UNIVERSITA' DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE ECONOMICHE E AZIENDALI
"MARCO FANNO"

CORSO DI LAUREA SPECIALISTICA IN ECONOMIA E FINANZA

TESI DI LAUREA

SAMPLING ERROR AND FIXED EFFECT ESTIMATION BIAS. A MONTECARLO SIMULATION

Relatore:

CH.MO.Prof. LUCA NUNZIATA

Laureando: CLEMENTE CORTILE

MATR: 1105253

ANNO ACCADEMICO 2016 - 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.

Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

INTRODUCTION 6

CHAPTER 1 - THEORETICAL BACKGROUND 9

1.1 Research design 9

1.2 Models overview. 12

1.2.1 Fixed effect model 12

1.2.2 Split sample instrumental variable model 13

1.3 Types of bias. 15

1.3.1 Omitted variable bias 15

1.3.2 Attenuation bias 16

CHAPTER 2 – SIMULATION RESULTS 19

2.1 Cell size and sampling rate 20

2.2 Baseline Population 21

2.2.1 Baseline FE model 22

2.2.2 Baseline FE vs SSIV model 24

2.3 Properties of SSIV 25

2.3 Optimal cell size 28

CHAPTER 3 – ALTERNATIVE POPULATIONS 35

3.1 Unexplained variance 37

3.2 Initial proportion of immigrants 39

3.3 Immigration shock's intensity 44

3.3.1 SSIV First stage quality 47

3.4 Panel data length 48

CHAPTER 4 – Pooled OLS 52

4.1 Model outline 52

4.1 POLS comparison 53

CONCLUSIONS 56

DATA SPEC – TECHNICAL SHEET 59

Simulation Stats 59

Master File Structure 59

BIBLIOGRAPHY 61

APPENDIX 65

INTRODUCTION

Economic science's main purpose is the efficient allocation of limited resources, but in our era of instantaneous communication, along with the management of classic resources, data and information play a pivotal role in any economic activity.

The exponential growth of modern computational power, and its ever increasing ease of access, is enabling researchers to use new approaches to statistical inference and data processing, without having to resort to extremely expensive mainframes. Even though statistic is always the best choice to convert data into information, if compared to the past, classic research and inference tools, such as regression models and statistical tests on datasets, have gained even more importance in every scientific field.

Following that trend, the trade-off between costs and data quality have increased in importance as well. Regardless of how well refined a statistical inference method might be, the quantity and quality of data is the source of the information that is to be processed and conveyed. Therefore, each and every research has to deal with attenuation biases in estimators and sampling size problems.

Also in recent years, as much more data for studies and researches becomes available, the topic of reproducibility of published results became more relevant; even without taking into consideration all the problem that could arise in the complex process of acquiring, preparing, setting up datasets and interpreting it, the simple length of the sample studied is an unquestionable and non-interpretable key factor.

Even in physics publications, where the requisite in testing for null hypothesis is extremely high: around six-sigma, or one in a three million and half chance of false positive, such false positives arose in several studies from more than ten published paper in regarding the false discovery of pentaquarks. The successive researches that didn't confirm the finding were simply featuring bigger samples. A similar case happened in astrophysics as well whereas the results of a nobel-prize study on supernovae stars from 1999, regarding the accelerating expansion of the universe, has been challenged this year by a replication of the same study with ten times the sample size of the original (Nielsen J.T. et al., 2015).

The research design for the present thesis is first introduced and explained in details, illustrating the literature underlying the tests and the simulations. The structure of the simulation is then analysed and motivated.

A brief summary about the endogeneity problem and attenuation bias is provided, with specific reference to the model considered, together with an overview of the literature used in the thesis.

The simulation's results will be divided in two parts: the first will cover various adaptations for each sampling rate, featuring different levels of focus, the second will test the robustness of the model by altering the population condition and by parameter's alterations.

The last section will cover the comparison of the fixed effect model (FE), the split sample instrumental variable implementation (denoted as SSIV in the text and as IV on the tables), and the pooled ordinary least square model (POLS)

CHAPTER 1 - THEORETICAL BACKGROUND

1.1 Research design

The aim of this work consists in providing a reference in choosing an effective sample size, and building a guideline on what kind of bias is to be expected for each given sample size and population conditions, when using fixed effect models methods. In particular, endogeneity problems in the independent variables, and attenuation bias, due to both the aforementioned sample size and sampling problems, will be addressed.

The present thesis will use the framework set up in the published paper *"Immigration and crime: evidence from victimization data"* (Nunziata, 2015), to build a simulated population of ten million individuals distributed over one hundred regions.

The measure of immigration is affected both by endogeneity and attenuation bias: the former due to the fact that the immigrants are not randomly assigned to regions; and latter is caused by sampling error due to specific reasons, such as the regional cell size being too small or the selection bias caused by the fact that immigrants might not take the survey as much as natives.

The simulated population will then feature a proportion of immigrants, which will be subject to yearly changes in forms of immigration waves. Such waves will be simulated as to represent a dynamic shock over time: one at a population level, and one at regional level. The strength of the waves will change each year according to its own variance. The shocks are only positive and will result in an increase in the overall population. Furthermore, the intensity of the waves will be increased in later testing; the idea is to reproduce immigration waves similar to those registered during the European migrant crisis, and also compare them with weaker types of immigration waves.

The studied relationship between the crime perception and the immigrant share of the population is be exploited to build an estimator normalized to one, as the relationship between the share of immigrant in the population and the perception of crime will be set accordingly to the findings in the paper.

The population will be so that a marginal increase in immigration will lead to a linear increase of the perceived probability of being a crime victim by 20%, controlling for fixed regional and time effects:

$$C_{rit} = m_{rt}\,\beta + \mu_t + \gamma_t + \varepsilon_{it} \qquad (1.1)$$

Where C stands for the crime perception, m for immigration share; μ and γ are regional and time effects accounted for in the population simulation.

This restriction reflects the data in the paper and allows for a coefficient normalized to one. With this setup, any immigration coefficient in a fixed effect model, built on a sample of the initial population, will immediately display its own accuracy in its deviation from one. This framework has been implemented with the use of the STATA software package. Use of Monte Carlo simulation on fixed effect models has already brought several results in the literature. Among the last, there's the measure of statistical power of instrumental variables in the presence of weak instruments (Semadeni et al, 2014), which was conducted on a simulated dataset of 500 observations. Population tables will list the characteristics of the different populations upon which the models are implemented:

POPULATION A	
Initial pop. count: 10'000'000	Regions: 100
Immigration Shocks Variance: 1x	Proportion of immigrants: 0,10
Time lenght: 4	Error term distr. : $N\sim$(0;0,5)

The work will present several variation of the initial population to simulate the results under different initial conditions.

The next step consists in building fixed effect models. The choice of FE is justified by the fact that the nation itself and each region have time invariant unobservable factors. On top of that a simpler pooled OLS will be evaluated to test the strength of the fixed effects upon the end results. The first model will be computed to check the population true parameters, which will result to be normalized to one; and several others will be computed using different sampling rates on the population. As in the paper, the type of fixed effect models build will be with and without instrumental variables for each level of sampling rate.

In short the analysed models will be:

(1.2)

$$\text{FE Model -} \qquad C_{rit} = m_{rt}\beta_{FE} + \mu_c + \mu_r + \varepsilon_{it}$$

Where μ represents time invariant regional and country effects, and X is a matrix of individual controls.

As for the instrumental variable model, a split sample instrumental variable one will be used by sampling the immigrant population variable twice, at different sampling rate. As a reference for the sampled variables, the work will denote variable as END (which was the European Social Survey in the paper) to indicate the endogenous independent measure of immigration, and INS (denoted as the Labour Force Survey in the paper) to indicate the instrument.

The spilt sample instrumental variable model is built and denoted as follows

:

(1.3)

IV 2nd Stage $$C_{rit} = \widetilde{m_{rt}^{END}}\beta_{IV} + \mu_c + \mu_r + \varepsilon_{it}$$

IV 1st Stage $$\widetilde{m_{rt}^{END}} = m_{rt}^{INS}\beta_{FSIV} + \eta_{it}$$

Each variable is sampled in a wide array of sampling rates; mainly the ones available in surveys, and then the model is estimated. Monte Carlo simulation are then used to randomly resample and recompute each model several times. Tiers of 50 and 100 hundred replications of the models are used, accordingly to the accuracy needed. The resulting coefficients and standard errors are then averaged out and presented for comparison.

The first group of results will focus on the attenuation bias problem: several sampling rates will be tested and compared to map the change in the bias. The simulations will be also conducted swapping the instrumental variables in the split-sample instrumental variable approach.

Robustness checks will involve modification of the above mentioned models to identify the most suited setup of it: changing in the length of the panel data, removal of time effects from the controls, and modification in the error term's distribution. Other checks will be performed by introducing changes in the population, the intensity of the shocks and the proportion of immigrants, as to evaluate the behaviour of the bias under different circumstances. Also the models will be tested against their POLS versions while varying the intensity of the fixed effects.

1.2 Models overview.

1.2.1 Fixed effect model

In econometrics, fixed effect models (henceforth: FE) represent a viable choice to analyse and study panel data, whenever the focus of the research is to explain the behaviour of a variable over time. FE models explore the relationship between the dependent and independent variables within a specific level of research: individual, regional, national, and so on, controlling for the influence of each dependent variable. In particular, the FE model allows to remove the effect of unobserved time invariant independent variables, simply by exploiting the observation at different times and cancelling them out.

(1.4) $$Y_{it} = \beta_1 X_{it} + \mu_i + u_{it}$$

$$t = 1, \dots, T$$
$$i = 1, \dots, N$$

In such a setup, FE, are able to pinpoint, over a certain time frame, the effect on a dependent variable, by an independent variable, within a group, while also controlling for other factors if needed. Time periods and groups can either be used as controlling factors, or as explanatory to seize their effect.

The main point, however, consists in the fact that any unobserved factor, can be removed as long as it is constant over time. If the assumption holds true, then any effect on the dependent variable is due to non-fixed influences (Stock and Watson, 2003, p.289-290), and the "within transformation" can be applied to the model.

(1.5) $$(Y_{it} - \bar{Y}_i) = \beta_1 (X_{it} - \bar{X}_i) + (\mu_i - \bar{\mu}_i) + (u_{it} - \bar{u}_i)$$

(1.6) $$\ddot{Y}_{it} = \beta_1 \ddot{X}_{it} + \ddot{u}_{it}$$

Beta coefficients in equation (2) can then be easily estimated using Ordinary Least Squares (OLS) methods, which can be also denoted as a "*within estimator*" in this particular case (Wooldridge J., 2015). Once in form as in eq (2), other several estimation methods could be used, but research as shown that OLS still retains the best results (Buddelmeyer et al, 2008).

Therefore, under the Gauss-Markov assumptions, and if the strict exogeneity conditions are met, the estimator for the independent variable is unbiased and consistent. As for the case of OLS estimation, using X to denote the matrix of the explanatory variable, the bias would be defined as

$$E(\hat{\beta}|X) - \beta = (X'X)^{-1}X'E(u|X) \quad (1.7)$$

In the specific case of our model the key condition for estimating the immigration coefficient would be:

$$Cov(m_{rt}, \varepsilon_{it}) = 0, \ \forall\, t. \quad (1.8)$$

$$C_{rit} = m_{rt}\beta_{FE} + X_{it}\lambda + \mu_c + \mu_r + \varepsilon_{it} \quad (1.9)$$

namely that the idiosyncratic error is independent from the explanatory variable for all times in the model.

In the model considered however, the immigration explanatory variable is not randomly assigned to region (Nunziata, 2015), and unlike random effect models, in the baseline FE model it is assumed that the explanatory variables in the dynamic equation are not random. Therefore, any correlation between them and the error term will lead to an endogeneity problem as condition (3) is violated. That will ultimately cause bias in the estimator.

1.2.2 Split sample instrumental variable model

As explained in the paper, one of the solution for endogeneity, and even for the attenuation bias would be to use split-sample instrumental variable method.

The SSIV estimator is relatively new as it was proposed for the first time in the paper "Split Sample Instrumental Variable" (Angrist J.D., Krueger A.B., 1993), as a solution to IV bias caused by weak instruments in finite samples.

In our case the SSIV method will feature two measure of immigration taken form two different surveys, ESS and LFS respectively, which are independent of each other. The setup can easily be reproduced by taking two different, independent samples from the population of immigrants in the simulation. If m is a correct, unbiased measurement of immigration, the two survey measures will still differ from that by an error. So the characteristics of the two measure of immigration will be:

$$(1.10) \qquad m_{ESS} - m = \varepsilon_1 \; ; \;\; m_{LFS} - m = \varepsilon_2$$

$$Cov(m_{LFS}, m_{LFS}) \approx m$$

Even if condition (3) is not met due to endogeneity, the first stage estimation using split-sample instrumental variable, will exploit the characteristics of the two measure of immigration to “partial out” the errors, and keep the variance, of the unbiased common part of m, intact.

$$(1.11) \qquad C_{rit} = \widetilde{m_{rt}^{ESS}} \beta_{IV} + X_{it}\lambda + \mu_c + \mu_r + \varepsilon_{it}$$

$$\widetilde{m_{rt}^{ESS}} = m_{rt}^{LFS} \beta_{FSIV} + X_{it}\gamma + \eta_{it}$$

The use of SSIV however doesn’t solve the problem of attenuation bias when the sample size is too small. The following section will investigate the underlying literature to explain how and why the bias is formed.

1.3 Types of bias.

1.3.1 Omitted variable bias

While trying to identify the effect and the dynamics of the attenuation bias, caused by smaller sampling rates, it is useful to underline its differences with respect to the type of bias that's the cause of endogeneity in many models: the omitted variable bias, and how it can affect both FE and SSIV's estimations.

Most of the econometric literature on the subject of FE models, studies how the two stage instrumental variable method, in the presence of endogeneity, is biased towards OLS estimations (Nagar, 1959; Hausman et al, 2002). Even fixed effect models with instrumental variable are biased towards OLS.

Using the matrix form for the IV models (Ebbes P et al, 2009), the bias for an IV model defined as:

(1.12) $$\begin{cases} Y = X\beta + \varepsilon \\ X = Z\pi + \eta \end{cases}$$

is quantified as follows:

(1.13) $$\frac{Plim\ \widehat{\beta_{IV}} - \beta}{Plim\ \widehat{\beta_{OLS}} - \beta} = \frac{\rho_{\ddot{Z},\varepsilon} / \rho_{X,\varepsilon}}{\rho_{X;Z}}$$

The omitted variable bias committed in IV models is then generally dependent on the intensity of the linear relationship between the instrument and the endogenous variable, and inversely correlated to the linear correlation between the regressors and the error term of the second stage.

The researchers are especially focused on two broad topics: the quality of the instruments (Bound J. et al, 1995), and the problem that arise in small sample sizes (Flores-Lagunes A., 2007).

As shown in the above bias, the explanatory power of the instrument is indirectly proportional to the bias, as well as the exogeneity of the instrument with respect to the first stage's error terms.

Plenty of solutions are available, including different types of estimators such as Higher Moments and Latent Instrumental Variable (Ebbes et al. 2009), to solve the problem; the downside being the heavier requirements in terms of sample size and computational power.

However, in spite of the similarities, SSIV (Split Sample Instrumental Variables) models are biased towards zero rather than OLS, if no exogenous regressors are included in the model (Angrist D., Krueger B. 1993).

Provided that the instrument used meets all the requirement for exogeneity constrictions, and are related to the endogenous variable, SSIV doesn't suffer from endogeneity bias and the expected quality of the results in the simulations will mainly depend on the quality of the instrument in the first stage, and the resulting explanatory power of the second stage.

1.3.2 Attenuation bias

The attenuation bias, instead, needs to be apprehended by considering more specific circumstances.

The attenuation bias has been studied first by taking the contradictory conclusion of the effect of immigration on wages (Aydemir A., Borjas J. 2011), elaborated from two different surveys: one with a sampling rate of a 33% of the population and the other with a less than 10% sampling rate. Following their work, $p_k$will be the regressor affected by attenuation bias, and $\pi_k$will be the unbiased regressor. They will differ by an error term u_k:

(1.14) $$p_k = \pi_k + u_k$$

The distribution of $u_k$depends on the characteristic of the phenomenon observed. Since immigration is the subject of the research both for the reference paper and our case study, the error term is geometrically distributed. It is possible to approximate it to a binomial distribution without changing the characteristic of the error.

At this point the authors define the sampling rate as $\frac{n_k}{N_k} = \tau$; where n is the cell size, and N is the total population considered by the phenomenon. The model considered in the paper is a fixed effect one, akin to the one used for the simulation.

And for such a model, the attenuation bias should be given by the equation:

(1.15) $$\frac{plim\,\widehat{\beta} - \beta}{\beta} = (1-\tau)\frac{\bar{p}\,(1-\bar{p})/\bar{n}}{(1-R^2)\sigma_p^2}$$

from which the implication that the difference between the regressor and the unbiased regressor must be always smaller than that of the “purged variance” of the model (Aydemir A., Borjas J. 2011):

(1.16) $$u_k = p_k - \pi_k < (1 - R^2)\sigma_p^2$$

Based on this conclusions, and considering the elements of Eq. (1.15), the attenuation bias that is expected to be found in the simulations will be proportional to: the sampling rate τ; the average cell size $\bar{n}$; the explained variance of the model used, and the variance of the affected regressor.

CHAPTER 2 – SIMULATION RESULTS

The following sections will report the results of the fixed effect models used, before focusing on the tests on the attenuation bias intensity itself, and then will move on to explore the various changes in bias caused by different characteristics of the population.

The total running time of the simulation is of 2304 hours, is comprised of a total of more than 300 million individuals divided upon 2100 different models. The total amount of data generated is of 2,27 TB.

All of the simulation results are expected to be explained by the relations highlighted in Eq. (4), therefore the relationship between the total bias and sampling rates and total population, or cell size, is expected to be negative: bigger cell sizes will lead to smaller biases; a larger purged variance should lead to smaller biases in the coefficient as well.

The goal for this second part will be to track the behaviour of the bias across different populations and different models, and to interpret it as shown in the Eq. (4) grouping them by the type of effect that they have on the bias: a change in bias' variance or a change in the numbers of cell.

2.1 Cell size and sampling rate

The purpose of this section is to examine the behaviour of the sampling rate, taking into consideration the effect of immigration on crime rate using the models described in the first part and to find an ideal cell size range to minimize the bias.

The baseline model from the reference paper will be replicated and each type of population will be catalogued. As previously explained the phenomenon of reference used to build the simulation is the effect of immigration on crime rate. The peculiarity of the framework implemented will then allow for a further manipulation of the population itself: by changing the population parameters in terms of initial size, initial proportion of immigrants, immigration waves, strength of fixed effects and length of time, several "population models" will be created. The goal is to track the behaviour of the bias as these changes occur in the population, and check on the performance of each model's estimator.

By taking into account the normalization of the coefficient at once, since the normalization was computed on average an effect of immigration on crime of 20%, it means that a 50% bias, or a coefficient of 0.5 for a beta, will register an effect of immigration on crime half as strong than it actually is.

The changes performed upon the key parameters in this section will solely affect the cell size indirectly by modifying the number of regions in the simulation and the number of total individual in the population model:

2.2 Baseline Population

For comparison we implemented the FE models to replicate the result from our reference paper (Nunziata, 2015), where the unbiased coefficient was normalized at one and in which the characteristic of the population were as follows:

BASELINE POPULATION SETUP	
Initial pop. count: 10'000'000	Regions: 100
Immigration Shocks Variance: 2% of total population	Average proportion of immigrants: 0,10
Time length: 4	Dependent variable error term distr. : $N \sim$(0;0,5)

The baseline population parameters are a good approximation of actual parameters for a country in terms of average proportion of immigrants and immigration waves. For this baseline setup the modification in population size, number of region and sampling rates covered a good range of cell sizes for both models:

TABLE 2.1 – POPULATION LIST BY PARAMETERS AND CELL SIZE RANGES COVERED

Key parameters	Parameter value	Cell range FE model	Cell range SSIV model Instrument variable	Cell range SSIV model Endogenous variable
Initial Population size	100'000 500'000; 1'000'000; 10'000'000;	(3, 209) (5, 522) (10, 1043) (10, 31299)	(3, 209) (5, 522) (10,1043) (104,10433)	(1, 14) (5, 52) (14, 104) (104, 1043)
Regions	50; 100	(209, 20866) (104, 10433)	(209, 20866) (104,10433)	(209, 280) (104, 1043)
SSIV Endogenous variable sampling rate (for immigration waves variance = 4%)			(104, 10433)	(10, 1020)

All the modification for the baseline population do not alter the unexplained variance of the regressor in the denominator of Eq. (1.15) and therefore can be considered as a starting point to track the bias given fixed conditions.

2.2.1 Baseline FE model

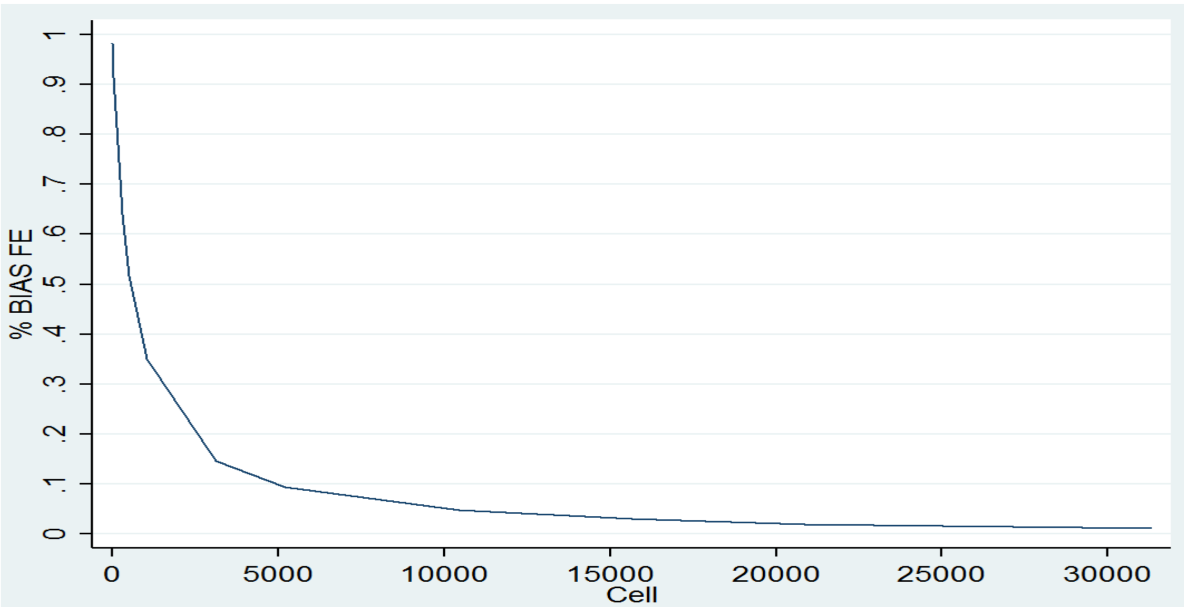
TABLE 2.2 – BASELINE POPULATION – FE MODEL

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20	30
$\widehat{\beta}^{FE}_{popul}$	.997	.998	.996	1	1	1	1	1	1	1	1	1
s.e.	.0721	.0322	.0228	.0131	.0102	.00719	.00415	.00322	.00228	.00186	.00161	.00131
$\widehat{\beta}^{FE}_{sample}$	.0184	.0862	.153	.36	.483	.65	.854	.905	.952	.97	.981	.989
s.e.	.00995	.00955	.00916	.008	.00713	.00585	.00386	.00306	.00222	.00183	.00159	.00131
Obs Cell	10.4	52.2	104	313	522	1043	3130	5216	10433	15649	20866	31299
s.e.	.53	2.14	4.28	12.8	21.4	42.8	128	214	428	642	856	1283

Table 1 refers to the fixed effect model for the baseline population: the first row reports the coefficient for the population model, as expected the value is one across the whole range; a minor divergence happens for very small cell sizes.

The beta "sample" coefficient refers to coefficient from the fixed effect model. As shown in the paper, the FE model results biased for low levels of sampling rate and as average the cell size and the sampling rate increase, the bias decreases. The key parameter in tracking the performance of the bias is the observation per Cell: the estimator is performing very poorly until the cell size hits 3000 observations. In other words, below 1000 the attenuation bias is so severe that the effect of immigration on crime rate is halved when considering a FE model.

GRAPH 2.1 - BIAS VS CELL SIZE FOR BASELINE FE



Graph # shows how the bias behaves across the considered cell range: for that population the bias becomes acceptable only when cell size is above 5000 cells per region. Considering the one hundred region used, the total amount of observation needed would be of half a million. The cell length for the fixed effect also features a standard error but since it is below 2% on average, it is to be considered solely for high values of the interval. It would be equal to 200 around a cell size of 5000.

An interesting fact is the relationship between bias and cell size which is exponential in the range studied: 90% of the bias is compressed in the first quarter of the interval. Also the marginal effect of adding a cell will become less and less meaningful after the first thousand cells.

2.2.2 Baseline FE vs SSIV model

TABLE 2.3 – BASELINE POPULATION – SSIV MODEL

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	.998	1	.999	.999	1	1
s.e.	.0198	.0114	.00886	.00626	.00362	.0028	.00198
$\widehat{\beta}^{FE}_{sample}$	.156	.354	.484	.651	.852	.908	.953
s.e.	.00801	.00693	.00628	.0051	.00335	.00268	.00194
$\widehat{\beta}^{IV}_{sample\ \ 0.1}$	.755	1.1	1.06	1.02	1.02	1.03	1.05
s.e.	.294	.0257	.0163	.00932	.00468	.00356	.0025
$\widehat{\beta}^{IVFS}_{sample\ \ 0.1}$	.154	.341	.482	.658	.849	.901	.936
s.e.	.00479	.00416	.00374	.00299	.00196	.00154	.0011
$\widehat{\beta}^{IV}_{sample\ \ 0.3}$	.981	.997	1.02	1.01	1.02	.993	1.01
s.e.	.0527	.0203	.0138	.00829	.0042	.00309	.00217
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.166	.361	.48	.65	.847	.921	.949
s.e.	.00313	.00261	.00232	.00182	.00115	.000914	.000648
$\widehat{\beta}^{IV}_{sample\ \ 0.5}$	1.06	.984	1.02	1.02	1.01	1.03	1.01
s.e.	.0553	.0196	.0134	.00821	.00411	.00313	.00211
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.151	.364	.481	.641	.844	.889	.95
s.e.	.00266	.0022	.00193	.00149	.000915	.000712	.000503
$\widehat{\beta}^{IV}_{sample\ \ 1}$	.997	1.01	1	.999	1	1	.998
s.e.	.0513	.0198	.013	.00788	.00399	.003	.00206
$\widehat{\beta}^{IVFS}_{sample\ \ 1}$	.158	.351	.485	.654	.853	.907	.957
s.e.	.00225	.00182	.00157	.00116	.000673	.000521	.00036
Obs Cell	104	313	522	1043	3130	5216	10433

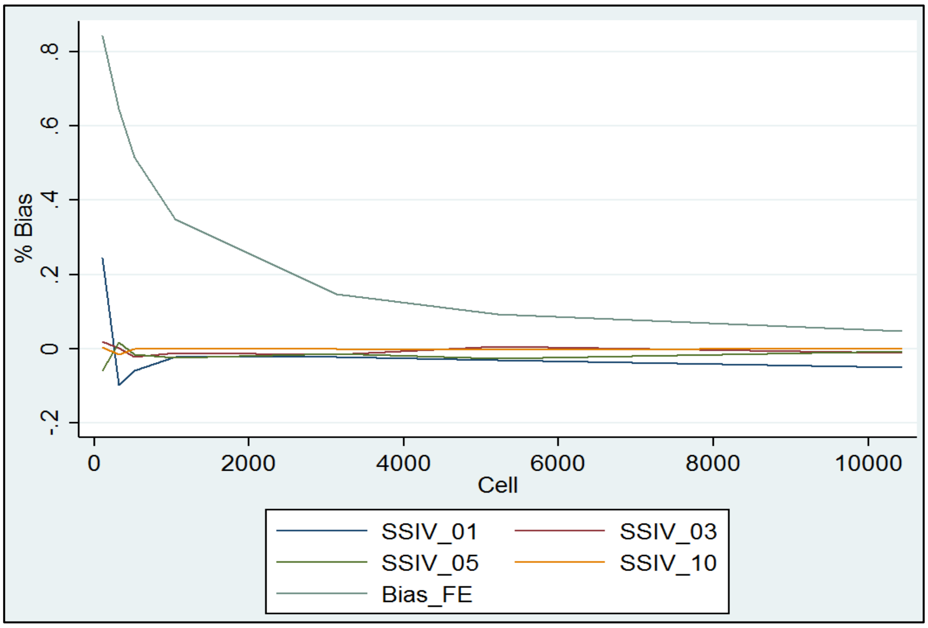
NB: Cell sizes value in the "Obs Cell" row apply to endogenous variable as well

In the SSIV model table the sampling rates of the instrumental variable are in the first row, while the first column reports the sampling rate for the endogenous regressor.

The 2nd stage coefficients are unbiased, even for smaller levels of sampling rate in the instrument.

A problem however lies in the coefficient for the first stage, which, affects the quality of the second stage estimates: the size of the standard error is smaller whenever the first stage estimates are higher. Raising the sampling rate of the endogenous variable does not increase the quality of the first stage, which raises only after an increase in the sampling rate of the instrument.

GRAPH 2.2 – BASELINE COMPARISON FE VS SSIV



In terms of cell size to bias, the SSIV is outperforming the FE model: the bias at one hundred cell per region is around 25% for the former and 85% for the latter. The comparison becomes meaningless as the sampling rate of the endogenous variable raises since the SSIV results almost unbiased.

Even if the gap in terms of performance between the two model is significant, as explained in chapter 1, the SSIV one needs to rely on two independent sample of the same measure, and therefore requires two different surveys. In the simulation the two surveys were independent as per construction, in real circumstances the simple fact that the surveys would come from different sources should be enough to guarantee their non-correlation.

2.3 Properties of SSIV

The peculiarity of SSIV consist in having two independent sample of the same endogenous variable. One sample is to be used as an instrument, generally the smaller one, and the other is to be used as a measure of the endogenous variable, as depicted in equation #.

Due to this this property, it is possible to swap instrument and endogenous variable in the model to achieve the same estimates. The only requirement for the two samples is simply to be independent of each other, and swapping them across the model doesn't affect their non-correlation in any way.

An application of such properties is presented in Table 3#, which is estimated with the same parameters as Table 2, and should, at least in principle, provide the same estimations.

TABLE 2.4 – SSIV INSTRUMENT AND ENDOGENOUS VARIABLE SWAP

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	.998	1	.999	.999	1	1
s.e.	.0198	.0114	.00886	.00626	.00362	.0028	.00198
$\widehat{\beta}^{FE}_{sample}$	.155	.361	.487	.654	.855	.914	.954
s.e.	.00806	.00697	.00632	.00511	.00336	.0027	.00194
$\widehat{\beta}^{IV}_{sample\ \ 0.1}$	1.27	.988	1.06	.964	1.01	1.01	1
s.e.	.134	.0309	.025	.0156	.00941	.00753	.00516
$\widehat{\beta}^{IVFS}_{sample\ \ 0.1}$	.16	.162	.154	.166	.155	.151	.155
s.e.	.00485	.00183	.0012	.000721	.000355	.000266	.000184
$\widehat{\beta}^{IV}_{sample\ \ 0.3}$	1.12	1.05	1.03	.98	1	1	.998
s.e.	.047	.0212	.0159	.0105	.00607	.00476	.0033
$\widehat{\beta}^{IVFS}_{sample\ \ 0.3}$	.347	.35	.351	.366	.361	.354	.362
s.e.	.00723	.00263	.0017	.001	.000481	.000353	.000243
$\widehat{\beta}^{IV}_{sample\ \ 0.5}$	1.04	.995	1.02	.997	1.01	1	.998
s.e.	.0354	.0173	.0135	.00916	.00528	.00405	.00282
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.487	.493	.484	.485	.477	.485	.49
s.e.	.00828	.00299	.00193	.0011	.000513	.000374	.000253
$\widehat{\beta}^{IV}_{sample\ \ 1}$	1.03	.995	1.01	.992	1	1.01	.999
s.e.	.0295	.0149	.0115	.00781	.0045	.00354	.00244
$\widehat{\beta}^{IVFS}_{sample\ \ 1}$	.648	.657	.645	.66	.652	.641	.656
s.e.	.00939	.00332	.00208	.00117	.000515	.000375	.000246
Obs Cell	104	313	522	1043	3130	5216	10433

However, there is a slight divergence in Table 3, even though the estimates are roughly the same, their values differ by a small degree of precision.

For example, the beta estimates for the sampling rate 100/100 are different across the two tables, but the % of bias is roughly around 30%. The differences between the two tables become smaller and smaller as we raise the cell size in both variables. Each increase in sampling rate, from 0.5 to 1, raises the precision from a decimal to a centesimal part.

In principle, if the precision of the simulation could be infinitely high, the estimates would be exactly the same. They get more and more precise as the number of repetition in the Monte Carlo simulation raises, or whenever the bias gets smaller and smaller.

This property can be useful to evaluate the consistency of the estimates across the various simulation because cell sizes should be diagonally mirrored on a table. For example, the pair 100 for the endogenous, 300 for the instrument and 300 for the instrument and 100 for the endogenous should have the same estimate or at least the same amount of bias. The only downside would be the time invested in actually estimating the same cell twice. Ultimately, by arranging the data diagonally and checking for the mirrored cells it would be possible to check the precision of the estimation: consistent cell ranges should be very close in terms of bias.

2.3 Optimal cell size

The criterion to select an optimal cell size would be to choose a cell size that:

reduces the bias by 90% (a pair of cell sizes in the case of SSIV)

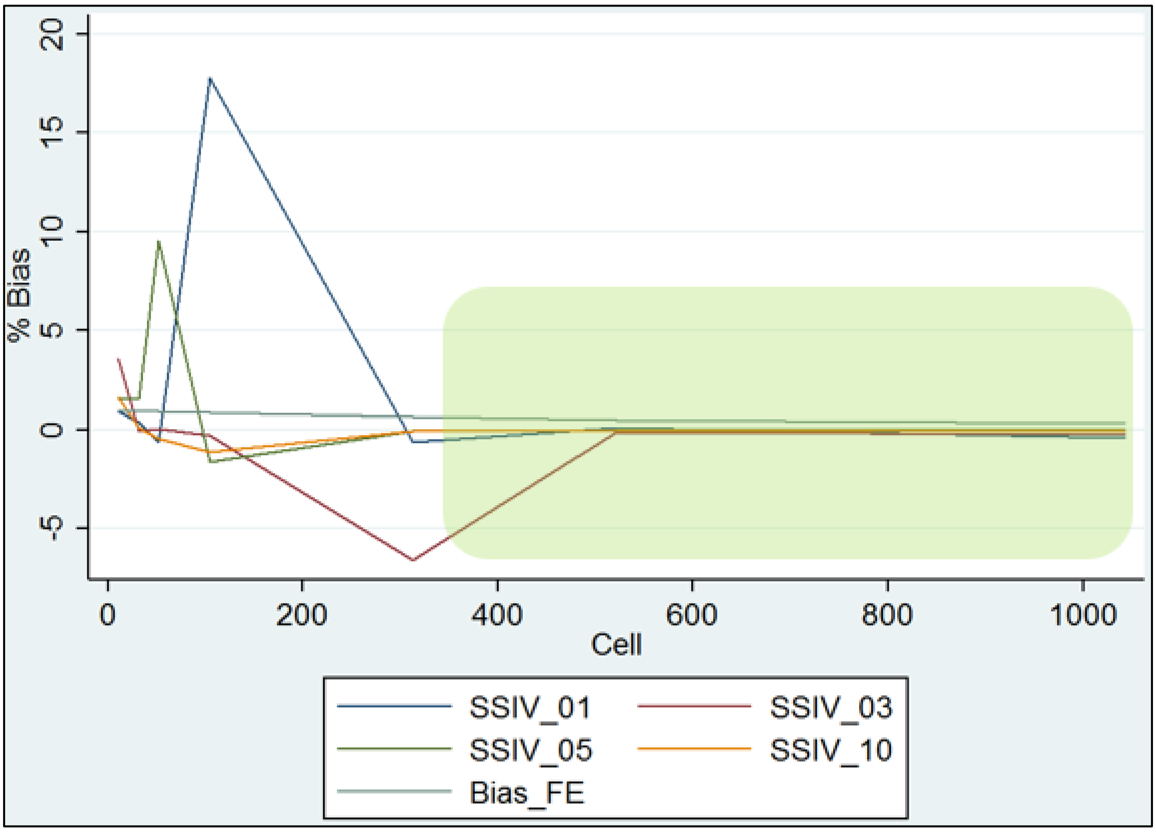
is consistent, namely that any higher value for the cell size has at least the same reduction

has a t-statistic lower than 2

The 90% reduction in bias is a good compromise for the best approximation attainable with the instrument at our disposal: more precision would require both longer computational times and bigger datasets. Even if the simulation as of now is capable to accurately estimate biases of less than 0.1% for larger cell sizes, the precision goes down as the cell size becomes smaller. For FE it is able to accurately predict up to a 1% reduction in bias due to the fact that the bias is still extremely high even for very large cell sizes. While for the SSIV the bias is actually sizable only for smaller cell sizes.

Moreover, in the case of SSIV there are two different cell size to select: one for the endogenous and one for the instrumental variable, the optimal pair would be the one that reduces the bias by 90%, and above such range the bias should be consistently lower, while all the estimates should be statistically significant. The endogenous/instrumental cell size pair could be interchanged due to the properties of SSIV. Since SSIV is extremely good for higher level of cell size, to find the optimal value it is necessary to explore the lower bounds of the interval.

GRAPH 2.3 – LOWER CELL SIZE RANGES SSIV



NB: Each line represents an Endogenous variable sampling rate

Graph # represents lower cell values for the SSIV, each line corresponds to a specific endogenous variable sampling rate from 0.1 to 1, in terms of cell size they range from 100 to 1000; the cell size of the instrument is on the x axis. The green area highlights a range in which the coefficients are more consistent: the bias is below 90% and the estimates are statistically significant. It corresponds roughly to the interval highlighted in table #, for which the coefficients are more consistent and do not assume extremely biased value for smaller cell sizes. The only endogenous sampling rate that remains consistent for lower levels is the one for a cell size of one thousand

Table 3 itself reports a lower range of sampling rates for the endogenous variable in the SSIV model: from 0.0001 to 0.01. The red section highlights the optimal range value which meet the required criteria.

TABLE 2.5 – BASELINE SSIV – LOWER CELL SIZE RANGE

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1.02	.997	.999	1	.996	.999	1
s.e.	.0629	.0362	.028	.0198	.0114	.00887	.00627
$\widehat{\beta}^{FE}_{sample}$	.0161	.0538	.0864	.156	.361	.495	.677
s.e.	.00876	.00851	.00833	.00816	.00711	.00635	.00523
$\widehat{\beta}^{IV}_{sample\ 0.1}$	.0737	.634	1.63	16.8	1.62	.89	1.39
s.e.	.832	7.96	8.23	2465	.251	.182	.0681
$\widehat{\beta}^{IVFS}_{sample\ 0.1}$	.0122	.0297	.0937	.137	.366	.499	.647
s.e.	.0158	.0152	.0149	.0145	.0129	.0115	.00936
$\widehat{\beta}^{IV}_{sample\ 0.3}$	-2.58	1.07	.948	1.27	7.59	1.14	1.22
s.e.	106	2.13	2.44	.262	32.6	.0309	.0178
$\widehat{\beta}^{IVFS}_{sample\ 0.3}$	.0164	.0481	.0783	.163	.328	.531	.638
s.e.	.0092	.00883	.00866	.0085	.00736	.00663	.00546
$\widehat{\beta}^{IV}_{sample\ 0.5}$	-.524	-.547	-8.52	2.62	1.04	.986	1.03
s.e.	14	4.76	513	3.78	.0281	.0163	.0106
$\widehat{\beta}^{IVFS}_{sample\ 0.5}$	.0239	.0473	.0739	.154	.376	.518	.702
s.e.	.00722	.00698	.00681	.00673	.00583	.00514	.0042
$\widehat{\beta}^{IV}_{sample\ 1}$	-.665	.893	1.46	2.1	1.07	1.06	1.08
s.e.	9.03	.585	.537	1.71	.0253	.016	.00986
$\widehat{\beta}^{IVFS}_{sample\ 1}$	.0157	.0463	.0794	.154	.361	.489	.648
s.e.	.00533	.00513	.005	.0049	.00418	.00375	.00306
Obs Cell	10.4	31.3	52.2	104	313	522	1043

NB: Cell sizes value in the "Obs Cell" row apply to endogenous variable as well

The values of the coefficients become less and less biased as the cell size in both variable increases. The optimal cell range seems to be located above a value of 300 for the independent and 50 for the endogenous variable.

Below 100 the coefficients are extremely unreliable in some cases: for the cell 100 and 10 the value of the coefficient has a 1600% bias; for the cell 10 and 30 instead, even if the coefficient itself has a low bias, the standard error is exceedingly large. The former doesn't meet the required criteria directly; the latter doesn't meet it in due to the inconsistency of its standard error.

TABLE 2.6 – BASELINE SSIV – COEFFICIENT BIAS BY EXTENDED CELL SIZE RANGE

Endogenous var cell size / Instrument cell size: Cell size	10	30	50	100	150	200	300	500	1000
10	92	36	- 63	177	86	7**	-62	10	-39
30	358	-7	5	-26	30	32	28	-13	-22
50	152	154	952	-162	-17	1	-4	1	-2
100	166	10	-45	25*	25	1	-7	-5	-8
150	45	95	1	13	17	10	2	-3	-3
200	40	-659	4	7	1	-1	1	-2	-2
300	3**	35	-3	2	-2	2	2	-2	-1
500	100	28	-3	6	-3	1	2	-2	-1
1000	150	1	6	1	1	1	1	0	1

**Inconsistent range*

***T-stat < 2*

NB: Optimal cell sizes are highlighted in green. Orange cell sizes are borderline values

Table # reports a much wider range for SSIV model, in which it is possible to identify the optimal range. The estimated coefficients for the second stage of SSIV can be classified in three categories: highly biased, borderline and optimal range. The ones in the first two columns and rows belong to the lowest ranges of cell sizes, the values are generally extremely biased or have a wide standard error. Some values have less bias but belong to an inconsistent range.

An explanation to justify the inconsistent behaviour of low range estimates can be found in how the Monte Carlo simulation operates. Picking up random samples in a larger population, that is by construction normally distributed, there is a chance that the sampling upon which that specific coefficient has been computed have very few extreme values and that a very small sample resembles the population distribution more than a bigger one. The problem is ultimately tied to the lack of precision due to the limited number of repeated sampling done during the

Monte Carlo simulation. A higher number of repeated sampling operation should highlight how all the values in the lower cell ranges from ten to fifty are actually biased.

A second tier of estimates is the borderline one: between 150 and 200 for both variables. This group features a bias below 15% on average, is fairly consistent in terms of bias reduction by increase in cell size but it doesn't show a consistent diagonal symmetry: same cell values repeated can yield different estimates. For example, the cell 200/100 and 100/200 diverge by a 6%; the cells 200/150 and 150/200 instead diverge by 9%.

The estimates tend to become more consistent around the diagonal, for a cell size of 200/200. This group comprehends the ranges 50/300-1000 and 300-1000/50. Even if the values are fairly consistent for these cell size and the bias is low, the very short sample of 50 in one the two variable, especially as seen in the 50/1000 cell, may be biased for higher level of precision of the simulation.

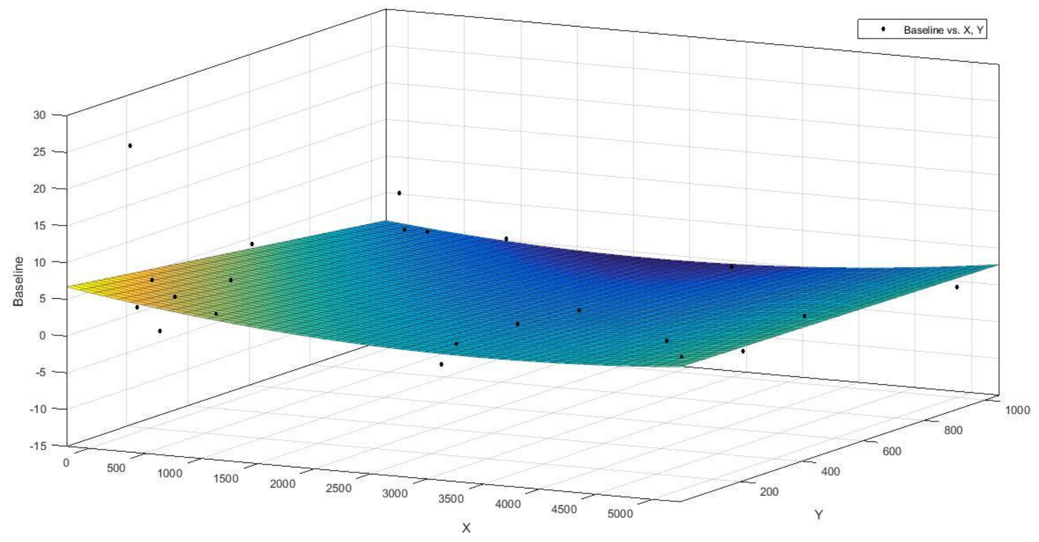
The last group above the 300/100 range, meets all the criteria for the optimal cell range: the bias reduction is way above 90-95%, it's consistent in terms of bias reduction by cell size increase, doesn't feature abnormal values, and it's mirrored up to a decimal precision. The biggest divergence, around 5%, is located around the limits of the cell range.

To measure the improvement of the alternate modification, the set of coordinates (bias, endogenous variable cell size, instrument cell size) has been fitted to a non-linear regression in table 2.7. The volume below the curved plane and inside the boundaries (0.5000) for X and (0.1000) for Y, will be used to compare the increase or decrease in bias w.r.t. the alternative populations.

TABLE 2.7 – SSIV BASELINE POPULATION –

BIAS VS ENDOGENOUS/ INSTRUMENT CELL SIZE

NON-LINEAR REGRESSION



$$f(x,y) = \alpha + \gamma_1 X + \gamma_2 Y + \gamma_3 X^2 + \gamma_4 XY$$

Coefficients (with 95% confidence bounds):

α = 2.06 (-19.33, 23.45)

γ_1 = -1.85 (-25.9, 22.2)

γ_2 = -1.349 (-13.49, 10.79)

γ_3 = 1.087 (-17.17, 19.34)

γ_4 = 0.5323 (-11.79, 12.85)

Goodness of fit:

SSE: 3.605e+04

R-square: 0.6678

CHAPTER 3 – ALTERNATIVE POPULATIONS

In this third section we'll use the methodology developed in chapter two to study how the attenuation bias behaves whenever the condition in the initial population change.

In this way, it would be possible not only to select an optimal cell size to build a fixed effect model of migration, but also to adjust the cell size to make it fit the population's condition more realistically.

Moreover, it would be interesting to understand which factor do affect the attenuation bias and how strong such effect is.

Table # reports the full list of all the changes and modifications done to the initial population, in order to track the bias's behaviour.

TABLE 3.1 – ALTERNATIVE POPULATION LIST BY PARAMETERS AND CELL SIZE RANGES

Key parameters	**Parameter value**	**Cell range FE model**	**Cell range SSIV model** Instrument variable	**Cell range SSIV model** Endogenous variable
Initial average immigrant share (as % of total population)	10%; 30%; 50%;	(10, 31299) (113, 5649) (122, 6082)	(104, 10433) (113, 5649) (122, 6082)	(104, 1043) (113,1130) (122, 1216)
Immigration waves variance (as % of tot increase in final population)	2%; 4%; 10%; 27%;	(10, 30599) (10, 31299) (11, 33062) (127, 12679)	(102, 10200) (104, 10433) (110, 11009) (127, 12679)	(102, 1020) (104, 1043) (110, 1101) (127,1268)
Panel data lenght (unit of time = 2 years)	2 4 8	(10, 20472) (10, 31299) (15, 30355)	(102, 10236) (104, 10433) (152, 15178)	(102, 1024) (104, 1043) (152, 1518)
SSIV Endogenous variable sampling rate (for immigration waves variance = 10%)			(104, 10433)	(10, 1043)
Smaller variance in the error term for dependent variable		(10, 31299)	(104, 10433)	(104, 1043)
Baseline parameters value				

The key parameters are chosen to simulate actual circumstances. The modifications will affect: initial average proportion of immigrants, the variability of immigration's shocks across time, and the length of time for the model.

The average initial proportion of immigrants was set at 10%. Which is around the average for western European countries. Whereas values such as 5% would represent the situation in eastern European countries. Bigger values, up to 30% can be used to replicate the condition of countries with a strong population of immigrants such as Australia and New Zealand (United Nations database, 2015).

The second parameter is immigration waves. It's measured by how much it affects the total increase in the final population after the observation period. For example, a 4%, which is the baseline parameter, would indicate an increase in population due to immigration of 0.5% per each year. As an upper limit, a 10% increase has been chosen, which corresponds to a 1.25% increase per year, and 2% as a lower limit, which corresponds to a 0.25% yearly increase in total population due to immigration.

For comparison, Germany's yearly immigration causes approximately an increase of 1% in the population, while the Italian immigration causes an increase of approximately 0.33% (Hawkins O, 2017). That'll make so that the variant with the lower bound parameter would be better to describe the bias on fixed effect models that use Italian data, meanwhile the upper bound variant would be better suited to track the bias of fixed effect models on German data.

The time length modification is performed to check the consistency of the time frame chosen and the effect of its alteration upon the bias. It is to be noted that the time unit used in the simulation corresponds to two years, since it is measured upon the bi-yearly release of Labour Force Survey data.

The following sections will then cover a brief explanation on the expectation of such modification, with a reminder of the theoretical basis; an in-depth look to the effects of the modification on each model, with the identification of optimal cell sizes for each modification. In addition, to compare all the model across the different modification, each and every model has been fit with a non-linear regression. The functions used to fit the models are:

(3.1) $$f(x) = \alpha_1 e^{\gamma_1 x} + \alpha_2 e^{\gamma_2 x}$$

For FE, where the dependent variable is the bias and x is the cell size.

(3.2) $$f(X,Y) = \alpha + \gamma_1 X + \gamma_2 Y + \gamma_3 X^2 + \gamma_4 XY$$

For SSIV, where the dependent variable is the bias as well and X and Y are the cell sizes used for the instrument and the endogenous variable. The same framework has been applied to all models.

3.1 Unexplained variance

The modifications in this section will not only affect the key parameter of cell size in Eq. (7), they will also have an effect on the denominator in the equation that defines the bias.

That is due to the reduction of unexplained variance in the regressor, or, in other words, to how migration becomes "better" at explaining the variance in crime rate. A stronger correlation between the regressor and the dependent variable leads to a smaller denominator for Eq. (7) and in return it means a smaller bias.

To test the relation directly in the simulation, the error term in the construction of the simulated dependent variable has been changed.

When the "crime rate" variable is simulated in the program, a normally distributed error with zero mean and 0.5 variance is included to its variability. By replacing that error with a smaller one that is lognormally distributed, and only has 0.1 variance, the explained variance of the regressor on the dependent variable should become bigger. It should be a consequence of the reduced "random" part of the crime rate variable.

The hypothesis is tested on both FE and SSIV models. The results for SSIV are on graph #, while table # displays the full range of results for both models:

GRAPH 3.1 – SSIV COEFFICIENT VALUES BY ERROR TERM

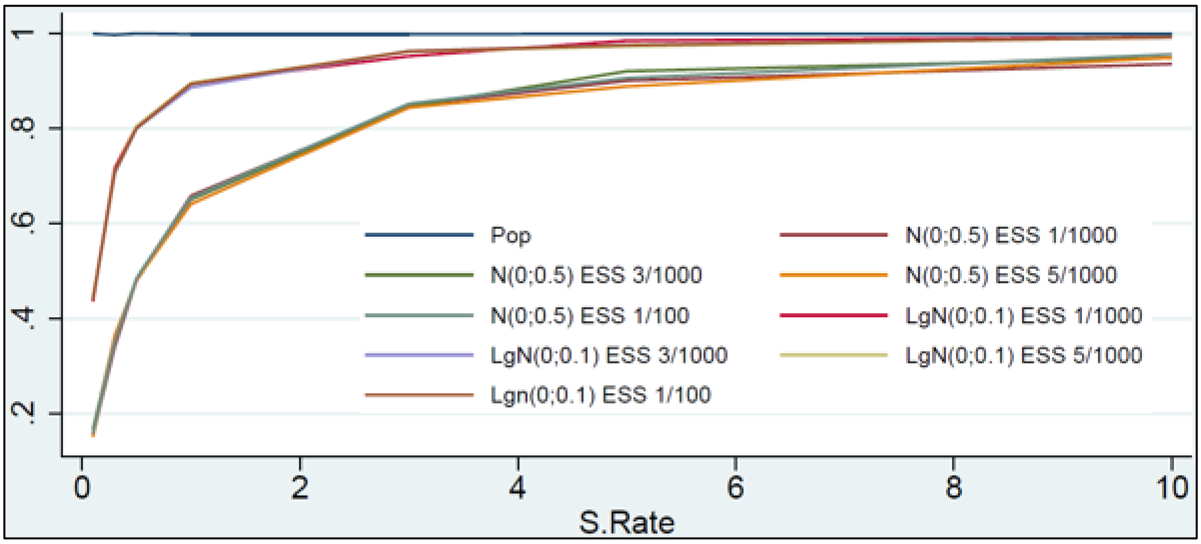


TABLE 3.2 – REDUCED ERROR TERM IN THE REGRESSAND

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	.999	1	1	1	1	1	1
s.e.	.00879	.00508	.00393	.00278	.00161	.00124	.00088
$\widehat{\beta}^{FE}_{sample}$	.439	.706	.804	.892	.962	.976	.989
s.e.	.00633	.00447	.00363	.00267	.00159	.00123	.000877
$\widehat{\beta}^{IV}_{sample\ \ 0.1}$	1.02	.993	1.01	1	1.02	.995	1
s.e.	.0161	.00706	.00519	.00347	.00196	.00147	.00104
$\widehat{\beta}^{IVFS}_{sample\ \ 0.1}$	.436	.717	.805	.895	.953	.986	.993
s.e.	.00426	.0029	.0023	.00167	.000969	.000755	.000535
$\widehat{\beta}^{IV}_{sample\ \ 0.3}$	1	1	1.01	1.01	1	1	.998
s.e.	.0142	.00646	.00467	.00315	.00174	.00134	.000936
$\widehat{\beta}^{IVFS}_{sample\ \ 0.3}$	.442	.705	.801	.887	.964	.976	.993
s.e.	.00314	.00196	.00149	.00104	.000583	.000444	.00031
$\widehat{\beta}^{IV}_{sample\ \ 0.5}$	.994	.998	1	.996	1	1	.997
s.e.	.0138	.00627	.00455	.00304	.0017	.00131	.000915
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.442	.708	.805	.897	.963	.973	.993
s.e.	.00287	.00172	.00128	.000852	.000461	.000352	.000244
$\widehat{\beta}^{IV}_{sample\ \ 1}$	.997	.996	1	.998	.999	1	.997
s.e.	.0136	.00615	.00449	.003	.00167	.00128	.000898
$\widehat{\beta}^{IVFS}_{sample\ \ 1}$	.441	.71	.801	.894	.963	.976	.993
s.e.	.00265	.00152	.00109	.000691	.000351	.000255	.000175
Obs Cell	110	330	551	1101	3303	5505	11009
s.e.	10.9	32.7	54.4	109	327	545	1089
N.of Obs.	400	400	400	400	400	400	400

In graph #, the estimated SSIV coefficients reported on table # are compared together with the ones on table #. The sampling rate, and consequently the cell size used for both endogenous variable and the instrument is the same in both groups.

The resulting graph shows how the estimated coefficient for the smaller error term, the one distributed as a lognormal with zero mean and 0.1 variance are closer to the population coefficient, for each level of sampling rate, meanwhile the coefficient simulated with the larger error are way lower and therefore way more biased.

The same results are to be expected from any modification that would reduce the unexplained variance of the regressor if the cell size remains unchanged.

3.2 Initial proportion of immigrants

As introduced before, the matter of reducing "the unexplained variability" in the regressor, will be covered in this section by simulating scenarios that will, ultimately, change the regressor's explanatory power. Each modification will be runned by both models

The "initial proportion of immigrants" represents the percentage of immigrants already included in the population before the immigration waves.

The different scenarios are compared in the table below:

TABLE 3.3 – FE BIAS W.R.T. CHANGE IN INITIAL PROPORTION OF IMMIGRANTS

Modified Populations	Cell size							
	Cell size	100	300	500	1000	3000	5000	Overall Improvement (in % w.r.t baseline bias)
	0.05	87	72	61	42	19	13	-33%
	0.1	84	64	51	34	14	09	-
	0.3	75	50	38	23	9	5	+ 46%
	0.5	77	52	39	24	9	6	+ 44%

The green represents the region in which the bias is reduced by 90 percentage points. The proportion of immigrants is listed in the column.

As the proportion is the reduced the bias becomes higher for the same cell size, meanwhile a bigger initial proportion of immigrants causes the bias to be reduced for the same cell size range. The ideal range in which the optimal criteria are met are way above 5000, exactly at 6250, for the lowest range of initial proportion of immigrants, 5000 for 0.1, and down to 3000 for 0.3. The graph 3.2 shows a direct comparison.

GRAPH 3.2 – FE BIAS VS CELL SIZE BY INITIAL PROPORTION OF IMMIGRANTS

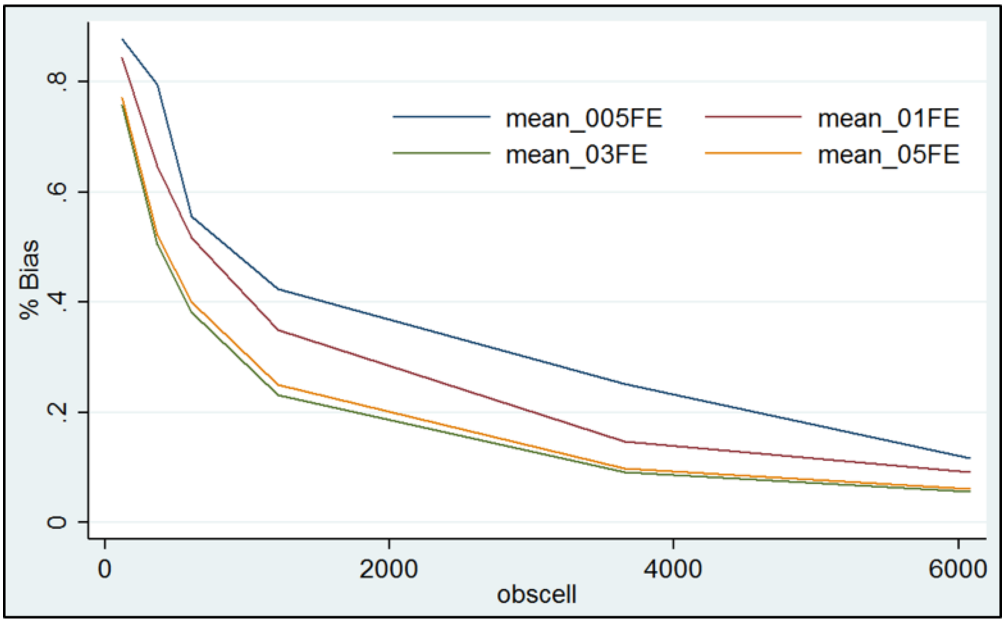


TABLE 3.4 –FE BIAS VS CELL SIZE BY INITIAL PROPORTION OF IMMIGRANTS

NON-LINEAR REGRESSION MODEL - FITTED VALUES

0.05

Coefficients (with 95% confidence bounds):
a = 0.6882 (0.5572, 0.8192)
b = -0.001337 (-0.001753, -0.000921)
c = 0.2805 (0.1347, 0.4263)
d = -0.000144 (-0.0002496, -3.848e-05)

Goodness of fit:
SSE: 8.307e-05
R-square: 0.9998

$$f(x) = 0.6882e^{-0.001337x} + 0.2805e^{-0.000144x}$$

0.1

Coefficients (with 95% confidence bounds):
a = 0.5971 (0.4156, 0.7787)
b = -0.001855 (-0.002829, -0.0008811)
c = 0.3768 (0.1675, 0.586)
d = -0.0002431 (-0.0003822, -0.000104)

Goodness of fit:
SSE: 0.0001488
R-square: 0.9997

$$f(x) = 0.5971e^{-0.001855x} + 0.3768e^{-0.0002431x}$$

0.3

Coefficients (with 95% confidence bounds):
a = 0.6513 (0.5016, 0.8009)
b = -0.002686 (-0.004088, -0.001283)
c = 0.2978 (0.1229, 0.4727)
d = -0.0003021 (-0.0004957, -0.0001085)

Goodness of fit:
SSE: 0.0002219
R-square: 0.9994

$$f(x) = 0.6513e^{-0.002686x} + 0.2978e^{-0.0003021x}$$

0.5

Coefficients (with 95% confidence bounds):
a = 0.6349 (0.4697, 0.8001)
b = -0.002667 (-0.004236, -0.001097)
c = 0.3226 (0.1293, 0.5159)
d = -0.0003015 (-0.0004981, -0.000105)

Goodness of fit:
SSE: 0.0002668
R-square: 0.9993

$$f(x) = 0.6349e^{-0.002667x} + 0.3226e^{-0.0003015x}$$

The overall improvement in Table 3.3 directly compares the area underneath the baseline's graph, with the areas of the other modifications. The function is obtained by a non-linear regression on the set of coordinates made up by bias in percentage points and cell size, and the area has been computed by calculating the definite integral of the function in the range (0, 5000).

Ultimately, a reduction of 0.05 in the initial proportion of immigrants leads to an increase of 33% in bias with respect to the bias of the baseline model. Instead, when the initial proportion of immigrant is at 0.3 or 0.5, the reduction in bias w.r.t. the baseline model is respectively 46% and 44%.

For SSIV we need to introduce another variable to measure the cell size of the instrument. The bias in percentage point is summed up in the following tables, one for the reduction and the other one for the increment:

TABLE 3.5 - SSIV BIAS W.R.T. INITIAL PROPORTION OF IMMIGRANTS(0.05)

Endogenous var cell size	Instrument cell size							
	Cell size	50	100	300	500	1000	3000	5000
	50	90	-73	-280	-39	-22	-3	-2
	100	60	72	-41	-16	6	-3	-4
	300	-99	91	-24	0	3	0	-1
	500	-25	-31	-9	-7	0	0	3
	1000	-22	-14	-8	-3	0	1	-1

The optimal cell size when the initial proportion of immigrants is reduced to 0.05 percent of the total population is located at 500 cells for the endogenous variable and 300 cells for the instrument. With respect to the baseline population the inconsistent values are more frequent for lower cell size ranges.

TABLE 3.6 - SSIV BIAS W.R.T. INITIAL PROPORTION OF IMMIGRANTS (0.3)

Endogenous var cell size	Instrument cell size						
Cell size	50	100	300	500	1000	3000	5000
50	48	-8	1	-7	-1	1	-1
100	-9	-11	-3	-2	-4	1	-2
300	1	-2	-1	-2	1	1	1
500	-6	0	-1	-1	1	1	0
1000	1	-1	0	1	-1	0	1

Table 3.6 reports of bias in percentage point, in case of an increase in initial proportion of immigrants for both endogenous and instrument in an SSIV model.

The results show how the ideal range is located at 300 cells for the endogenous and 50 for the instrument. The 100/50 range shows a bias lower than 10, but the borderline result at 100/100 is an indicator of low consistency for this range. The green area instead is perfectly consistent in reducing the bias below 10 percentage points.

The same method for comparing the two models' improvement has been applied, this time in three dimensions. The results of the non-linear regression are as reported in table 3.4

Table 3.7 – SSIV Bias vs Cell size by initial proportion of immigrants

Non-linear regression model - Fitted values

$$f(x,y) = \alpha + \gamma_1 X + \gamma_2 Y + \gamma_3 X^2 + \gamma_4 XY$$

Initial proportion of immigrants = 0.05	Initial proportion of immigrants = 0.3
Coefficients (with 95% confidence bounds):	Coefficients (with 95% confidence bounds):
α = 11.68 (-5.926, 29.29)	α = 1.267 (-1.852, 4.387)
γ_1= -24.01 (-43.81, -4.206)	γ_1= -0.4953 (-4.003, 3.012)
γ_2= -9.788 (-19.78, 0.2069)	γ_2= -0.7163 (-2.487, 1.054)
γ_3 = 9.68 (-5.345, 24.71)	γ_3= 0.1794 (-2.482, 2.841)
γ_4 = 5.163 (-4.979, 15.3)	γ_4= 0.3535 (-1.443, 2.15)
Goodness of fit: R-square: 0.7403	Goodness of fit: R-square: 0.6659

The non-linear model has been used to compute the volume below the functions, which would indicate the level of bias, and it has been compared to the volume of the baseline population in table 2.7. The result indicates that the lower level of initial proportion of immigrants causes a 1037% increase in bias w.r.t the baseline population, meanwhile the higher level causes a decrease of 77% in bias.

3.3 Immigration shock's intensity

Immigration shock 's intensity refers to the strength of the immigration waves in the simulation. Their utility consists in showing how the bias fares whenever there's a period of peak immigration. As previously explained they do affect the explained variability of the migration regressor on the dependent variable.

TABLE 3.8 – FE BIAS W.R.T. CHANGE IN IMMIGRATION SHOCK'S INTENSITY

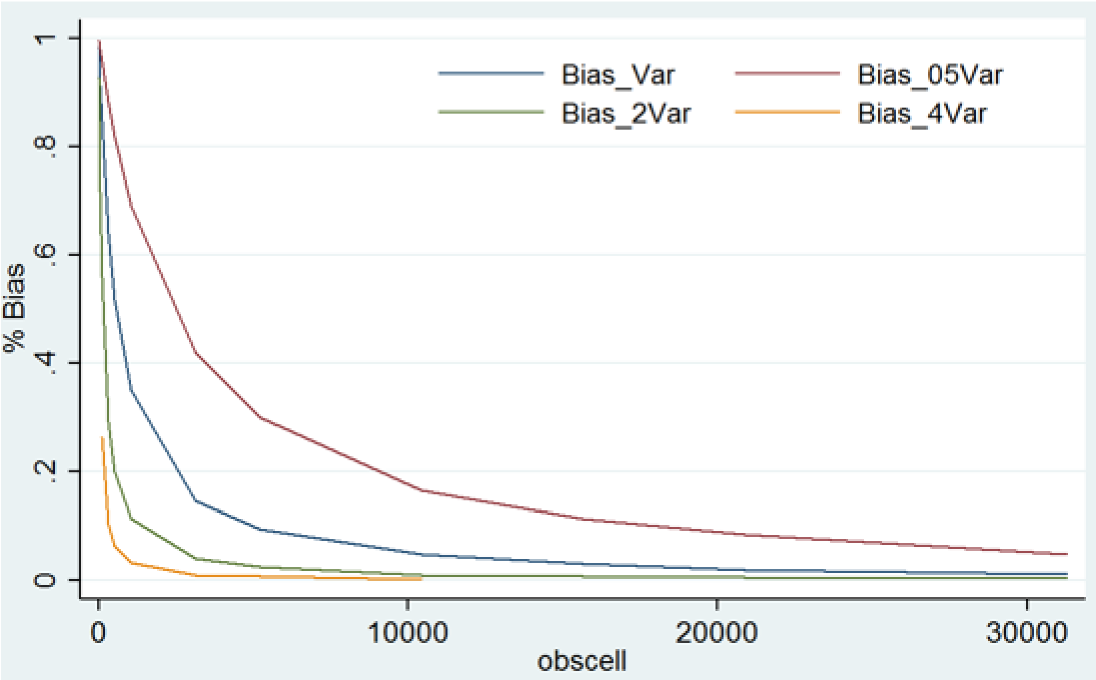
Modified Populations	Cell size								
Cell size	10	30	50	100	300	500	1000	5000	Overall Improvement (in % w.r.t baseline bias)
0.5%	99	98	97	95	89	83	70	30	- 206 %
1%	96	94	91	85	66	53	33	11	-
2.5%	92	82	73	57	28	20	13	0.008	+ 64%
6.75%	43	39	34	26	10	0.06	0.03	0.004	+ 87%

NB: Immigration shock's intensity is measured as percentage of yearly increase in total population

Change in Immigration shock's intensity seems to cause a substantial increase or decrease in bias for FE models. The reduction of the shock down to 0.5% increase per year, causes the optimal cell size to ramp up to 15000 cells. An increase up to 2.5% drops the cell size, required to reduce the bias by almost completely, down to 5000. It is to be noted that at around 1000 cells the bias is already reduced by 85 percentage points.

The biggest reduction is obtained by increasing the yearly shock to a 6.75% yearly: the optimal cell size will be obtained with only 300 cells. Such a result is comparable to the performance of a SSIV model.

GRAPH 3.3 – FE BIAS VS CELL SIZE BY INCREASE IN IMM. SHOCK INTENSITY



The average improvement by change in intensity shock is computed by comparing the area below each graph with the baseline population: the 0.5% reduction causes a 200% increase in bias, while increasing the intensity to 2.5% and 6.75% cause a decrease in bias by 64% and 87% respectively.

Table 3.9 and 3.10 report the result for the SSIV model. The reduction in immigration shock's intensity causes the optimal cell size to rise up to 1000 cell for the endogenous variable and 300 for the instrument.

TABLE 3.9 - SSIV BIAS W.R.T. CHANGE IN IMMIGRATION SHOCK'S INTENSITY (0.05)

Endogenous var cell size	Instrument cell size							
	Cell size	50	100	300	500	1000	3000	5000
	50	76	130	97	98	-48	-307	3
	100	160	54	-15	4	2	-28	-18
	300	99	-89	14	-79	-12	-4	-2
	500	99	-26	-23	-88	-4	-4	-2
	1000	-27	53	-3	-5	2	-1	-3

A raise in immigration shock intensity, up to 2.5% increase in total population per year, causes the optimal cell size to drop down to 50 cells for both variables. This scenario brings the best results so far in terms of reduction in attenuation bias.

TABLE 3.10 - SSIV BIAS W.R.T. CHANGE IN IMMIGRATION SHOCK'S INTENSITY (2.5)

Endogenous var cell size	Instrument cell size						
Cell size	50	100	300	500	1000	3000	5000
50	6	2	1	0	1	-3	-1
100	3	-2	1	-1	0	-2	1
300	1	0	0	-1	-1	0	0
500	4	1	1	-1	1	-1	-1
1000	0	1	1	-1	1	1	-1

Table 3.11 shows the fitted values for non-linear regression on the SSIV models regarding the change in immigration shock. The change on average with respect to the baseline simulation is of an increase of 1214% of the bias, in the case of the weaker shock, and a reduction up to 97% of the bias when the more intense shock occurs.

TABLE 3.11 – SSIV BIAS VS CELL SIZE BY CHANGE IN IMMIGRATION SHOCK'S INTENSITY

NON-LINEAR REGRESSION MODEL - FITTED VALUES

Immigration shock intensity = 0.5%

$$f(x,y) = \alpha + \gamma_1 X + \gamma_2 Y + \gamma_3 X^2 + \gamma_4 XY$$

Coefficients (with 95% confidence bounds):

$\alpha =$ 23.43 (-0.3843, 47.24)
$\gamma_1 =$ -32.58 (-59.35, -5.807)
$\gamma_2 =$ -15.28 (-28.79, -1.76)
$\gamma_3 =$ 11.1 (-9.214, 31.42)
$\gamma_4 =$ 7.539 (-6.174, 21.25)

Goodness of fit: R-square: 0.6673

Immigration shock intensity = 2.5%

$$f(x,y) = \alpha + \gamma_1 X + \gamma_2 Y + \gamma_3 X^2 + \gamma_4 XY$$

Coefficients (with 95% confidence bounds):

$\alpha =$ 0.9291 (-0.997, 2.855)
$\gamma_1 =$ -0.3301 (-2.047, 1.386)
$\gamma_2 =$ -0.7622 (-2.413, 0.8883)
$\gamma_3 =$ -0.08476 (-2.189, 2.02)
$\gamma_4 =$ 0.3422 (-1.185, 1.87)

Goodness of fit: R-square: 0.6933

3.3.1 SSIV First stage quality

The modification regarding shock's intensity has been used as a test to track the behaviour of the first stage in the SSIV model. The literature regarding instrumental variable is often focused on the quality of the first stage, especially on the explanatory power that comes from choosing a good instrument (Bound J. et al., 1995). Even if the SSIV builds the instrument using a different approach, as shown in section 1.2.2, it is still worth to examine the relationship between the first stage estimate and the resulting bias in the coefficient for the second stage.

TABLE 3.12 – SSIV – FIRST STAGE ESTIMATES BY SHOCK'S INTENSITY

S.Rate	0.1	0.3	0.5	1	3	5	10
0.5xVar FSIV ESS 1/1000	.0566	.119	.192	.328	.571	.688	.849
0.5xVar FSIV ESS 3/1000	.0317	.124	.181	.312	.578	.725	.842
0.5xVar FSIV ESS 5/1000	.0398	.115	.181	.307	.582	.714	.837
0.5xVar FSIV ESS 1/100	.0407	.121	.179	.314	.586	.699	.833
1xVar FSIV ESS 1/1000	.154	.341	.482	.658	.849	.901	.936
1xVar FSIV ESS 3/1000	.166	.361	.48	.65	.847	.921	.949
1xVar FSIV ESS 5/1000	.151	.364	.481	.641	.844	.889	.95
1xVar FSIV ESS 1/100	.158	.351	.485	.654	.853	.907	.957
2xVar FSIV ESS 1/1000	.436	.717	.805	.895	.953	.986	.993
2xVar FSIV ESS 3/1000	.442	.705	.801	.887	.964	.976	.993
2xVar FSIV ESS 5/1000	.442	.708	.805	.897	.963	.973	.993
2xVar FSIV ESS 1/100	.441	.71	.801	.894	.963	.976	.993
4xVar FSIV ESS 1/1000	.739	.897	.932	.968	.988	.988	.996
4xVar FSIV ESS 3/1000	.736	.896	.933	.969	.986	.994	.997
4xVar FSIV ESS 5/1000	.734	.896	.938	.972	.995	.989	.999
4xVar FSIV ESS 1/100	.736	.895	.935	.969	.989	.995	.994

Table 3.12 reports the coefficients for the first stage. As expected the first stage estimates approach one as the sampling rate for the instrument increases. Another expected consequence of this is the reduction of the bias in the second stage.

3.4 Panel data length

TABLE 3.13 – FE BIAS BY PANEL DATA LENGTH

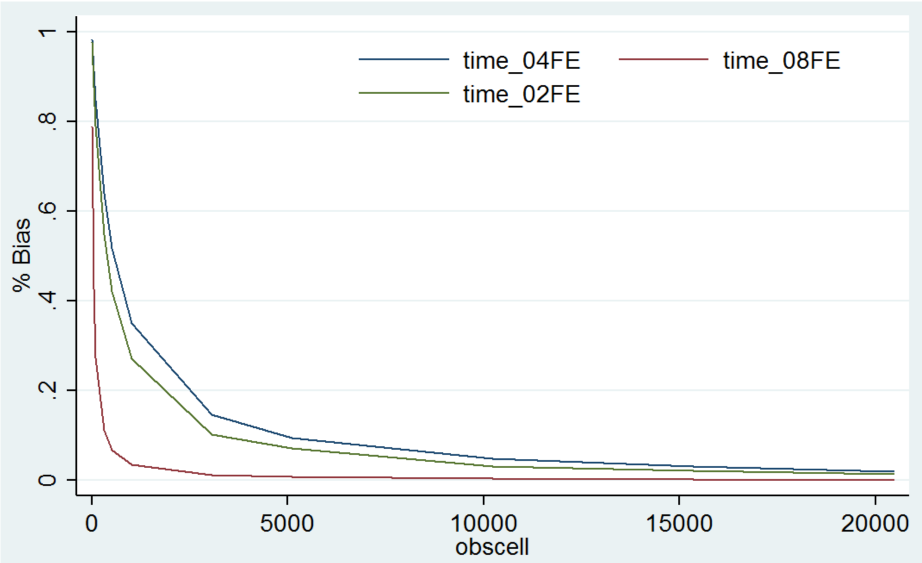
Modified Populations	Cell size										
Cell size	10	50	100	300	500	1024	3000	5000	10000	15000	Overall Improvement (in % w.r.t 1 baseline bias)
T = 2	97	87	79	54	42	27	10	7	3	2	+ 28%
T = 4	98	91	84	63	51	34	14	9	4	3	-
T = 8	78	42	27	11	6	3	1	0	0	0	+ 63%

The last population modification examined regards the change in panel's data length. Table 3.13 reports the bias of FE coefficients by level of panel data length. The lower level and the baseline offer almost the same level of unbiasedness for 5000 observations, within a close 2 percentage point margin. The resulting average improvement (refer to appendix [table] for non-linear regression estimates) is of a 28% reduction in bias with respect to the baseline population. In this case however the 28% value is not a good indication due to the fact that both level have similar estimates. As shown in graph 3.5 the confidence intervals do overlap for several intervals in the range considered.

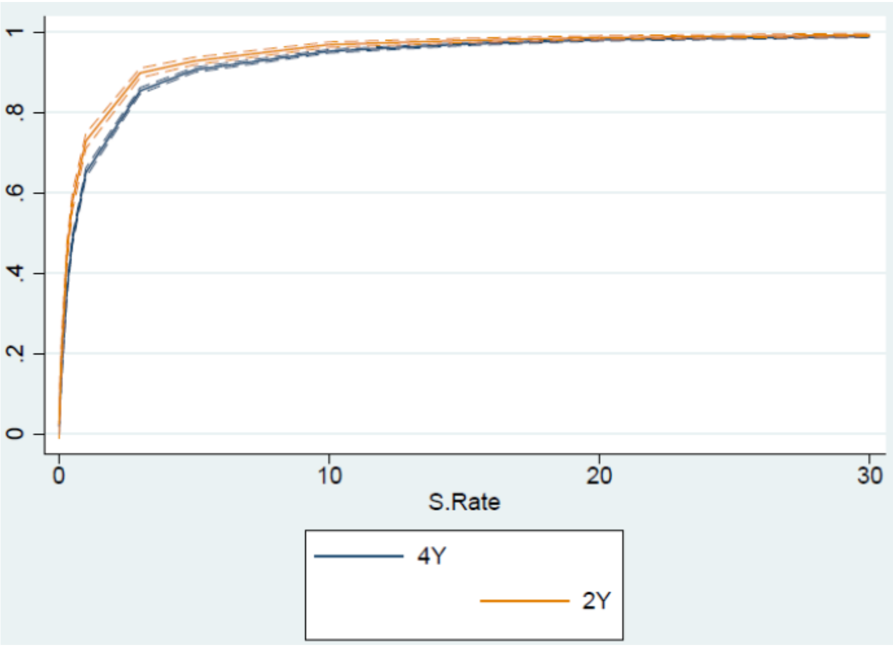
The estimates appear less biased for the lower level of panel length due to the results of cell 10, 50 and 100. In graph 3.4 it is shown how the line are overlapping for the lower cells, as a consequence the area below the graph for low and baseline level is approximately the same.

The notable result, however, if found when raising the panel data length up to 8. The optimal cell size for the FE model is already obtained around 500 cells. The resulting overall improvement with respect to the baseline population is of 63%. The lower ranges of the simulation still have high level of bias.

GRAPH 3.4 – FE BIAS BY PANEL DATA LENGTH



GRAPH 3.5 - FE ESTIMATES BY PANEL DATA LENGTH WITH CONFIDENCE INTERVALS



The SSIV results show less similarities with the FE ones in this case. When reducing the panel length data down to 2 the estimates are more biased: the optimal cell size is located at 500 cells for the endogenous variable and 1000 cells for the instrument. The 100 range, in the first row and first column, appears to be biased even for higher values of the instrument as in 100/3000. With respect to the baseline SSIV, in the cell ranges considered, there's an increase in bias of 564% on average.

TABLE 3.14 – SSIV BIAS BY PANEL DATA LENGTH (T = 2)

Endogenous var cell size	Instrument cell size					
Cell size	100	300	500	1000	3000	5000
100	-19	-120	-11	-6	-12	-1
300	-7	-4	-5	-4	-1	1
500	-4	-3	-3	1	-2	-3
1000	-1	-3	-1	-1	-1	-1

Raising the panel data length up to 8 lowers the bias almost to zero as shown in table 3.15. The decrease in bias is almost of 100% with respect to the baseline model. This result is the best scenario possible for a SSIV model.

TABLE 3.15 – SSIV BIAS BY PANEL DATA LENGTH (T = 2)

Endogenous var cell size	Instrument cell size					
Cell size	100	300	500	1000	3000	5000
100	0	-1	0	1	0	0
300	-1	-1	0	0	0	0
500	0	-1	0	-1	0	0
1000	0	0	0	0	1	0

CHAPTER 4 – Pooled OLS

4.1 Model outline

The last part will analyse how the conclusion reached so far for SSIV and FE can be applied to other models such as pooled ordinary least square (henceforth: POLS).

To test the behaviour of the attenuation bias, and its effect on a POLS model, the fixed effects have been removed from equation 1.4. The POLS model is then computed to estimate the coefficient of the baseline population from equation 1.1.

The same model is then used again on the population from 1.1 when the fixed effects for the population are removed

TABLE 4.1 – POLS ON POPULATION WITH FIXED EFFECTS

S.Rate	0.1	0.3	0.5	1	3	5
$\hat{\beta}^{FE}_{popul}$	.997	1	.999	1	.999	.999
s.e.	.0191	.011	.00854	.00604	.00349	.0027
$\hat{\beta}^{POLS}_{sample}$	.119	.285	.394	.568	.805	.872
s.e.	.0067	.00605	.00551	.00462	.00316	.00254
Obs Cell	69.7	209	348	696	2088	3480
s.e.	2.84	8.74	14.7	29.1	87.5	146

$$y = 56{,}52 e^{-0{,}009956x} - 63{,}05\, e^{-0{,}0003687x}$$

Goodness of fit:
SSE: 688.1
R-square: 0.8503

As described in table 4.1 the POLS is extremely biased whenever the population contains fixed effects as in the original baseline model. The numbers of observed cell by region have been changed a little to cover a wider range. An interesting consideration can be made by checking the fitted model of the regressor on the bias: the usual inverse exponential function has an extremely high sum of residual squares and as a result doesn't fit the data quite as well.

TABLE 4.2 – POLS ON POPULATION WITHOUT FIXED EFFECTS

S.Rate	0.1	0.3	0.5	1	3	5
$\hat{\beta}^{FE}_{popul}$	.996	.996	.996	.997	.997	.997
s.e.	.00814	.0047	.00364	.00257	.00149	.00115
$\hat{\beta}^{POLS}_{sample}$	.524	.762	.848	.912	.967	.978
s.e.	.00636	.00428	.00345	.0025	.00147	.00114
Obs Cell	104	313	522	1043	3130	5216
s.e.	4.28	12.8	21.4	42.8	128	214

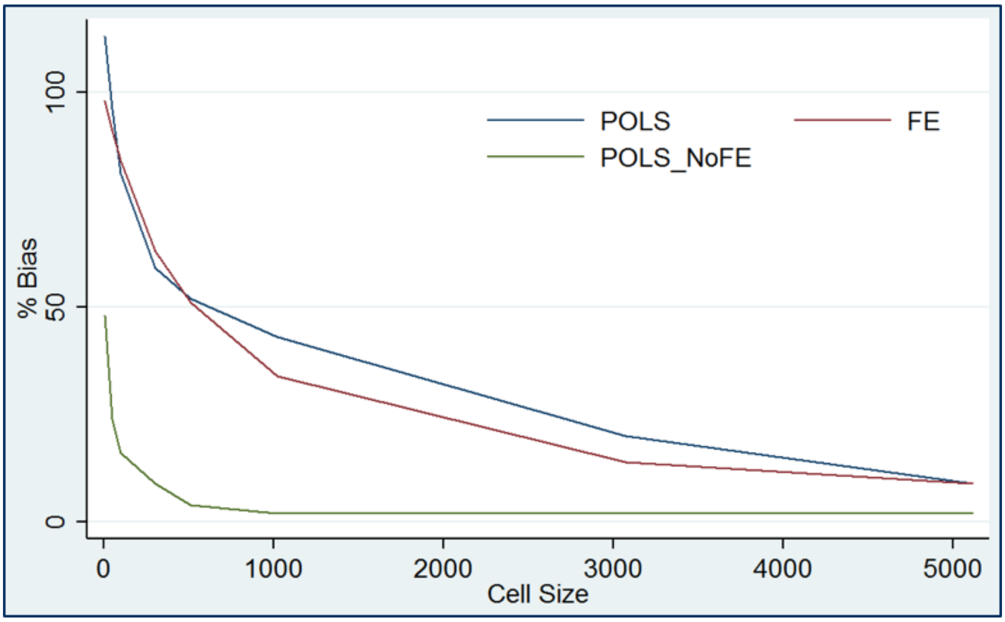
$y =$
$58{,}96e^{-0{,}004737x} - 12{,}3\, e^{-0{,}0003304x}$

Goodness of fit:
SSE: 3.475
R-square: 0.9979

Table 4.2 presents the result whenever we remove fixed effects in the population. As a consequence, the POLS is less biased than in table 4.1 and the results are a better fit for the usual function that correlates bias and number of cells.

4.1 POLS comparison

GRAPH 4.1 – POLS COMPARISON (WITH AND WITHOUT POPULATION FIXED EFFECT)



Graph 4.1 presents the bias in percentage points versus the cell size of for baseline FE model, and the POLS model from tables 4.1 and 4.2.

POLS model, even in presence of time invariant effects in the population, has a performance similar to that of FE. It is to be noted that POLS remains however more biased than FE. On a side note, POLS model, by averaging out all the observation across times, often fail to predict the actual effect of a regressor for a single time unit, but since the bias on migration can only be positive there is much less room for error.

TABLE 4.4 – POLS COMPARISON (WITH AND WITHOUT POPULATION FIXED EFFECT)

Modified Populations	Cell size								
Cell size	10	50	100	300	500	1024	3000	5000	Overall Improvement (in % w.r.t POLS With FE)
POLS With FE	113	96	81	59	52	43	20	9	-
POLS No FE	48	24	16	9	4	2	2	2	+350%
F.E.	96	94	85	66	53	33	11	9	-

Table 4.4 quantifies the bias per cell size and also identifies the optimal cell size for each model. The green area shows how the POLS without fixed effect in the population is much less biased, and the optimal cell size per region is 300. Meanwhile, for the POLS with fixed effect on the population model the optimal cell size rises up to at least 5000 per region. The performance is however close to the FE model. The improvement obtained by removing the fixed effect in the population and then computing the POLS is of a 350% reduction in bias w.r.t. the bias in POLS with FE, namely the POLS with no FE performs 4,5 times better due to the lack of bias from fixed effects in the population.

CONCLUSIONS

A solution to the estimation of the attenuation bias, as it has been presented in chapter one, has been elaborated by using the theoretical framework that identifies the characteristics of the bias itself in FE models.

Exploiting those characteristics, summed up in equation 1.15, the MCS simulation framework has been used to define and explore the relationship between attenuation bias and cell size in FE models. The SSIV application of the model has been considered as well.

The population model used as a reference for the simulation examines the effect of regional migration on crime rate. Since the final objective is to determine the optimal cell size for each FE model, any pair of regressor and regressand pair is viable, as long as the theoretical framework can be applied to the regressor.

The baseline population model simulated a population of ten million individuals with characteristics of immigration comparable to those of western Europe in terms of average initial proportion of migrants and yearly immigration shock equal to a 1% increase in total population. The optimal cell size to achieve a 90% reduction in attenuation bias for such a population is of 5000 observations per region if a FE model is used.

A SSIV model only requires 300 observations per region for the second stage regressor and 100 observations per region for the instrument.

Alteration of the baseline population parameter have been considered to study how the bias behaves whenever there's change in the regressor's information happens.

A raise in the initial proportion of migrants from 0.1 to 0.3 percent of total population causes in the optimal cell size to 3000 observations per region for FE models, and 300 and 50 for SSIV's second stage regressor and instrument. A reduction down to 0.05 for the initial proportion of immigrants in the population will lead to an increase in the optimal cell size to 6650 observations per region for FE models, and 500 and 300 for SSIV's second stage regressor and instrument.

An increase in yearly intensity of migration's shocks, up to a 2.5% yearly increase in total population, will cause the optimal cell size per region to be reduced down to a 1000 for FE models and 50 for both SSIV's second stage regressor and instrument. A decrease in the same parameter, down to 0.5% will lead to an increase for the optimal regional cell size up to 15000 for FE and 1000 and 300 for SSIV's second stage regressor and instrument.

The last test conducted on the length of the panel data shows how an increase in the unit of time observed up to 8 will lead to a decrease in terms of optimal cell size down to 500 cells per region for FE and 50 for both SSIV's second stage regressor and instrument. A decrease in the same parameter will lead to an increase in optimal regional cell size up to 5000 for FE and 300 for both SSIV's second stage regressor and instrument.

The simulation has been repeated for a POLS model using two different populations: one with time invariant characteristics and one without. The POLS estimates resulted less susceptible to attenuation bias than FE, and its estimate are more reliable if there are no time invariant characteristics in the population. That results renders the optimal cell sizes, for the reduction of attenuation bias, found for FE viable for POLS as well, if there are no fixed effect to account for in a population model.

The tests show how the information, added to the regressor by the shock's intensity over time, is stronger than the information added by the initial proportion of migrants. That might be explained by the dynamic nature of the shock, which adds information during each time unit rather than providing it at the beginning of the observed period. The conclusion seems to be consistent with the increase in panel data length, which by providing more comparisons over time for the "within estimator" to be computed, adds enough information to the regressor to reduce the attenuation bias down to a negligible size

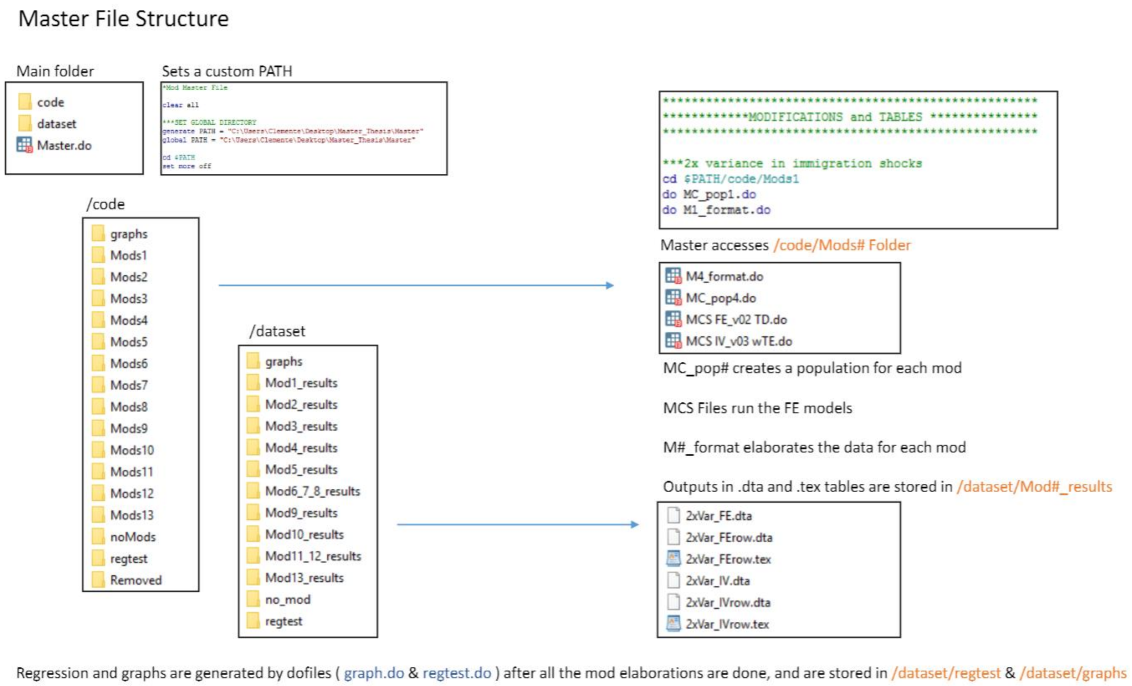
DATA SPEC – TECHNICAL SHEET

Simulation Stats

Total simulation computing time	***2544 hrs (106 days)
Total number of individuals	321'477'100
Models simulated	2400
Amount of data generated	2.27 TB

******Runtimes on a desktop IntelCorei5-4430 Processor at 3.20 GHz - 16GB RAM - Win10 x64 by STATA MP14***

Master File Structure



BIBLIOGRAPHY

Aydemir, A. and Borjas, G.J., 2010. Attenuation bias in measuring the wage impact of immigration (No. w16229). National Bureau of Economic Research.

Angrist, J.D. and Krueger, A.B., 1995. Split-sample instrumental variables estimates of the return to schooling. Journal of Business & Economic Statistics, 13(2), pp.225-235.

Angrist, J.D. and Krueger, A.B., 1990. Does compulsory school attendance affect schooling and earnings? (No. w3572). National Bureau of Economic Research.

Basu, A. and Chan, K.C.G., 2014. Can we make smart choices between OLS and contaminated IV methods?. Health economics, 23(4), pp.462-472.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80(6), pp.2369-2429.

Bound J., Brown C., and Mathiowetz N., 2001. Measurement error in survey data. Handbook of econometrics, 5, pp.3705-3843.

Bound, J., Jaeger, D.A. and Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American statistical association, 90(430), pp.443-450.

Bollen, K.A., Kirby, J.B., Curran, P.J., Paxton, P.M. and Chen, F., 2007. Latent variable models under misspecification: two-stage least squares (2SLS) and maximum likelihood (ML) estimators. Sociological Methods & Research, 36(1), pp.48-86.

Buddelmeyer, H., Jensen, P.H., Oguzoglu, U. and Webster, E., 2008. Fixed effects bias in panel data estimators. Available at SSRN 1136288.

Dufour, J.M., 2003. Identification, weak instruments, and statistical inference in econometrics. Canadian Journal of Economics/Revue canadienne d'économique, 36(4), pp.767-808.

Flores-Lagunes, A., 2007. Finite sample evidence of IV estimators under weak instruments. Journal of Applied Econometrics, 22(3), pp.677-694.

Ebbes, P., Wedel, M. and Böckenholt, U., 2009. Frugal IV alternatives to identify the parameter for an endogenous regressor. Journal of Applied Econometrics, 24(3), pp.446-468.

Hahn J, Hausman J. 2002a. Notes on bias in estimators for simultaneous equations models. Economics Letters 75: 237–241.

Hahn J, Hausman J. 2002b. A new specification test for the validity of instrumental variables. Econometrica 70: 163–189.

Hahn J, Hausman J. 2003. Weak instruments: diagnosis and cures in empirical econometrics. American Economic Review Papers and Proceedings 93: 118–125.

Hendry, D.F. and Harrison, R.W., 1974. Monte Carlo methodology and the small sample behaviour of ordinary and two-stage least squares. Journal of Econometrics, 2(2), pp.151-174.

Nagar, A.L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica: Journal of the Econometric Society, pp.575-595.

Nielsen, J.T., Guffanti, A. and Sarkar, S., 2015. Marginal evidence for cosmic acceleration from Type Ia supernovae. arXiv preprint arXiv:1506.01354.

Nunziata, L., 2015. Immigration and crime: evidence from victimization data. Journal of Population Economics, 28(3), pp.697-736.

Hawkins, O, 2017. Migration statistics. House of commons, Briefing Paper Number SN06077

Sawa, T., 1969. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. Journal of the American Statistical association, 64(327), pp.923-937.

United Nations, Department of Economic and Social Affairs, Population Division (2015). Trends in International Migrant Stock: The 2015 Revision. (United Nations database, POP/DB/MIG/Stock/Rev.2015)

Zivot, E., Startz, R. and Nelson, C.R., 1998. Valid confidence intervals and inference in the presence of weak instruments. International Economic Review, pp.1119-1144.

APPENDIX

Model and Population list of parameters mods 66

Baseline Population – 100 Reps 67

Immigration Shocks on population 68

Alternative error component for the population model 70

Panel data length tables 72

ESS Sampling rate – Baseline 74

Instrument and instrumented variable swap 76

Alternative Population – Raise in shock's intensity by ESS Sampling rate – 77

Alternate Population - Raise in Initial proportion of immigrants 79

Population size 81

Immigration Shock graphs 81

Alternative error terms for population graphs 84

Panel data length graphs 85

Population size graphs 87

Model and Population list of parameters mods

1. Population model (“Mcrime” Line 142) - 2x Variance in random increase in population model
2. Population model (“Mcrime” Line 249 & 254) - Pdf of probcrimevictim variable switched from uniform to lognormal
3. Population model (“Mcrime” Line 142) - 0.5x Variance in random increase in population model
4. Population, IV and FE model (“Mcrime” Line 45, “FE” Line 32, “IV” Line 31) – Year count Y raised from 4 to 8
5. Population, IV and FE model (“Mcrime” Line 45, “FE” Line 32, “IV” Line 31) – Year count Y reduced from 4 to 2
6. IV Model (“IV” from Line 56 to 89) Change in ESS Sampling rate from [0.1; 0.3; 0.5; 1] to [0.05; 0.20; 0.40; 0.80]
7. IV Model (“IV” from Line 56 to 89) Change in ESS Sampling rate from [0.1; 0.3; 0.5; 1] to [0.001; 0.01; 0.03; 0.07]
8. IV Model (“IV” from Line 56 to 89) Change in ESS Sampling rate from [0.1; 0.3; 0.5; 1] to [0.15; 0.25; 0.35; 0.45]
9. IV Model (“IV” from Line 149 to 191) Swapping instrumented and instrumental variable for each ESS sampling rate
10. Population Model (“Mcrime” from Line 232) Removing time effects by setting them from 0.01 to 0.00
11. IV Model (“IV” from Line 56 to 89)- 2xVar Pop - Change in ESS Sampling rate from [0.1; 0.3; 0.5; 1] to [0.05; 0.1; 0.3; 0.5]
12. IV Model (“IV” from Line 56 to 89)- 2xVar Pop - Change in ESS Sampling rate from [0.1; 0.3; 0.5; 1] to [1; 3; 5; 10]
13. Population model (“Mcrime” Line 142) - 4x Variance in random increase in population model
14. 1x Var - ESS(end) SR [0.05;0.1;0.3;0.5] 100 Reps
15. 1x Var - ESS (end) SR [1;3;5;10] 100 Reps
16. 1x Var - ESS(Instr. swap) SR [0.05;0.1;0.3;0.5] 100 Reps
17. 1x Var - ESS(Instr. swap) SR [1;3;5;10] 100 Reps
18. Change in prop. of immigrants (0.30) – Pop Model Line 52
19. Change in prop. of immigrants (0.50) – Pop Model Line 52
20. Population, IV model (“Mcrime” Line 45, “IV” Line 31) – Year count Y = 1
21. Population 2xVar, IV model (“Mcrime” Line 45, “IV” Line 31) – Year count Y = 1
22. Population reduction – (down to 1 000 000) - Pop Model Line 57
23. Region reduction – (down to 50) - Pop Model Line 41
24. POLS – Var pop 1x (50) – FE included (0.01), mu (0.01) ***mu = regional FE
25. POLS – Var pop 0.02x (1) – FE excluded (0.00), mu (0.01)
26. POLS - Var pop 1x (50) – FE excluded (0.00), mu (0.01)
27. POLS – Var pop 1x (50) – FE included (0.01), mu (0.01)
28. POLS – Var pop 1x (50) - FE excluded (0.00), mu (0.00)
29. POLS – Var pop 1x (50)- FE included (0.1), mu (0.1)
30. Pop reduction – (down to 500 000) – SR [0.3 – 0.5 – 1 – 3 – 5 -10 – 20]
31. Pop reduction – (down to 100 000) - SR [0.3 – 0.5 – 1 – 3 – 5 -10 – 20]

Baseline Population – 100 Reps

Table 1:No Mod Population Shock - FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20	30
$\widehat{\beta}^{FE}_{popul}$	.997	.998	.996	1	1	1	1	1	1	1	1	1
s.e.	.0721	.0322	.0228	.0131	.0102	.00719	.00415	.00322	.00228	.00186	.00161	.00131
$\widehat{\beta}^{FE}_{sample}$	.0184	.0862	.153	.36	.483	.65	.854	.905	.952	.97	.981	.989
s.e.	.00995	.00955	.00916	.008	.00713	.00585	.00386	.00306	.00222	.00183	.00159	.00131
Obs Cell	10.4	52.2	104	313	522	1043	3130	5216	10433	15649	20866	31299
s.e.	.53	2.14	4.28	12.8	21.4	42.8	128	214	428	642	856	1283
N.of Obs.	400	400	400	400	400	400	400	400	400	400	400	400
I.Pop	.137	.137	.137	.137	.137	.137	.137	.137	.137	.137	.137	.137
I.Pop.Sam	.138	.138	.138	.137	.137	.137	.137	.137	.137	.137	.137	.137

Table 2: No Mod- IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	.998	1	.999	.999	1	1
s.e.	.0198	.0114	.00886	.00626	.00362	.0028	.00198
$\widehat{\beta}^{FE}_{sample}$	.156	.354	.484	.651	.852	.908	.953
s.e.	.00801	.00693	.00628	.0051	.00335	.00268	.00194
$\widehat{\beta}^{IV}_{sample}$ 1/1000	.755	1.1	1.06	1.02	1.02	1.03	1.05
s.e.	.294	.0257	.0163	.00932	.00468	.00356	.0025
$\widehat{\beta}^{IVFS}_{sample}$ 1/1000	.154	.341	.482	.658	.849	.901	.936
s.e.	.00479	.00416	.00374	.00299	.00196	.00154	.0011
$\widehat{\beta}^{IV}_{sample}$ 3/1000	.981	.997	1.02	1.01	1.02	.993	1.01
s.e.	.0527	.0203	.0138	.00829	.0042	.00309	.00217
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.166	.361	.48	.65	.847	.921	.949
s.e.	.00313	.00261	.00232	.00182	.00115	.000914	.000648
$\widehat{\beta}^{IV}_{sample}$ 5/1000	1.06	.984	1.02	1.02	1.01	1.03	1.01
s.e.	.0553	.0196	.0134	.00821	.00411	.00313	.00211
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.151	.364	.481	.641	.844	.889	.95
s.e.	.00266	.0022	.00193	.00149	.000915	.000712	.000503
$\widehat{\beta}^{IV}_{sample}$ 1/100	.997	1.01	1	.999	1	1	.998
s.e.	.0513	.0198	.013	.00788	.00399	.003	.00206
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.158	.351	.485	.654	.853	.907	.957
s.e.	.00225	.00182	.00157	.00116	.000673	.000521	.00036
Obs Cell	104	313	522	1043	3130	5216	10433
s.e.	4.28	12.8	21.4	42.8	128	214	428
N.of Obs.	400	400	400	400	400	400	400
I.Pop	.137	.137	.137	.137	.137	.137	.137
I.Pop.Sam	.137	.137	.137	.137	.137	.137	.137

Immigration Shocks on population

Table 3 - 2xVar Population Shock - FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20	30
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	.999	1	1	.999	1	1	1	1
s.e.	.0319	.0143	.0101	.00583	.00452	.0032	.00185	.00143	.00101	.000825	.000715	.00
$\widehat{\beta}^{FE}_{sample}$	.0741	.282	.44	.708	.799	.887	.96	.974	.99	.993	.995	.997
s.e.	.0095	.00819	.00713	.00509	.00414	.00304	.00182	.00141	.00101	.000823	.000713	.00
Obs Cell	11	55	110	330	551	1101	3303	5505	11009	16514	22018	330
s.e.	1.12	5.45	10.9	32.7	54.4	109	327	545	1089	1634	2179	326
N.of Obs.	400	400	400	400	400	400	400	400	400	400	400	400
I.Pop	.183	.182	.182	.182	.182	.182	.182	.182	.182	.182	.182	.18
I.Pop.Sam	.184	.182	.183	.182	.182	.183	.182	.183	.182	.182	.182	.18

Table 4 - 2xVar Population Shock - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	.999	1	1	1	1	1	1
s.e.	.00879	.00508	.00393	.00278	.00161	.00124	.00088
$\widehat{\beta}^{FE}_{sample}$	.439	.706	.804	.892	.962	.976	.989
s.e.	.00633	.00447	.00363	.00267	.00159	.00123	.000877
$\widehat{\beta}^{IV}_{sample}$ 1/1000	1.02	.993	1.01	1	1.02	.995	1
s.e.	.0161	.00706	.00519	.00347	.00196	.00147	.00104
$\widehat{\beta}^{IVFS}_{sample}$ 1/1000	.436	.717	.805	.895	.953	.986	.993
s.e.	.00426	.0029	.0023	.00167	.000969	.000755	.000535
$\widehat{\beta}^{IV}_{sample}$ 3/1000	1	1	1.01	1.01	1	1	.998
s.e.	.0142	.00646	.00467	.00315	.00174	.00134	.000936
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.442	.705	.801	.887	.964	.976	.993
s.e.	.00314	.00196	.00149	.00104	.000583	.000444	.00031
$\widehat{\beta}^{IV}_{sample}$ 5/1000	.994	.998	1	.996	1	1	.997
s.e.	.0138	.00627	.00455	.00304	.0017	.00131	.000915
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.442	.708	.805	.897	.963	.973	.993
s.e.	.00287	.00172	.00128	.000852	.000461	.000352	.000244
$\widehat{\beta}^{IV}_{sample}$ 1/100	.997	.996	1	.998	.999	1	.997
s.e.	.0136	.00615	.00449	.003	.00167	.00128	.000898
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.441	.71	.801	.894	.963	.976	.993
s.e.	.00265	.00152	.00109	.000691	.000351	.000255	.000175
Obs Cell	110	330	551	1101	3303	5505	11009
s.e.	10.9	32.7	54.4	109	327	545	1089
N.of Obs.	400	400	400	400	400	400	400
I.Pop	.182	.182	.182	.182	.182	.182	.182
I.Pop.Sam	.182	.182	.182	.183	.183	.182	.182

Table 5 - 0.5xVar Population Shock - FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20	30
$\widehat{\beta}^{FE}_{popul}$	.989	1	.996	1	.999	.999	.999	1	1	1	1	1
s.e.	.156	.0697	.0492	.0284	.022	.0156	.00899	.00697	.00493	.00402	.00348	.00284
$\widehat{\beta}^{FE}_{sample}$	.00481	.02	.043	.117	.181	.31	.581	.699	.834	.886	.915	.951
s.e.	.0104	.0103	.0104	.00981	.00949	.00878	.00686	.00582	.0045	.00379	.00333	.00277
Obs Cell	10.1	51	102	306	510	1020	3060	5100	10200	15300	20399	30599
s.e.	.287	.989	1.9	5.66	9.42	19	56.8	94.6	189	284	379	568
N.of Obs.	400	400	400	400	400	400	400	400	400	400	400	400

Table 6 - 0.5xVar Population Shock - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	.999	1	1	1	1	1	1
s.e.	.0429	.0248	.0192	.0136	.00783	.00606	.00429
$\widehat{\beta}^{FE}_{sample}$	.0388	.118	.183	.305	.583	.711	.832
s.e.	.00896	.00865	.00825	.00756	.00602	.00514	.00391
$\widehat{\beta}^{IV}_{sample}$ 1/1000	.468	1.15	.967	.985	1.28	1.18	1.05
s.e.	1.09	.659	.302	.0355	.0201	.011	.00586
$\widehat{\beta}^{IVFS}_{sample}$ 1/1000	.0566	.119	.192	.328	.571	.688	.849
s.e.	.00496	.00475	.00456	.00415	.00332	.00279	.00215
$\widehat{\beta}^{IV}_{sample}$ 3/1000	1.89	.863	1.79	1.12	1.04	1.02	1.03
s.e.	7.94	.272	.228	.0315	.0114	.00778	.00511
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.0317	.124	.181	.312	.578	.725	.842
s.e.	.003	.00284	.00273	.00247	.00194	.00166	.00126
$\widehat{\beta}^{IV}_{sample}$ 5/1000	1.26	1.23	1.88	1.04	1.04	1.02	1.01
s.e.	.717	.102	.569	.0269	.0111	.00761	.00491
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.0398	.115	.181	.307	.582	.714	.837
s.e.	.00239	.0023	.00218	.00198	.00154	.00128	.000972
$\widehat{\beta}^{IV}_{sample}$ 1/100	.472	1.03	1.05	.989	1.01	1.03	1.01
s.e.	1.03	.0774	.0485	.0248	.0105	.00754	.00482
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.0407	.121	.179	.314	.586	.699	.833
s.e.	.00184	.00175	.00166	.00148	.00112	.000942	.000701
Obs Cell	102	306	510	1020	3060	5100	10200
s.e.	1.9	5.66	9.42	19	56.8	94.6	189
N.of Obs.	400	400	400	400	400	400	400

Table 7: 4xVar Population Shock - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	1	1	1
s.e.	.00425	.00246	.0019	.00135	.000777	.000602	.000426
$\widehat{\beta}^{FE}_{sample}$	.736	.896	.936	.968	.99	.992	.997
s.e.	.00413	.00245	.00191	.00135	.000778	.000601	.000426
$\widehat{\beta}^{IV}_{sample}$ 1/1000	.997	1	1	1	1	1.01	1
s.e.	.00579	.00305	.00233	.00161	.000922	.000715	.000502
$\widehat{\beta}^{IVFS}_{sample}$ 1/1000	.739	.897	.932	.968	.988	.988	.996
s.e.	.00296	.00164	.00123	.000855	.000484	.000374	.000262
$\widehat{\beta}^{IV}_{sample}$ 3/1000	.999	1	1	.999	1	.998	1
s.e.	.00525	.00276	.0021	.00145	.000833	.00064	.000452
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.736	.896	.933	.969	.986	.994	.997
s.e.	.00232	.00114	.000816	.000535	.000287	.000218	.000152
$\widehat{\beta}^{IV}_{sample}$ 5/1000	1	1	.999	.995	.995	1	.998
s.e.	.00515	.00269	.00204	.00141	.000806	.000628	.000441
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.734	.896	.938	.972	.995	.989	.999
s.e.	.00219	.00101	.000703	.000446	.000229	.000171	.00012
$\widehat{\beta}^{IV}_{sample}$ 1/100	1	1	1	.998	1	.997	1
s.e.	.00505	.00265	.00201	.00139	.000796	.000613	.000435
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.736	.895	.935	.969	.989	.995	.994
s.e.	.00207	.000902	.000608	.000356	.000173	.000127	.0000856
Obs Cell	127	380	634	1268	3804	6339	12679
s.e.	33.7	101	168	337	1010	1684	3368
N.of Obs.	400	400	400	400	400	400	400

NB: 2xVar = 2.5% yearly increase. 0.5xVar = 0.5% yearly increase. 4xVar = 27% yearly increase

Alternative error component for the population model

Table 8:Lognormal distribution(0,0.1)- FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20	30
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	.999	1	1	.999	1	1	1	1
s.e.	.0319	.0143	.0101	.00583	.00452	.0032	.00185	.00143	.00101	.000825	.000715	.000
$\widehat{\beta}^{FE}_{sample}$	.0741	.282	.44	.708	.799	.887	.96	.974	.99	.993	.995	.997
s.e.	.0095	.00819	.00713	.00509	.00414	.00304	.00182	.00141	.00101	.000823	.000713	.000
Obs Cell	11	55	110	330	551	1101	3303	5505	11009	16514	22018	3302
s.e.	1.12	5.45	10.9	32.7	54.4	109	327	545	1089	1634	2179	3268
N.of Obs.	400	400	400	400	400	400	400	400	400	400	400	400

Table 9:Lognormal distribution(0,0.1) - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	.999	1	1	1	1	1	1
s.e.	.00879	.00508	.00393	.00278	.00161	.00124	.00088
$\widehat{\beta}^{FE}_{sample}$	.439	.706	.804	.892	.962	.976	.989
s.e.	.00633	.00447	.00363	.00267	.00159	.00123	.000877
$\widehat{\beta}^{IV}_{sample \ 1/1000}$	1.02	.993	1.01	1	1.02	.995	1
s.e.	.0161	.00706	.00519	.00347	.00196	.00147	.00104
$\widehat{\beta}^{IVFS}_{sample \ 1/1000}$	.436	.717	.805	.895	.953	.986	.993
s.e.	.00426	.0029	.0023	.00167	.000969	.000755	.000535
$\widehat{\beta}^{IV}_{sample \ 3/1000}$	1	1	1.01	1.01	1	1	.998
s.e.	.0142	.00646	.00467	.00315	.00174	.00134	.000936
$\widehat{\beta}^{IVFS}_{sample \ 3/1000}$	.442	.705	.801	.887	.964	.976	.993
s.e.	.00314	.00196	.00149	.00104	.000583	.000444	.00031
$\widehat{\beta}^{IV}_{sample \ 5/1000}$	.994	.998	1	.996	1	1	.997
s.e.	.0138	.00627	.00455	.00304	.0017	.00131	.000915
$\widehat{\beta}^{IVFS}_{sample \ 5/1000}$	.442	.708	.805	.897	.963	.973	.993
s.e.	.00287	.00172	.00128	.000852	.000461	.000352	.000244
$\widehat{\beta}^{IV}_{sample \ 1/100}$	.997	.996	1	.998	.999	1	.997
s.e.	.0136	.00615	.00449	.003	.00167	.00128	.000898
$\widehat{\beta}^{IVFS}_{sample \ 1/100}$	.441	.71	.801	.894	.963	.976	.993
s.e.	.00265	.00152	.00109	.000691	.000351	.000255	.000175
Obs Cell	110	330	551	1101	3303	5505	11009
s.e.	10.9	32.7	54.4	109	327	545	1089
N.of Obs.	400	400	400	400	400	400	400

Panel data length tables

NB: Value for years are doubled

Table 10: 8 Years - FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20
$\widehat{\beta}^{FE}_{popul}$	1	1	.999	1	1	1	1	1	1	1	1
s.e.	.0099	.00441	.00312	.0018	.0014	.000988	.00057	.000442	.000312	.000255	.000221
$\widehat{\beta}^{FE}_{sample}$	.212	.573	.727	.888	.932	.964	.989	.993	.997	.998	.999
s.e.	.0059	.00388	.00295	.00178	.00139	.000985	.00057	.000441	.000312	.000255	.000221
Obs Cell	15.2	75.9	152	455	759	1518	4553	7589	15178	22766	30355
s.e.	6.97	34.8	69.6	209	348	696	2087	3478	6956	10434	13911
N.of Obs.	800	800	800	800	800	800	800	800	800	800	800

Table 11: 8 Years - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	1	1	1
s.e.	.00272	.00157	.00122	.00086	.000496	.000385	.000272
$\widehat{\beta}^{FE}_{sample}$	.726	.889	.931	.964	.99	.994	.997
s.e.	.00265	.00157	.00122	.00086	.000497	.000385	.000272
$\widehat{\beta}^{IV}_{sample}$ 1/1000	1	1.01	1	.999	1	1	1
s.e.	.00381	.00201	.0015	.00104	.000596	.000459	.000324
$\widehat{\beta}^{IVFS}_{sample}$ 1/1000	.724	.881	.927	.965	.986	.993	.998
s.e.	.00195	.00109	.000818	.000567	.000323	.000249	.000175
$\widehat{\beta}^{IV}_{sample}$ 3/1000	.999	.999	1	1	1	1	.999
s.e.	.0034	.00177	.00135	.000936	.000535	.000412	.00029
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.727	.89	.93	.961	.986	.993	.997
s.e.	.00154	.000754	.000542	.000353	.000193	.000147	.000101
$\widehat{\beta}^{IV}_{sample}$ 5/1000	1	.999	1	.999	1	1	1
s.e.	.00332	.00173	.00131	.00091	.00052	.000402	.000284
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.725	.889	.931	.964	.988	.992	.996
s.e.	.00145	.000675	.000469	.000294	.000153	.000116	.000079
$\widehat{\beta}^{IV}_{sample}$ 1/100	1	1	1	1	.999	1	1
s.e.	.00325	.0017	.00129	.000892	.000509	.000394	.000278
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.727	.887	.932	.964	.991	.993	.995
s.e.	.00136	.000601	.000402	.000238	.000115	.0000847	.0000576
Obs Cell	152	455	759	1518	4553	7589	15178
s.e.	69.6	209	348	696	2087	3478	6956
N.of Obs.	800	800	800	800	800	800	800

Table 12: 2 Years - FE

S.Rate	0.01	0.05	0.1	0.3	0.5	1	3	5	10	15	20
$\widehat{\beta}^{FE}_{popul}$	.991	1.01	1.01	.998	.995	1	1	1	1	1	1
s.e.	.109	.0487	.0345	.0199	.0154	.0109	.00629	.00487	.00344	.00281	.00244
$\widehat{\beta}^{FE}_{sample}$	.0237	.123	.209	.452	.579	.729	.898	.928	.969	.979	.986
s.e.	.0178	.0173	.0161	.0137	.0118	.00935	.00598	.00467	.00339	.00278	.00242
Obs Cell	10.2	51.2	102	307	512	1024	3071	5118	10236	15354	20472
s.e.	.422	1.6	3.13	9.32	15.6	31.1	93.4	156	311	467	623
N.of Obs.	200	200	200	200	200	200	200	200	200	200	200

Table 13: 2 Years - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	.996	1	.999	.996	1	.998	.999
s.e.	.03	.0173	.0134	.00947	.00547	.00424	.003
$\widehat{\beta}^{FE}_{sample}$	.208	.458	.58	.73	.896	.936	.971
s.e.	.014	.0118	.0106	.00818	.00518	.00412	.00296
$\widehat{\beta}^{IV}_{sample\ 1/1000}$	1.19	2.21	1.11	1.06	1.12	1.01	1.05
s.e.	.124	.471	.0231	.0131	.00743	.00492	.0036
$\widehat{\beta}^{IVFS}_{sample\ 1/1000}$	.225	.431	.565	.713	.849	.954	.96
s.e.	.00679	.00553	.00492	.00374	.00237	.00184	.00135
$\widehat{\beta}^{IV}_{sample\ 3/1000}$	1.07	1.04	1.05	1.04	1.01	.995	1.04
s.e.	.0763	.0274	.0199	.0121	.00602	.00452	.00329
$\widehat{\beta}^{IVFS}_{sample\ 3/1000}$	.209	.451	.576	.714	.901	.95	.943
s.e.	.00448	.00359	.0032	.00233	.00139	.00109	.000772
$\widehat{\beta}^{IV}_{sample\ 5/1000}$	1.04	1.03	1.03	.994	1.02	1.03	1.02
s.e.	.0721	.0268	.019	.0113	.006	.00461	.00318
$\widehat{\beta}^{IVFS}_{sample\ 5/1000}$	.206	.453	.573	.742	.884	.915	.956
s.e.	.00397	.00303	.00262	.00186	.00108	.000862	.000599
$\widehat{\beta}^{IV}_{sample\ 1/100}$	1.01	1.03	1.01	1.01	1.01	1	1.01
s.e.	.0687	.0265	.0184	.0114	.0059	.00444	.00311
$\widehat{\beta}^{IVFS}_{sample\ 1/100}$	.209	.447	.577	.725	.888	.938	.967
s.e.	.00341	.00254	.00215	.00151	.00082	.000633	.000433
Obs Cell	102	307	512	1024	3071	5118	10236
s.e.	3.13	9.32	15.6	31.1	93.4	156	311
N.of Obs.	200	200	200	200	200	200	200

ESS Sampling rate – Baseline

NB: ESS = second stage regressor's sampling rate

Table 14: ESS sampling rate - IV - part 1

S.Rate	0.1	0.3	0.5	1	3	5	10
$\hat{\beta}^{FE}_{popul}$	1	.998	1	.999	.999	1	1
s.e.	.0198	.0114	.00886	.00626	.00362	.0028	.00198
$\hat{\beta}^{FE}_{sample}$	.156	.354	.484	.651	.852	.908	.953
s.e.	.00801	.00693	.00628	.0051	.00335	.00268	.00194
$\hat{\beta}^{IV}_{sample}$ 1/10000	1.07	1.02	1	1.04	1.04	.997	1.01
s.e.	.061	.0213	.0139	.00885	.00442	.00317	.00221
$\hat{\beta}^{IVFS}_{sample}$ 1/10000	.158	.357	.5	.641	.832	.921	.959
s.e.	.00364	.00307	.00273	.0022	.00139	.00111	.000795
$\hat{\beta}^{IV}_{sample}$ 3/10000	1.01	1	1.01	1.02	1.01	1.01	.992
s.e.	.054	.0199	.0134	.00828	.00412	.0031	.00209
$\hat{\beta}^{IVFS}_{sample}$ 3/10000	.161	.36	.489	.647	.851	.903	.964
s.e.	.00285	.00236	.00209	.00164	.00102	.000787	.000557
$\hat{\beta}^{IV}_{sample}$ 5/10000	2.12	1.12	1.16	1.09	1.04	1.06	1.09
s.e.	1.33	.0312	.0222	.0121	.00543	.00423	.00307
$\hat{\beta}^{IVFS}_{sample}$ 5/10000	.153	.352	.461	.648	.844	.885	.923
s.e.	.00654	.00572	.00514	.00417	.00275	.00218	.00158
$\hat{\beta}^{IV}_{sample}$ 7/10000	1.02	1.01	.986	1	1	.998	1
s.e.	.0533	.0198	.0128	.00797	.00401	.003	.00207
$\hat{\beta}^{IVFS}_{sample}$ 7/10000	.157	.354	.497	.656	.854	.911	.953
s.e.	.00239	.00193	.00168	.00127	.000743	.000574	.000403
$\hat{\beta}^{IV}_{sample}$ 1/1000	.755	1.1	1.06	1.02	1.02	1.03	1.05
s.e.	.294	.0257	.0163	.00932	.00468	.00356	.0025
$\hat{\beta}^{IVFS}_{sample}$ 1/1000	.154	.341	.482	.658	.849	.901	.936
s.e.	.00479	.00416	.00374	.00299	.00196	.00154	.0011
$\hat{\beta}^{IV}_{sample}$ 1,5/1000	1.13	1.03	1.02	1.01	1.02	1.02	1.02
s.e.	.0665	.0224	.0146	.00876	.00444	.00334	.00229
$\hat{\beta}^{IVFS}_{sample}$ 1,5/1000	.152	.355	.49	.655	.846	.901	.956
s.e.	.00407	.0035	.00313	.00248	.00161	.00127	.000902
$\hat{\beta}^{IV}_{sample}$ 2/1000	1.04	1.01	1.03	1.03	1.03	1	1.02
s.e.	.0599	.0211	.0143	.00865	.00437	.0032	.00225
$\hat{\beta}^{IVFS}_{sample}$ 2/1000	.163	.359	.48	.644	.841	.915	.945
s.e.	.00363	.00304	.00272	.00216	.00139	.00111	.000792
$\hat{\beta}^{IV}_{sample}$ 2,5/1000	.992	1	1.02	1.02	1.03	.996	1.02
s.e.	.054	.0206	.014	.00842	.0043	.00313	.00221
$\hat{\beta}^{IVFS}_{sample}$ 2,5/1000	.166	.361	.481	.647	.839	.918	.944
s.e.	.00332	.0028	.00248	.00196	.00125	.000999	.00071

Table 15: ESS sampling rate - IV - part 2

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{IV}_{sample}$ 3/1000	.981	.997	1.02	1.01	1.02	.993	1.01
s.e.	.0527	.0203	.0138	.00829	.0042	.00309	.00217
$\widehat{\beta}^{IVFS}_{sample}$ 3/1000	.166	.361	.48	.65	.847	.921	.949
s.e.	.00313	.00261	.00232	.00182	.00115	.000914	.000648
$\widehat{\beta}^{IV}_{sample}$ 3,5/1000	1.11	.998	1.01	1.03	1.03	1.01	1.01
s.e.	.0593	.0202	.0136	.0084	.00423	.00312	.00215
$\widehat{\beta}^{IVFS}_{sample}$ 3,5/1000	.146	.362	.484	.638	.836	.905	.948
s.e.	.00295	.00247	.0022	.00171	.00107	.000842	.000597
$\widehat{\beta}^{IV}_{sample}$ 4/1000	1.09	.995	1.01	1.02	1.02	1.02	1.01
s.e.	.0577	.02	.0135	.00827	.00417	.00313	.00214
$\widehat{\beta}^{IVFS}_{sample}$ 4/1000	.148	.362	.485	.644	.841	.896	.947
s.e.	.00284	.00236	.00209	.00162	.00101	.000792	.000559
$\widehat{\beta}^{IV}_{sample}$ 4,5/1000	1.01	1.03	1	1.01	.996	1	.995
s.e.	.053	.0205	.0133	.00816	.00404	.00306	.00209
$\widehat{\beta}^{IVFS}_{sample}$ 4,5/1000	.158	.349	.486	.648	.859	.91	.962
s.e.	.00275	.00227	.002	.00153	.00095	.000754	.000531
$\widehat{\beta}^{IV}_{sample}$ 5/1000	1.06	.984	1.02	1.02	1.01	1.03	1.01
s.e.	.0553	.0196	.0134	.00821	.00411	.00313	.00211
$\widehat{\beta}^{IVFS}_{sample}$ 5/1000	.151	.364	.481	.641	.844	.889	.95
s.e.	.00266	.0022	.00193	.00149	.000915	.000712	.000503
$\widehat{\beta}^{IV}_{sample}$ 8/1000	.994	1.01	.998	.995	.997	1	.998
s.e.	.0513	.0198	.013	.00788	.00399	.00302	.00206
$\widehat{\beta}^{IVFS}_{sample}$ 8/1000	.159	.353	.487	.657	.856	.907	.958
s.e.	.00236	.00192	.00166	.00125	.000736	.000577	.0004
$\widehat{\beta}^{IV}_{sample}$ 1/100	.997	1.01	1	.999	1	1	.998
s.e.	.0513	.0198	.013	.00788	.00399	.003	.00206
$\widehat{\beta}^{IVFS}_{sample}$ 1/100	.158	.351	.485	.654	.853	.907	.957
s.e.	.00225	.00182	.00157	.00116	.000673	.000521	.00036

Instrument and instrumented variable swap

Table 16: Instrument and instrumented variable swap - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	.998	1	.999	.999	1	1
s.e.	.0198	.0114	.00886	.00626	.00362	.0028	.00198
$\widehat{\beta}^{FE}_{sample}$	.155	.361	.487	.654	.855	.914	.954
s.e.	.00806	.00697	.00632	.00511	.00336	.0027	.00194
$\widehat{\beta}^{IV}_{sample\ 1/1000}$	1.27	.988	1.06	.964	1.01	1.01	1
s.e.	.134	.0309	.025	.0156	.00941	.00753	.00516
$\widehat{\beta}^{IVFS}_{sample\ 1/1000}$	.16	.162	.154	.166	.155	.151	.155
s.e.	.00485	.00183	.0012	.000721	.000355	.000266	.000184
$\widehat{\beta}^{IV}_{sample\ 3/1000}$	1.12	1.05	1.03	.98	1	1	.998
s.e.	.047	.0212	.0159	.0105	.00607	.00476	.0033
$\widehat{\beta}^{IVFS}_{sample\ 3/1000}$	.347	.35	.351	.366	.361	.354	.362
s.e.	.00723	.00263	.0017	.001	.000481	.000353	.000243
$\widehat{\beta}^{IV}_{sample\ 5/1000}$	1.04	.995	1.02	.997	1.01	1	.998
s.e.	.0354	.0173	.0135	.00916	.00528	.00405	.00282
$\widehat{\beta}^{IVFS}_{sample\ 5/1000}$	.487	.493	.484	.485	.477	.485	.49
s.e.	.00828	.00299	.00193	.0011	.000513	.000374	.000253
$\widehat{\beta}^{IV}_{sample\ 1/100}$	1.03	.995	1.01	.992	1	1.01	.999
s.e.	.0295	.0149	.0115	.00781	.0045	.00354	.00244
$\widehat{\beta}^{IVFS}_{sample\ 1/100}$	.648	.657	.645	.66	.652	.641	.656
s.e.	.00939	.00332	.00208	.00117	.000515	.000375	.000246
Obs Cell	104	313	522	1043	3130	5216	10433
s.e.	4.28	12.8	21.4	42.8	128	214	428
N.of Obs.	400	400	400	400	400	400	400

Alternative Population – Raise in shock's intensity by ESS Sampling rate –

Table 19: ESS sampling rate - 2xVar - IV

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	1	1	1
s.e.	1.15e-08	6.61e-09	5.12e-09	3.62e-09	2.09e-09	1.62e-09	1.14e-09
$\widehat{\beta}^{FE}_{sample}$	.16	.359	.48	.654	.854	.907	.953
s.e.	.00179	.00135	.0011	.000726	.00032	.000203	.000102
$\widehat{\beta}^{IV}_{sample\ 5/10000}$	-.534	1.41	.282	-.298	1.36	.0567	.324
s.e.	1.78	6.38	1.14	.401	.475	.552	.119
$\widehat{\beta}^{IVFS}_{sample\ 5/10000}$	.098	.448	.671	.555	.952	.98	1.17
s.e.	.0463	.0398	.0368	.0292	.0199	.0156	.0114
$\widehat{\beta}^{IV}_{sample\ 1/1000}$	.291	2.94	.616	2.66	.917	1.03	1.13
s.e.	.384	6.78	.142	.955	.114	.00653	.00651
$\widehat{\beta}^{IVFS}_{sample\ 1/1000}$	.11	.405	.479	.765	.756	.98	.948
s.e.	.0145	.0125	.0113	.00921	.00607	.00485	.00351
$\widehat{\beta}^{IV}_{sample\ 3/1000}$	-.0251	.744	1.19	1.13	1.16	1.09	1.07
s.e.	.931	2.11	.0328	.0116	.00694	.00406	.00266
$\widehat{\beta}^{IVFS}_{sample\ 3/1000}$	.139	.39	.518	.64	.855	.895	.967
s.e.	.0085	.00734	.00657	.00538	.00354	.00281	.00203
$\widehat{\beta}^{IV}_{sample\ 5/1000}$	.622	1.11	1.21	1.02	1.14	1.01	1.01
s.e.	1.9	.0228	.0181	.00694	.00444	.00254	.00176
$\widehat{\beta}^{IVFS}_{sample\ 5/1000}$	.169	.369	.445	.681	.796	.922	.982
s.e.	.00668	.00573	.00517	.00413	.00272	.00219	.00157
$\widehat{\beta}^{IV}_{sample\ 1/100}$	1.01	.999	1	1	1	1	1
s.e.	.0137	.00617	.00447	.00301	.00167	.00128	.000903
$\widehat{\beta}^{IVFS}_{sample\ 1/100}$	.437	.708	.804	.89	.961	.976	.987
s.e.	.00264	.00152	.00108	.00069	.000349	.000255	.000175
$\widehat{\beta}^{IV}_{sample\ 3/100}$	.999	1	1	1	.996	.999	1
s.e.	.0134	.00611	.00441	.00298	.00164	.00127	.000892
$\widehat{\beta}^{IVFS}_{sample\ 3/100}$	.44	.706	.804	.889	.966	.977	.987
s.e.	.00247	.00135	.000928	.000559	.00024	.000168	.000107
$\widehat{\beta}^{IV}_{sample\ 5/100}$	.998	.997	1	.997	1	1	1
s.e.	.0134	.00606	.0044	.00295	.00165	.00126	.000888
$\widehat{\beta}^{IVFS}_{sample\ 5/100}$	.44	.709	.804	.894	.962	.976	.989
s.e.	.00244	.00132	.000897	.000519	.000215	.000146	.0000886
$\widehat{\beta}^{IV}_{sample\ 1/10}$	.998	1	1	1	1	1	1
s.e.	.0134	.00608	.0044	.00296	.00164	.00126	.000886
$\widehat{\beta}^{IVFS}_{sample\ 1/10}$	.44	.705	.804	.891	.962	.976	.989
s.e.	.00241	.00129	.00087	.000497	.000193	.000124	.0000715
Obs. Cell	110	330	551	1101	3303	5505	11009
Cell s.e.	10.9	32.7	54.4	109	327	545	1089

Table 20: FSIV Betas by ESS S.R. and Pop Shock

S.Rate	0.1	0.3	0.5	1	3	5	10
0.5xVar FSIV ESS 1/1000	.0566	.119	.192	.328	.571	.688	.849
0.5xVar FSIV ESS 3/1000	.0317	.124	.181	.312	.578	.725	.842
0.5xVar FSIV ESS 5/1000	.0398	.115	.181	.307	.582	.714	.837
0.5xVar FSIV ESS 1/100	.0407	.121	.179	.314	.586	.699	.833
1xVar FSIV ESS 1/1000	.154	.341	.482	.658	.849	.901	.936
1xVar FSIV ESS 3/1000	.166	.361	.48	.65	.847	.921	.949
1xVar FSIV ESS 5/1000	.151	.364	.481	.641	.844	.889	.95
1xVar FSIV ESS 1/100	.158	.351	.485	.654	.853	.907	.957
2xVar FSIV ESS 1/1000	.436	.717	.805	.895	.953	.986	.993
2xVar FSIV ESS 3/1000	.442	.705	.801	.887	.964	.976	.993
2xVar FSIV ESS 5/1000	.442	.708	.805	.897	.963	.973	.993
2xVar FSIV ESS 1/100	.441	.71	.801	.894	.963	.976	.993
4xVar FSIV ESS 1/1000	.739	.897	.932	.968	.988	.988	.996
4xVar FSIV ESS 3/1000	.736	.896	.933	.969	.986	.994	.997
4xVar FSIV ESS 5/1000	.734	.896	.938	.972	.995	.989	.999
4xVar FSIV ESS 1/100	.736	.895	.935	.969	.989	.995	.994

Alternate Population - Raise in Initial proportion of immigrants

Initial proportion of immigrants: 0.3 - IV

S.Rate	0.1	0.3	0.5	1	3	5
$\widehat{\beta}^{FE}_{popul}$	1	1	1	1	1	1
s.e.	.0107	.00619	.00479	.00339	.00196	.00152
$\widehat{\beta}^{FE}_{sample}$	.241	.496	.619	.768	.909	.945
s.e.	.00566	.00452	.00392	.00304	.00188	.00148
$\widehat{\beta}^{IV}_{sample\ \ 0.1}$	1.11	1.03	1.02	1.04	.998	1.02
s.e.	.0337	.0117	.00805	.00526	.0026	.00204
$\widehat{\beta}^{IVFS}_{sample\ \ 0.1}$	.231	.493	.618	.751	.919	.934
s.e.	.00455	.00361	.00311	.0024	.00146	.00113
$\widehat{\beta}^{IV}_{sample\ \ 0.3}$	1.02	1.01	1.02	.995	.999	.992
s.e.	.0248	.00965	.00685	.00422	.00225	.00169
$\widehat{\beta}^{IVFS}_{sample\ \ 0.3}$	.239	.495	.61	.775	.913	.955
s.e.	.00314	.00235	.00197	.00148	.000877	.000677
$\widehat{\beta}^{IV}_{sample\ \ 0.5}$	1	1.01	1.01	.993	.997	1
s.e.	.0234	.00933	.00652	.00407	.00217	.00165
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.242	.493	.617	.775	.914	.944
s.e.	.00271	.00199	.00167	.00122	.000693	.000529
$\widehat{\beta}^{IV}_{sample\ \ 1}$	1.01	1	.996	1.01	1	.996
s.e.	.023	.00899	.00628	.00404	.00212	.0016
$\widehat{\beta}^{IVFS}_{sample\ \ 1}$	.241	.497	.622	.761	.909	.949
s.e.	.00241	.0017	.0014	.000965	.000517	.000394
Obs Cell	113	339	565	1130	3390	5649
s.e.	12.8	38.5	64.1	128	385	642
N.of Obs.	400	400	400	400	400	400
I.Pop	.38	.38	.38	.38	.38	.38
I.Pop.Sam	.38	.38	.38	.381	.38	.38

Initial proportion of immigrants: 0.5 - IV

S.Rate	0.1	0.3	0.5	1	3	5
$\widehat{\beta}^{FE}_{popul}$	.998	.998	1	1	1	1
s.e.	.0109	.00627	.00486	.00344	.00198	.00154
$\widehat{\beta}^{FE}_{sample}$	.228	.478	.6	.75	.902	.939
s.e.	.00554	.00452	.00392	.00303	.0019	.0015
$\widehat{\beta}^{IV}_{sample\ \ 0.1}$	1.01	1.04	1.02	1.01	1.03	.995
s.e.	.0298	.0122	.00828	.0051	.00277	.00199
$\widehat{\beta}^{IVFS}_{sample\ \ 0.1}$	.234	.471	.596	.751	.889	.955
s.e.	.00438	.00355	.00306	.00235	.00147	.00114
$\widehat{\beta}^{IV}_{sample\ \ 0.3}$	1.04	1.03	1.02	1	1	.994
s.e.	.0264	.0102	.00704	.00434	.00228	.00172
$\widehat{\beta}^{IVFS}_{sample\ \ 0.3}$	.225	.468	.593	.753	.906	.95
s.e.	.003	.00229	.00193	.00146	.000864	.00067
$\widehat{\beta}^{IV}_{sample\ \ 0.5}$	1.01	.99	.992	.99	.994	1
s.e.	.0246	.00947	.00661	.00413	.00219	.00168
$\widehat{\beta}^{IVFS}_{sample\ \ 0.5}$	.23	.484	.608	.76	.91	.939
s.e.	.00264	.00195	.00165	.00118	.000687	.000526
$\widehat{\beta}^{IV}_{sample\ \ 1}$	.997	.997	.996	1.01	1	1
s.e.	.0236	.0093	.00646	.0041	.00215	.00164
$\widehat{\beta}^{IVFS}_{sample\ \ 1}$	.23	.48	.604	.744	.902	.936
s.e.	.00229	.00167	.00136	.000943	.000515	.00039
Obs Cell	122	365	608	1216	3649	6082
s.e.	21.4	64.1	107	214	642	1069
N.of Obs.	400	400	400	400	400	400
I.Pop	.589	.589	.589	.589	.589	.589
I.Pop.Sam	.589	.589	.589	.589	.589	.589

Population size

Population: 1'000'000 - IV - POP A (1xVar)

S.Rate	0.1	0.3	0.5	1	3	5	10
$\widehat{\beta}^{FE}_{popul}$	1.02	.997	.999	1	.996	.999	1
s.e.	.0629	.0362	.028	.0198	.0114	.00887	.00627
$\widehat{\beta}^{FE}_{sample}$	.0161	.0538	.0864	.156	.361	.495	.677
s.e.	.00876	.00851	.00833	.00816	.00711	.00635	.00523
$\widehat{\beta}^{IV}_{sample\ 0.1}$	.0737	.634	1.63	-16.8	1.62	.89	1.39
s.e.	.832	7.96	8.23	2465	.251	.182	.0681
$\widehat{\beta}^{IVFS}_{sample\ 0.1}$	.0122	.0297	.0937	.137	.366	.499	.647
s.e.	.0158	.0152	.0149	.0145	.0129	.0115	.00936
$\widehat{\beta}^{IV}_{sample\ 0.3}$	-2.58	1.07	.948	1.27	7.59	1.14	1.22
s.e.	106	2.13	2.44	.262	32.6	.0309	.0178
$\widehat{\beta}^{IVFS}_{sample\ 0.3}$	.0164	.0481	.0783	.163	.328	.531	.638
s.e.	.0092	.00883	.00866	.0085	.00736	.00663	.00546
$\widehat{\beta}^{IV}_{sample\ 0.5}$	-.524	-.547	-8.52	2.62	1.04	.986	1.03
s.e.	14	4.76	513	3.78	.0281	.0163	.0106
$\widehat{\beta}^{IVFS}_{sample\ 0.5}$	.0239	.0473	.0739	.154	.376	.518	.702
s.e.	.00722	.00698	.00681	.00673	.00583	.00514	.0042
$\widehat{\beta}^{IV}_{sample\ 1}$	-.665	.893	1.46	2.1	1.07	1.06	1.08
s.e.	9.03	.585	.537	1.71	.0253	.016	.00986
$\widehat{\beta}^{IVFS}_{sample\ 1}$	.0157	.0463	.0794	.154	.361	.489	.648
s.e.	.00533	.00513	.005	.0049	.00418	.00375	.00306
Obs Cell	10.4	31.3	52.2	104	313	522	1043
s.e.	.53	1.33	2.14	4.28	12.8	21.4	42.8
N.of Obs.	400	400	400	400	400	400	400

Immigration Shock graphs

TABLE A1: ESTIMATION VS SAMPLING RATE BY IMMIGRATION SHOCK'S INTENSITY

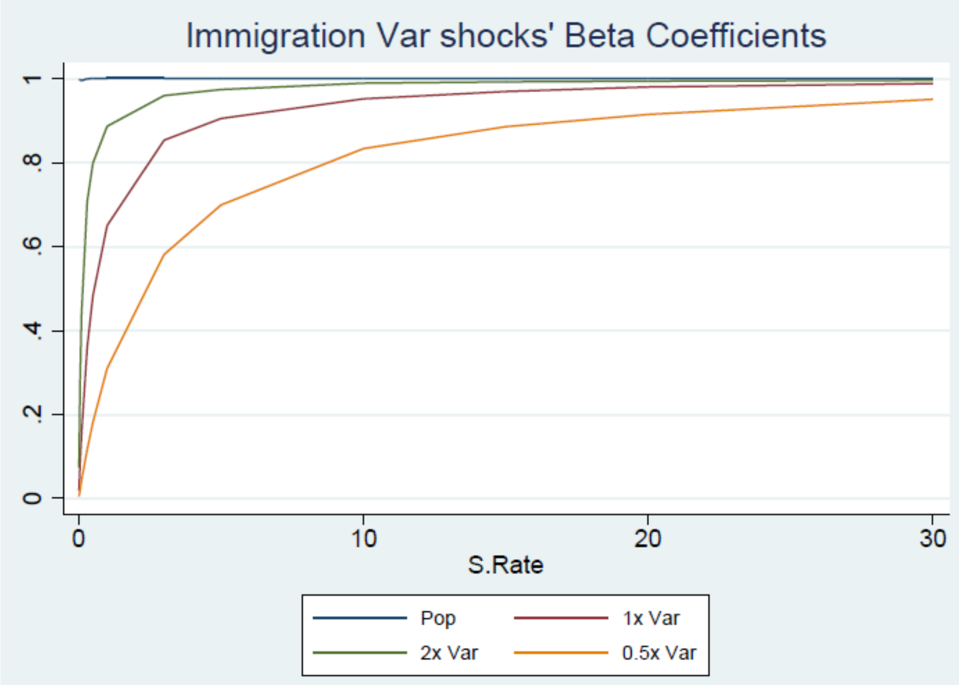


TABLE A2: SSIV FS ESTIMATION VS SAMPLING RATE BY SECOND STAGE SAMPLING RATES – GROUP 1

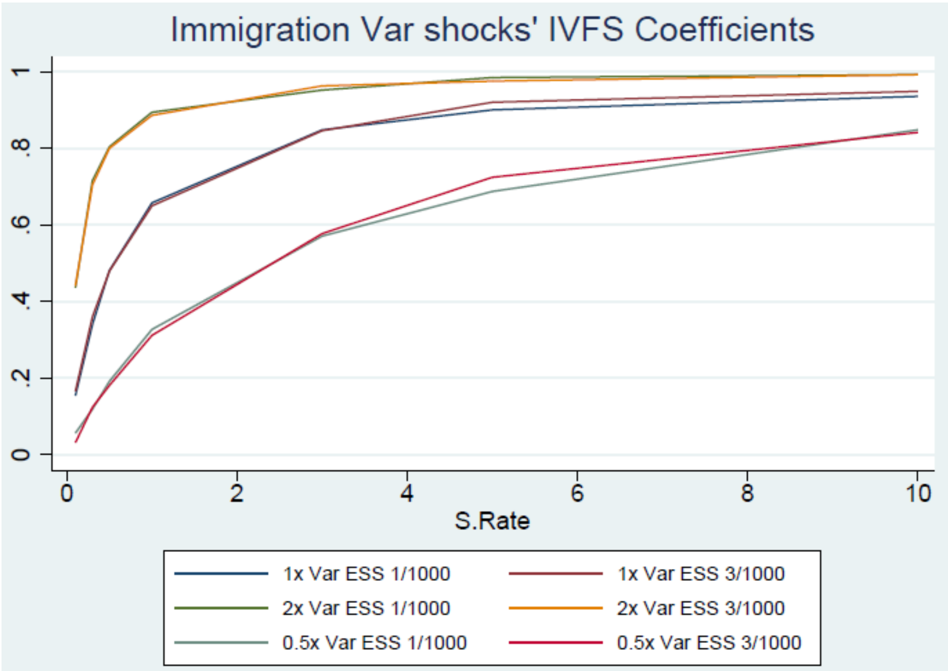


TABLE A3: SSIV FS ESTIMATION VS SAMPLING RATE BY SECOND STAGE SAMPLING RATES – GROUP 2

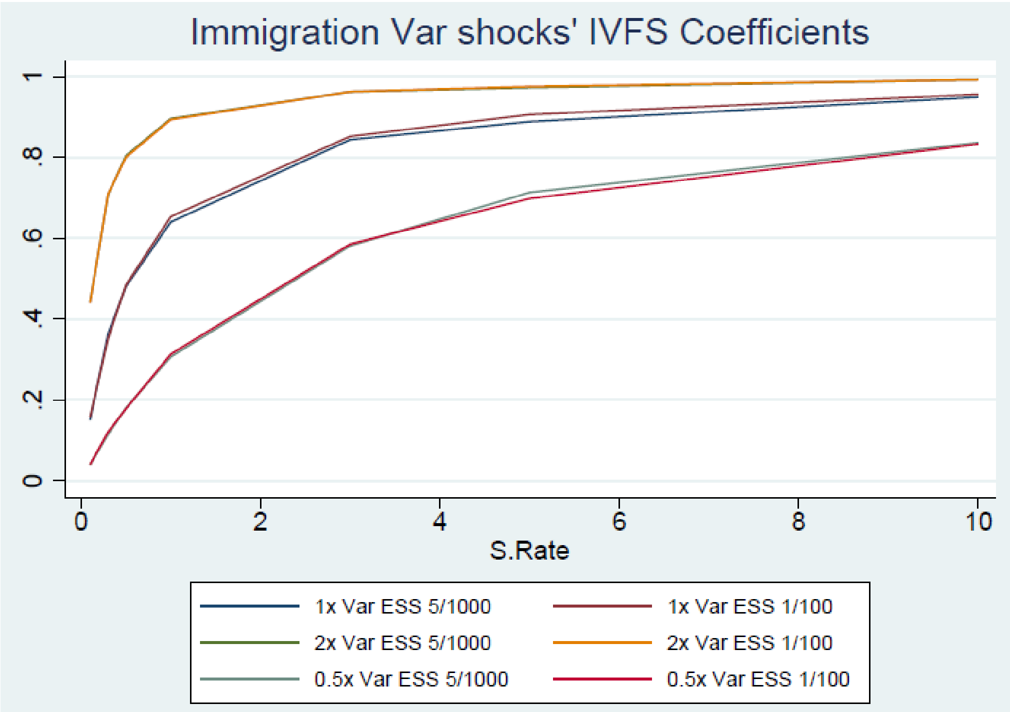


TABLE A4: SSIV FS ESTIMATION VS SAMPLING RATE BY SECOND STAGE SAMPLING RATES – GROUP 3

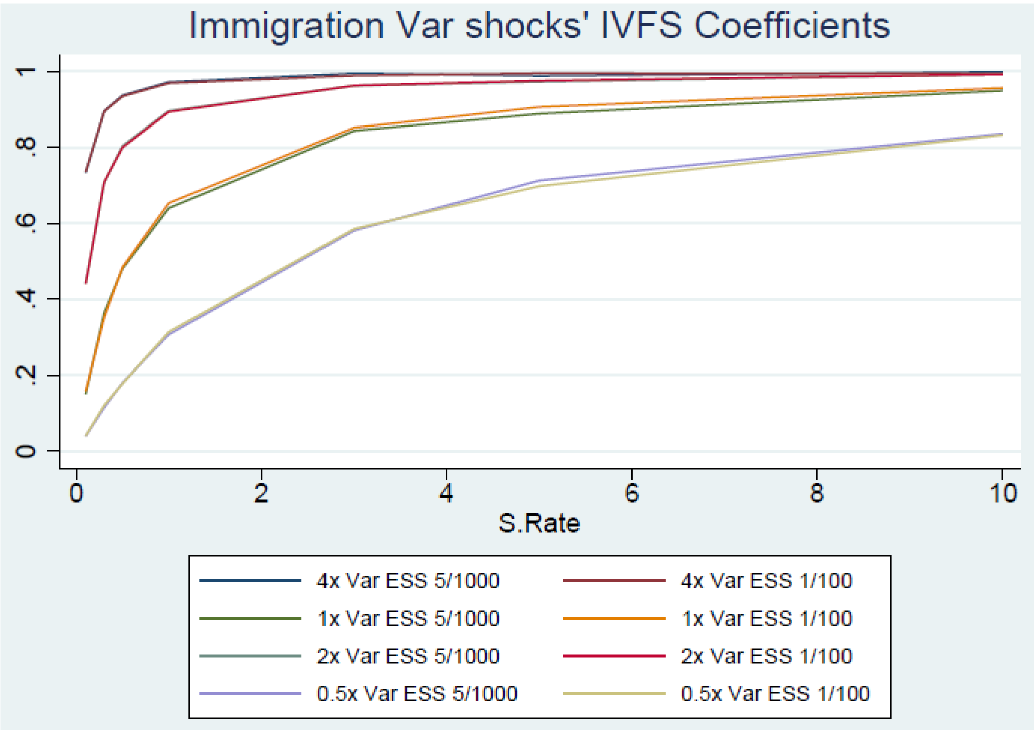


TABLE A5: SSIV FS ESTIMATION VS SAMPLING RATE BY SECOND STAGE SAMPLING RATES – GROUP 4

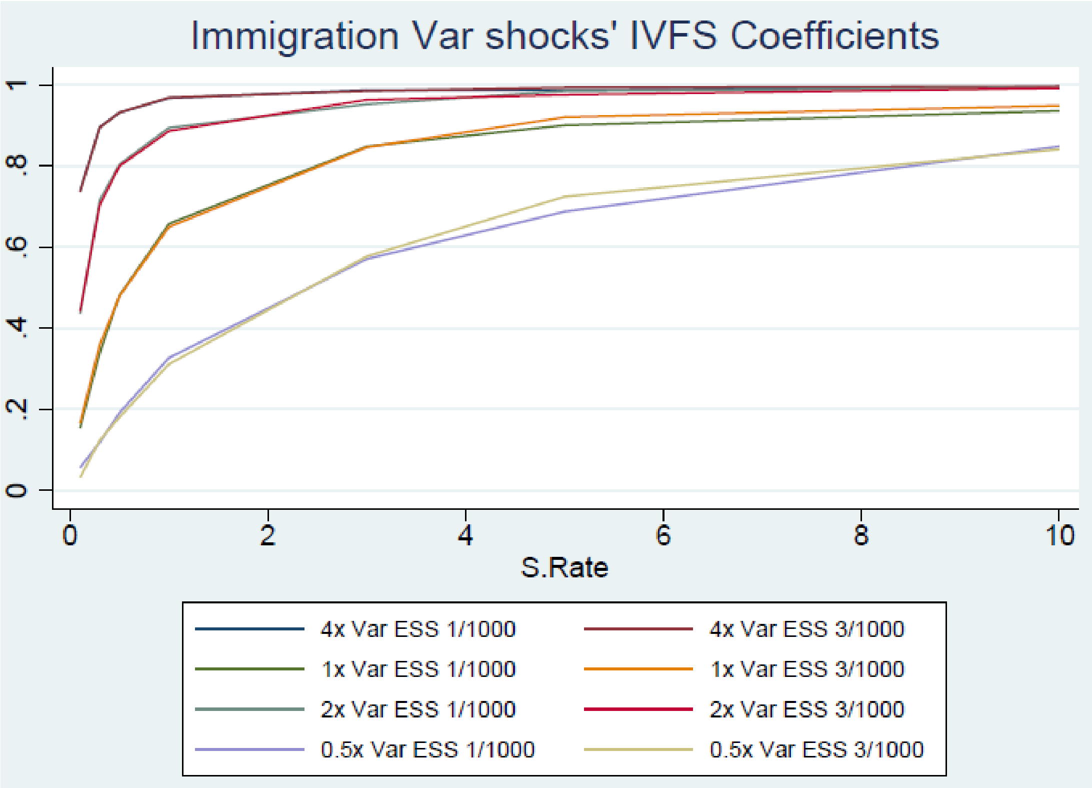
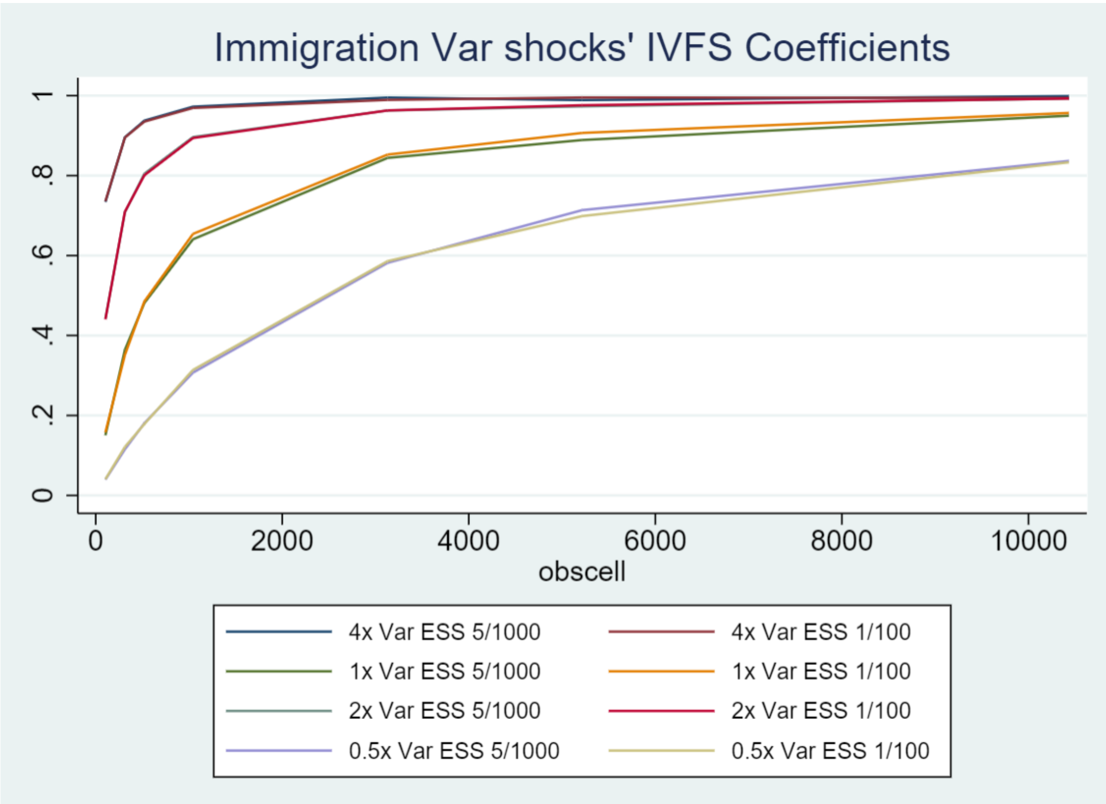


TABLE A5B: SSIV ESTIMATION VS SAMPLING RATE BY CELL SIZE – GROUP 5



Alternative error terms for population graphs

TABLE A6: FE COEFFICIENT BY REGRESSAND'S ERROR PROBABILITY DISTRIBUTION FUNCTION

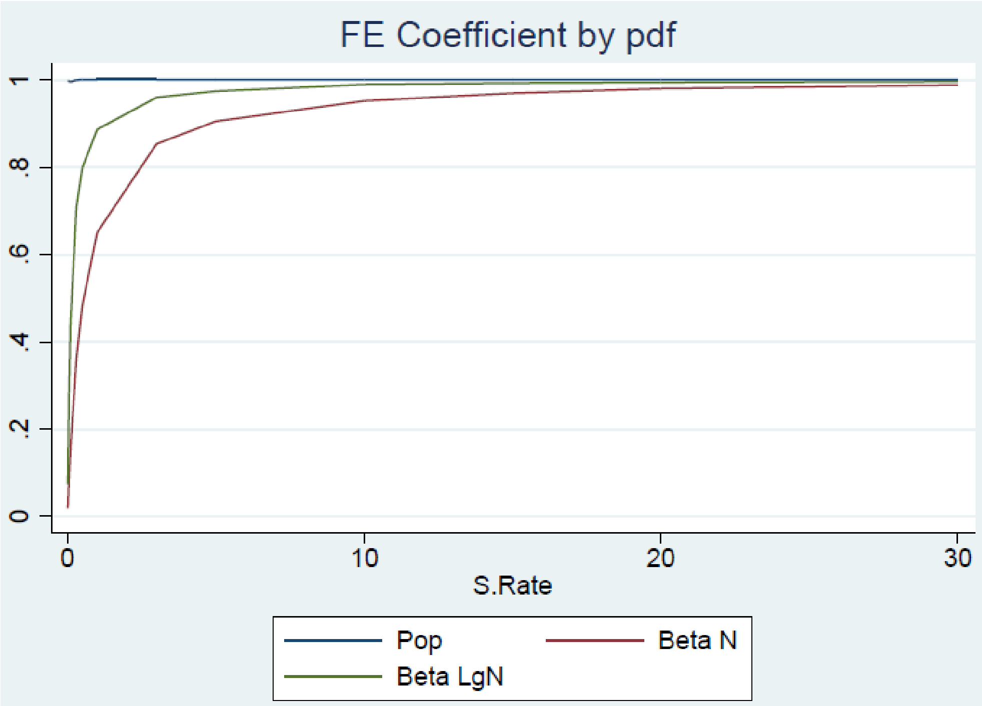
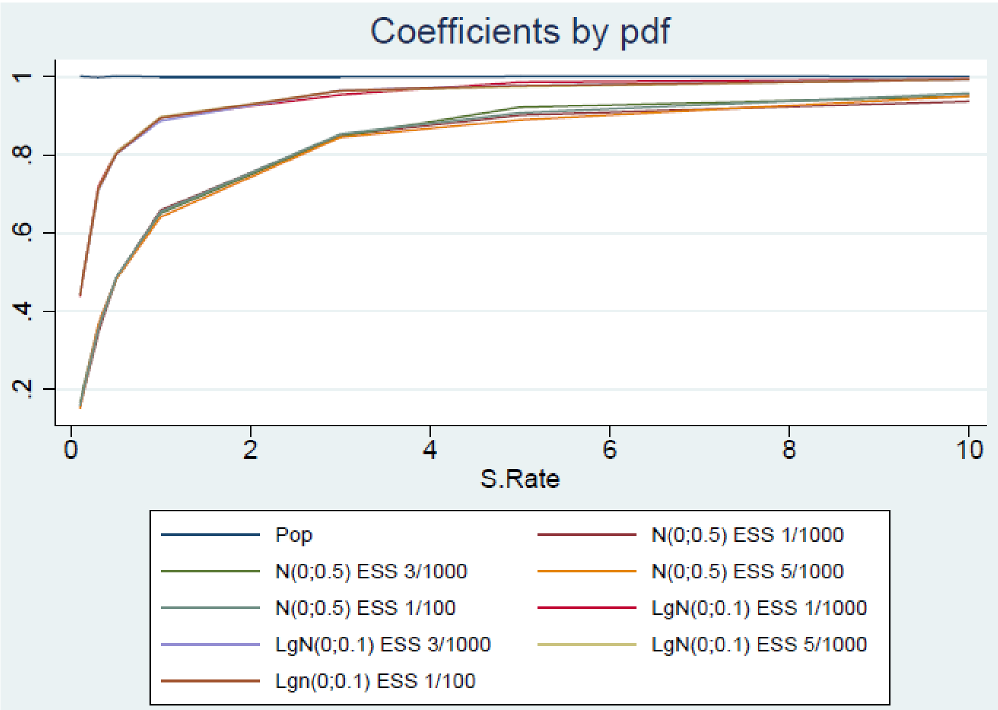
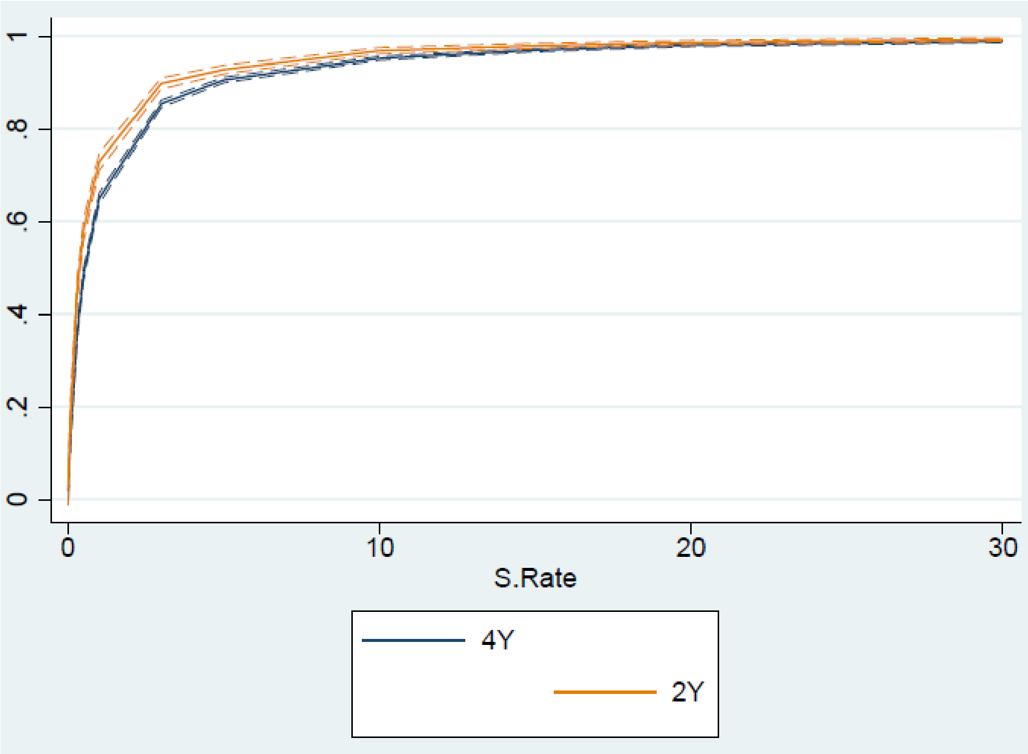


TABLE A7: FSIV COEFFICIENT BY REGRESSAND'S ERROR PROBABILITY DISTRIBUTION FUNCTION



Panel data length graphs

TABLE A8: FE COEFFICIENT BY PANEL DATA LENGTH. Y = 2 YEAR


1
.8
.6
.4
.2
0
0
10
20
30
S.Rate
4Y
2Y

Population size graphs

TABLE A9: SECOND STAGE COEFFICIENT BY POPULATION SIZE AND SAMPLING RATE

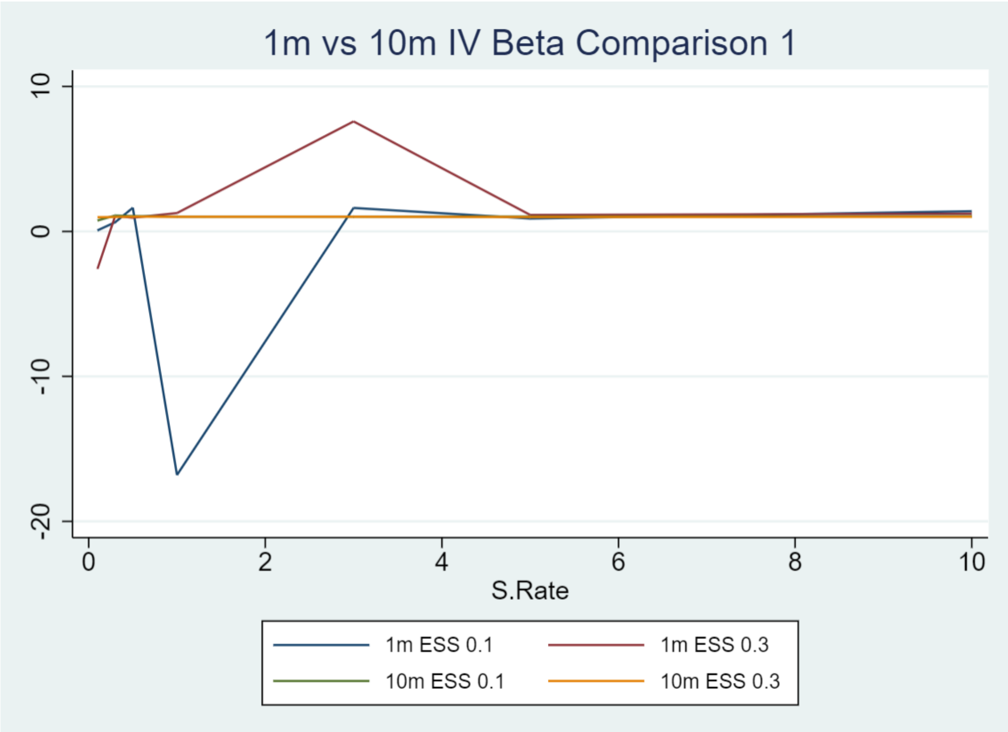


TABLE A10: SECOND STAGE COEFFICIENT BY POPULATION SIZE AND SAMPLING RATE

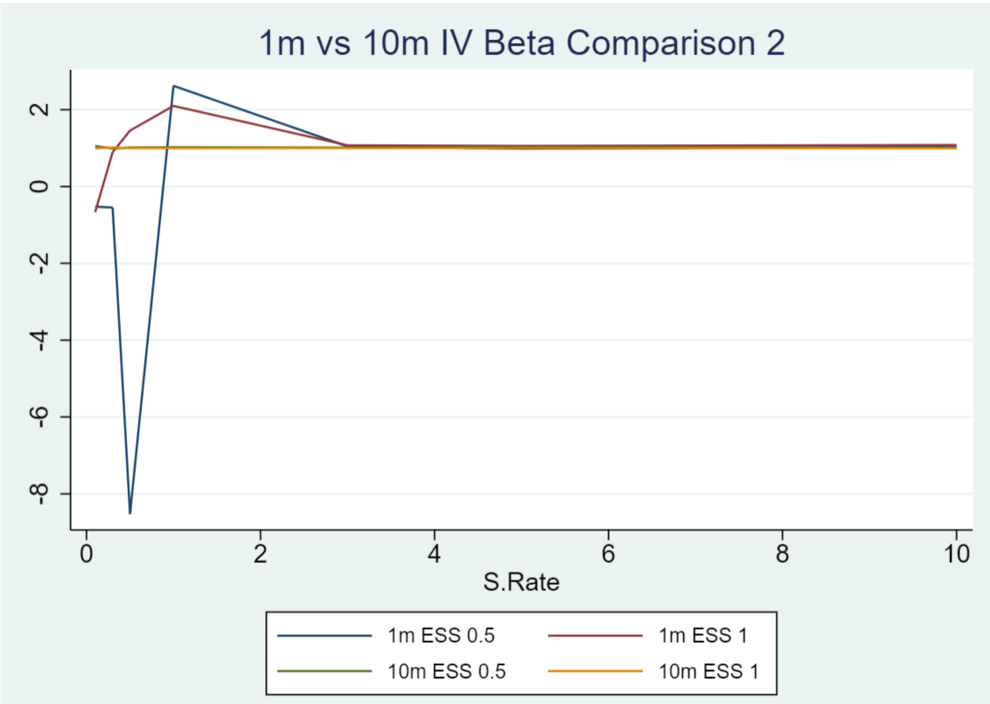


TABLE A11: SECOND STAGE COEFFICIENT BY POPULATION SIZE AND SAMPLING RATE – GROUP 1

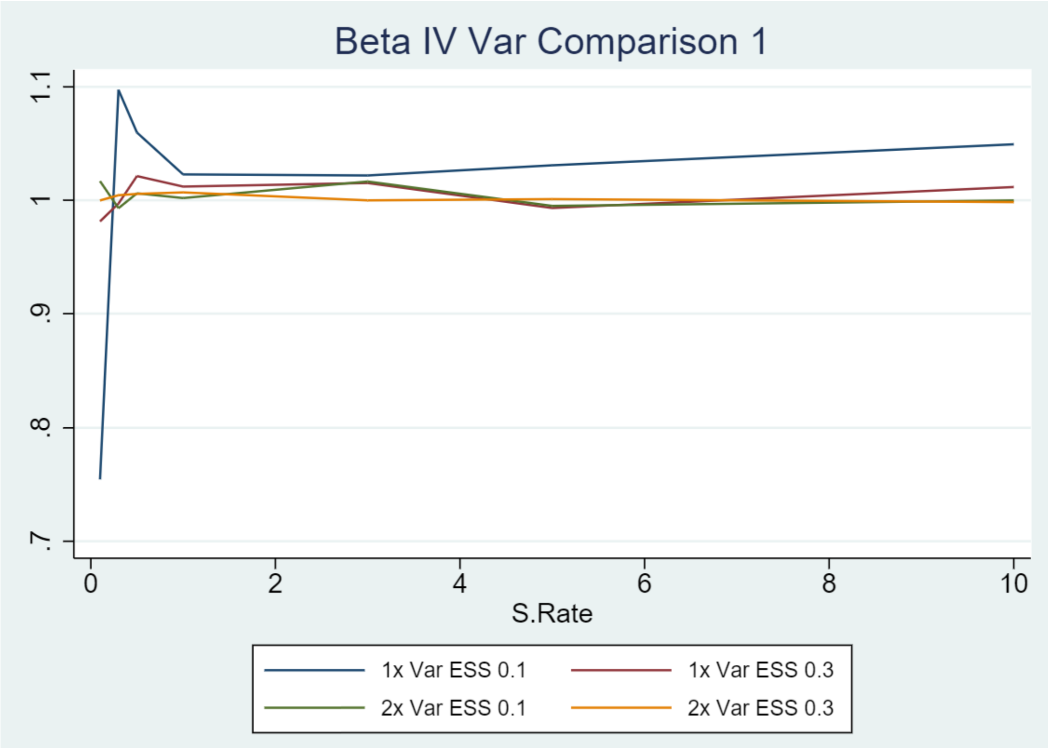


TABLE A12: SECOND STAGE COEFFICIENT BY POPULATION SIZE AND SAMPLING RATE – GROUP 2

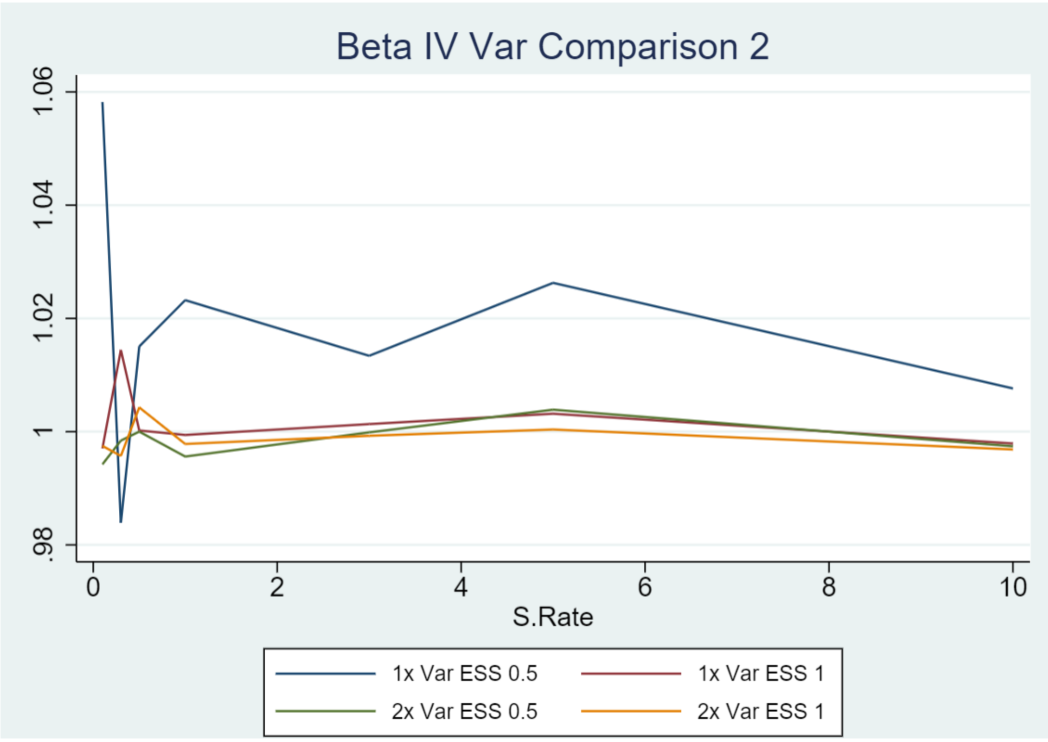
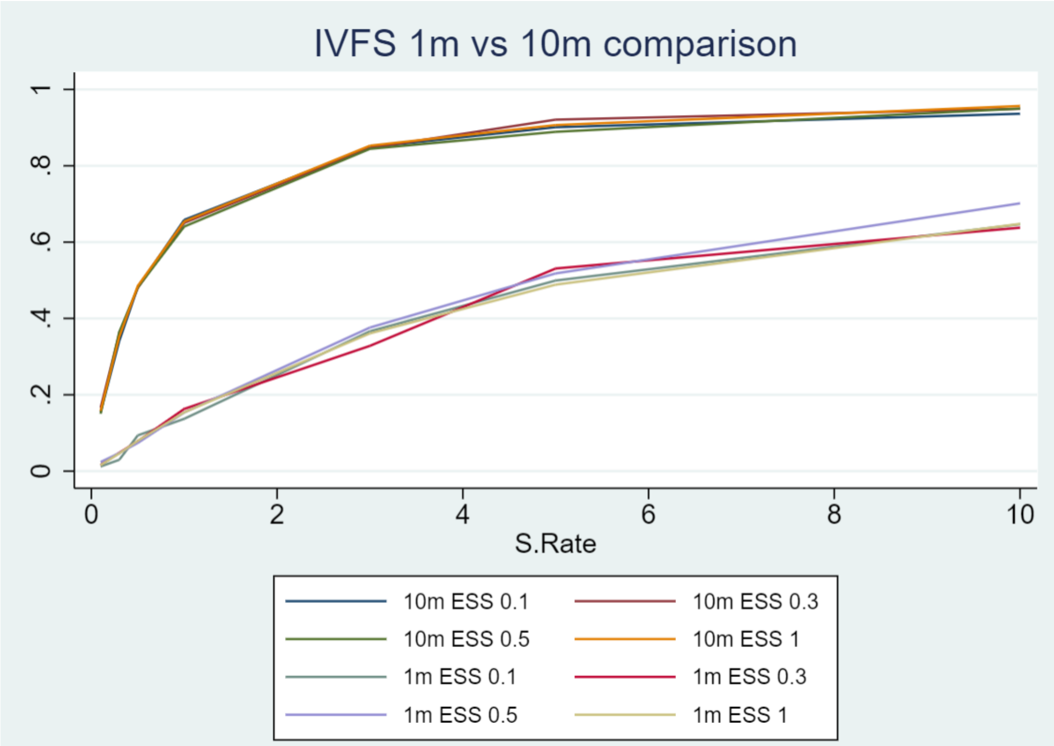


TABLE A13: FIRST STAGE COEFFICIENT BY POPULATION SIZE AND SAMPLING RATE


IVFS 1m vs 10m comparison
0
.2
.4
.6
.8
1
0
2
4
6
8
10
S.Rate
10m ESS 0.1
10m ESS 0.3
10m ESS 0.5
10m ESS 1
1m ESS 0.1
1m ESS 0.3
1m ESS 0.5
1m ESS 1