Immigration and Crime Across U.S. Cities:
An Analysis of The Great American Crime Decline of the 1990s

Jessa Wang, Leah Ross, Denise Upton, Pieter Fildes

I. Introduction

Anti-immigration fervor has been growing over the past few decades[1]. The proliferate depiction of immigrants as violent criminals, drug dealers, rapists, and terrorists has raised the fear for crimes and terror brought by foreigners[2]. However, little evidence has found that increased immigration leads to higher rates of crime. Instead, many studies discovered a negative association between immigration and crime rate[3]. To further understand the relationship between immigration population and crime, our study looks at census panel data across American cities from two years — 1990 and 2000.

1990 and 2000 are particularly interesting years to look at due to the drastic changes in both crime and immigration rates that occurred during the 1990s. Crime sharply declined in this period, with violent crimes plummeting by around 51% and property crimes decreasing by about 43% [4]. During the same period, the foreign-born population hit a peak, taking in 1.2 million immigrants every year on average[5]. The majority of the inflows have settled in metro areas and have changed the demographics of cities considerably. Nevertheless, the rise of immigration does not appear to result in more crimes. To refute the stereotypes about immigrants, we think it's

---

[1] (2018). Mainstreaming Hate: The Anti-Immigrant Movement in the US. *Center on Extremism,* https://www.adl.org/the-anti-immigrant-movement-in-the-us

[2] Harris, C., and Gruenewald, J. (2020). News Media Trends in the Framing of Immigration and Crime, 1990–2013. *Social Problems*, 67(3), 452–470. https://doi.org/10.1093/socpro/spz024

[3] Martinez, R., Stowell, J., and Iwama, J. (2016).. The Role of Immigration: Race/Ethnicity and San Diego Homicides Since 1970. *Journal of Quantitative Criminology,* 32, 471–488. https://doi.org/10.1007/s10940-016-9294-9 ; Kim, Y., Hipp, J., and Kubrin, C. (2019). Where they live and go: Immigrant ethnic activity space and neighborhood crime in Southern California. *Journal of Criminal Justice,* 64, 1-12. https://doi.org/10.1016/j.jcrimjus.2019.05.004.

[4] Gramlich, J. (2020). What the data says (and doesn't say) about crime in the United States. *Pew Research Center,* https://www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/

[5] Passel, J., and Suro, R. (2005). Rise, peak and decline: Trends in U.S. immigration 1992 – 2004. *Pew Research Center,* https://www.pewresearch.org/wp-content/uploads/sites/5/reports/53.pdf

important to analyze this relationship, taking into consideration other changing factors between 1990 and 2000, such as economic growth and education[6].

II. Data

Our sample was drawn from U.S. Census Bureau data and consists of 110 cities across 37 states in the years 1990 and 2000. Our dependent variable is *total_crime*, which includes the total number of violent and property crimes. The variable of interest is *pctforeign* and stands for the proportion of the population that is not a U.S. citizen at birth, including those who later become U.S. citizens through naturalization. Other population variables included in our model are *citypop,* which represents city population, and age variables that represent the proportion of the population between various age groups (*p0to17, p18to24, p18-24, p25to44, p45to64, p65*). The demographic variable *pctblack*, which represents the percent of the population that is black, is also included.

The raw education attainment data gives the number of adults (aged 25+) that earned a high school diploma, a bachelor's degree, or have some college education. In order to prevent multicollinearity, we included the number of adults as proportions. *Edu_hsgrad* represents the number of adults (25+) with a high school diploma. This was then divided by the total number of adults (25+) in the population for each observation, to generate *pct_edu_hsgrad*. Similarly, *edu_somecoll_orhigher* represents the number of adults with some college or higher, divided by the total number of adults in the population for each observation. This created *pct_edu_somecol_orhigher*.

[6] Levitt, S., (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives*, 18(1), 163–190. https://pricetheory.uchicago.edu/levitt/Papers/LevittUnderstandingWhyCrime2004.pdf

Other variables include city median income *(income*), annual unemployment rate (*unemployment*), and average daily temperature in July (*julyavgdlytemp*).

Our study has in total 220 observations, with no missing data in our sample. The descriptive statistics (Table 1) show large variances across cities in *total_crime*, *pctforeign,* and *citypop*. Our sample range from large cities with millions of individuals to smaller urban towns. The foreign-born population is very skewed in our sample as well. It takes up nearly 60% of the population in cities like Miami and is insignificant in smaller cities like Macon city, GA.

In our sample, the average total crime fell from 39,846 in 1990 to 26,894 in 2000, marking a nearly 33% decrease. The foreign population grew roughly 37%. The huge changes in crime and immigration population in the sample cities confirm our hypothesis that crime and immigration are negatively correlated.

III. Model and Empirical Strategy

Our regression model was estimated using OLS with fixed and time effects and integrated various control variables and functional form transformations, as described below.

$$\begin{aligned} ln\_total\_crime_i &\\ = \beta_0 &+ \beta_1 pctforeign_{1,it} + \beta_2 pctblack_{2,it} + \beta_3 p18to24_{3,it} + \beta_4 p25to44X_{4,it} + \beta_5 p45to64_{5,it} \\ &+ \beta_6 p65_{6,it} + \beta_7 ln\_income_{7,it} + \beta_8 unemployment_{8,it} + \beta_9 pct\_edu\_hsgrad_{9,it} \\ &+ \beta_{10} pct\_edu\_somecol\_orhigher_{10,it} + \beta_{11} ln\_citypop_{11,it} + a_i + T_t + \epsilon_{it} \end{aligned}$$

<u>Base Model with Time Effects</u>

Model #1 in Table 2 represents our base model, where total crime is a function of *pctforeign* and our year dummy, *year2000*. As expected, the coefficient on our year variable is negative, representing the large decreases in total crime across cities in the 1990s[7]. Based on our theoretical understanding of crime and immigration, we expect the sign of the coefficient on

[7] Ibid.

pctforeign to be negative. However, in Model #1, we find that *pctforeign* is positive and statistically significant. Model #1 is too simplified to measure the many complex factors contributing to the relationship between *total_crime* and *pctforeign* and suffers from specification errors, including omitted variable bias. This is evident as only 7.8% of the variation in *total_crime* is explained by *pctforeign* and *year2000*.

Controls

In Model #2, we add additional control variables that are important in explaining *total_crime*. In terms of race, the variable *pctback* was added, as literature shows that this measure is related to increased rates of crime. This is due to institutional racism and biased policing policies, such as stop and frisk, which disproportionately target black communities[8].

There is also evidence that age is correlated with crime. Literature shows that those in their teens to mid-20s are most likely to commit a crime, while older individuals (60+) are much less likely to commit a crime[9]. To control for changes in age across the different cities and over time, we included all the possible age groups, except for one to avoid perfect multicollinearity, including *p18to24, p25to44, p45to64,* and *p65*.

To control for economic effects, the variables *income* and *unemployment* are included. Literature shows these two variables are positively correlated to property crime[10].

In terms of education, increased higher education is generally associated with lower overall crime. We found that obtaining a high school degree is largely related to decreased crime, so we added *pct_educ_hsgrad* to our model[11]. We also found that obtaining additional education

[8] Barker, V. (2010). Explaining the great American crime decline: A review of Blumstein and Wallman, Goldberger and Rosenfeld, and Zimring. *Law & Social Inquiry*, 35(2), 489-516, https://www.jstor.org/stable/pdf/40783025.pdf?refreqid=excelsior%3A8cce0f778fa072da11832c55ccd2e067

[9] Ulmber, J., and Steffensmeier, D. (2014). The nurture versus biosocial debate on criminology. *SAGE,* 377-396, https://sk.sagepub.com/books/the-nurture-versus-biosocial-debate-in-criminology/n23.xml

[10] Raphael, S., and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law and Economics,* 44(1), 259-283, https://doi.org/10.1086/320275

[11] Gentry, B., Mokkapati, R., and Rampersad, K. (2016). Impact of educational attainment on crime in the United States: A cross-metropolitan analysis. *Georgia Institute of Techn*ology, http://hdl.handle.net/1853/56029

after high school has marginal effects on total crime. We know that education is an important variable that encompasses many additional socio-economic factors, this further justified creating the new variable *edu_somecoll_orhigher*.

We include city population as a variable, as we know more densely populated areas are more prone to crime and that large cities across the U.S. saw a major decrease in crime in the 90s. We also added a dummy variable for temperature, *julyavgdlytemp,* because the literature theorizes that more crimes are committed in the warmer months, perhaps because people spend more time outdoors[12].

The addition of these various controls in Model #2 provides a more well-rounded analysis compared to Model #1—for example, our R-squared increases to 82.9% compared to 7.8%. However, there are still problems with this model. In Model #2, we find that only a handful of our variables are statistically significant, including *pctblack, citypop*, *julyavgdlytemp*, and *year2000.* Crucially, our variable of interest, *pctforeign*, continues to be positive and statistically significant, suggesting that Model #2 may be biased.

<u>Heteroskedasticity and Multicollinearity</u>

Following the addition of the controls, a White Test was conducted to determine whether there was variance in the residuals of our variables. The White's test resulted in a p-value of 0.000, indicating heteroskedasticity in Model #2. The residuals for each of the variables were graphed and *total_crime* and *citypop* were found to have the greatest variance (see Figures 1 and 2).

Next, a test of multicollinearity was conducted. While there was some multicollinearity present in the year dummy variable (Variance Inflation Factor=9.88) and in one of the age

[12] Field, S. (1992). The Effect of Temperature on Crime. *The British Journal of Criminology,* 32(2), pp. 340-351. https://www.jstor.org/stable/23637533

variables, *p45to64* (VIF=7.16), the variable of interest, *pctforeign*, was not severely affected by multicollinearity (VIF=2.72). Meaning, we are not greatly concerned about multicollinearity being present in the model.

<u>Functional Form Transformations</u>

To address the heteroskedasticity present in the residuals for the dependent variable *total_crime*, it was transformed into a logarithm. The transformed variable was integrated into Model #3 and the residuals of the logged variable were graphed. As evidenced by Figure 3, logging the independent variable helped minimize the variance in its residuals.

To address the heteroskedasticity present in the residuals for *citypop*, it was transformed into a logarithm. The transformed variable was integrated into Model #3 and the residuals of the logged variable were graphed. As evidenced by Figure 4, logging *citypop* helped minimize the variance in its residuals.

To facilitate interpretation of the model and to control for slight variance in its residuals, the *income* variable was also transformed into a logarithm.

Next, it was hypothesized that the relationship between temperature and crime was not linear, and the positive association between these variables would decrease in magnitude as temperature increases. As such, the temperature variable was included as a quadratic. The joint F-test of *julyavgdlytemp* and *julyavgdlytemp_sq* resulted in a p-value of 0.0079, demonstrating the significance of this polynomial.

The functional form transformations discussed above were integrated into Model #3 and the model was retested for multicollinearity and heteroskedasticity. The White's test indicated a p-value of 0.7878, indicating that the functional form transformations helped minimize heteroskedasticity in the model.

A retest of multicollinearity showed that, as expected, our temperature polynomial was highly multicollinear. Multicollinearity remained present in the year dummy variable (VIF=10.36) and in one of the age variables, *p45to64* (VIF=7.32). However, the variable of interest, *pctforeign*, was not severely affected by multicollinearity (VIF=2.80).

<u>City Fixed Effects</u>

In the final model, fixed effects were added to control for unobserved time-invariant effects that are unique to each of the cities, such as city-specific public policies or cultural characteristics. The introduction of fixed effects to the model eliminates any bias present due to the omission of time-invariant variables. As such, the use of fixed effects requires that time-invariant variables be removed from the specification, so the temperature polynomial was omitted from the final model.

IV. Results

While the initial relationship between our dependent variable, *total_crime,* and our variable of interest, *pctforeign,* was found to be positive and insignificant (Model #1), following the addition of time effects, controls, functional form transformations, and city fixed effects, the coefficient on *pct_foreign* in Model #4 is negative and significant, with a p-value of 0.066.

The R-squared for our final specification indicated that 74.54% of the variation in total crime can be explained by the variables used in Model #4. Additionally, an F-test of the overall significance of the model results in a p-value of 0.000, demonstrating that the model is statistically significant at the 1% level.

Turning specifically to our variable of interest, Model #4 shows that on average, a 10% increase in the proportion of the city population that is foreign is associated with a decrease in total crime of approximately 0.19%, *ceteris paribus*. This relationship is significant at the 10% level.

Additionally, we found *pctblack* and our year dummy to have interesting results. Though not our primary variable of interest, *pctblack* was found to be highly statistically significant at the 1% level in Models #2-4. This is an interesting finding which we believe can be explained by racial bias, specifically related to policing and convictions that occurred across cities[13]. Similarly, *year2000* was found to be statistically significant in Model #1 and Model #2 at the 10% level and highly statistically significant at the 1% level in Model #4. This reflects the various changes that occurred at both the city and federal levels during the 1990 and 2000s.

V. Conclusion

In conclusion, our analysis revealed several policy implications concerning U.S. crime rates in cities and communities with sizable foreign-born populations. Our findings refute commonly held stigmas towards immigrants, where anti-immigration politicians and residents perceived an association between increases in foreign-born residents and increases in a city's crime rate. In actuality, though, areas and cities with higher foreign-born populations have lower crime rates, and from 1990-2000 when the United States experienced one of its largest increases in immigration, a significant decrease in crime occurred.[14]

In the future, we think it would be interesting to include some additional variables to better understand the relationship between foreign-born residents and crime, such as alcohol or drug consumption on the community level, allocation of community services, policing or incarceration data, and the legality of abortion in the cities surveyed. The literature indicates that increased alcohol or drug consumption in a community corresponds with increases in most types

[13] Barker, 2010
[14] Passel and Suro, 2005

of crime. Additionally, increased spending on community services, better-policing strategies, increased use of capital punishment, and access to legalized abortion are often associated with decreases in crime[15].

Other issues with our dataset include the income variable, consistency of survey questions, and the lack of detailed information by location (such as neighborhood or zip code). Concerning income, the variable in our dataset only focused on city-wide median income, which does not account for income disparities. The income variable also does not account for inflation over the study period. Additional data that considers granular level factors would give better insights toward areas where income inequality played a role in crime levels. Since our data was collected from U.S. Census questions, we did have to keep in mind that we were unable to control what was asked and that the survey is typically updated or revised every ten years. Though this helps from a relevancy standpoint, it may hinder some of the consistency in the data collected, as the conditions for questions and responses over multiple censuses might not be identical.

Going forward, we hope to see additional research that accounts for the previously mentioned missing variables and examines the various inequalities between cities. Doing so will help us better understand the relationship between immigrants and total crime in cities across the US.

---

[15] Levitt, 2004 ; Sampson, R. (2008). Rethinking crime and immigration. *Contexts*, 7(1), 28-33. https://journals.sagepub.com/doi/pdf/10.1525/ctx.2008.7.1.28

VI. Appendix

Table 1: Descriptive Statistics of All the Variables

	N	Mean	Std. Dev.	min	max
total_crime	220	33370.027	60536.089	3837	710222
pctforeign	220	.118	0.115	.008	.597
citypop	220	411498.74	816235.009	94911	8008278
income	220	3.748	0.925	1.972	7.481
unemployment	220	5.561	2.181	2.1	16.1
pct_edu_hsgrad	220	.254	0.054	.079	.386
pct_edu_somecol_orhigher	220	.346	0.085	.156	.569
julyavgdlytemp	220	76.901	5.512	62.1	93.5
pctblack	220	.243	0.198	.006	.84
p0to17	220	.259	0.020	.178	.362
p18to24	220	.102	0.010	.077	.145
p25to44	220	.313	0.020	.265	.362
p45to64	220	.203	0.022	.143	.246
p65	220	.123	0.019	.087	.182

Table 2: Regression Results for Linear Model

	(1) total_crime	(2) total_crime	(3) ln_total_crime	(4) ln_total_crime
pctforeign	136,830**	14,961	0.0207	-1.870*
	(59,038)	(18,150)	(0.200)	(1.008)
pctblack		20,037***	0.536***	2.562***
		(7,194)	(0.118)	(0.531)
p18to24		-562.0	0.357	-3.772
		(188,842)	(3.047)	(3.714)
p25to44		-110,390	-0.874	-1.171
		(75,225)	(1.471)	(3.501)
p45to64		15,910	-0.883	0.703
		(121,873)	(1.835)	(3.788)
p65		-25,928	4.109***	-15.11*
		(78,032)	(1.273)	(8.234)
income		-3,612		
		(2,808)		
ln_income			-0.713***	0.311
			(0.155)	(0.420)
unemployment		775.7	0.0263***	0.0174
		(625.1)	(0.00957)	(0.0155)
pct_edu_hsgrad		-26,867	-0.858*	0.236
		(22,631)	(0.496)	(1.686)
pct_edu_somecol_orhigher		32,777	1.555***	0.375
		(21,476)	(0.375)	(1.336)
citypop		0.0656***		
		(0.0174)		
ln_citypop			0.978***	1.132***
			(0.0243)	(0.231)
julyavgdlytemp		422.2*	-0.127*	
		(222.3)	(0.0684)	
julyavgdlytemp_sq			0.000867**	
			(0.000438)	
year2000	-17,938*	-14,939*	-0.160	-0.489***
	(9,211)	(8,333)	(0.104)	(0.148)
Constant	26,259***	14,074	2.672	-2.551
	(4,397)	(49,398)	(2.652)	(4.376)
Observations	220	220	220	220
R-squared	0.078	0.826	0.926	0.745
City Fixed Effects	No	No	No	Yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Residuals for *total_crime*

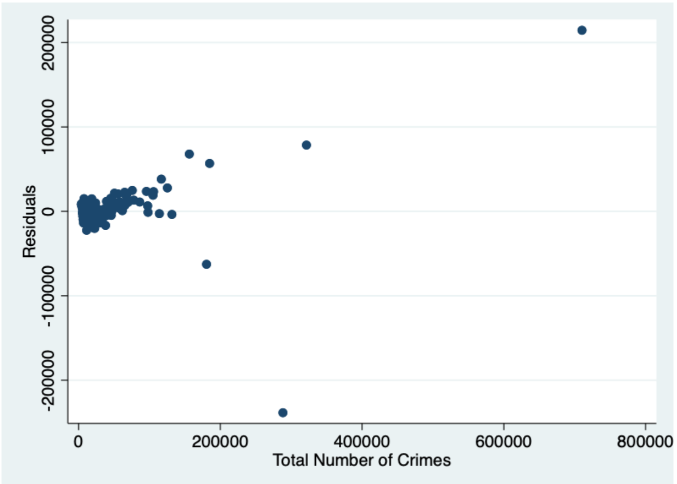


Figure 2: Residuals for *citypop*

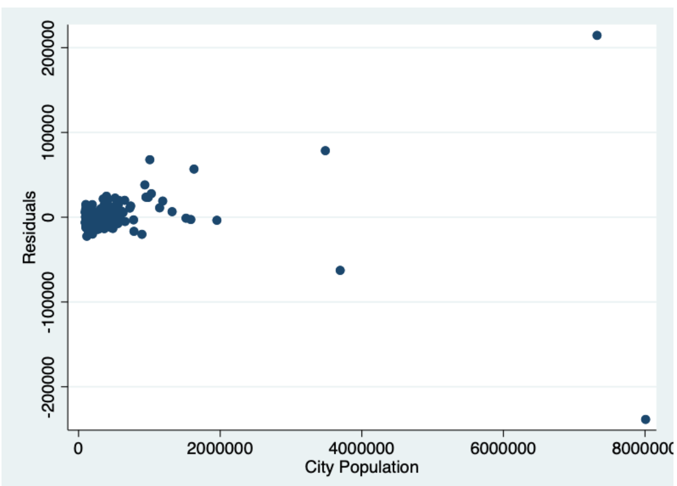


Figure 3: Residuals for *ln_total_crime*

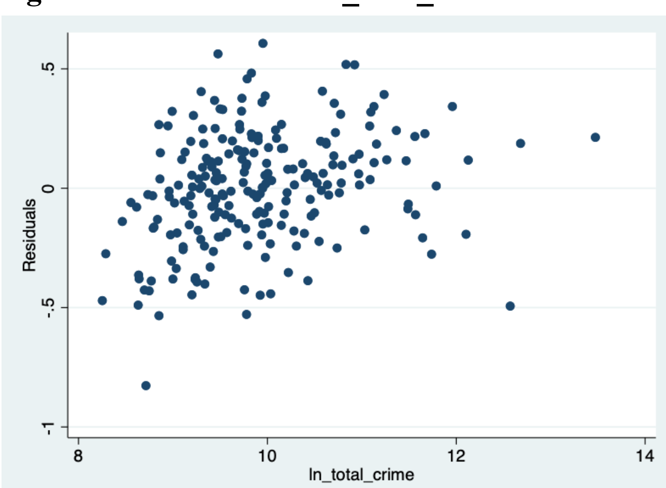
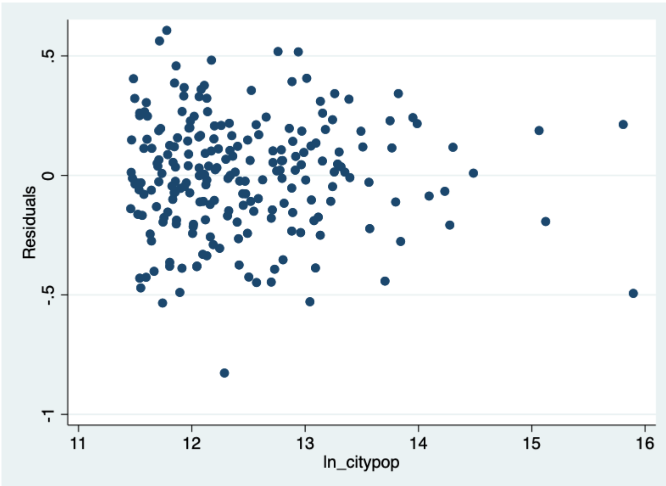


Figure 4: Residuals for *ln_citypop*



```
---------------------------------------------------------------------------------------------

. use "$datadir/dataset_large_city_final - Copy.dta", clear

.
. *FUNCTIONAL FORM TRANSFORMATIONS
.
. *Year Dummy
.
. gen year2000 = (year==2000)

.
. *Logarithms
.
. gen ln_total_crime = ln(total_crime)

.
. gen ln_citypop = ln(citypop)

.
. gen ln_income = ln(income)

.
. gen ln_pctforeign = ln(pctforeign)

.
. *Polynomial
.
. gen julyavgdlytemp_sq = julyavgdlytemp*julyavgdlytemp

.
. *Education variables
.
. gen edu_somecol_orhigher = edu_somecol+edu_bachelors

.
. gen p25andover = p25to44+p45to64+p65

.
. gen citypop_25andover = citypop*p25andover

.
. gen pct_edu_hsgrad = edu_hsgrad/citypop_25andover

.
. gen pct_edu_somecol_orhigher = edu_somecol_orhigher/citypop_25andover

.
. *DESCRIPTIVE STATISTICS
.
. asdoc sum total_crime pctforeign citypop income unemployment pct_edu_hsgrad
pct_edu_somecol_orhigher julyavgdlytemp pctblack p0to17 p18to24 p25to44 p45to64 p6
> 5, stat(N mean sd min max) replace

             |          N       Mean         SD        Min        Max
-------------+-------------------------------------------------------
 total_crime |        220   33370.03   60536.09       3837     710222
  pctforeign |        220   .1175162   .1153671    .008374   .5972087
```

```
     citypop |        220    411498.7      816235      94911     8008278
      income |        220    3.747634    .9254112     1.9725      7.4813
unemployment |        220    5.560909    2.181118        2.1        16.1
pct_edu_hs~d |        220    .2538238    .0544144   .0786601    .3861768
pct_edu_so~r |        220    .3457523    .0851292   .1563747    .5686861
julyavgdly~p |        220    76.90091    5.512369       62.1        93.5
    pctblack |        220    .2426943    .1976309   .0060205    .8402955
     p0to17  |        220    .2586227     .019722       .178        .362
     p18to24 |        220    .1021136     .010354       .077        .145
     p25to44 |        220    .3133591    .0203699       .265        .362
     p45to64 |        220    .2027091    .0217205       .143        .246
         p65 |        220    .1234045     .019151       .087        .182

Click to Open File:  Myfile.doc

.
. *REGRESSION 1: Base model with time effects
.
. reg total_crime pctforeign year2000, robust

Linear regression                               Number of obs     =        220
                                                F(2, 217)         =       2.80
                                                Prob > F          =     0.0631
                                                R-squared         =     0.0778
                                                Root MSE          =      58401

------------------------------------------------------------------------------
             |               Robust
 total_crime | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  pctforeign |     136830   59038.26     2.32   0.021     20468.22    253191.9
    year2000 |  -17937.87   9211.105    -1.95   0.053    -36092.56    216.8123
       _cons |   26259.22   4397.206     5.97   0.000     17592.52    34925.92
------------------------------------------------------------------------------


.
. *Graph initial relationship
.
. twoway scatter total_crime pctforeign || lfit total_crime pctforeign

.
. *REGRESSION 2: Controls
.
. reg total_crime pctforeign pctblack p18to24 p25to44 p45to64 p65 income unemployment
pct_edu_hsgrad pct_edu_somecol_orhigher citypop julyavgdlytemp year2000, r
> obust

Linear regression                               Number of obs     =        220
                                                F(13, 206)        =      13.95
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8255
                                                Root MSE          =      26073

------------------------------------------------------------------------------------------
                         |               Robust
             total_crime | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------------+----------------------------------------------------------------
              pctforeign |   14960.65   18150.01     0.82   0.411    -20822.95    50744.24
```

```
               pctblack |   20037.07   7194.416     2.79   0.006     5852.938    34221.19
                p18to24 |  -562.0471   188842.3    -0.00   0.998    -372873.4    371749.3
                p25to44 |  -110390.2   75224.61    -1.47   0.144      -258699    37918.64
                p45to64 |   15910.49   121873.5     0.13   0.896    -224368.8    256189.8
                    p65 |  -25927.66   78031.52    -0.33   0.740    -179770.4    127915.1
                 income |  -3612.438   2808.011    -1.29   0.200    -9148.562    1923.686
           unemployment |   775.6784   625.0658     1.24   0.216    -456.6679    2008.025
         pct_edu_hsgrad |  -26867.34   22630.72    -1.19   0.237    -71484.87    17750.18
pct_edu_somecol_orhigher |   32777.16   21476.12     1.53   0.128    -9564.016    75118.33
                citypop |   .0656326   .0173901     3.77   0.000     .0313472     .099918
          julyavgdlytemp |   422.2465   222.3303     1.90   0.059      -16.088     860.581
               year2000 |  -14938.52   8332.585    -1.79   0.074     -31366.6    1489.565
                  _cons |   14074.38   49397.51     0.28   0.776    -83315.12    111463.9
------------------------------------------------------------------------------------------

.
. *Multicollinearity
.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
    year2000 |      9.88    0.101227
     p45to64 |      7.16    0.139683
      income |      4.56    0.219126
     p25to44 |      3.82    0.261716
pct_edu_so~r |      3.67    0.272380
         p65 |      3.44    0.290619
     p18to24 |      3.18    0.314452
  pctforeign |      2.72    0.367772
pct_edu_hs~d |      2.20    0.455524
unemployment |      2.10    0.475922
    pctblack |      2.01    0.496364
julyavgdly~p |      1.29    0.775724
     citypop |      1.15    0.866125
-------------+----------------------
    Mean VIF |      3.63

.
. *Heteroskedasticity
.
. estat imtest, white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

  chi2(103) = 219.55
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

--------------------------------------------------
              Source |       chi2     df         p
---------------------+----------------------------
  Heteroskedasticity |     219.55    103    0.0000
            Skewness |      10.46     13    0.6557
            Kurtosis |  -1.13e+07      1    1.0000
```

```
---------------------+----------------------------
               Total |  -1.13e+07    117    1.0000
--------------------------------------------------

.
. *Graph residuals to determine which specific variables are heteroskedastic
.
. predict residuals, residuals

.
. twoway scatter residuals total_crime

.
. twoway scatter residuals pctforeign

.
. twoway scatter residuals pctblack

.
. twoway scatter residuals p0to17

.
. twoway scatter residuals p18to24

.
. twoway scatter residuals p25to44

.
. twoway scatter residuals p45to64

.
. twoway scatter residuals p65

.
. twoway scatter residuals income

.
. twoway scatter residuals unemployment

.
. twoway scatter residuals pct_edu_hsgrad

.
. twoway scatter residuals pct_edu_somecol_orhigher

.
. twoway scatter residuals citypop

.
. twoway scatter residuals julyavgdlytemp

.
. *REGRESSION 3: Functional Form
.
. reg ln_total_crime pctforeign pctblack p0to17 p18to24 p25to44 p45to64 ln_income
unemployment pct_edu_hsgrad pct_edu_somecol_orhigher ln_citypop julyavgdlytemp
>  julyavgdlytemp_sq year2000, robust
```

```
Linear regression                               Number of obs     =        220
                                                F(14, 205)        =     175.72
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9256
                                                Root MSE          =     .24533

------------------------------------------------------------------------------------------
                         |               Robust
          ln_total_crime | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------------+----------------------------------------------------------------
              pctforeign |   .0277137   .2005118     0.14   0.890     -.367616    .4230435
                pctblack |   .5383644   .1186381     4.54   0.000     .3044571    .7722717
                  p0to17 |  -3.853634   1.274446    -3.02   0.003    -6.366337   -1.340932
                 p18to24 |  -3.760242   2.501152    -1.50   0.134    -8.691521    1.171037
                 p25to44 |  -4.893955   1.278579    -3.83   0.000    -7.414806   -2.373105
                 p45to64 |  -4.625764   2.308625    -2.00   0.046    -9.177458   -.074071
               ln_income |  -.7103559   .1560468    -4.55   0.000    -1.018018   -.4026934
            unemployment |   .0259017   .0095948     2.70   0.008     .0069846    .0448187
            pct_edu_hsgrad |  -.8555629   .4968835    -1.72   0.087     -1.83522    .1240943
pct_edu_somecol_orhigher |   1.539905   .3775574     4.08   0.000     .7955119    2.284299
              ln_citypop |     .97765   .0242534    40.31   0.000      .929832    1.025468
          julyavgdlytemp |   -.128838   .0684235    -1.88   0.061     -.263742     .006066
       julyavgdlytemp_sq |   .0008792    .000438     2.01   0.046     .0000156    .0017428
                year2000 |  -.1703657   .1043325    -1.63   0.104    -.3760681    .0353367
                   _cons |   6.684832    2.72748     2.45   0.015     1.307323    12.06234
------------------------------------------------------------------------------------------

.
. *Regraph residuals for heteroskedastic variables
.
. drop residuals

.
. predict residuals, residuals

.
. twoway scatter residuals ln_total_crime

.
. twoway scatter residuals ln_citypop

.
. *Significance test for polynomial
.
. test julyavgdlytemp julyavgdlytemp_sq

 ( 1)  julyavgdlytemp = 0
 ( 2)  julyavgdlytemp_sq = 0

       F(  2,   205) =    4.95
            Prob > F =    0.0079

.
. *Retest Heteroskedasticity
.
. estat imtest, white

White's test
```

```
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

  chi2(117) = 104.66
Prob > chi2 = 0.7862

Cameron & Trivedi's decomposition of IM-test

--------------------------------------------------
              Source |       chi2     df         p
---------------------+----------------------------
  Heteroskedasticity |     104.66    117    0.7862
            Skewness |      11.48     14    0.6478
            Kurtosis |       0.10      1    0.7468
---------------------+----------------------------
               Total |     116.24    132    0.8339
--------------------------------------------------

.
. *Retest Multicollinearity
.
. vif

    Variable |       VIF       1/VIF  
-------------+----------------------
julyavgdly~q |    440.73    0.002269
julyavgdly~p |    440.04    0.002273
     p45to64 |     13.72    0.072876
    year2000 |     10.38    0.096342
   ln_income |      5.49    0.182109
pct_edu_so~r |      4.01    0.249606
      p0to17 |      3.68    0.271836
     p25to44 |      3.01    0.332628
  pctforeign |      2.79    0.357944
     p18to24 |      2.30    0.433880
pct_edu_hs~d |      2.22    0.449641
unemployment |      2.18    0.459471
    pctblack |      2.13    0.469291
  ln_citypop |      1.28    0.780037
-------------+----------------------
    Mean VIF |     66.71

.
. *Regression 4: Fixed Effects
.
. xtset city year

Panel variable: city (strongly balanced)
 Time variable: year, 1990 to 2000, but with gaps
         Delta: 1 unit

.
. xtreg ln_total_crime pctforeign pctblack p0to17 p18to24 p25to44 p45to64 ln_income
unemployment pct_edu_hsgrad pct_edu_somecol_orhigher ln_citypop year2000, fe
>  robust

Fixed-effects (within) regression               Number of obs     =        220
Group variable: city                            Number of groups  =        110
```

```
R-squared:                                      Obs per group:
     Within  = 0.7445                                         min =          2
     Between = 0.6712                                         avg =        2.0
     Overall = 0.6703                                         max =          2

                                                F(12,109)         =      31.08
corr(u_i, Xb) = -0.6071                         Prob > F          =     0.0000

                                   (Std. err. adjusted for 110 clusters in city)
------------------------------------------------------------------------------------------
                         |               Robust
          ln_total_crime | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------------+----------------------------------------------------------------
              pctfor
eign | -1.769463   .9932426    -1.78   0.078    -3.738037    .1991118
                pctblack |   2.528971   .5340924     4.74   0.000     1.470417    3.587524
                  p0to17 |   14.95942    8.85047     1.69   0.094    -2.581927    32.50076
                 p18to24 |   11.29701    7.29836     1.55   0.125      -3.1681    25.76212
                 p25to44 |   14.04262   8.067964     1.74   0.085    -1.947827    30.03306
                 p45to64 |   16.08824   7.836178     2.05   0.042      .557187    31.61929
               ln_income |   .2936805   .4162335     0.71   0.482    -.5312808    1.118642
            unemployment |   .0177232   .0154318     1.15   0.253     -.012862    .0483085
          pct_edu_hsgrad |   .4424582   1.682268     0.26   0.793    -2.891742    3.776658
pct_edu_somecol_orhigher |   .4506178   1.321057     0.34   0.734    -2.167675     3.06891
              ln_citypop |   1.129486   .2265182     4.99   0.000     .6805339    1.578438
                year2000 |  -.5033741   .1478864    -3.40   0.001    -.7964801   -.2102681
                   _cons |  -17.73398   7.220254    -2.46   0.016    -32.04429   -3.423674
-------------------------+----------------------------------------------------------------
                 sigma_u |    .617261
                 sigma_e |  .15109825
                     rho |  .94346628   (fraction of variance due to u_i)
------------------------------------------------------------------------------------------

.
. log close
      name:  <unnamed>
       log:  /Volumes/GoogleDrive/My Drive/_Wagner/2021-09/PADM-GP 2902 Multiple Regression
and Introduction to Econometrics/Final Project/Final Project.smcl
  log type:  smcl
 closed on:  19 Dec 2021, 13:29:57
---------------------------------------------------------------------------------------------
```