UNDERSTANDING BRAIN DRAIN:
THE INFLUENCE OF EMIGRATION ON CORRUPTION

by

GIOVANNI RIGHI

(Under the Direction of Laura V. Zimmermann)

ABSTRACT

This paper studies one method by which the emigration of skilled workers contributes to the emergence of corruption. It first builds on a political agency model and connects that to a simple model of media investigation. In doing so, it formalizes a conception of public scrutiny and helps explain the remarkable convergence in rankings of corruption and output. The paper then takes a step towards estimating the aggregate effects of emigration on corruption with the Ferraz and Finan Brazilian audit data. The estimation and theoretical contribution are incomplete because I do not test the media investigation mechanism. Nonetheless, the estimated effect points in the expected direction.

Index words: Corruption, Emigration, Journalism
UNDERSTANDING BRAIN DRAIN:
THE INFLUENCE OF EMIGRATION ON CORRUPTION

by

GIOVANNI RIGHI

Approved:

Major Professor:  Laura V. Zimmermann
Committee: Ian Schmutte
              John Turner

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016
Acknowledgments

Thanks to the professors that have helped through the thesis process and commented on this work: Laura Zimmermann, Ian Schmutte, John Turner, Nicholas Magnan. Thanks also to John Drake, and for the support of parents and friends. I wouldn’t be where I am today without all of you.
# Contents

Acknowledgments ........................................ iv
List of Figures ........................................ vi
List of Tables ......................................... vii
1 Introduction ........................................ 1
   1.1 Brain Drain Literature .......................... 4
   1.2 The Setting in Brazil ......................... 7
2 Theory .............................................. 8
   2.1 Politician’s Career Model .................... 9
   2.2 Solution to the Model ......................... 12
3 Empirics ............................................ 17
   3.1 Data ........................................... 17
   3.2 Empirical Strategy ............................ 20
   3.3 Results ....................................... 23
   3.4 Extensions .................................... 26
4 Conclusion .......................................... 28
5 Appendix ........................................... 31
   5.1 An alternative solution ...................... 31
   5.2 Investigation Model ......................... 32
   5.3 Note on ryacas ................................ 34
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CPI versus migration</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Audit map</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Comparative statics</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Irregularity Distribution</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Emigration versus distance</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>Causal Diagram</td>
<td>35</td>
</tr>
</tbody>
</table>
# List of Tables

1. OLS Results ................................................................. 23
2. Emigration across FPM levels ........................................... 24
3. IV Results ................................................................. 25
1 Introduction

Akerlof’s seminal paper predicted that information asymmetries could be employed to uncover features of many markets and social settings (Akerlof 1970). One such setting in which dishonesty and information asymmetries may drive honest individuals out of the pool of eligible candidates is in the election of politicians. Extending Akerlof’s analogy of dishonest vendors, when honest candidates are not remunerated for their virtue, the quality of the pool of candidates that chooses to run will decline. This under-recognition would occur if voters do not have enough information about candidates, either of their inherent type or about their actions. This paper leverages the explanation of information and argues why information is endogenously lacking in places from which there is emigration. In doing so, it proposes one explanation for the emergence of corruption.

To leverage this explanation, we must define how information is generated in political markets. Most second-hand information about candidates and public scrutiny reaches the public through the media. Glaeser, Gentzkow, and Shapiro (2006) model how media has evolved in response to market forces more generally, and their model is useful because it includes a variable that can be interpreted to change with emigration. They show that decreasing print costs, expanding markets, and increased competition fostered the rise of an independent and informative press in the US at the turn of the 20th century. The last sentence of that paper reads, “It seems a reasonable hypothesis that the rise of the informative press was one of the reasons why the corruption of the Gilded Age was reduced during the subsequent Progressive Era.” I provide evidence for this hypothesis.
This paper makes both a theoretical and an empirical contribution to the literature. Since it is not immediately clear how emigration might affect corruption, I present a model extending on Brollo et al. (2013) to formalize the hypothesis that increased public scrutiny reduces the rents that incumbents extract. I suggest how emigration may reduce public scrutiny which is understood as investigation by the media. The form of investigation I model, which is about incumbents’ actions, has been shown to increase voter welfare (Besley 2006, p. 128-131). I argue that investigation decreases because emigration shrinks the pool of paying news subscribers, decreasing the payoff for media companies to investigate. It does not consider the skill-bias in emigration that is typical in the brain drain literature, and so my hypothesis measures a lower bound of the effect of emigration. I then test this hypothesis with recent randomly collected data from Brazil. These are the beginnings of some theoretical and empirical arguments, and I comment on specific data and theoretical analyses that can support the hypothesis.

The empirical section of the paper tests the basic prediction that corruption is a result of emigration, but it does not begin to test the mechanism. This would require adding some covariates regarding media investigation, but I do not have data on these. Some theoretical justification about the direction of cross-partial effects is also warranted, but I do not yet have this, so I measure only the aggregate effect of emigration. Naive regressions are promising in that they show a positive correlation between emigration and corruption, but there are endogeneity problems. I propose an instrument for emigration based on proximity to large urban centers to remedy this. Those results further support the hypothesis of the link between emigration and corruption that I propose, but the results are interesting for more reasons than simply establishing this link.

The explanation I propose also sheds light on the remarkable correspondence between corruption and output across countries\(^1\) because people are expected to emigrate

\(^1\)See, for example, Shao et al. 2007
in a direction that reinforces corruption. Indeed, the correspondence is what inspired studying the mechanism of emigration. Emigration has been proposed as major reason for why core-periphery patterns emerge (Krugman 1991), and the global periphery are also areas with more endemic corruption. The connection between emigration and output is explained by the endogenous concentration of manufacturing activity, while this model relies on changes in information that happen as a result. This is not the only effect that emigration may have on corruption, but it is one of the channels through which it acts. When that corruption imposes social costs (Shleifer and Vishny 1993), it may further encourage the emigration of workers. This idea could be pursued to develop a model where corruption emerges endogenously as more people move, but my theory is not yet this developed.

This connection to corruption geography distinguishes this paper from, and adds to, the substantial brain drain literature. Fifty years of studies in that direction have mostly focused on the link between corruption and output. Much of the literature even focuses on the effect in a single direction, that is, the effect that emigration of skilled workers has on poverty, poor institutions, etc. (Docquier and Rapoport 2011). I focus on an effect that is in the same direction but which focuses on corruption, a rather different topic.

Randomized controlled trials have also recently focused on questions of political economy. In particular, Benjamin Olken (2007) conducted an RCT in Indonesia to monitor corruption, showing the effectiveness of independent audits in combating corruption. The policy prescriptions are identical to ones that might come from positive results within this paper, but one should be cautious and avoid a red herring in interpreting Olken’s alternative scheme of “grassroots participation.” It may lead one to believe that participation by the citizenry is unimportant and to prioritize audits over other forms of public participation (although, of course, Olken has not tested all possible alternative grassroots schemes), in the way chastised by Acemoglu and Robinson (2012). The
mechanism I present instead allows for audits to facilitate public participation by providing information to the public that would not otherwise be discovered.

The rest of this paper will be organized as follows. The next subsection gives a background on the setting in Brazil. In section 2 I lay out two models, one of political agency and one of media investigation, and I show that they are connected by an important parameter, $q$, the probability of discovery of an accounting irregularity. In section 3, I briefly describe the empirical strategy I use in section 4, which describes the estimation results. That section also comments on some possible extensions to this study that can improve the results.

1.1 Brain Drain Literature

It may not be surprising that there is so little analysis of corruption in the brain drain literature. At first glance, the data don’t show much. Note in Figure 1 that there is almost no significant relationship between emigration and corruption. This is not, however, problematic for the hypotheses in this paper. Corruption is certainly a problem with many-dimensional causes, and so brain drain may be sufficient - by the channel I propose - but not necessary. Consider for example the old Soviet republics, which show very little emigration but considerable corruption (Docquier and Rapoport 2011). The story which explains corruption in those places is more likely related to political economy after the disintegration of the Soviet Union. However, in the countries that do witness brain drain, one might expect more corruption through decreasing investigation. Even this does not control for enough factors, however. In the upper right-hand of Figure 1, there are countries like the UK, Austria, Cyprus, and Portugal, which are periphery countries of the EU. It is unsurprising that they also have quite good corruption scores. Clearly a cross-section analysis of this kind is difficult, even leaving aside the fact that the Corruption Perceptions Index is a poor measure of corruption.
Figure 1: Comparison of corruption perceptions index and emigration rate across countries. Emigration data are those from Docquier and Rapoport (2011, available at http://perso.uclouvain.be/frederic.docquier/filePDF/DLM_PDR.pdf). Note that rate is the stock of skilled workers abroad, rather than the flow (see the descriptive note on the website).

It has been shown before that having citizens who are educated in democratic foreign countries increases a measure of democracy\(^2\) (Spilimbergo 2009), but foreign-educated individuals do not always return home. In countries where they stay away, freedoms that facilitate investigation, like freedoms of conscience and press, are more likely to be curtailed (Brunetti and Weder 2003).\(^3\) What happens to the home countries in that case, when those individuals remain away? This paper attempts to answer that question using a strategy that is different from the cross-section strategies that are frequently used in brain-drain literature.
Figure 2: Map of irregularity counts by Brazilian municipalities. The width of each circle matches the count, and black points are major population centers.
1.2 The Setting in Brazil

In 2003 the Lula da Silva government started an anticorruption program under the Controladoria Geral do Uniao (CGU), through which it randomly audits municipal governments. Fifty to sixty cities are chosen per year through lotteries. Auditors are then sent to those cities to examine accounts and documents, meet with members of the community and municipal councils, and inspect the quality of public work construction. A detailed finding of reports with itemized irregularities is then sent to the federal CGU office, the Tribunal de Contas da União, to public prosecutors, and to the municipal legislative body. Summaries of the reports are also published on the internet and disseminated to media (Ferraz and Finan 2008).

The funds which are under audit are directly under the responsibility of the mayor and are largely federal funds. For cities with fewer than 50,000 residents - most of the cities in the dataset used in this paper - 40 percent of the budget is a transfer from the federal Fundo de Participação do Municipios (FPM), and a much smaller share comes directly from taxes (Brollo et al. 2013). The FPM is allocated according to a series of cutoffs that increase with population size. Brollo et al. have shown that corruption increases around the cutoffs for different FPM transfer amounts; consequently, I control for FPM cutoffs in estimations for this paper.

Those authors also describe a few features of municipal institutions. Control of the funds is typically left to mayors rather than city councils, which act as oversight authorities. Thus mayors are the appropriate political actors whose actions we need to model. Nonetheless, with the success of the anticorruption program, the oversight role of city councils has strengthened. For example, some cities have created independent

---

2Specifically the Freedom House Political Rights Index
3Brunetti and Weder (2003) show the correlation between free press and corruption
anti-corruption commissions or impeached mayors. I presume that for the data used here, the mayor is still the relevant political agent.

The irregularities counted here consider many possible definitions of corruption. Ferraz and Finan describe that

“most corruption schemes used by local politicians to appropriate resources are based on a combination of frauds in procurements, the use of fake receipts or ‘phantom’ firms, and over-invoicing the value of products or services. In addition, the audit reports also suggest that some politicians simply divert resources for personal purposes... Illegal procurement practices typically consist of benefiting friendly or family firms with insider information on the value of a project, or imposing certain restrictions to limit the number of potential bidders.”

The emigration data used here are net migration counts from Brazil. Docquier and Rapoport show that brain drain was in fact quite low in Brazil in the early 2000s; in fact, the country had one of the lowest rates in the world (2% of college graduates). Since mine are municipality-level data, we are considering the movement of emigrants within rather than across countries. Migration within the country still permits emigrants to remain much more connected to the media market of their home country than if they moved across countries, thus the political implications of internal migration are plausibly different from international emigration. Keeping this caveat in mind, this analysis may still be able to shed some light on what effect emigration might have.

2 Theory

In Brazilian municipalities, the crucial political actor is the mayor. He or she has considerable power in shaping the budget and implementing expenditure programs (Brollo
et al. 2013). This supports the use of a (career concerns) political agency model like these, but Brollo et al. do not consider the degree of scrutiny to which mayors are subject. I will try to add that in this analysis.

This work follows a small portion of the large literature in political agency models, drawing on work by Brollo et al. (2013). That paper shows that additional transfers from the federal government to municipalities increase corruption. They show that this is due to weaker electoral punishment, and that for a given transfer there is more corruption when electoral challengers are less educated. This first point shows the importance of their modeling. It is not intuitively clear what the effect of increased transfers from the federal government should be. On the one hand, funds might elicit increased scrutiny thus decreasing rents, but on the other, they increase the incentives for a politician to extract more rents. I will point out at what point my model departs from theirs and my reasons for doing so (otherwise, I make all the same assumptions). In particular, it would be helpful for this analysis to show some peculiar cross-effects that may be estimated through this model.

2.1 Politician’s Career Model

Consider an incumbent municipal politician who sets some policy at $t = 1$, subsequently faces elections, and sets policy once more at $t = 2$. Let us assume that this politician is the one with complete control over municipal contracts and the only one who extracts rents from the city budget. The government has constant $\tau$ in federal transfers which it allocates between public goods $g_t$ and rents $r_t$. Its competency in providing the public good is $\theta$, so that

$$g_t = \theta(\tau - r_t). \quad (1)$$

Competence $\theta$ is a random individual feature that is realized only after carrying out a policy. It is drawn from one of two alternative uniform distributions with density $\xi$ (for
both) and mean $1 + \sigma$ (with $0 < \sigma < 1, \sigma^H = \sigma = -\sigma^L$). These distributions are always known. Thus the two values of $\sigma$ correspond to some variable that voters can use to know something about their candidates before observing their policy outcomes. This captures one important feature of political agency problems. Notice also that individuals of type $H$ are more competent than those of type $L$ on average, but in some instances a less educated individual may be more competent than one that is more educated (assuming that the densities are sufficiently low for the distributions to overlap). Another important feature is that voters’ imperfect information about the incumbent’s competence creates an incentive to appear competent so as to please voters.

Having voters choose politicians solely based on competence is imperfect. There is a whole camp in political theory that thinks politicians should be elected on the basis of having shared motivations, not for being experts on what their job will entail, say, construction projects. More importantly, one might expect that voters want a candidate’s choice of $g$ to conform to their preferences and that this value might differ across candidates. An open question is how much variation there is in proposed $g$.\footnote{This may be an interesting challenge to the hypothesis of Brollo et al. It is possible that their same predictions could be generated by considering that less competent candidates will offer more $g$ to be politically relevant.} Leaving this caveat aside, I assume that competence is the only parameter over which voters choose candidates.

During the incumbent’s reelection, he or she will face challengers that are also of unknown competency but of a known type. Let the challenger be of type $L$ with probability $\pi$ and of type $H$ with probability $(1 - \pi)$. Then the incumbent will expect that $\hat{\sigma} = \pi \sigma_L + (1 - \pi) \sigma_H$ at the time policy is set and before any challengers have stepped up.

Now that we have described candidates and how they capture rents, we should also consider that they can be punished. Let $d(r_t) = q \cdot \frac{r_t}{r}$ be the probability that the government official is caught. This is where my model departs from that of Brollo et al.
because those authors consider a probability of capture as an unrelated constant 
\(d(r_t) = q\). On the contrary, it appears plausible to that larger rents will be more obvious
to the public and so more likely to be discovered and disseminated in the media. Indeed,
Brollo et al. suggest they will use the value \(d(r_t) = q \cdot \frac{r_t}{\tau}\) but drop the \(q\) later in the paper,
probably for analytical tractability. Keeping \(q\) also gives the attractive functional form
\(qr_t \cdot (-r_t) = -qr_t^2\) to the expected loss from being discovered when discovery only requires
the candidate to pay back the rent. In this case, the expected loss will be \(-qr_1r_2\): we will
assume that they are uncovered with a certain probability from their actions in period 1,
but they lose the rent from the following election (in this case, they do not pay back rents
from the first election).

In fact, loss from being discovered might be more like some function \(qr_1\lambda\) with
\(\lambda > 1\). This reflects that there are likely to be costs beyond having to return stolen money,
such as a sanction in the labor market. Brollo et al. use a model of this form to compare
across candidates with different types \(\lambda\), but we can ignore this factor for simplicity.

Politicians will want to maximize their utility over the two periods,

\[ U_2 = (1 - qr_1)(r_2 + R) \]
\[ U_1^J = (1 - qr_1)r_1 + R + p^J U_2. \]

\(p^J\) gives the probability of reelection. Notice that it is the only variable here that depends
on the politician’s type. This will be useful because the probability of reelection is modeled
as the probability that the particular realization of the incumbent’s type, \(\theta\), is greater than
the expected type of his or her challengers. In the first period, rents are either \(r_1\) or 0, so
the expected rents are \((1 - qr_1)r_1\). If the incumbent arrives to the second period, however,
he or she takes all of \(r_2\) without penalty. To reflect the fact that utility drops to zero if
discovered, this is scaled by the probability of remaining under the radar.

In summary, then:
• Incumbents draw, through their actions (distributing \( g_t \)), a value of competence, \( \theta \), from a known probability distribution after having set policy. This competence is only known to the incumbent.

• They claim their rents while setting policy according to their utility maximization problem.\(^5\)

• The opponent’s type - but not his or her competence - is revealed and elections are held. Voters observe \( g_1 \) but not \( r_1 \) while voting.

• When the politician is reelected, he sets \( r_2 \) and potentially faces an audit.

### 2.2 Solution to the Model

The model is solved backwards, so that a politician sets the optimal rent in period 2 and accordingly sets the proper rent for period 1. Assume that rents cannot exceed a given upper bound, \( r_t \leq \psi \tau \). The politician will set \( r_2 \) to this level so as to maximize expected utility. We want to find, then, \( \frac{\partial U}{\partial r_{t}} \) for each period. Starting in period 2, the agent sets

\[
r_2 = \psi \tau
\]

because \( \frac{\partial U_2}{\partial r_1} \) since this is the maximum possible rent. By construction, the politician doesn’t face a penalty for rents in period two once they are reelected. This means that

\[
U_2 = (1 - q \frac{r_1}{\tau}) (\psi \tau + R).
\]

To solve the first-period problem, the politician will set

\[
\frac{\partial U_1}{\partial r_1} = (1 - 2q \frac{r_1}{\tau}) + \frac{\partial p^J}{\partial r_1} U^J_2 + \frac{\partial U^J_2}{\partial r_1} p^J = 0.
\]

\(^5\)Many of the variables that politicians consider for this problem are uncertain, and I do not analyze what the implications of that uncertainty are. Introducing uncertainty may considerably change the results.
This diverges from Brollo et al. (2013) since those authors considered only a probability of discovery $q$, while I keep a probability of discovery $qr_1$. The next step, then, is to solve for probability of reelection, $p^J$.

Let us consider that an incumbent would like that (rather simply)

$$E(\theta|g_1, J) \geq 1 + \sigma^O = 1 + \hat{\sigma},$$

that is, he or she wants for voters to expect that their competence is greater than the mean competence of her challengers, given that they know her public good expenditure $g_1$ and type $J$. Certainly, it is not realistic that competence in providing a public good is the only relevant criterion on which politicians are measured, but we are sacrificing realism for analytical tractability. It is sufficient that we are considering all the ways in which $r_1$ influences their reelection. If for example, $r_1$ was disbursed to voters as bribes, then this model would not be sufficient.

The reason for the expectation in the equation (3) is because voters do not have perfect information about $\theta$; they need to estimate it. Using (1), we can rewrite this

$$E(\theta|g_1, J) = \frac{g_1}{\tau - r_1^{cJ}}$$

where $r_1^{cJ}$ is the rent that voters expect a politician to extract. Recall that $g_1$ is actually set by the politician according to $g_1 = \theta(\tau - r_1^{cJ})$, so

$$E(\theta|g_1, J) = \frac{\theta \tau - r_1^{cJ}}{\tau - r_1^{cJ}}$$

describes the competence that voters expect from their politicians. The quantity the politician is then interested in using for her maximization problem is

$$P\left(\frac{\theta \tau - r_1^{cJ}}{\tau - r_1^{cJ}} \geq (1 + \hat{\sigma})\right) = P\left(\frac{g_1}{g_1^{cJ}} \geq (1 + \hat{\sigma})\right),$$

---

6Though by doing this they contradict themselves; it seems that originally meant to use the same form I use, but they lose the $r_1$. 
and $\theta$ is the random variable of interest. The right hand side of this equality is written out to point out an attractive feature of this formulation. What the politician needs is not that their competence is greater than the average competence of her competitors. He or she considers, instead, competence scaled by extracted rents: for a given competence, if he or she extracts a bit more than expected, this probability falls and vice versa.

Note that

$$P\left(\theta \frac{\tau - r^j_1}{\tau - r^j_1} \geq (1 + \hat{\sigma})\right) = P\left(\theta \geq \frac{\tau - r^j_1}{\tau - r^j_1} (1 + \hat{\sigma})\right),$$

which is straightforward recalling that $\theta$ follows a uniform distribution,

$$P(x < \theta) = \frac{1}{2} + \xi(1 + \sigma^J - x)$$

$$\Rightarrow P\left(\theta \geq \frac{\tau - r^j_1}{\tau - r^j_1} (1 + \hat{\sigma})\right) = \frac{1}{2} + \xi \left(1 + \sigma^J - \frac{\tau - r^j_1}{\tau - r^j_1} (1 + \hat{\sigma})\right) = p^J.$$

Notice that we are considering the politician’s problem as the reverse of what one might typically expect. I am considering that the average competence of candidates is known, and that the incumbent wants her perceived level of competence to be above that level on her distribution of $\theta$ (which is known given the her type).

We can return now to examine equation (2). Notice that we now have all the component for the expression\(^7\) from which we can derive the effects of our variables on $r_1$: \(^\tag{4}\)

$$\frac{\partial U_1}{\partial r_1} = (1 - 2q\frac{r_1}{\tau}) + \frac{\xi(1 + \hat{\sigma})}{\tau - r^j_1} (1 - q\frac{r_1}{\tau}) (\psi \tau + R) +$$

$$\left( -q \left( \psi + \frac{R}{\tau} \right) \right) \left[ \frac{1}{2} + \xi \left(1 + \sigma^J - \frac{\tau - r^j_1}{\tau - r^j_1} (1 + \hat{\sigma})\right) \right] = 0$$

This can be solved for $r_1$ from which we can derive predictions. It is important that new predictions still match the old predictions because Brollo et al. provide empirical evidence to support each of their predictions. Then we want to check how the tweak I’ve added influences the rent as a function of $q$. Comparative statics are difficult to derive

\(^7\)An alternative specification is given in the appendix, and the statics are not very different.
analytically, so I use yacas to calculate the optimal $r_1$ and perform simulations in R setting reasonable parameter values.\footnote{$\xi = .75, \tilde{\sigma} = .25, \psi = .3, R = 8, \tau \in (1 : 100), q \in (0 : 1)$.}

Simulation results to observe marginal effects are shown in Figure 3. The results for rents increase with transfers quite linearly. Scales are not plotted on the y-axis because I am only presenting the direction and not the magnitude of the effect. Thus, the effect of transfers on rents agrees with past literature. However, I leave at least one discrepancy unresolved. In my model, there is a nonlinearity in how the quality of challengers $\hat{\sigma}$ decreases rents. This seems to disagree with the predictions of Brollo et al. Because I have not yet resolved this important prediction, I have not even analyzed the comparative statics of Predictions 4-6 in (Brollo et al. 2013). The prediction of the effect of $\hat{\sigma}$ would be important to resolve first.

Figure 3: Simulation results for visualizing comparative statics.
Next, as expected, we see that investigation strength decreases rents. This result is in line with another political agency model by Besley (2006), which predicts that investigation revealing politicians’ actions (but not their type) increases voter welfare. And it is here that the political agency model is connected with emigration. Glaeser, Gentzkow, and Shapiro (2006) show how an increased pool of readership increases investment for investigation of politically salient matters (I replicate their model without modification in the Appendix). It is in this way that emigration enters my model: as emigrants leave, media companies realize a smaller payoff to investigating political matters and so devote fewer resources to it.9

Adding the y-axis scale (for rents) reveals that the magnitude of the effect from increasing \( q \) is quite large, but rents do not approach 0 as \( q = 1 \). It would make more sense for every level of rent to decline to 0 as \( q \) approaches 1 - no mayor would take a rent if they are sure to be discovered. The rents also absorb 98% of the transfer pretty consistently, and this is unrealistic. Therefore this model does not conform well enough to reasonable expectations about how rents should respond to the different variables. The model needs revisions, and this is clear from seeing simulations more than it is clear from the comparative statics, which otherwise point in the appropriate direction.10

There are, of course, several avenues through which emigration might affect rent-taking by politicians, and I am only considering one. First, the decrease in population might decrease the average quality of available challengers.11 This is particularly true if migration is biased towards high-skill occupations, which is a reasonable hypothesis but

---

9This decreases the information about candidates that is available to the public, but the story is not identical to the one in *The Market for Lemons* (Akerlof 1970). The decreased information causes incumbents to prefer higher rents, but I have not yet shown why “the bads drive out the good.” The model discussed in Brollo et al. (2013) suggest that it is possible, however.

10This begs the question of whether the same problem exists with the magnitude of the effects in the Brollo et al. model results.

11This is inspired by Federalist no. 10.
It has also been shown that emigration will discourage corruption (in Brazil) if that flow of emigrants pushes a municipality below a population threshold that decreases the amount of federal transfers the municipality receives (Brollo et al. 2013). I will only consider the effect of emigration on the probability of discovery of irregularities through changes in media investigation and assume that emigration decreases the amount of media investigation. Specifically, firms will invest more where there is a large population and in places to which people are moving because there is a higher payoff for discoveries from their investigation, so I assume that \( q \) increases with immigration. The reason for this assumption is formalized with a small model in the appendix. Now I turn to estimation of the effect.

### 3 Empirics

Two primary sources of data are used for this analysis. First, the Ferraz and Finan data from Brazil count the number of irregularities uncovered in audits of municipal administrations. These count data are plotted in Figure (2). The second important source of data is the Brazilian Institute of Geography and Statistics (IBGE, Instituto Brasileiro de Geografia e Estatística), from which population counts are pulled. Let’s explore these two data sources in some detail.

#### 3.1 Data

The CGU reports have been compiled with irregularity counts and quantities (Ferraz and Finan 2008). The data used here were assembled in 2005, at which point 669 municipalities had been audited. The full data set has not yet been released online, but I use the

---

\(^{12}\) It is possible that the flow and skill level of migrants are correlated. I am not aware of any literature pointing in one direction or another, but if it is true, areas with stronger emigrant flows may be losing more skilled migrants and potential politicians, or vice versa.

\(^{13}\) This could be verified with data on media investment in Brazil or in other contexts, which I do not have.
restricted sample from Ferraz and Finan (2008) of the 373 municipalities which had mayors available for reelection. Because this study leverages changes in population size over time, removing the municipalities that were created in the study period reduces the sample size to 357.\textsuperscript{14}

The frequency distribution of irregularity counts over all 360 municipalities seems to follow a Poisson distribution as expected (see Fig. 4). I do not leverage this feature of the data in this study, but it may be worthwhile in subsequent studies. For example, it may be interesting to estimate the probability of discovering corruption violations in a certain city, and modeling this outcome as a Poisson point process along demographic features of those cities may be promising.

\textsuperscript{14}This is another small effect that I do not control: the shrinking of some municipalities might appear as emigration in my data. Unfortunately, I don’t have information about which municipalities changed to control for this.
On the other hand, the major source for demographic data in Brazil is IBGE. Population counts are available for years during which there was a census, and the institute publishes population estimates for years in between. I use population counts from 1992 and 2010 to have sufficient distance in the past so that emigration has had time to influence media companies’ investment decisions (using 2000 census data may not leave companies with enough time).\textsuperscript{15} With more time to download and clean data,\textsuperscript{16} it may be best to find the population estimates for the year of the audit and subtract these from the 1992 population. This also requires more thinking about how to control for FPM cutoffs, which I discuss later. One should also note that I make the critical assumption of demographic stability, and so I do not account for birth and death rates. Thus the measure I have collected actually combines internal migration, external migration, and demographic variables. The first is what I should measure, the latter two are unaccounted for.

Brazilian census data are quite rich. Several other variables would be available with a better constructed dataset of IBGE or IPUMS census data, such as per capita income, income inequality, population density, share of the population that lives in urban areas, literacy rate, and so on. These are kept in SQL databases that can be downloaded from IBGE servers, but I was not able to completely access the data. With this type of data, it would be possible to search for better instruments, such as literacy rate. Though it is not immediately clear which instruments might work well, data tabulated by municipality could be used with a learning method to discover appropriate instruments. Boosted regression trees are an obvious option.

\textsuperscript{15}I welcome better options for year choices, comments on the standard practice for measuring emigration, and comments on alternative strategies that may be used to search for appropriate years.

\textsuperscript{16}Font encodings are troublesome. see http://bit.ly/1SwLz1g
3.2 Empirical Strategy

To begin with, I run some naïve OLS regressions. Estimation is done in R (R Core Team 2015). OLS regressions will be useful because they provide one simple check for subsequent instrumental variable regressions, by checking whether the estimates are above or below OLS estimates. These regressions take the form

\[ y_i = \beta_1 x_i + \beta_2 \gamma_i + \epsilon_i, \]

where \( y_i \) is the irregularity count for a municipality \( i \), \( x \) is a measure of emigration, and \( \gamma \) is a matrix containing a constant and a couple of additional covariates. The emigration measure is simple net migration, the difference between 1992 and 2010 population sizes for each municipality, with positive measures indicating more emigration. This value is divided by 10,000 to find reasonable coefficients. \( \gamma \) is instead a vector containing the 2010 population to control for population size and FPM cutoff covariates. This permits us to control for the relative size of the city, because emigration values should be different for larger and smaller cities. The other covariate, the list of FPM cutoffs, is included because Brollo et. al show that the amount of federal transfers a municipality receives significantly affects the amount of corruption.

There is, however, very likely a reverse causality problem with OLS. Through anecdotal evidence, corruption and other correlated weak institutions may themselves cause emigration. Immigrants frequently cite corruption at least as evidence that their home countries do not reward merit. I use anecdotal evidence here because this is a connection that, to my knowledge, has not been rigorously justified in the economics literature. As Docquier and Rapoport (2011) state, this reverse effect has not been sufficiently studied in the brain drain literature. At any rate, I proceed with an IV estimation of the form

\[ y_i = \beta_1 \hat{x}_i + \beta_2 \gamma_i + \epsilon_i \]
\[ \hat{x} = \alpha z_i + \alpha_2 z_i^2 + u_i \]

where \( y \) and \( x \) are vectors of irregularity counts and emigration values across municipalities, and \( \gamma \) is a matrix of additional covariates, including 2010 population sizes and dummies for FPM cutoffs. This study leverages randomness in the discovery of corruption incidences for those other covariates to be plausibly exogenous.

The central strategy is to instrument for emigration with haversine distance to major population centers in the spirit of Card 1993. I consider as population centers only those cities with more than 1 million people (this is simply to reduce the combinations and could be improved). With these, using six cities gives the highest first-stage F-statistic. Among all possible six-city combinations, the cities to which proximity yields the highest predictive power for emigration are Fortaleza, Belo Horizonte, Manaus, Belém, Goiania, and Campinas.

For the instrument to work, it needs to be sufficiently strong to predict emigration, and it needs to be exogenous in the structural model. There are, therefore, two serious threats to this estimation strategy. First, the instrument may be weak, even if it is exogenous to the structural equation. We will see that the first-stage F-statistics are fairly robust, but the Wald statistics remain low. Secondly, this strategy could fail if distance to major cities influences corruption through other channels than its effect on emigration. For example, this distance may affect the travel costs for media companies, which I have not included in my model. In that case, more remote municipalities may see less investigation if media companies send fewer correspondents to those areas, and the predicted direction of the effect would remain the same. This would result in significant estimation, but it would be incorrect insofar as that estimation would not justify the model. It is for this reason that alternative predictions would be helpful for justifying the model.
Figure 5: Instrument reduced form. It is not difficult to see that distance does not predict emigration well over the range of distances that we have.

In fact, the first stage regressions will show that emigration is relatively flat for the range of distances we calculate (see Figure 5). It follows that the estimation has a low Wald statistic, which on a $\chi^2(10)$ has a p-value of nearly 1 for nearly all of my specifications, like the one below. However, the first-stage F-statistics all point to the instruments being appropriate.

```r
> tsls <- ivreg(ncorrupt ~ pop10 + factor(cutoff) + emig | dist + dist2 + pop10 + factor(cutoff), data=data)
> summary(tsls)

Estimate Std. Error t value Pr(>|t|)
...
emig   -0.000588  0.001528  -0.385   0.701

Multiple R-Squared: -0.1051, Adjusted R-squared: -0.1084
Wald test: 0.1058 on 10 and 346 DF, p-value: 0.9998
```
### Table 1: OLS Results

<table>
<thead>
<tr>
<th></th>
<th>Incidence Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>emig</td>
<td>0.183***</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
</tr>
<tr>
<td>pop10</td>
<td>0.077**</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
</tr>
<tr>
<td>emig:pop10</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>prop</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>(0.168)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.582***</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
</tr>
<tr>
<td>Cutoff dummies?</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>357</td>
</tr>
<tr>
<td>R^2</td>
<td>0.020</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.015</td>
</tr>
<tr>
<td>F Statistic</td>
<td>3.645**</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01  
prop is the number of emigrants between 1992 and 2010 as a proportion of the 2010 population  
emig the number of emigrants  
pop10 the population in 2010

### 3.3 Results

I report two main sets of results to observe what changes between the naïve OLS and IV specifications. This is because the naïve regressions are, at first, promising, even if the magnitudes of the effects they suggest are quite small. These results turn out to be masking some of the effect of the FPM cutoffs discussed by Brollo et al. Instrumenting for emigration makes the significance of the results disappear.
Table 2: Emigration across FPM levels

<table>
<thead>
<tr>
<th>FPM cutoff</th>
<th>Mean emigration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.129</td>
</tr>
<tr>
<td>2</td>
<td>-0.181</td>
</tr>
<tr>
<td>3</td>
<td>-0.148</td>
</tr>
<tr>
<td>4</td>
<td>-0.336</td>
</tr>
<tr>
<td>5</td>
<td>0.062</td>
</tr>
<tr>
<td>6</td>
<td>-0.239</td>
</tr>
<tr>
<td>7</td>
<td>-1.845</td>
</tr>
<tr>
<td>8</td>
<td>0.231</td>
</tr>
<tr>
<td>9</td>
<td>-2.608</td>
</tr>
</tbody>
</table>

Even among the naïve OLS specifications, simply adding exogenous covariates causes the statistical significance to quickly decline. Specification (1) of Table 1 suggests that an outflow of 10,000 migrants increases the counted incidences of corruption by 0.183, and likewise, that an increase in the 2010 population size increases corruption instances by 0.077. That population size might increase irregularity counts is not surprising; there is a much larger budget from which to extract rents in a larger city. The FPM cutoff controls also decrease the effect of emigration. This is because there is known to be a discontinuous increase in rents extracted as federal transfers increase in stages along a set of cutoffs (Brollo et al. 2013). Table 2 shows that emigration masks the fifth FPM cutoff: emigration was not only more likely from that level of population, but that cutoff is the one that drove the significance of Brollo et al. results. It was the FPM cutoff that caused the most statistically significant increase in corruption. In summary, preferred specification is the third because it includes the appropriate interactions and excludes the unnecessary interaction. Nevertheless, the covariates included in that regression explain a meager 1% of the variation in irregularity counts.
Table 3: IV Results

**Dependent variable:** Irregularity Count

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>emigration</td>
<td>0.610**</td>
<td>1.045***</td>
<td>4.634</td>
<td>3.295</td>
<td>1.159**</td>
</tr>
<tr>
<td></td>
<td>(0.283)</td>
<td>(0.397)</td>
<td>(4.329)</td>
<td>(3.160)</td>
<td>(0.510)</td>
</tr>
<tr>
<td>2010 population</td>
<td></td>
<td></td>
<td>2.762</td>
<td>1.957</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.604)</td>
<td>(1.901)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>1.998***</td>
<td>1.578***</td>
<td>0.045</td>
<td>0.452</td>
<td>1.589***</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(0.232)</td>
<td>(1.370)</td>
<td>(1.002)</td>
<td>(0.250)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutoff dummies?</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>341</td>
<td>341</td>
<td>357</td>
<td>357</td>
<td>341</td>
</tr>
<tr>
<td>R²</td>
<td>−0.524</td>
<td>−1.317</td>
<td>−5.720</td>
<td>−2.786</td>
<td>−1.649</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>−0.529</td>
<td>−1.380</td>
<td>−5.915</td>
<td>−2.895</td>
<td>−1.721</td>
</tr>
</tbody>
</table>

*Note:* *p*<0.1; **p*<0.05; ***p*<0.01

Columns 2, 4, and 5 include squared distance as an instrument

Notice also in Table 1 that from column (3) to column (4), the variable changes from `emig` to `prop`. This changes our covariate of interest from a gross measure of emigration to a measure of emigration as a proportion of the 1992 population. Again from columns (4) to (5) we see that adding the FPM cutoffs decreases the point estimate of emigration measure. This proportion measure is not the best way to control for city size, however, because it lacks any measure of size with even minimal exogeneity.

The results from instrumented regressions are, oddly, much larger than the OLS effect. In Table 3, the specifications from left to right add complexity to the model. (1) and (2) do not include a population covariate, (3) and (4) bring it in, and (5) uses the Google Maps distance API to get a more accurate measure of distance (in minutes of travel time). Columns (2), (4), and (5) include squared distance as an instrument. We should
expect that all of the emigration coefficient would increase as compared to OLS. This is because the reverse causality problem should bias the coefficients upwards: for any level of corruption, there is less emigration when we remove corruption as an incentive to emigrate. This relationship would mean that the pure effect of emigration would have a higher slope, increasing the point estimates as we see in Table 3. The first specification excludes dummies because it is the only one for which the first-stage $F$ statistic surpasses the Staiger-Stock Rule of Thumb of $F = 10$ (Staiger and Stock 1997). The (unreported) wald statistics are all significant, suggesting that OLS is indeed inconsistent and that these instrumental variable regressions improve our estimation.

One problem with this estimation is that the inclusion of a population covariate washes away the significance of the effect. This is because population estimates emigration values better than any other covariate or instrument, so the non-independence of covariates muddles the effect. This covariate was originally introduced because larger cities are expected to have a larger media-consuming population. Excluding this covariate is not problematic because it is proxied by the FPM cutoff. Another solution to control for the size of cities would be to add a covariate of municipality area, considering that city boundaries change less frequently than the number of residents, but this may still covary too strongly with emigration.

### 3.4 Extensions

There may be alternative methods to improve the instrument. The best alternative is boosted regression trees on sufficiently exogenous demographic features of the municipalities. This statistical learning technique is able to compute relationships across a long list of potentially correlated variables to estimate how they might be related to emigration. So long as the variables entered into the list from which the algorithm chooses
are exogenous, the method can search for relevant predictors quickly. It importantly also considers variable interaction. The completion of the IBGE dataset would help to this end.

Two more extensions with more local Atlanta-Journal Constitution data may provide other relevant contributions to this analysis. Clearly, data from corruption discovered in the metro-Atlanta area is not appropriate for these types of regressions because corruption has been endogenously discovered. It would, however, allow us to see whether the AJC has decreased investigation expenditures, whether this is related to emigration, and with what kind of lag. Another measure that would be useful is the percent of Atlanta corruption scandals that have been uncovered by journalists rather than the state, city, or county government. An even better answer to these questions would extend this analysis across cities.

A separate issue these extensions have not yet addressed is a test of the mechanism. I do not have enough data to test the specific mechanism that this model proposes. The mechanism itself is not straightforward because it relies on the abstract parameter $q$, which stands for the probability of discovery of an irregularity in the municipal books by a media organization. But all is not lost. A first way to check the mechanism would be to add a covariate of the number of radio stations and newspapers in each municipality (which could be treated differently if they are local or national in scope). If the coefficient on emigration remains positive, this means that for a given number of radio stations, an increase in the number of emigrants increased corruption. Of course, this would not prove the mechanism because the two covariates could be acting independently. To provide support for the mechanism it will be useful to have some cross-partial derivatives of function used to set $q$. First, however, I develop a robust skeleton that supports the aggregate pattern I am postulating in this paper.
4 Conclusion

This paper argues that one of the effects through which “brain drain” can increase corruption is the decrease of investigative investment by media companies. First, a theoretical model is developed to explain how emigration may effect corruption in this direction. Instrumental variable estimation is then used to search for the expected outcome using proximity to large cities as an instrument for emigration. The basic estimation results provide support for this hypothesis, but the mechanism needs to be further substantiated.

The interesting question is whether there are market configurations that are able to prevent the emergence of corruption. The argument I present suggests that there are some settings in which investigation by private media is inherently insufficient. Of course, there may be other variables that affect the discovery of corrupt actions by politicians. But understanding that market-based investigation is already insufficient in some settings supports the rationale government intervention in the form of audits. RCTs have already shown that audits are useful (Olken 2007), but I argue that they may also be a necessary catalyst to public participation.
References


5 Appendix

5.1 An alternative solution

An alternative solution to the probability distribution should give the same first order conditions on \( U_1 \) (Equation 4). Since candidate competence is along a uniform distribution, rather than scaling the average challenger’s type, we can scale the distribution of candidate competence. The incumbent would thus like that

\[
P(f(\theta) \geq (1 + \hat{\sigma}))
\]

This yields that the probability of having a competence greater than the challenger’s average competence is

\[
p^J = \frac{1}{2} + \xi(1 + \sigma^J) \left( \frac{\tau - r^J_1}{\tau - r^e_1} \right) - \xi(1 + \hat{\sigma})
\]

and so

\[
\frac{\partial p^J}{\partial r^J_1} = -\frac{\xi(1 + \sigma^J)}{(\tau - r^e_1)} < 0
\]

whereas previously this value was

\[
\frac{\xi(1 + \hat{\sigma})}{\tau - r^e_1}.
\]

Brollo et al. justify this incongruence by claiming that these two quantities are equal “in equilibrium,” so that the incumbent’s type is equal to the average type of the challengers and incumbents set their rents to what is expected by the public. These assumptions are contrary to what we discovered earlier in the paper, so I stick with what I’ve presented but describe this solution for full disclosure.
5.2 Investigation Model

Suppose that a newspaper decides a level of investment that determines the probability it will acquire an interesting true story, \( q \). Let cost be a function of this probability, \( K(q) \). That story has ideological content \( \omega \in \{0, \infty\} \), reflecting the degree to which it implicates a politician in a scandal, and this content returns \( r \). Otherwise, media companies earn profit from advertising revenues net of their printing costs, \( a - c \). These companies have a dedicated group of listeners, readers, or subscribers, denoted \( P \), who always purchase their content. There is another group, \( C \), who buys content only if there is a new and interesting story with \( \omega > 0 \). Firm profits are, therefore,

\[
(C + P)(a - c) + r\omega - K(q)
\]

Since these companies anyway earn \( P(a - c) - K(q) \), the payoff for printing ideological content is \( \pi = C(a - c) + r\omega \). It follows that there is a level \( \omega^* = \frac{-C(a - c)}{r} \) at which a firm is indifferent between printing a piece because it yields zero profit. The company’s expected profits are then

\[
(a - c)[P + Cq(1 - F(\omega^*))] + qr \int_{\omega^*}^{\infty} \omega f(\omega)d\omega - K(q)
\]

This equation accounts for the fact that type \( C \) consumers buy in to the media with probability \( q \), that stories are profitable with probability \( (1 - F(\omega^*)) \), and that returns from a story, \( r\omega \), fall along the distribution \( f(\omega) \). Firms choose \( q \) by the first order condition

\[
K'(q) = (a - c)[P + Cq(1 - F(\omega^*))] + qr \int_{\omega^*}^{\infty} \omega f(\omega)d\omega = G(C)
\]

I’ve denoted the right-hand-side of this condition \( G(C) \), because this function can be derived with respect to \( C \) to understand how firms change \( q \) in response. Using the chain
rule and the fundamental theorem of calculus, one can show that

\[
\frac{\partial G(C)}{\partial C} = (a - c) \left[ (1 - F(\omega^*)) + \left( \frac{-C(a - c)}{r} \right) \left( f \left( \frac{-C(a - c)}{r} \right) \right) \right] \\
= (a - c) \left[ (1 - F(\omega^*)) + (\omega^*) \left( f \left( \omega^* \right) \right) \right] > 0
\]

The solution to this first order condition allows us to analyze how a firm may set \( q \), but it does not yield an analytically simple expression for \( q \). Otherwise, a simple expression could be put into the political agency model from the previous section and more detailed predictions could be derived. This equation does show that the optimal value at which to set \( K'(q) \) increases with \( C \), and higher values of \( K'(q) \) are set by increasing \( q \). Thus the larger the market of casual readers, the more a firm will invest to increase the probability of discovering some ideological content.

This suggests that an increase in the number of casual readers will increase the amount that media companies invest to provide information. An appealing feature of this model is that it allows us to formalize the idea of public scrutiny that is often discussed in political conversation. Gentzkow, Glaeser, and Goldin (2006) provide support for this model and show that one of the factors contributing to the decline of ideological content in the early 20th century in the United States was the increase in market size. To be sure, the three mechanisms they consider all move in the same direction, so they jointly provide support for their hypotheses, but it is not clear whether any of these mechanisms individually pushed the ideological content of newspapers in the other direction. This would only be problematic for us if it the rise in market size for newspapers at that time actually had no effect on investigation strength. Leaving this problem aside, there are a few more conditions which must hold for us to expect that investigation will increase with population size. It requires that casual readers do exist and really buy media - newspapers,
or tune in to the radio - when a piece is published implicating a politician in a certain degree of corruption.

Certainly this model is too basic to fully capture the decisions of media companies, but it is illustrative of the fact that media companies will be willing to invest more in investigation when their markets are growing, increasing the profits to capture. That investigation is denoted \( q \) because I equate it with the \( q \) we considered earlier, that is, the probability that a politician’s rents are uncovered to the public. So this leads us to expect that emigration, which decreases market size, will cause a decline in the \( q \) set by media and in turn increase the rents that a politician will capture. Let’s test whether this is true.\(^\text{17}\)

### 5.3 Note on ryacas

Without the `ryacas` package in R, here is an example of how to solve expressions for the variable of interest in terminal. The hashed lines should be run in terminal and the output needs to be copy-pasted to an R function.

```
# In Terminal:
# `yacas`
# Solve( (1-q*r)*(1-c) - d - h*(1/(t-r)), r)

r <- function(d, g, h, t, q){
  (sqrt((t*q*g-t*q+g+d-1)^2-4*(q-q*g)*(t-t*d-t*g-h))- 
   (t*q*g-t*q+g+d-1)) / 
   (2*(q-q*g))
}
```

\(^{17}\)More theoretical predictions are still needed.
Figure 6: Causal Diagram