

### **Summary of the article**

The article examines the relationship between the percentage of migrants to the US which were apprehended by the police and the level of corruption in migrants' origin countries. The authors find a positive and statistically significant correlation between corruption in the country of origin and the percentage of migrants from that country apprehended by the US police, controlling for other country of origin characteristics.

### **General assessment**

The article contributes to shedding some light on an interesting question: to what extent is migrants' behavior in destination countries, in this case the US, influenced by values and norms acquired in their origin countries.

The authors could however do a better job in explaining the choice of their right hand side variables, possibly by providing some theoretical background for their hypotheses, in justifying the way some of the control variables were constructed, and in interpreting the results. Writing could also be significantly improved. See comments below.

### **Major Comments**

The paper examines the probability to be apprehended for immigrants from different origin countries but no literature is cited on the determinants of the probability to be apprehended, not necessarily for immigrants, but in general. What is known about the impact of laws and regulations, education, socio-economic background, cultural background, etc, on the probability to be apprehended? Insights from this literature could be used for explaining the choice of the explanatory variables.

Section 2 Opinions and Facts is in my opinion too broad and insufficiently focused. It could instead briefly present immigration figures for the US, possibly by origin countries and skill levels, as well as a brief review of studies examining immigrants' illegal behavior.

The authors should justify their choice of the sample period, 2009-2011, as it seems to me that most of the data is available for a larger number of years.

The authors should better explain, based on what criteria, Mexico, Honduras and Guatemala were excluded from the sample and how results change when these countries are kept in the sample.

I did not understand why the authors use GDP and not GDP per capita for measuring wealth. GDP depends on the size of the population, so not only is it an inaccurate measure of wealth, but it can capture the effect of the population size in the origin country, which may be correlated with emigration rates and selection of migrants.

It is not clear why labor force growth in the origin country should affect emigrants' criminal behavior in the US. The authors should explain this clearly. It is also not clear why the stock of immigrants in the origin country should affect emigrants' illegal behavior in the US. The authors mention familiarity with immigration laws as a possible justification, but immigration laws are obviously not the same in origin countries and in the US and moreover, it is not clear why familiarity with those specific laws should affect illegal behavior in general. It could be that the level of immigration in origin countries captures other origin country characteristics, such as rule of law: countries with better rule of law attract more immigrants and citizens of such are less likely to commit crimes in the US?

The vocabulary used in the paper should be made more academic. Many sentences/phrases seem inappropriate for a scientific journal, such as: "corruption...puts every possible obstacle towards humanity to confront with....leads nations into the darkness"... "there is no perfect answer...because every nation has people who are honest and dishonest....it is difficult to measure the ration of honesty and dishonesty of people".... "...corruption should not be regarded as something justifiable because every single illegal act has its negative effect on the economy and the society"... "fortunately we found somehow convincing solution to this puzzle" and many others. Such sentences should be replaced with more neutral, academic terminology.

Some sentences are misleading, such as this one, in the first paragraph of the introduction: "One may feel curious why the gap among countries is so wide in terms of economic performance and lifestyle"...this question is too far from the point of the paper to be in the first paragraph of the introduction. The introduction could be made sharper and more to the point.

When interpreting the results, some rather unclear extrapolations are made, such as: "If El Salvador would reduce its corruption level from 65 to 55, the average number of apprehended Salvatorians in the US could be reduced to about 89 people"...does this extrapolation assume that when corruption in El Salvador is reduced, the same number and type of people would leave the country? Is it 89 people or is it 89 out of 100.000 people?

The authors transform many control variables into dummies, with the justification that this reduces the collinearity problem. This choice should be better justified, as, from the collinearity matrix, correlations do not seem to be excessively high and a lot of information is lost by doing so. Moreover, the authors should justify the precise way in which the dummies were constructed: why does the dummy take values 1 and 10 instead of 1 and 0 as it is usually done, and why do they chose the median value for defining the values of the dummy? This implies that a country with a homicide rate slightly higher than the median would take the value 10 for the homicide dummy and a country with a homicide rate slightly lower than the median would take a value one. This is a very specific

choice and should be better justified. Robustness of results to using all information in the variables instead of dummies should be discussed.

I do not understand what the authors mean by: “We expect that people coming to the United states from countries where prostitution is legally allowed by law are more likely to practice prostitution **due to probability of higher earning chances**”. As for other control variables, the authors should be more clear about the intuition or theoretical reason for including this variable in the right hand side of the equation

The authors divide origin countries in two groups, more and less corrupt, based on the median corruption level in origin countries. The authors should justify this choice. An alternative classification could be countries with higher corruption levels than the US and those with lower corruption levels than the US. Moreover, the sample size when dividing the sample into more and less corrupt countries is really small, only 52, putting some doubts on the robustness of the regression results. Instead, the authors could include a dummy for very corrupt countries, or countries with corruption higher than the US and estimate the model on the whole sample.

The authors do not provide any convincing explanation for the puzzling result that in the sample of more corrupt countries, corruption in the country of origin is not a significant predictor of apprehensions in the US.

The explanation given for the puzzling result that migrants from countries with higher human capital are more likely to be apprehended is unconvincing. The authors cite the results from Ariu and Squicciarini (2013), that highly skilled people are more likely to move abroad if their origin country is highly corrupt in their explanation, but Ariu and Squicciarini (2013) in fact also show that unskilled migrants are even more responsive to corruption, i.e. even more likely to leave corrupted origin countries. An alternative explanation could be that it is easier to emigrate to the US from countries with higher levels of human capital, because of both budget and visa constraints. Could it be that migrants from these countries are less selected than migrants from poor countries with low levels of human capital, where only a small and highly selected minority is able to migrate to the US?

In order to better understand and interpret the results, it could be useful to present regression results when the main control variables are added one by one in the right hand side of the regression equation.

The authors state that their results contradict the results of Dimant (2013) that immigrant flows from more corrupt countries is associated with increases in OECD countries' corruption levels, but that is not correct: the authors do not show anything about corruption in the US so there is no contradiction.

As a robustness check, the authors could include in the right hand side of the regression equation the total number of migrants from each origin country.

The result that the high income and low income revenue group dummies are both significant with negative signs, while the upper and lower middle income dummies are not significant is not very intuitive and a theoretical or intuitive explanation should be provided for it.

In the conclusion, it seems to me that the authors push the policy implications of their analysis a bit too far when stating “we encourage immigrant receiving countries to develop specific screening tests based on the degree of corruption in immigrants’ home country due to the probability of immigrants being subject for carrying their immoral attitudes to the destination country”.

I do not understand the meaning of the last sentence in the conclusion.

#### **Minor comments**

I find the title a bit misleading. It is not clear from the title that it is the corruption in migrants’ origin countries (and not corruption in the US) that will be analyzed in relation with apprehensions in the US.

Some of the terminology used is non standard, ex: “multi-front regression analysis”, “sub-variables”, “an entirely clean country”, “unlawful homicides” (are there also lawful homicides?).

What is meant by “apprehension level is not directly observable phenomenon”?

Some information given in the paper seems irrelevant, such as explaining the objectives of the Department of Homeland Security , of the Migration Policy Institute and of Transparency International.

I do not understand the following statement: “The ratio of APPR is multiplied by 100000 for the sake of using the data in LOG form”.

I did not understand why transposition of the scale in (2) is necessary for making interpretation of the results easier.

I think regression results based on standardized variables could be included in the Appendix instead of being discussed in an entire section.

The variable PRST is included as a dummy. I do not understand why the authors interpret the size of its coefficient in terms of standard deviation increases in this variable.