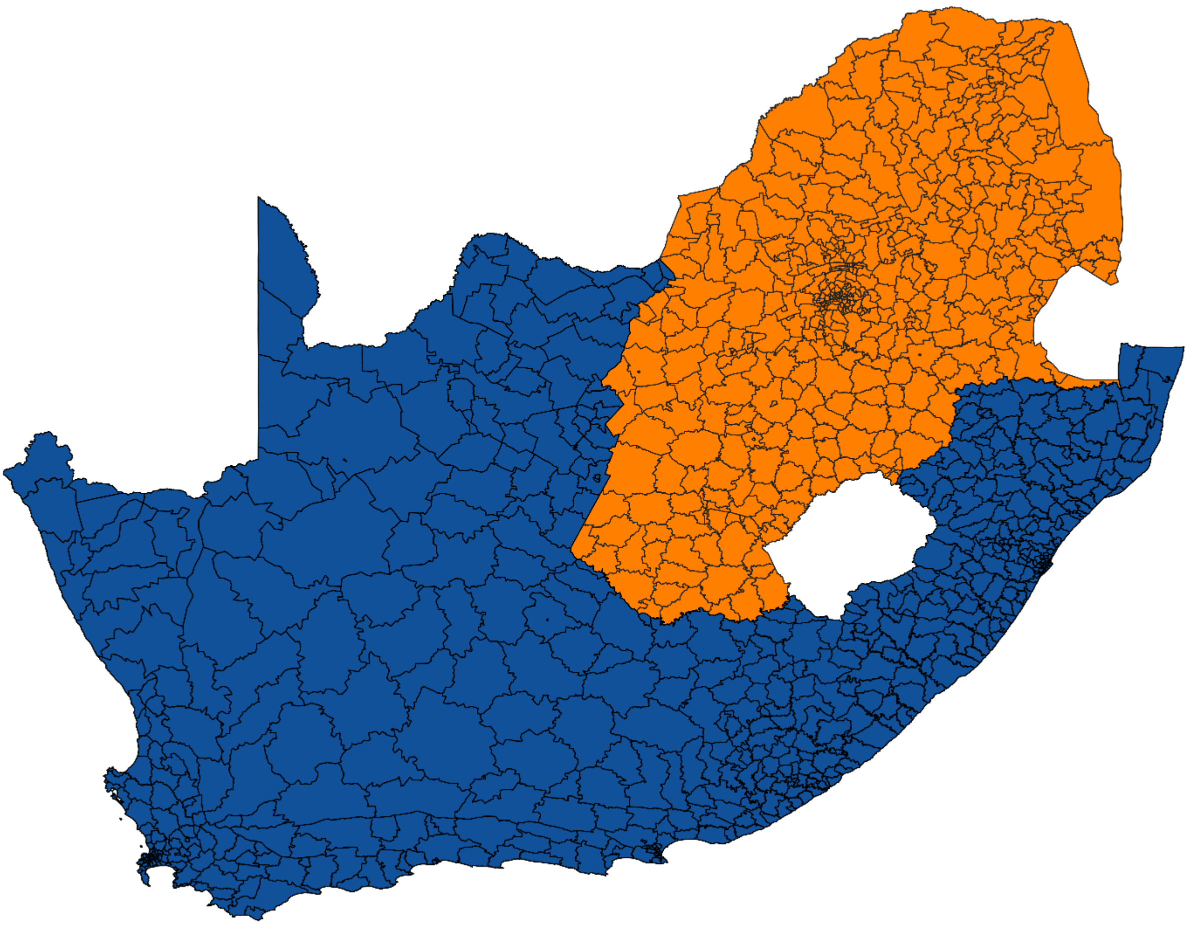
Figure S6: Locations of South Africa's eight metropolitan municipalities



Source: Own calculations using data from SAPS and StatsSA.
Notes: Metropolitan municipalities are filled green and outlined red. Black lines denote police precinct boundaries.

Figure S7 below classifies precincts according to the dominant former colonial power, i.e., whether the precincts were colonized the Dutch or the British Empire before the formation of the Union of South Africa in 1910. Blue precincts are formerly British dominated, i.e., they belonged to the Cape Colony or the Colony of Natal, and orange precincts are formerly Dutch dominated, i.e., they belonged to the Transvaal Republic or the Orange Free State.

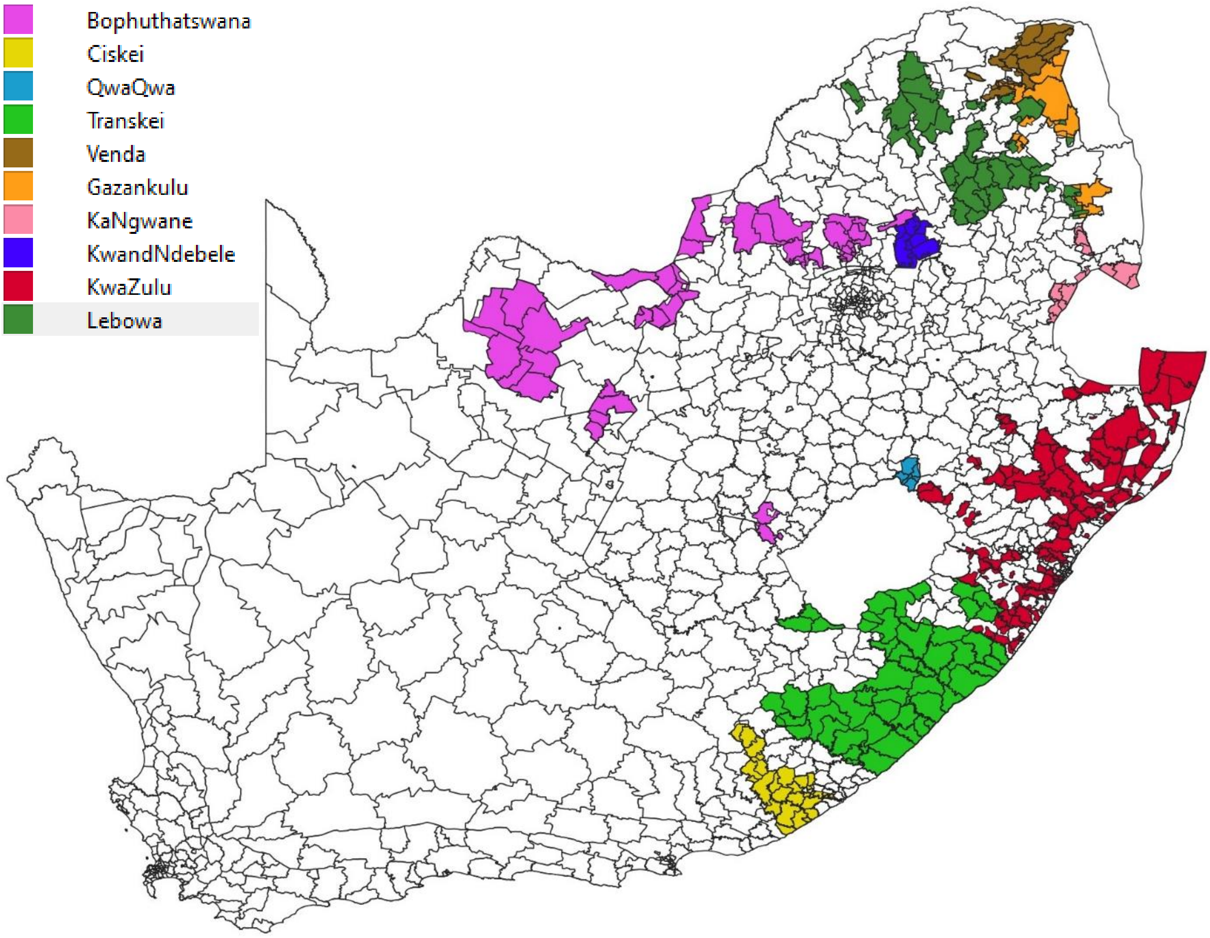
Figure S7: Division of police precincts by dominant former colonial power



Source: Own calculations using data from SAPS.
Notes: Blue precincts are classified as formerly British dominated and orange precincts are classified as formerly Dutch dominated.

Figure S8 below maps the locations and boundaries of the former Bantustans (or Homelands) under Apartheid that were created to concentrate the various black ethnic groups of South Africa and to keep the majority of the country's land for the white minority population. There were in total ten Bantustans for nine ethnic groups: Bophuthatswana (Tswana), Ciskei (Xhosa), Gazankulu (Tsonga), KaNgwane (Swazi), KwaNdebele (Ndebele), KwaZulu (Zulu), Lebowa (Northern Sotho), QwaQwa (Southern Sotho), Transkei (Xhosa), Venda (Venda).

Figure S8: Boundaries of former Bantustans under Apartheid



Source: Own calculations using data from openAFRICA (https://open.africa/en/dataset/former-tbvc-states-shape-file/resource/4ff0ad91-b247-46b2-a08e-0a86badc73b0).
Notes: Colored areas denote the ten former Bantustans for black South Africans under Apartheid. Blank areas were areas mostly reserved for the white minority population under Apartheid.

G. Robustness checks

In this section, I present the results from four types of robustness checks. On the next page, Table S11 shows bias-adjusted coefficients and degree of selection required for null effects, based on Oster (2019). Figure S9 shows the results from the randomization inference. Tables S12 and S13 show the results from regressions with the reduced and the expanded set of control variables, respectively. Lastly, Tables S14 to S16 show the results from regressions with log-transformed crimes, alternative inequality indices, and alternative grouping of education and housing categories, respectively.

A test of unobservable selection and coefficient stability

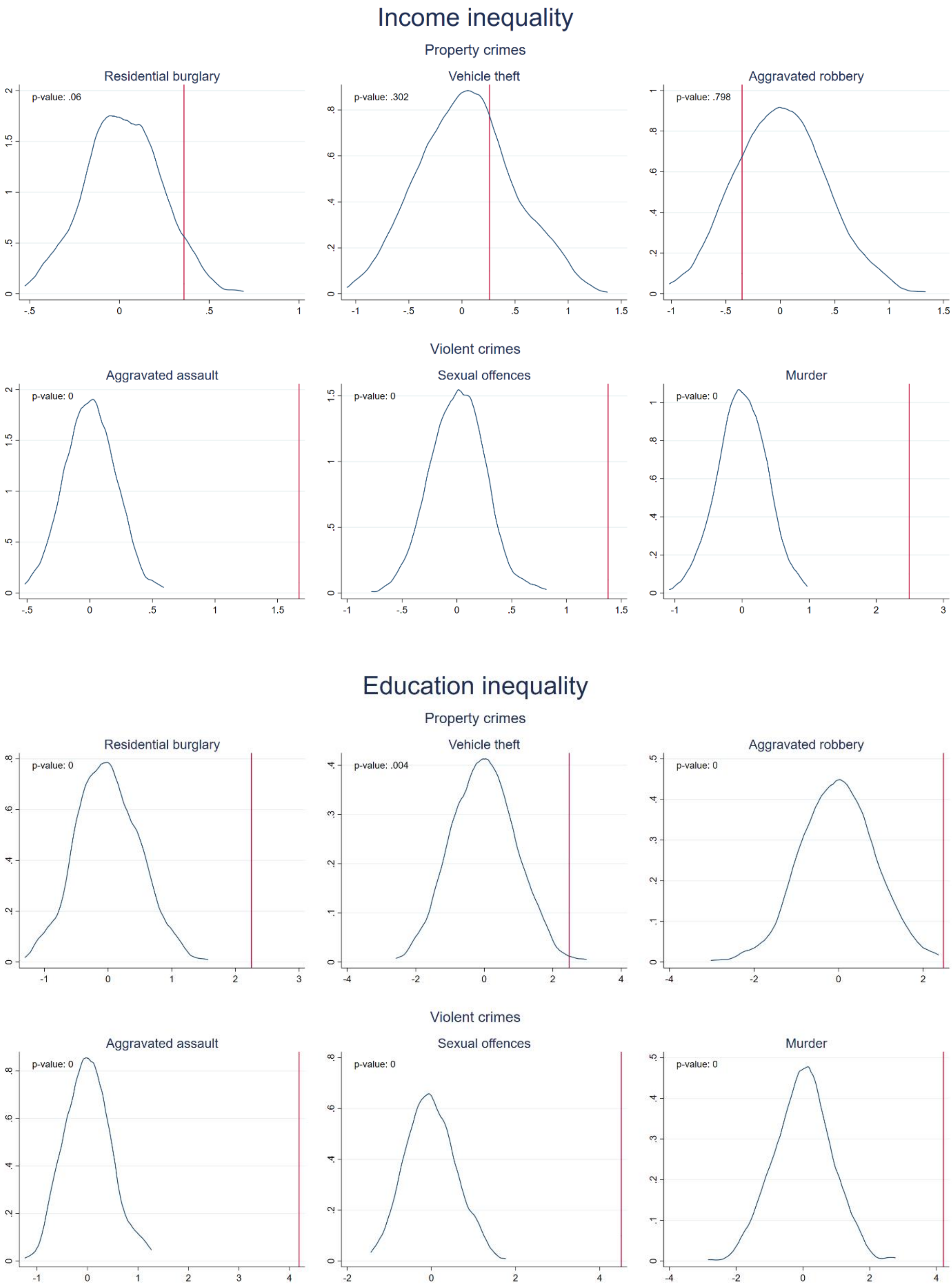
Table S11: Regressions of local crime rates on inequality within police precincts – bias-adjusted coefficients (β) and degree of selection required for null effects (δ)

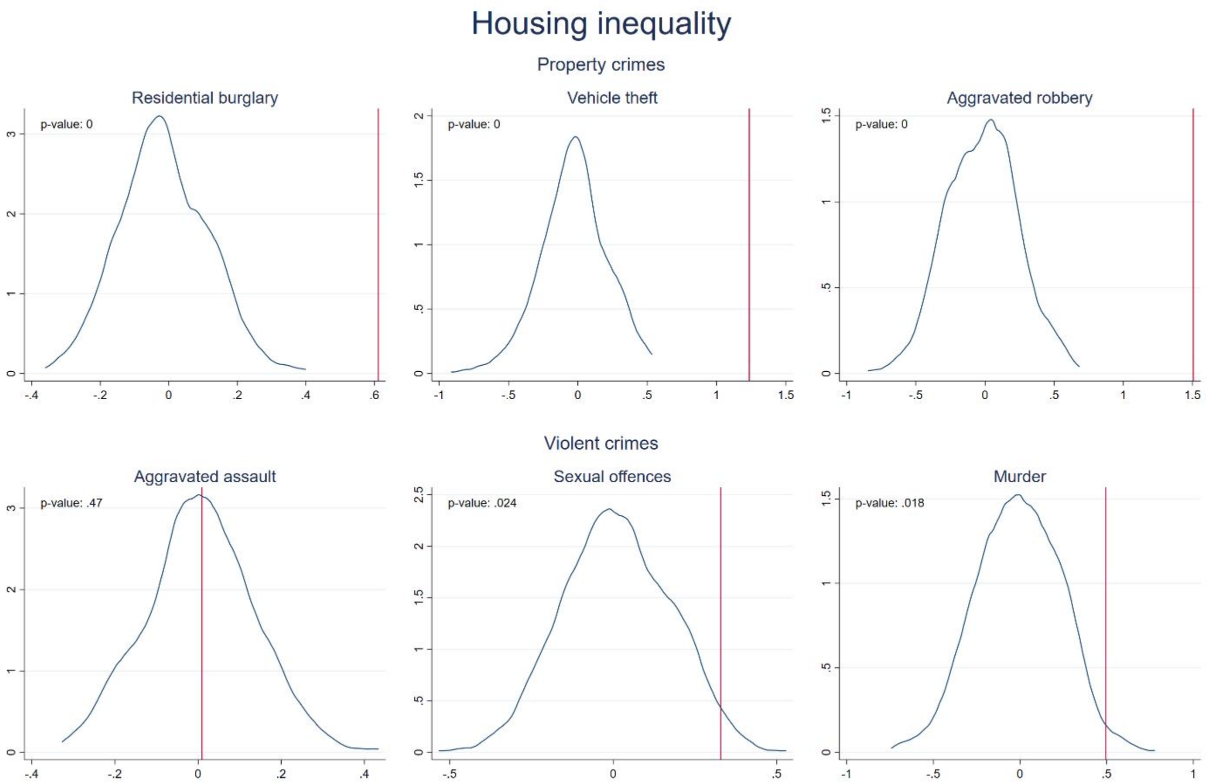
	Residential burglary		Vehicle theft		Aggravated robbery		Aggravated assault		Sexual offences		Murder	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Income inequality	-0.409	-0.315	0.948	0.681	12.497	2.972	2.663†	1.995†	3.530†	2.490†	10.920†	4.919†
(Gini index)	(0.610)	(0.660)	(-2.225)	(-3.134)	(0.105)	(0.225)	(2.716)	(4.610)	(-2.981)	(-4.408)	(-0.797)	(-1.732)
Education inequality	1.265†	1.353†	-0.745	0.305†	-5.159	-0.502	-0.540	2.187†	1.041†	2.967†	-2.748	2.134†
(Jenkins index)	(2.095)	(2.289)	(0.791)	(1.126)	(0.401)	(0.865)	(0.904)	(1.806)	(1.230)	(2.234)	(0.658)	(1.767)
Housing inequality	-0.256	-0.147	-1.306	-0.424	-0.724	0.559†	-0.643	-0.360	-0.184	0.022	-0.522	0.124
(Jenkins index)	(0.712)	(0.811)	(0.497)	(0.750)	(0.701)	(1.535)	(-0.012)	(-0.022)	(0.665)	(1.067)	(0.523)	(1.308)
Racial heterogeneity	-5.116	-1.971	-16.146	-3.535	120.276†	27.109†	9.906†	4.152†	8.747†	4.011†	11.788†	3.704†
(Gen. var., norm.)	(0.092)	(0.195)	(0.144)	(0.396)	(0.428)	(0.906)	(-2.954)	(-4.871)	(2.221)	(3.334)	(-0.343)	(-0.785)
Assumption on R_{max}	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$	$1.3\tilde{R}$	$\tilde{R} + (\tilde{R} - \dot{R})$
Observations	2,170	2,170	2,170	2,170	2,170	2,170	2,170	2,170	2,170	2,170	2,170	2,170

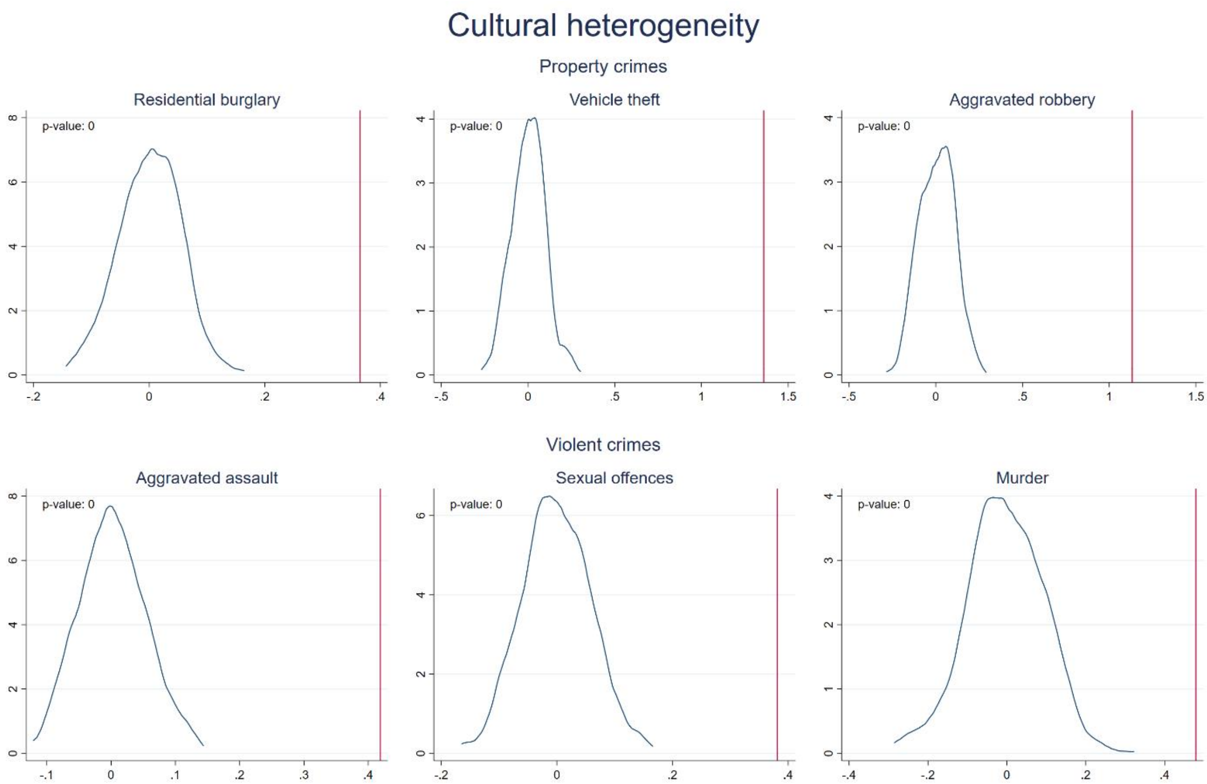
Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: For each combination of crime type and inequality dimension, i.e., for each regression, I calculated a value for the bias-adjusted coefficient (β) and for the degree of selection required for null effects (δ). β is presented in the first row and δ in the row below in parentheses. The values are calculated based on the assumption that $R_{max} = 1.3\tilde{R}$ and alternatively $R_{max} = \tilde{R} + (\tilde{R} - \dot{R})$. † indicates robustness to unobservable selection. The reduced set of controls only comprises the spatial lag, cluster-fixed effects, and province-specific time-effects. The full set of controls additionally includes ln mean p.c. household income, ln mean p.c. household income, ln population density, unemployment rate, share of adolescents, share of recently moved, share of female-headed hh, share of urban hh, share of coloureds. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate was not instrumented, as the Stata command used for the calculation of β and δ, psacalc, does not support instrumental variable regressions. However, this should not influence the results in a meaningful way, as the results from the main regressions barely differ between instrumenting or not instrumenting the spatial lag of the crime rate.

Randomization inference

Figure S9: Randomization inference







Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: The red line indicates the coefficient obtained with the true sample and the blue line the distribution of the coefficients from the permuted samples.

Robustness to reduction and expansion of the set of controls

Excluding potentially endogenous control variables

Table S12: Regressions of local crime rates on inequality within police precincts, excluding potentially endogenous controls

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(1)	(2)	(3)	(4)	(5)	(6)
Income inequality						
Gini index	0.781*	0.652	0.595	1.689***	1.605***	2.560***
	(0.4059)	(0.7447)	(0.7688)	(0.3916)	(0.4727)	(0.7117)
Education inequality						
Jenkins index	2.587***	2.609**	2.668**	4.885***	4.852***	4.659***
	(0.7271)	(1.3016)	(1.1813)	(0.8726)	(0.9627)	(1.3152)
Housing inequality						
Jenkins index	0.604***	1.220***	1.423***	-0.033	0.305	0.445
	(0.2134)	(0.3733)	(0.3815)	(0.1783)	(0.2609)	(0.3466)
Racial heterogeneity						
Generalized variance (norm.)	0.283***	1.204***	0.816***	0.476***	0.365***	0.458**
	(0.1061)	(0.2036)	(0.2075)	(0.1053)	(0.1230)	(0.1911)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Cluster-fixed effects	yes	yes	yes	yes	yes	yes
Reduced set of controls	yes	yes	yes	yes	yes	yes
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column contains four regressions; one for each inequality dimension. The reduced set of controls includes ln mean p.c. household income, ln population density, share of adolescents, share of recently moved, share of urban hh. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Adding further control variables

Table S13: Regressions of local crime rates on inequality within police precincts, additional controls

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(1)	(2)	(3)	(4)	(5)	(6)
Income inequality						
Gini index	0.319	-0.679	-1.609**	1.728***	1.210**	1.669**
	(0.4095)	(0.7764)	(0.7801)	(0.4008)	(0.4865)	(0.7655)
Education inequality						
Jenkins index	2.380***	1.645	1.756	4.222***	4.313***	3.153**
	(0.7153)	(1.3324)	(1.1868)	(0.8627)	(0.9531)	(1.3219)
Housing inequality						
Jenkins index	0.603***	0.991***	1.304***	0.047	0.265	0.364
	(0.2129)	(0.3633)	(0.3598)	(0.1817)	(0.2606)	(0.3470)
Racial heterogeneity						
Generalized variance (norm.)	0.340***	1.163***	0.905***	0.430***	0.345***	0.375*
	(0.1162)	(0.2135)	(0.2077)	(0.1059)	(0.1338)	(0.1971)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Cluster-fixed effects	yes	yes	yes	yes	yes	yes
Expanded set of controls	yes	yes	yes	yes	yes	yes
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column contains four regressions; one for each inequality dimension. The full set of controls includes ln mean p.c. household income, ln population density, unemployment rate, share of adolescents, share of recently moved, share of female-headed hh, share of urban hh, share of coloureds, dummy for coastal precincts, dummy for border precincts, ln distance to nearest metropolitan municipality, ln precinct area, dummy for split precincts. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Further robustness checks

Log-transformed crime rates

Table S14: Regressions of local crime rates on inequality within police precincts, log crime rates

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(1)	(2)	(3)	(4)	(5)	(6)
Income inequality						
Gini index	0.154 (0.3621)	0.023 (0.7116)	-0.386 (0.5953)	1.317*** (0.3091)	0.810** (0.3312)	1.748*** (0.5637)
Education inequality						
Jenkins index	1.992*** (0.5756)	2.144* (1.1944)	1.543* (0.8642)	3.687*** (0.6129)	3.257*** (0.6035)	3.382*** (0.9632)
Housing inequality						
Jenkins index	0.604*** (0.1728)	1.132*** (0.3329)	1.064*** (0.2683)	0.140 (0.1326)	0.242 (0.1598)	0.398 (0.2506)
Racial heterogeneity						
Generalized variance (norm.)	0.386*** (0.0958)	1.294*** (0.1911)	0.890*** (0.1555)	0.332*** (0.0870)	0.327*** (0.0907)	0.372*** (0.1443)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Cluster-fixed effects	yes	yes	yes	yes	yes	yes
Full set of controls	yes	yes	yes	yes	yes	yes
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column contains four regressions; one for each inequality dimension. The full set of controls includes ln mean p.c. household income, ln population density, unemployment rate, share of adolescents, share of recently moved, share of female-headed hh, share of urban hh, share of coloureds. Dependent variables are log-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Alternative inequality indices

Table S15: Regressions of local crime rates on inequality within police precincts, alternative indices

Independent variables	Property crimes			Violent crimes		
	Residential burglary	Vehicle theft	Aggravated robbery	Aggravated assault	Sexual offences	Murder
	(1)	(2)	(3)	(4)	(5)	(6)
Income inequality						
Mean log deviation	-0.032 (0.1179)	-0.089 (0.2446)	-0.392 (0.2431)	0.333** (0.1577)	0.365*** (0.1410)	0.659*** (0.2369)
Theil index	-0.068 (0.1176)	-0.139 (0.2170)	-0.304 (0.2274)	0.218* (0.1187)	0.207 (0.1337)	0.453** (0.2252)
Education inequality						
Cowell-Flachaire-index (α=0)	1.303*** (0.4187)	1.444* (0.7481)	1.889*** (0.6935)	2.036*** (0.4858)	2.302*** (0.5160)	1.872** (0.7676)
Cowell-Flachaire-index (α=0.25)	0.712*** (0.2272)	0.714* (0.4295)	0.642* (0.3807)	1.343*** (0.2675)	1.459*** (0.3058)	1.423*** (0.4285)
Cowell-Flachaire-index (α=0.5)	1.154*** (0.3594)	1.202* (0.6638)	1.245** (0.5975)	2.086*** (0.4218)	2.278*** (0.4701)	2.127*** (0.6691)
Cowell-Flachaire-index (α=0.75)	0.712*** (0.2272)	0.714* (0.4295)	0.642* (0.3807)	1.343*** (0.2675)	1.459*** (0.3058)	1.423*** (0.4285)
Cowell-Flachaire-index (α=0.9)	0.316*** (0.1034)	0.309 (0.1986)	0.250 (0.1744)	0.610*** (0.1218)	0.662*** (0.1415)	0.661*** (0.1968)
Housing inequality						
Cowell-Flachaire-index (α=0)	0.728*** (0.2202)	1.657*** (0.3780)	1.676*** (0.3802)	0.019 (0.1658)	0.407 (0.2678)	0.267 (0.3501)
Cowell-Flachaire-index (α=0.25)	0.187*** (0.0717)	0.376*** (0.1256)	0.507*** (0.1270)	0.012 (0.0586)	0.077 (0.0861)	0.187 (0.1191)
Cowell-Flachaire-index (α=0.5)	0.378*** (0.1326)	0.788*** (0.2310)	0.972*** (0.2337)	0.022 (0.1060)	0.177 (0.1604)	0.309 (0.2174)
Cowell-Flachaire-index (α=0.75)	0.187*** (0.0717)	0.376*** (0.1256)	0.507*** (0.1270)	0.012 (0.0586)	0.077 (0.0861)	0.187 (0.1191)
Cowell-Flachaire-index (α=0.9)	0.074** (0.0300)	0.146*** (0.0527)	0.208*** (0.0532)	0.005 (0.0248)	0.028 (0.0359)	0.082 (0.0502)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Cluster-fixed effects	yes	yes	yes	yes	yes	yes
Full set of controls	yes	yes	yes	yes	yes	yes
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.

Notes: Each column contains twelve regressions; two for income inequality, and five for education and housing inequality, respectively. The full set of controls includes ln mean p.c. household income, ln population density, unemployment rate, share of adolescents, share of recently moved, share of female-headed hh, share of urban hh, share of coloureds. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Alternative grouping/ordering of education and housing categories

Table S16: Regressions of local crime rates on inequality within police precincts, alternative grouping

Independent variables	Property crimes Residential burglary (1)	Vehicle theft (2)	Aggravated robbery (3)	Violent crimes Aggravated assault (4)	Sexual offences (5)	Murder (6)
Education inequality						
Jenkins index (8 cat.)	2.685*** (0.8140)	4.287** (1.7088)	4.056*** (1.5254)	5.457*** (1.1562)	5.188*** (1.2153)	5.083*** (1.6897)
Jenkins index (5 cat.)	1.778** (0.7365)	4.474** (1.8777)	0.916 (1.6529)	4.340*** (1.1564)	3.254*** (1.1694)	2.506 (1.7315)
Housing inequality						
Jenkins index (7 cat., alt. order)	0.617*** (0.2040)	1.227*** (0.3697)	1.472*** (0.3772)	0.007 (0.1731)	0.334 (0.2570)	0.500 (0.3497)
Jenkins index (5 cat.)	0.680*** (0.2049)	1.263*** (0.3854)	1.471*** (0.3958)	-0.025 (0.1889)	0.423 (0.2632)	0.585 (0.3627)
Province-specific time-effects	yes	yes	yes	yes	yes	yes
Spatial lag	yes	yes	yes	yes	yes	yes
Cluster-fixed effects	yes	yes	yes	yes	yes	yes
Full set of controls	yes	yes	yes	yes	yes	yes
Observations	2,170	2,170	2,170	2,170	2,170	2,170

Source: Own calculations using data from ISS, SAPS, StatsSA, and WorldPop.
Notes: Each column contains four regressions; two for education and housing inequality, respectively. The full set of controls includes ln mean p.c. household income, ln population density, unemployment rate, share of adolescents, share of recently moved, share of female-headed hh, share of urban hh, share of coloureds. Dependent variables are IHS-transformed crime rates (crimes per 100,000 inhabitants). The spatial lag of the crime rate is instrumented with the spatial lags of first, second, and third order of all independent variables. Robust standard errors in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

References

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

Elbers, C., J. O. Lanjouw and P. Lanjouw (2003), Micro-level estimation of poverty and inequality, *Econometrica*,, 71(1): 355-364.

Li, X., Y. Zhou, M. Zhao and X. Zhao (2020), A harmonized global nighttime light dataset 1992–2018, *Scientific Data*, 7(168): 1-9.

Oster, E. (2019), Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics*, 37(2): 187-204.

Zhao, Q. and P. Lanjouw (n.d.), *Using PovMap2: A user's guide*, World Bank, Washington, DC.