MASTERARBEIT | MASTER'S THESIS

Titel | Title

Corruption and Development: Evidence from Brazilian Federative Units

verfasst von | submitted by

Fernanda Gomes da Silva BSc (WU)

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Applied Economics
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Christine Zulehner Privatdoz.

Abstrakt

Korruption ist ein häufiges Problem in Entwicklungsländern und hat weitreichende Auswirkungen auf das Wirtschaftswachstum und die soziale Entwicklung. Diese Masterarbeit untersucht die Beziehung zwischen staatlicher Korruption und Entwicklung in den brasilianischen Bundesstaaten, wobei der CIPM (irreguläre Konten pro Million Einwohner) als objektives Maß für Korruption verwendet wird, das aus den Urteilen des brasilianischen Rechnungshofs (TCU) abgeleitet wird. Ein Instrumentalvariablen-Ansatz (IV) wird verwendet, um potenzielle Endogenität zwischen Korruption und Entwicklungsindikatoren zu berücksichtigen. Die Ergebnisse zeigen eine komplexe Beziehung: Während Korruption eine positive und statistisch signifikante Korrelation mit dem BIP pro Kopf aufweist, was darauf hindeutet, dass sie wirtschaftliche Aktivitäten durch die Umgehung bürokratischer Ineffizienzen erleichtern kann, steht sie gleichzeitig in einem negativen Zusammenhang mit der menschlichen Entwicklung, wie sich in niedrigeren HDI-Werten in korrupteren Staaten zeigt. Dies deutet darauf hin, dass Korruption zwar das kurzfristige Wirtschaftswachstum fördern könnte, jedoch langfristig die sozialen Ergebnisse, insbesondere in den Bereichen Bildung und Gesundheitswesen, untergräbt. Darüber hinaus zeigt die Analyse, dass Korruption in einigen Kontexten die Einkommensungleichheit verringern kann, möglicherweise durch informelle Umverteilung, was jedoch nicht zwangsläufig zu einer nachhaltigen Entwicklung führt. Die Arbeit untersucht auch regionale Unterschiede und stellt fest, dass stärker korrupte Regionen höhere Armutsraten aufweisen könnten. Diese Ergebnisse unterstreichen die Notwendigkeit kontext spezifischer Anti-Korruptionsstrategien, die sich auf die Förderung von Transparenz, die Stärkung von Institutionen und die Verbesserung der Regierungsführung konzentrieren, um die negativen Auswirkungen der Korruption wirksam zu mindern.

Abstract

Corruption is a common issue in developing economies, with far-reaching implications for economic growth and social development. This master thesis explores the relationship between governmental corruption and development across Brazil's federative units, utilising the CIPM (irregular accounts per million inhabitants) as an objective measure of corruption derived from Brazil's Federal Court of Accounts (TCU) judgments. An Instrumental Variable (IV) approach addresses potential endogeneity between corruption and development indicators. The findings reveal a complex relationship: while corruption shows a positive and statistically significant correlation with Gross Domestic Product (GDP) per capita, suggesting that it may facilitate economic activity by bypassing bureaucratic inefficiencies, it simultaneously has a negative relationship with human development, as reflected in lower Human Development Index (HDI) scores in more corrupt states. This indicates that while corruption might boost short-term economic output, it undermines long-term social outcomes, particularly in education and healthcare. Additionally, the analysis shows that corruption can reduce income inequality in some contexts, possibly through informal redistribution, though this does not necessarily lead to sustainable development. The thesis also explores regional variations and finds that more corrupt regions might experience higher poverty rates. These findings highlight the need for context-specific anti-corruption strategies focused on enhancing transparency, strengthening institutions, and improving governance to mitigate the adverse effects of corruption effectively.

Table of Contents

Introduction

Common sense often suggests that corruption hampers economic growth and development, but the relationship between these variables is highly complex. Studies over the years have highlighted the corrosive effects of corruption on government expenditure, the provision of public services, and the attraction of foreign investment, which are particularly damaging for developing economies where accountability is weaker (Mauro (1995), Tanzi and Davoodi (1997)). On the other hand, a second stream of studies suggests a positive relationship between corruption and economic development and growth, arguing that corruption increases efficiency and growth by channelling more investments to overlooked areas of the economy (Leff (1964), Dreher & Gassebner (2013)).

This thesis analyses the relationship between governmental corruption and development across the Brazilian federative units, drawing on the existing international literature on the subject and the specific characteristics of the Brazilian experience. The research question "How does governmental corruption relate to economic and social development indicators across the Brazilian federative units?" is addressed with an Instrumental Variable (IV) strategy and a fixed effects model. The designed econometric specification addresses the potential endogeneity of corruption, offering a robust analysis that controls for both observed and unobserved heterogeneity.

A central feature of the thesis is using the CIPM (irregular accounts per million inhabitants) as a measure of corruption. Unlike more commonly used subjective indicators, the CIPM is a recently proposed objective measure based on judicial rulings by Brazil's Federal Court of Accounts. It provides a more concrete basis for analysis when contrasted with other subjective indicators used in the literature. Additionally, the instrumental variable used, which is the average remuneration of federal government public servants, isolates exogenous variation in corruption, helping ensure the validity of the causal inferences drawn from the data.

Brazil provides an optimal case study to investigate the relationship between development and economic and social development indicators. The country scored 36 points on the 2023 Corruption Perception Index (CPI) on a scale that goes from 0 to 100, where 0 is "very corrupt" and 100 is "very clean". Brazilians have suffered from corruption at all levels of government, from the Operation Car Wash (Lava Jato) scandal that revealed widespread corruption involving state-owned enterprises and high-level politicians to more endemic issues of bureaucratic corruption affecting local governance. Additionally, the country's federal structure comprises 26 states and a federal district with varying levels of autonomy, resources, and challenges. The disparities between these states reflect a broader narrative of inequality and uneven development across the country. The Brazilian national context provides a robust framework to explore how corruption at the state level might affect development outcomes within the same country. The work on the thesis not only adds to the scholarly debate but also informs policy, providing insights that can guide efforts to combat corruption and promote development not only in Brazil but also in other developing economies.

Objectives and Scope

The primary objective of this thesis is to investigate the relationship between governmental corruption and development across the Brazilian federative units. It aims to (a) construct an econometric model that captures the effect of corruption on development outcomes, using an instrumental variable strategy tailored to the socioeconomic and political landscape of Brazil, and (b) identify the specific levels of corruption that are associated with the best and worst development rates across different Brazilian states.

This thesis's central research question is: *How does governmental corruption, as measured through objective indicators, relate to economic and social development across the Brazilian federative units?* The question is designed to quantify the influence of corruption on key socio-economic development metrics, namely Gross Domestic Product (GDP) per capita, Human Development Index (HDI), poverty rates, and the Gini coefficient, which measures inequality. Using an objective measurement of corruption and deploying an IV strategy, the study aims to provide robust evidence of how corruption relates to development in Brazilian states.

Literature Review

Corruption, economic development and growth

Corruption has long been a challenge in developing economies, undermining governance and posing a barrier to sustainable development. In a communique issued by the United Nations in December 2023[1], it was warned that progress towards achieving the United Nations Sustainable Development Goal (SDG) 16, which calls for peace, justice and strong institutions, is dangerously off track. Recent research by Bazie ad et al. (2024) provides evidence of corruption's corrosive effects, demonstrating how it distorts the allocation of public resources. The study reveals that corruption diverts funds towards sectors where resources can be easily concealed, often at the expense of areas where oversight is stronger. While the harmful impact of corruption on governance and public investment is well documented, the academic literature presents mixed findings regarding its broader effect on economic growth and development.

The earlier literature on the impact of corruption on economic growth suggested that corruption has a positive effect on economic growth, the so-called “grease the wheels” hypothesis. Leff (1964) argued that corruption could promote development by channelling investments into overlooked areas of the economy, fostering innovation and increasing overall efficiency. This perspective suggested that in rigid bureaucratic systems, corruption can be a catalyst for economic activity, allowing businesses to navigate bureaucracy more efficiently and access necessary services that might otherwise be severely affected by procedural delays.

Supporting the "grease the wheels" hypothesis, Dreher et al. (2013) conducted an empirical analysis using data from 43 countries from 2003 to 2005. Their findings indicated that governmental corruption could promote economic growth by helping entrepreneurs overcome bureaucratic hurdles. This research provided quantitative backing to the idea that corruption can benefit the economy, particularly in contexts where governmental inefficiencies hamper business operations.

[1] Available at: https://unis.unvienna.org/unis/en/pressrels/2023/uniscp1172.html, accessed on 25 August 2024.

On the other side of the debate is the "sand the wheels" theory, which states that corruption negatively affects economic growth. This hypothesis is based on the argument that corruption leads to inefficiency in production and hampers innovation by distorting market mechanisms and misallocating resources. Mauro (1995), using subjective indices of corruption, identified a clear negative correlation between corruption and investment and economic growth, highlighting the detrimental effects of corrupt practices. Further adding to the literature, Méon and Sekkat (2005) reviewed over 60 studies and concluded that corruption impedes growth not only by reducing capital accumulation but also through other less explored mechanisms. A more recent study by Nur-tegin and Jakee (2020) found that while the impact of corruption varies by type, the overall effect on the economy is predominantly negative.

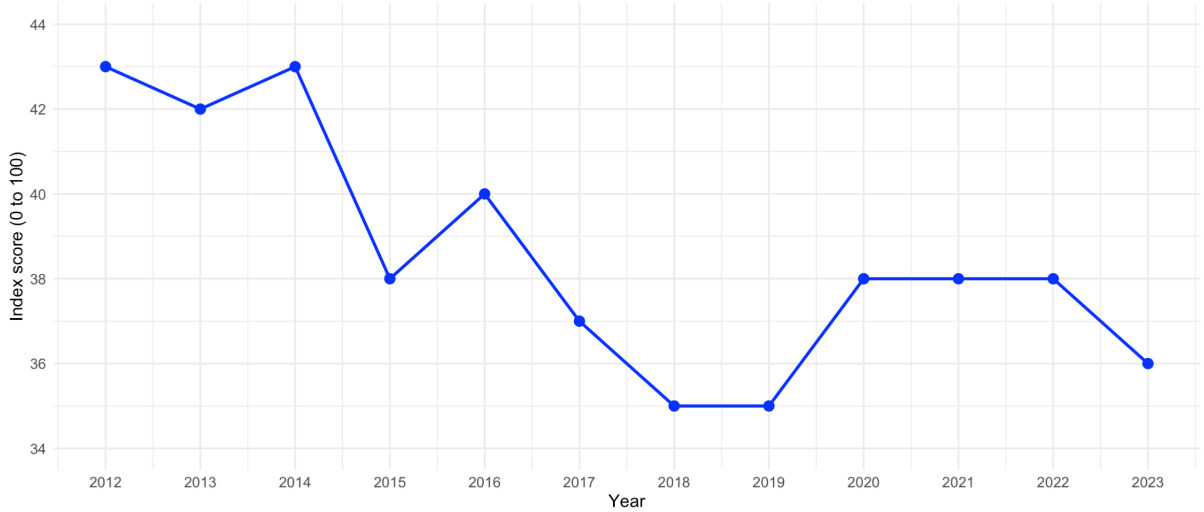
Regarding socioeconomic indicators of economic development, a large share of the literature agrees that corruption has largely negative impacts. Expanding the discussion beyond the “sand” and “grease” hypothesis and more closely looking at the socioeconomic effects of corruption, Li et al. (2000) demonstrated that corruption influences income distribution in a quadratic manner, implying that countries with moderate levels of corruption experience more severe income inequality compared to those with very high or very low levels. This suggests that the impact of corruption on income distribution is complex and highly dependent on the degree of corruption. Additionally, Gyimah-Brempong and Munoz de Camacho (2006) used panel data from 61 countries to show that corruption worsens income inequality significantly more in regions like Latin America and Africa than in OECD countries. They found that reductions in corruption could lead to substantial improvements in income equality, particularly in Latin America.

Corruption and development in the Brazilian context

Corruption has long been an issue in Brazil, affecting its governance and development trajectory. The CPI and the CPIM indexes offer a comprehensive view of how corruption has evolved over the past decade. As shown in Figure 1, Brazil’s CPI score has steadily declined, dropping from 43 points in 2012 to just 36 in 2023 on a scale where 0 represents "very corrupt" and 100 indicates "very clean." This downward trend reflects a growing concern over the quality of governance and accountability of institutions in the country. In the most recent survey, Brazil ranked 104th out of 180 countries, highlighting the challenges it faces in combating corruption.

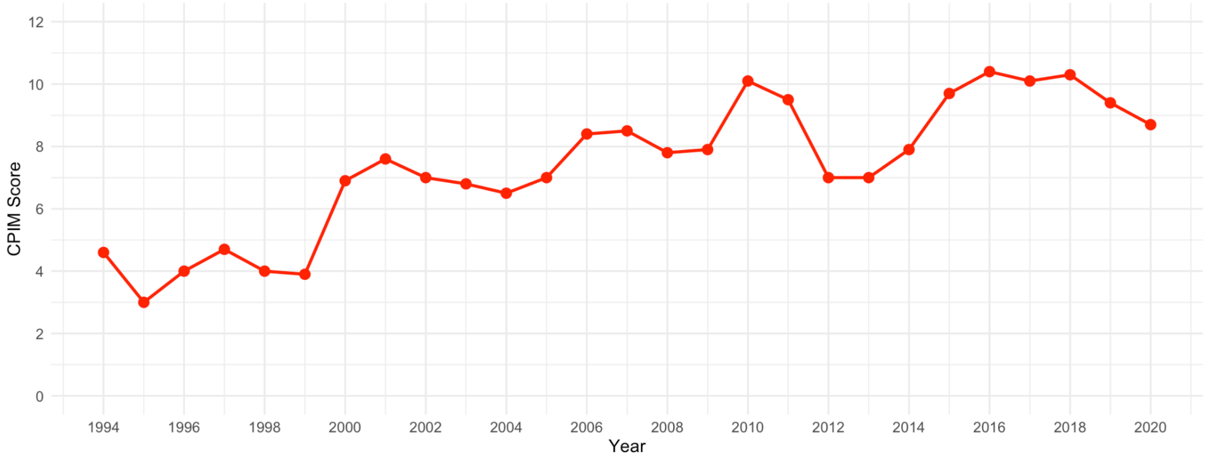
In parallel, Figure 2 illustrates the trajectory of corruption as measured by the CPIM, the objective corruption indicator that lies at the centre of this thesis's analysis. Unlike the CPI, which is based on perceptions, the CPIM quantifies corruption through documented instances of irregularities. The CPIM index operates on an inverse scale to the CPI, where a higher score indicates a higher level of corruption. The data reveals a trend that mirrors the CPI's findings, showing a consistent rise in corruption between 2012 and 2020, the years for which both indicators are available. The alignment between subjective perceptions and objective measures further emphasises the depth and persistence of corruption issues within Brazil's political and administrative landscape.

Figure 1: CPI Score of Brazil from 2012 to 2023



Source: Designed by the author using R programming language. Data available at: https://www.transparency.org/en/countries/brazil, accessed on 25 August 2024.

Figure 2: CPIM Score of Brazil from 1994 to 2020



Source: Designed by the author using R programming language. Data from Garcia (2022).

Corruption is widely acknowledged as a systemic issue within Brazilian society. This means that it is a practice that does not only happen regularly but also simultaneously and interconnectedly in different areas of society, involving multiple individuals and institutions (Trombini et al., 2024). This systemic nature means corruption is not an isolated or occasional occurrence but a persistent issue and part of the cultural and institutional framework of the country.

The concept of "Brazilian *jeitinho*" exemplifies how corrupt behaviour is present not only at the institutional level but also in the citizens' everyday life. A popular concept, *jeitinho* has been previously defined as a great capacity to get around the rules, which can easily be through transgression of the law or blind neglect (Castor, 2002). An example of *jeitinho* given by Castor (2002) is that of a bureaucrat who interprets the rules in a friendly way to advance his interest or that of his family and friends. Rocha (2019) argues that this normalisation of bending rules undermines the rule of law, eroding trust in public institutions and reducing the efficacy of legal frameworks designed to combat corruption.

Corrupt activities in Brazil became a central political issue and gained international media attention in 2014 with Operation Carwash. This investigation targeted allegations that the country's largest construction firms had overcharged Petrobras, a state-owned oil company, for building contracts. Operation Carwash uncovered some of the most extensive corruption schemes in history, involving billions of dollars and implicating politicians and businesses worldwide. These schemes included procurement fraud, illicit campaign financing, money laundering, and influence peddling. Initially focused on Petrobras, the investigation expanded to other parts of the Brazilian federal government, major state-owned enterprises, and several state governments (France, 2020). Financially, corruption was estimated to account for 17.84% of construction costs, totalling BRL 673 million out of BRL 3.77 billion. Corrupt agents personally benefited by BRL 92.3 million, or 13.7% of the diverted amount, highlighting that personal gain was much smaller than the overall social damage, with 26.9% of funds used for money laundering (Sallaberry et al., 2020).

Many studies have observed the relationship between corruption and economic and social economic indicators in the Brazilian context. Onody et al. (2022) found a statistically significant and positive relationship between corruption and Foreign Direct Investment (FDI). The authors also identified that the actual level of regional corruption affects FDI intensity in different ways, with investors being potentially indifferent to very low levels of corruption but likely to avoid regions with high levels of corruption.

Azevedo et al. (2017) explored the impact of corruption on the HDI of the Brazilian geographical regions, using the General Index of Corruption (IGC) for Brazilian states. The authors found that every unit increment of the IGC led to a reduction of 0,162 on the HDI, which suggests that corruption directly undermines the quality of health, education, and standard of living in the region. Melo et al. (2015) additionally explored the impact of corruption on Entrepreneurship using the IGC as the corruption indicator. Their study identified a positive relationship between corruption and entrepreneurial activity, suggesting an institutional framework with barriers to starting businesses, where bureaucratic corruption becomes an alternative and way to circumvent excessive regulation.

Methodology

The methodology of this thesis is a quantitative approach to analyse the relationship between corruption and socio-economic development using balanced panel data. The corruption indicator is “Irregular accounts per million inhabitants” (CPIM). This indicator has been recently proposed by Garcia (2022). It reflects the results of the division of public state accounts that have been ruled as irregular by Brazil’s Federal Court of Accounts (TCU) by the state’s population size. This indicator is proposed as a successor to the General Index of Corruption (IGC), developed by Boll (2010). The CPIM is an objective measure that offers a population-adjusted indicator of irregular state accounts and reduces reliance on subjective assessments, which dominate much of the existing corruption research.

Socioeconomic development indicators include GDP per capita, HDI, poverty rates, and the Gini coefficient. To account for the skewness and scale differences across the variables, the natural logarithm of each variable is used in the econometric specification of the model. Taking the logarithm normalises the distribution of the variables, reducing the impact of outliers and allowing a more consistent and meaningful comparison across states and over time.

This thesis uses an IV strategy to address the endogeneity problem inherent in analysing the relationship between corruption and development indicators. The key concern is that corruption may be endogenous due to omitted variable bias or reverse causality. Unobserved factors such as institutional quality or transparency could influence both corruption and development variables, leading to biased estimates if not controlled for in the model specification. Additionally, development outcomes could themselves influence levels of corruption, creating a reverse causality issue. This endogeneity could pose a threat to the validity of the estimates.

To address these issues, the IV strategy was chosen as the most appropriate method for this analysis, as it isolates exogenous variation in corruption through the use of an instrument that is correlated with corruption but uncorrelated with the error term. The use of IVs to study the relationship between corruption and economic growth and development is common practice in the literature and, for example, was used by Ribeiro and Gomes (2022) to explore the impact of corruption on poverty in a sample of Brazilian municipalities.

In this thesis, the chosen IV is the average remuneration of federal government public servants. This instrument captures variations in corruption influenced by remuneration policies that are exogenously determined by federal administrative decisions and not by the economic or developmental characteristics of the states themselves. The instrument can, therefore, capture variations in corruption influenced by public servant remuneration without being confounded by direct effects on state-level development outcomes. Its validity, including relevance and exogeneity, will be further discussed and tested in the results section.

To further improve the robustness of the analysis, a unit and time-fixed effects model is used. The model, therefore, controls for unobserved, time-invariant characteristics specific to each unit (in this case, Brazilian states), as well as time-specific effects that may impact all units simultaneously, such as policy changes or economic shocks. The combination of the IV strategy with unit and time-fixed effects ensures that the estimated effects of corruption on development are not confounded by unobserved heterogeneity across states or by temporal shocks that affect all states. This dual approach strengthens the validity of the causal inference by addressing both the potential endogeneity of corruption and the unobserved factors that could bias the results.

To ensure comparability and robustness, this thesis employs multiple model specifications alongside the IV model with fixed effects, including OLS models (with and without fixed effects) and an IV model without fixed effects. These alternative specifications allow for systematic comparisons, highlighting how endogeneity and unobserved heterogeneity influence the estimated relationships. By comparing the results across these models, the robustness of the findings can be evaluated, and the impact of model restrictions on the interpretation of the results can be better understood.

The econometric specification for the **first stage regression** estimates the relationship between corruption and the explanatory variables. It is given by the following equation:

(1.1)

$$ln(CPIM_{it}) = \beta_0 + \beta_1 ln(IV) + \beta_2 ln(Education_{it}) + \beta_3 ln(Population_{it}) + \alpha_i + \lambda_t + v_{it}$$

In the equation, $ln()$ denotes the natural logarithm. The variable IV_{it} represents the average remuneration of federal government public servants and serves as the IV in this model. $CPIM_{it}$ captures the level of corruption for unit i at time t, while $Education_{it}$ and $Population_{it}$, account for education and population factors, respectively. The coefficients β_1,β_2 and β_3 are estimated in this stage, and the terms α_i and λ_t represent unit-specific (state) and time-specific (year) fixed effects. Finally, v_{it} represents the error term.

The **second stage regression** builds on the results of the first stage to explore the relationship between predicted corruption levels and the development indicators. Its specification is as follows:

(1.2)

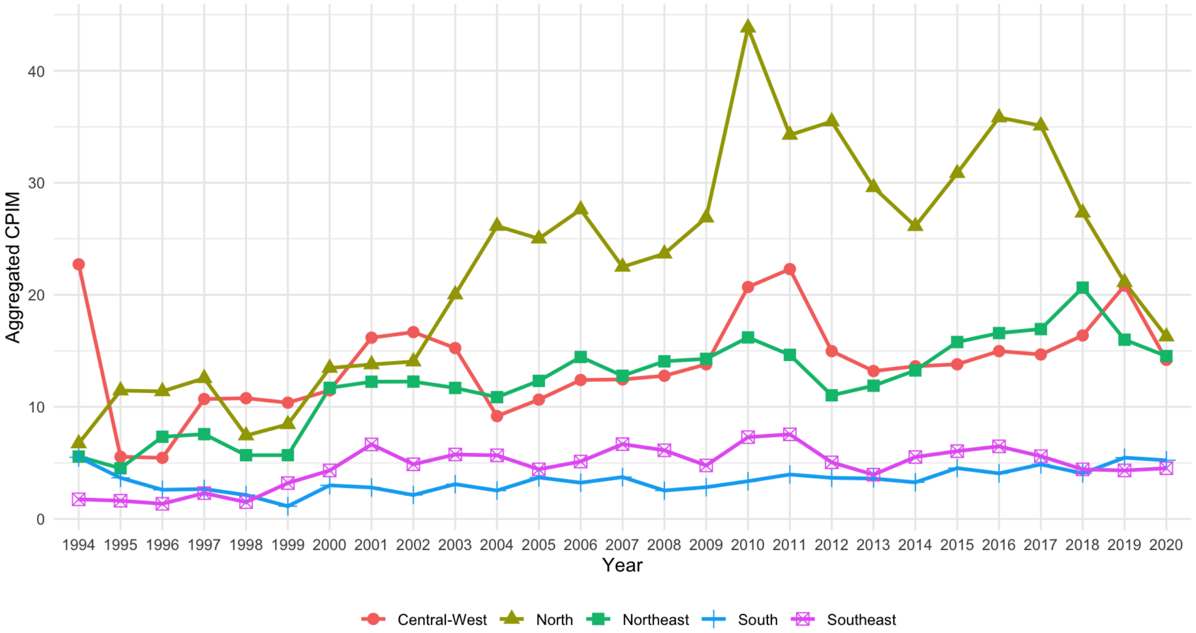
$$ln(Dev_{it}) = \theta_0 + \theta_1 ln(\widehat{CPIM}_{it}) + \theta_2 ln(Education_{it}) + \theta_3 ln(Population_{it}) + \alpha_i + \lambda_t + \mu_{it}$$

In the specification, Dev_{it} denotes the development indicators for unit i at time t, while $\widehat{CPIM}_{it}$ represents the predicted values of the corruption variable obtained from the first stage regression. The coefficients θ_1,θ_2 and θ_3 are estimated in this stage, and the terms α_i and λ_t again represent unit-specific (state) and time-specific (year) fixed effects. The error term is denoted by μ_{it}.

Data

The data for the corruption indicator is taken from the study "Proposal for a corruption indicator in Brazil based on data from the Federal Court of Accounts" by Gilson Piqueras Garcia. The indicator, CIPM (irregular accounts per million inhabitants), is calculated from 1998 to 2020 by dividing the number of accounts judged irregular by the TCU by the state's population. An 'irregular account', as defined by the Organic Law of the Federal Court of Accounts (LOTCU), refers to public accounts that are judged as irregular under specific circumstances, including omission in the duty to render accounts, illegal or illegitimate management acts, damage to the public treasury due to such acts, and embezzlement or diversion of public funds. The CPIM indicator provides an objective and regional perspective on corruption, offering a concrete measure based on legal rulings rather than subjective perceptions or opinions.

Figure 3: Evolution of the average CPIM by region



Source: Designed by the author using R programming language.

The CIPM indicator is preferable to subjective indicators commonly used in the corruption literature. Its objective foundation contrasts with subjective indicators, which often rely on analysts' and experts' perceptions, opinions, or surveys. The CPI, published by Transparency International (TI), is one such indicator, frequently criticised yet widely used by academics in literature exploring the connection between corruption and a range of issues.

As Gilma (2018) describes, the CPI faces fundamental issues for academic researchers, primarily because it is not a direct survey by TI but a composite of 13 different perception surveys. These sub-indicators focus more on the business environment, economic development, and risks of doing business rather than directly on corruption. They often address related topics like public trust, government transparency, and accountability and sometimes only pose one or two questions specifically about corruption. The definition of corruption also varies among respondents, and the surveys are typically designed for business clients. CPI scores might also be influenced by media coverage, as the opinion of experts and other analysts is highly influenced by when a corruption scandal is investigated and becomes public. Additionally, the CPI's ranking system can be misleading, as small changes in scores can result in significant shifts in rankings. Given the existing critique of subjective indicators, the use of the CPIM in this paper is based on the understanding that objective data provides a more concrete and verifiable basis for measuring corruption.

The table below outlines the control, instrumental and dependent variables used in the model. Given the data availability from CPIM, the Brazilian Institute of Geography and Statistics (IBGE), the Institute of Applied Economic Research (IPEA), and the Brazilian National Household Sample Survey (PNAD), the final dataset includes data for the 27 Brazilian states from 2012 to 2019, totalling 216 observations. The indicators that utilise data from the PNAD, in most cases, were also available for the years of 1991, 2000 and 2010 from the national census. These years were excluded from the sample and are not included in the analysis for two key reasons. One, as the goal consists of the analysis of trends over time, excluding non-continuous years provides a clearer picture. Two, the methodology and instrument for data collection changed between these earlier years and the latter, which could affect the reliability of findings. With that in mind, the increase in sample size by 81 observations, 27 states times three extra years, is not considered to outweigh the potential downsizes.

Table 1: Model variables

#	Variable Name	Type	Description	Source
1	Gini Coefficient	Dependent	Measures the degree of inequality in the distribution of individuals according to the household per capita income. Its value is 0 when there is no inequality (household per capita income of all individuals is the same), and it increases towards one as inequality increases	IBGE, PNAD
2	% of people in poverty	Dependent	The proportion of individuals with a per capita household income equal to or less than BRL 140.00 per month, in reais as of August 1, 2010	IBGE, PNAD
3	% of people in extreme poverty	Dependent	The proportion of individuals with a per capita household income equal to or less than BRL 70.00 per month, in reais as of August 1, 2010	IBGE, PNAD
4	GDP per capita	Dependent	GDP divided by total population, in thousands of reais as of August 1, 2010	IBGE
5	Human Development Index (Municipal) (HDI-M)	Dependent	It is an adaptation of the Human Development Index (HDI) calculated by the United Nations Development Programme and encompasses the same three dimensions as the Global HDI: longevity, education, and income	IBGE, PNAD
6	Education	Control	Percentage of the population aged 18 and over who have completed primary education in any of its forms	IBGE, PNAD
7	Population	Control	Total population, with July 1st as reference for each calendar year	IBGE
8	Average Remuneration of Federal Government Public Servants	Instrument	The average monthly wage paid to employees working in the federal government sector of Brazil. This figure is calculated by summing	IPEA

			the total remuneration of all federal government public servants and dividing it by the number of employees. The data was available in reais of 2019, but for the purpose of this thesis, it was adjusted to reais of 2010 using the Consumer Price Index (CPI) as published by the World Bank.	

Descriptive Statistics

The table below offers a comprehensive overview of the dataset to help readers better understand trends. The summary statistics include the number of observations (N), the mean, standard deviation (SD), and minimum and maximum values for each variable used in the models.

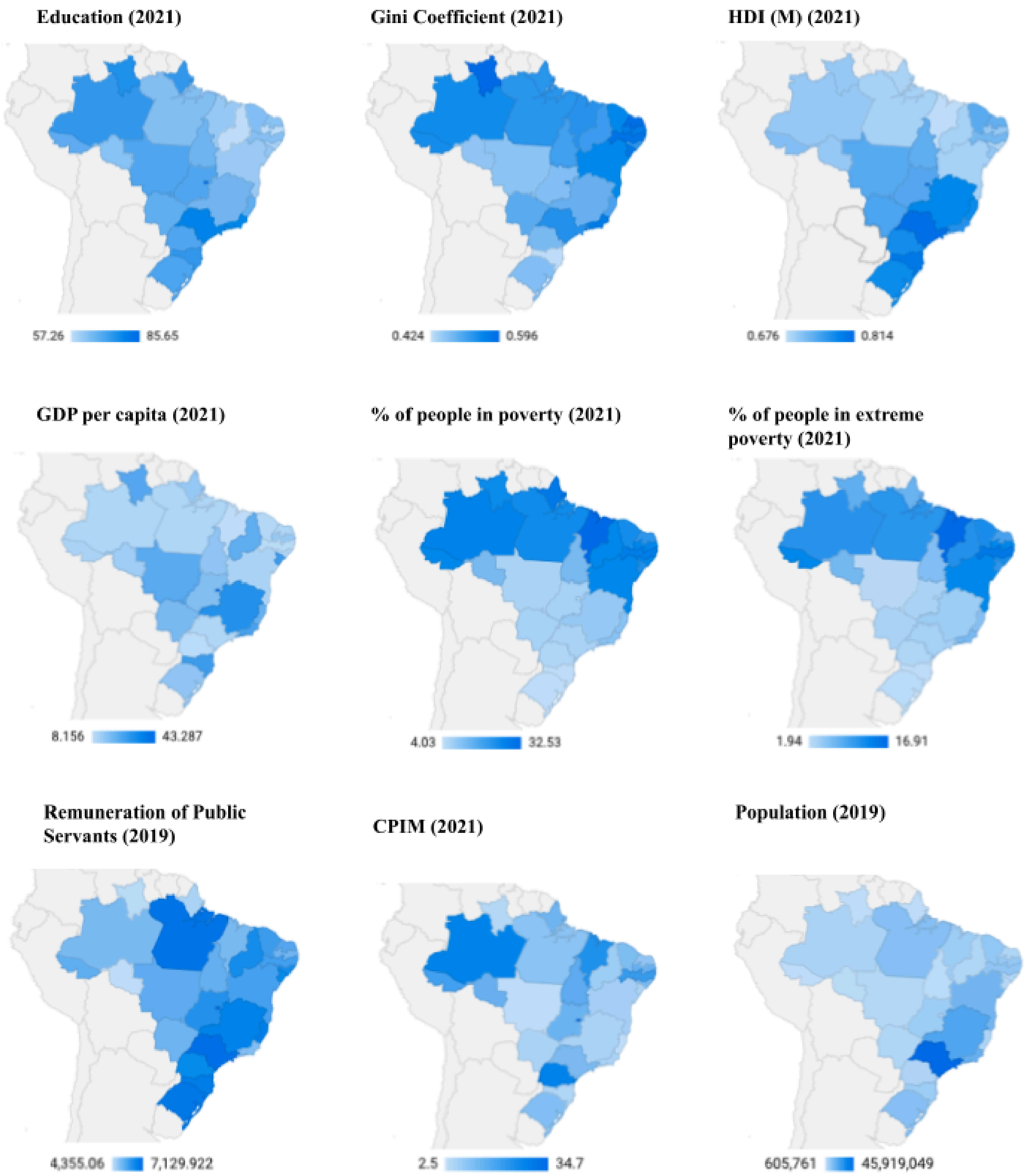
Table 2. Summary statistics of dependent, independent, instrumental and control variables

Statistic	N	Mean	St. Dev.	Min	Max
CPIM	216	16.414	15.524	1.300	97.600
Gini	216	0.516	0.041	0.408	0.601
HDI(M)	216	0.741	0.048	0.648	0.859
GDP_pc	216	17.209	9.080	7.705	52.992
% of people in poverty	216	13.544	8.208	1.830	30.260
% of people in extreme poverty	216	5.510	3.767	0.740	17.350
Remuneration	216	5,479.671	672.323	3,784.427	7,129.922
Education	216	61.900	7.674	46.430	81.260
Population	216	7,567,512.000	8,848,500.000	469,524	45,919,049

Source: Calculated by the author using R programming language.

The maps below illustrate the heterogeneity of the nine variables used in the models. The southeastern and southern states generally show higher development levels as displayed by socio-economic and development indicators, such as education, HDI, and GDP per capita, along with lower levels of poverty and corruption. In contrast, the northern and northeastern states exhibit lower education levels, higher income inequality (Gini coefficient), greater poverty rates, and higher corruption. The visualisations suggest the need to analyse the dynamics between corruption not only at a state level but also at a regional level, given the clear uneven distribution of resources.

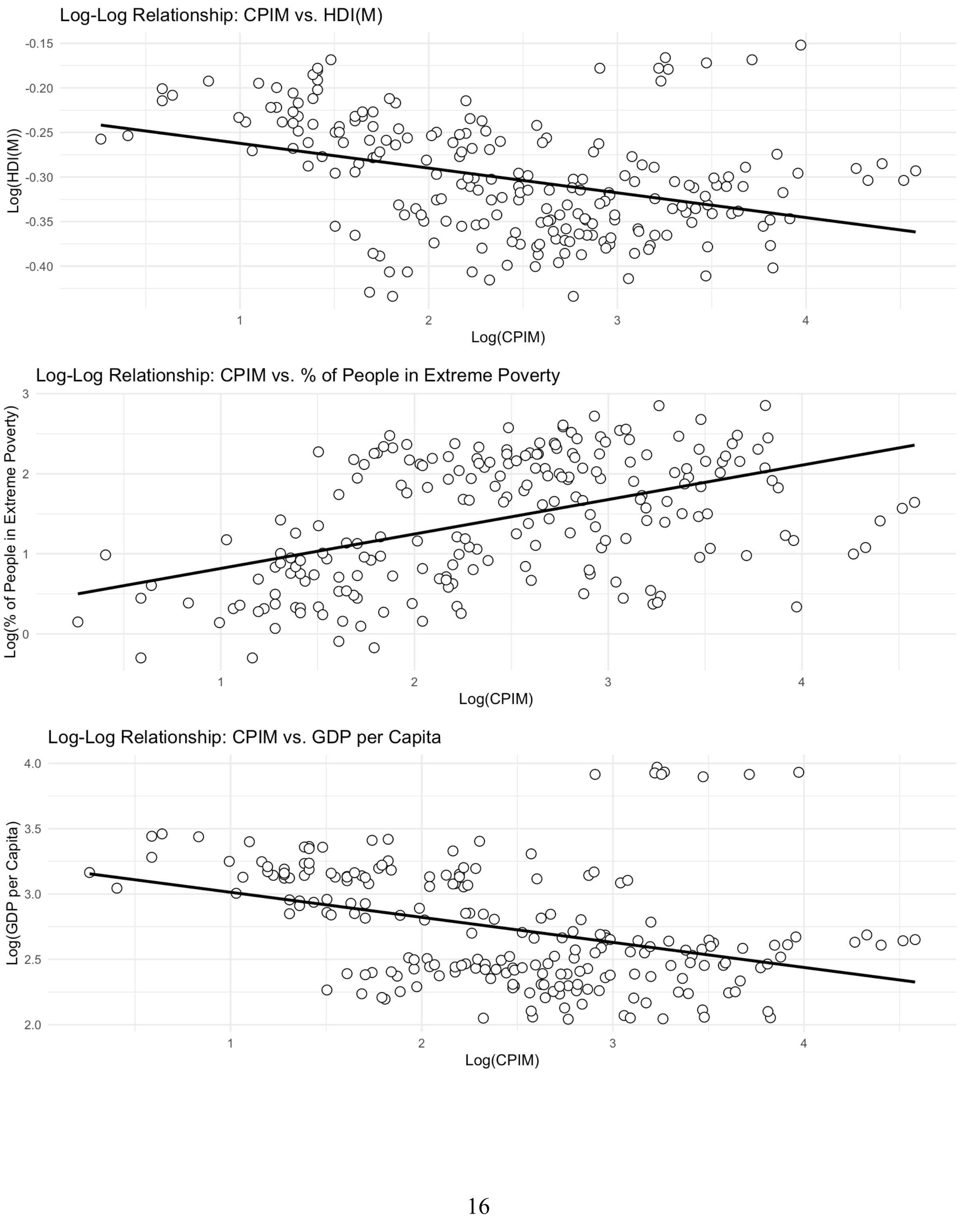
Figure 4: Distribution of variables across states for the latest year in the sample

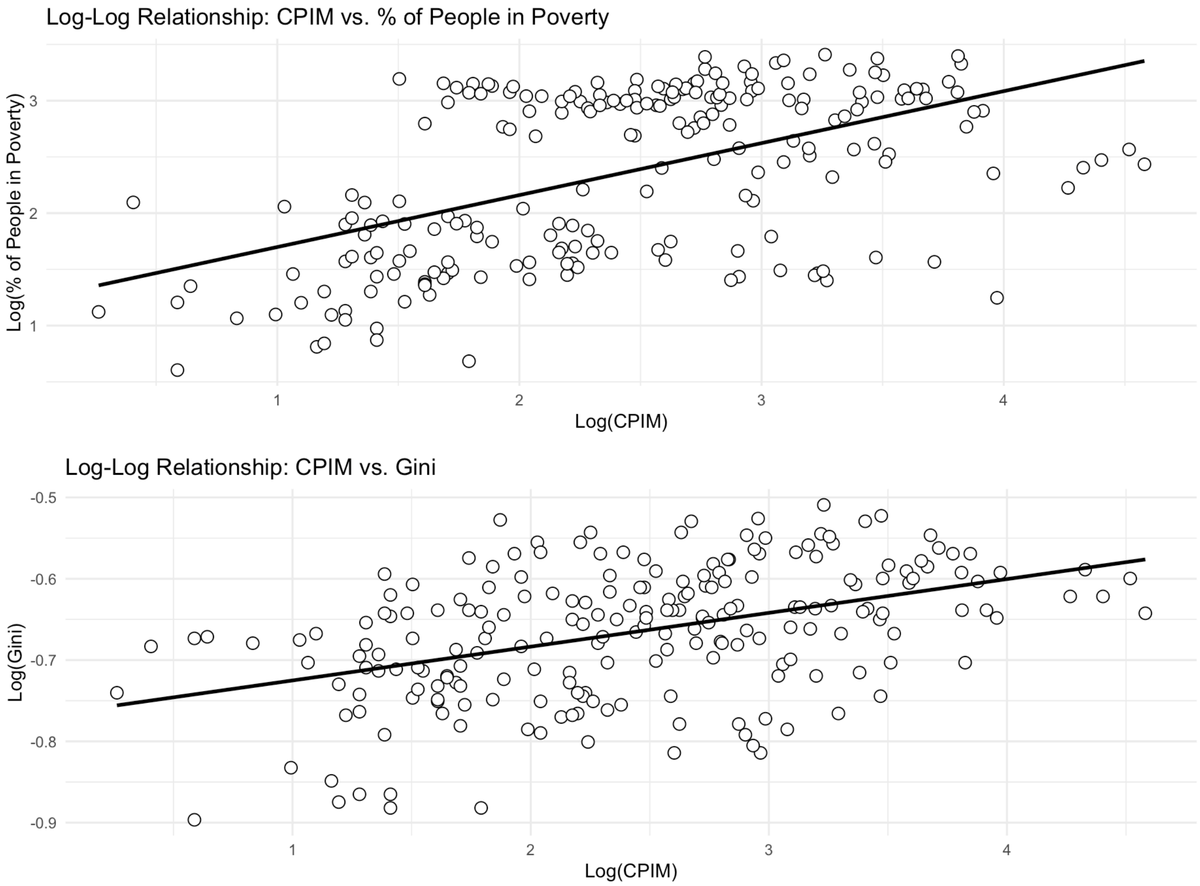


Source: Designed by the author using Google's Looker Studio.

Figure 5 visually examines the relationship between corruption and development outcomes, suggesting that higher corruption levels are generally linked to poorer development indicators. The graphs show that lower corruption (as indicated by lower CPIM values) is associated with better socio-economic development outcomes, such as reduced inequality, lower poverty rates, higher HDI(M), and increased GDP per capita. Conversely, higher inequality and poverty are correlated with higher CPIM values.

Figure 5: Relationship between the natural logarithm of the CPIM and that of the dependent variables.





Source: Designed by the author using R programming language.

Results

Instrumental Variable Assumption Checks

This thesis utilises an IV approach to explore the relationship between corruption and socio-economic development indicators. This section thoroughly examines three essential assumptions required for a valid IV design. First, the instrument's relevance is assessed to ensure it has adequate predictive power for the endogenous variable. Second, the exogeneity condition is evaluated to confirm that the instrument is uncorrelated with the error term in the outcome equation.

Instrument relevance condition

The relevance condition states that the instrument must be correlated with the endogenous explanatory variable (Wooldridge, 2010). In the context of this thesis, this equates to the average remuneration of federal government public servants being correlated with the CPIM. Formally, the instrument must be relevant to the independent variable, meaning that the correlation between the IV and the CPIM is different than 0, as shown in equation 1.3. This ensures that the instrument can explain some of the variation in the CPIM, allowing for proper identification of the causal effect.

(1.3)

$$Cov(IV,\ CPIM)\ \neq\ 0$$

We can test this assumption with the results of the first stage OLS, directly assessing the instrument's relevance by quantifying the strength of the relationship between the instrument (average remuneration of federal government public servants) and the endogenous variable (CPIM). To capture the robustness of the instrument across different levels of aggregation and specifications, four first-stage models were run: two with state-level data and two with aggregated regional-level data. Each level includes a pooled model, which does not account for fixed effects, and a fixed effects model, which controls for state- and time-specific unobserved heterogeneity.

The results in Table 3 demonstrate the outcomes of the first stage OLS regressions with both pooled and fixed effects models at the state and regional levels. In the pooled model with state-level data (Column 1), the coefficient of log(Remuneration) is -2.431, significant at the 1% level. This result indicates that a 1% increase in remuneration is associated with a 2.431% decrease in corruption levels. The control variables, log(Education) and log(Population), have coefficients of -0.272 and -0.415, respectively, with log(Population) being statistically significant at the 1% level. The model has strong explanatory power, with an adjusted R-squared of 0.584 and an F-statistic of 101.410, which is highly significant ($p < 0.01$). These results validate the relevance of log(Remuneration) as an instrument under the pooled state-level specification.

The fixed effects model with state-level data (Column 2) controls for state- and time-specific unobserved heterogeneity. In this model, the log(Remuneration) coefficient is 1.943, significant at the 1% level, implying that a 1% increase in remuneration is associated with a 1.943% increase in corruption levels. Unlike the pooled model, the control variable log(Population) has a stronger, negative relationship with corruption, with a coefficient of -5.370, significant at the 1% level. Including fixed effects reduces the model's explanatory power, as reflected by the adjusted R-squared of -0.018, which is expected due to the increased constraints of fixed effects. The F-statistic of 10.720 ($p < 0.01$) reinforces the strength of log(Remuneration) as an instrument even when accounting for unobserved heterogeneity.

In the pooled model with regional-level data (Column 3), the coefficient for log(Remuneration) is -4.948, significant at the 1% level. This indicates that a 1% increase in remuneration is associated with a 4.948% decrease in corruption levels. The model shows high explanatory power, with an adjusted R-squared of 0.744 and an F-statistic of 38.872 ($p < 0.01$). These results confirm that log(Remuneration) remains a strong and valid instrument at the regional level under the pooled specification. Finally, the fixed effects model with regional-level data (Column 4) provides further insights. Here, the coefficient for log(Remuneration) is 7.828, significant at the 1% level, indicating that a 1% increase in remuneration is associated with a 7.828% increase in corruption levels. The control variables log(Education) and log(Population) have coefficients of 4.617 and -4.343, respectively, with both being statistically significant at the 5% and 1% levels. Although the adjusted R-squared is lower at 0.236, the F-statistic of 8.690 ($p < 0.01$) reflects that log(Remuneration) remains a relevant instrument under this specification.

Overall, the results across all models confirm that log(Remuneration) satisfies the relevance condition for an instrumental variable. The consistent statistical significance of log(Remuneration) across pooled and fixed effects models and at both state and regional levels establishes it as a robust instrument for addressing endogeneity in subsequent stages of this analysis.

Table 3: Results of the first stage OLS for regional- and state-level models

First Stage OLS Results

	Dependent variable:			
	log(CPIM)			
	States		Regions	
	(1)	(2)	(3)	(4)
log(Remuneration)	-2.431***	1.943***	-4.948***	7.828***
	(0.531)	(0.516)	(0.769)	(1.123)
log(Education)	-0.272	1.161	-1.723**	4.617***
	(0.366)	(1.058)	(0.711)	(1.272)
log(Population)	-0.415***	-5.370***	-0.239**	-4.343**
	(0.048)	(1.143)	(0.107)	(1.907)
Constant	30.833***		56.339***	
	(4.905)		(4.615)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	40	40
R^2	0.589	0.152	0.764	0.510
Adjusted R^2	0.584	-0.018	0.744	0.236
Residual Std. Error	0.553 (df = 212)		0.388 (df = 36)	
F Statistic	101.410*** (df = 3; 212)	10.720*** (df = 3; 179)	38.872*** (df = 3; 36)	8.690*** (df = 3; 25)

Note: This table tests the relevance of remuneration as an IV for corruption (log(CPIM)). The results demonstrate strong statistical significance for log(Remuneration) in all models. The coefficients for remuneration are consistent across both state- and regional-level models, with high levels of significance (p<0.01), confirming its predictive power over corruption levels.
Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

To further strengthen the evidence that the relevance condition is fulfilled, the Pearson correlation coefficient between the logarithm of the IV and that of the CPIM was calculated. The calculation uses the logarithm of the variables, which is consistent with the models specified in this thesis. The resulting correlation coefficient is -0.653, indicating a strong negative correlation between the instrument and the CPIM. This significant correlation further supports the relevance condition by demonstrating the instrument has substantial predictive power for the endogenous variable.

Instrument exogeneity condition

The exogeneity condition postulates that the instrument must be correlated with the endogenous explanatory variable (in this case, corruption) but uncorrelated with the error term in the outcome equation (Wooldridge, 2010). In mathematical form:

(1.4)

$$Cov(IV,\ \mu) = 0$$

The average remuneration of federal government public servants fulfils this criterion. The determination of federal public servant remuneration is subject to institutional rules and regulations implemented uniformly across the country. These rules standardise compensation across different regions, irrespective of local economic conditions or development levels. The remuneration is revisited yearly and published on the government's official platform.[2] As such, the salary levels are isolated from state-level development indicators, further supporting the exogeneity condition.

That said, the difference in the yearly average remuneration of federal government public servants across states is determined by the composition and distribution of roles of different natures and salary brackets within each state. Federal government positions encompass various roles, from administrative and clerical positions to more specialised or senior roles, such as judges, auditors, and senior management, each with its own distinct salary bracket. The concentration of higher-paying roles in certain states, often those with larger federal administrative centres or federal institutions, naturally leads to higher average remuneration in those states. While up until the early 2000s, public servants tended to be more concentrated in the richer federative units, this started changing in 2007 as positions began to be more equally distributed across the country (Palotti et al., 2015). As the data observed in this thesis is captured from 2012, this does not create biases in the sample.

This variation in remuneration reflects the organisational and functional needs of the federal government rather than any direct influence of local economic conditions or development outcomes. For instance, a state that hosts numerous federal ministries, agencies, or military bases would likely have a higher average federal salary due to the presence of more senior positions. In states with less pronounced federal presence, the average salary would be lower, as the roles may predominantly be in lower-paying brackets. The current variations are driven by the operational and logistical requirements of the federal government and do not directly relate to the state's economic performance or development level. On top of that, the distribution is influenced by historical and political factors, such as the placement of federal institutions and the regional allocation of federal resources, which are typically decided at the national level.

This structural characteristic of federal remuneration distribution reinforces the validity of using average federal government remuneration as an instrument in the analysis. It highlights that while federal salaries may vary across states, these variations are determined by federal administrative needs and job roles and not by the economic or developmental characteristics of the states themselves. Given this context, the instrument can reliably capture variations in corruption influenced by public servant remuneration without being confounded by direct effects on state-level development outcomes.

The remuneration of federal public servants, while affecting their level of motivation (Corrêa et al., 2020) and, by extension, their engagement in corrupt activities, should not directly influence broader development indicators. The salary structure of federal employees is not designed to alter development outcomes but instead reflects compensation for services rendered. The number of federal government public servants across states is also not significant enough to generate meaningful effects on local wages and, consequently, development levels. Therefore, any impact on development would be indirect, operating solely through the corruption channel.

[2] The table of remuneration of federal government public servants is published on an yearly basis at: https://www.gov.br/servidor/pt-br/observatorio-de-pessoal-govbr/tabela-de-remuneracao-dos-servidores-publicos-federais-civis-e-dos-ex-territorios. It is important to note that there were no salary adjustments or changes in 2022. This does not affect the results of the analysis, as the data sample concludes in 2019.

Given that the model only uses one instrument, tests like Hansen's J test or Sargan for overidentifying restrictions can't be applied to test the exogeneity condition. With that in mind, the theoretical justification should serve as evidence that the exogeneity condition is fulfilled.

State Analysis

This section presents the state-level results of the second-stage models, which examine the relationship between corruption and key socio-economic indicators: the Gini coefficient, GDP per capita, HDI, poverty rates, and extreme poverty rates. Each model incorporates controls for education and population to account for their influence on the dependent variables.

The second-stage results will include OLS models (both with and without fixed effects) alongside the IV models with and without fixed effects to provide a more comprehensive analysis. This approach enables a comparison of the baseline associations observed in OLS models with the causal estimates from the IV strategy. Including both fixed-effects and non-fixed-effects specifications highlights the role of unobserved heterogeneity at the state and temporal levels, offering additional insights into how these factors shape the relationship between corruption and socio-economic development indicators.

Gini Coefficient

Table 4 provides second-stage regression results regarding the relationship between corruption and inequality, with inequality measured using the Gini coefficient and controlling for education and population. In the OLS model (Column 1), the results suggest that corruption is positively associated with inequality, with a 1% increase in corruption associated with a 0.061% increase in inequality (Column 1). However, the relationship becomes negative and significant when fixed effects are introduced (Column 2). The coefficient is also substantially smaller, with a 1% increase in corruption associated with a 0.008% decrease in inequality.

The IV models (Columns 3 and 4) provide similar results. The IV model without fixed effects (Column 3) indicates a positive and significant relationship between corruption and inequality, with a 1% increase in corruption associated with a 0.085% increase in inequality. However, when fixed effects are added (Column 4), the relationship turns negative, with a 1% increase in corruption predicting a 0.044% decrease in inequality.

Regarding explanatory power, the OLS model (Column 1) achieves the highest R-squared (0.236), indicating that approximately 23.6% of the variation in inequality is explained by corruption, education, and population. However, this drops substantially to 0.018 in the fixed-effects OLS model (Column 2), likely due to the inclusion of fixed effects absorbing much of the variation. Similarly, the IV model without fixed effects (Column 3) has an R-squared of 0.201, while the fixed-effects IV model (Column 4) shows a much lower R-squared of 0.012. These reductions in R-squared when fixed effects are added suggest that much of the variation in inequality is driven by unobserved heterogeneity, which is controlled for in these models.

The control variables, log(Education) and log(Population), also exhibit significant associations with inequality. In models with fixed effects, higher education is positively associated with inequality, while higher population is negatively associated.

Table 4: Second-Stage Regression Results: CPIM and the Gini Coefficient (State-level Data)

Second Stage Results with State-level Data - Gini Coefficient

	Dependent variable:			
	log(Gini)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.061***	-0.008***	0.085***	-0.044***
	(-0.001)	(0.00004)	(-0.005)	(0.001)
log(Education)	-0.019***	0.112***	-0.021***	0.148***
	(-0.004)	(-0.0001)	(-0.002)	(-0.0002)
log(Population)	0.022***	-0.095***	0.037***	-0.282***
	(-0.001)	(0.0001)	(-0.003)	(0.003)
Constant	-1.073***		-1.355***	
	(0.028)		(0.066)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	216	216
R^2	0.236	0.018	0.201	0.012
Adjusted R^2	0.225	-0.180	0.190	-0.187
Residual Std. Error (df = 212)	0.071		0.072	
F Statistic	21.778*** (df = 3; 212)	1.077 (df = 3; 179)		3.971

Note: This table presents second-stage regression results on the relationship between corruption (log(CPIM)) and inequality (log(Gini)). The OLS and IV models include controls for education and population, with and without fixed effects. In OLS and IV models with fixed effects, there is a consistent negative association between corruption and inequality.

Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

GDP per Capita

Table 5 presents the second-stage regression results on the relationship between corruption and GDP per capita. The results of the OLS model (Column 1) suggest a significant negative association between the two variables. A 1% increase in corruption is associated with a 0.198% decrease in GDP per capita. When fixed effects are introduced in Column 2, the relationship becomes positive and statistically significant, with a 1% increase in corruption associated with a 0.017% increase in GDP per capita. The IV models (Columns 3 and 4) also yield contrasting results. The IV model without fixed effects (Column 3) estimates a strong negative association, with a 1% increase in corruption leading to a 0.383% decrease in GDP per capita. When fixed effects are introduced in the IV framework (Column 4), the relationship shifts again, with corruption exhibiting a small but positive and significant effect, where a 1% increase in corruption predicts a 0.012% increase in GDP per capita.

Higher education is significantly associated with higher GDP per capita without fixed effects. However, in fixed-effects models, the relationship between education and GDP per capita turns negative, and the coefficient becomes substantially smaller. Population, on the other hand, exhibits a negative association with GDP per capita in all models.

The R-squared values show significant differences in explanatory power. The OLS model (Column 1) exhibits the highest R-squared (0.712). Still, including fixed effects in both OLS and IV models significantly reduces the R-squared values (0.133 in Column 2 and 0.132 in Column 4, respectively). This drop reflects the absorption of variation by fixed effects, leaving less variation to be explained by the remaining covariates. The IV model without fixed effects (Column 3) shows an intermediate R-squared of 0.650.

Table 5: Second-Stage Regression Results: CPIM and GDP per Capita (State-level Data)

Second Stage Results with State-level Data - GDP per Capita

	Dependent variable:			
	log(GDP pc)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	-0.198***	0.017***	-0.383***	0.012***
	(-0.019)	(0.0004)	(-0.082)	(0.004)
log(Education)	2.750***	-0.019***	2.760***	-0.014***
	(-0.145)	(0.0003)	(-0.126)	(-0.006)
log(Population)	-0.006	-0.836***	-0.116**	-0.860***
	(-0.010)	(0.002)	(-0.048)	(0.028)
Constant	-8.015***		-5.910***	
	(0.783)		(1.450)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	216	216
R^2	0.712	0.133	0.650	0.132
Adjusted R^2	0.708	-0.041	0.645	-0.042
Residual Std. Error (df = 212)	0.241		0.266	
F Statistic	174.586*** (df = 3; 212)	9.147*** (df = 3; 179)		25.253***
Note:	This table presents second-stage regression results on the relationship between corruption (log(CPIM)) and GDP per capita. Both OLS and IV models indicate a significant negative association, with fixed effects altering this relationship to positive and significant. Significance levels: * 90%, ** 95%, *** 99%.			

Source: Calculated by the author using R programming language.

HDI (M)

Table 6 presents the second-stage regression results with HDI as the dependent variable. Across all models, corruption is negatively and significantly associated with HDI. The OLS model (Column 1) shows that a 1% increase in corruption is associated with a 0.025% decrease in HDI, a result that remains significant in all specifications. When fixed effects are added (Column 2), the negative relationship persists but is reduced in magnitude. In the IV models, corruption's negative impact on HDI is stronger without fixed effects (Column 3), with a 1% increase in corruption leading to a

0.082% decrease in HDI. Adding fixed effects (Column 4) again reduces the magnitude of the association, though the relationship remains significant, with a 0.009% decrease in HDI for every 1% increase in corruption.

The R-squared values reveal similar trends to those seen in earlier tables. The OLS model (Column 1) has a high R-squared of 0.837, indicating strong explanatory power. However, as with GDP per capita and inequality, including fixed effects in Columns 2 and 4 significantly reduces the R-squared. The IV model with fixed effects (Column 4) has a lower R-squared (0.374).

Table 6: Second-Stage Regression Results: CPIM and HDI (M) (State-level Data)

Second Stage Results with State-level Data - HDI (M)

	Dependent variable:			
	log(HDI)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	-0.025***	-0.002***	-0.082***	-0.009***
	(-0.0001)	(0.00000)	(-0.002)	(0.0001)
log(Education)	0.435***	0.225***	0.438***	0.232***
	(-0.001)	(0.00000)	(-0.002)	(-0.0001)
log(Population)	0.003***	-0.090***	-0.031***	-0.126***
	(-0.0001)	(0.00001)	(-0.001)	(0.0005)
Constant	-2.078***		-1.435***	
	(0.007)		(0.039)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	216	216
R^2	0.837	0.447	0.557	0.374
Adjusted R^2	0.835	0.336	0.551	0.248
Residual Std. Error (df = 212)	0.026		0.043	
F Statistic	364.114*** (df = 3; 212)	48.309*** (df = 3; 179)		123.996***
Note:	This table presents second-stage regression results on the relationship between corruption (log(CPIM)) the HDI for municipalities. Across all models, corruption is generally negatively associated with HDI. Fixed effects reduce the magnitude of this relationship, but it remains significant. Significance levels: * 90%, ** 95%, *** 99%.			

Source: Calculated by the author using R programming language.

Poverty and Extreme Poverty Rates

Tables 7 and 8 present second-stage regression results examining the relationship between corruption (log(CPIM)) and poverty outcomes, specifically the rates of people living in poverty (Table 7) and extreme poverty (Table 8). Across both tables, the results highlight significant but nuanced relationships that depend on model specification.

In Table 7, the OLS model (Column 1) suggests a positive and significant relationship between corruption and the rate of people in poverty, with a 1% increase in corruption associated with a 0.542% increase in poverty rates. However, when fixed effects are introduced (Column 2), the magnitude of the association is drastically reduced, with a 1% increase in corruption associated with only a 0.014% increase in poverty rates. This shift underscores the potential influence of unobserved state-level and time-varying factors that may confound the relationship in the OLS model.

The IV models provide further insights. Without fixed effects (Column 3), the IV model estimates a much larger positive relationship, with a 1% increase in corruption leading to a 1.024% rise in poverty rates. However, when fixed effects are introduced (Column 4), the association becomes statistically insignificant, with corruption no longer meaningfully impacting poverty rates. This finding aligns with the idea that unobserved heterogeneity plays a significant role in shaping this relationship.

Table 7: Second-Stage Regression Results: CPIM and the Rate of People in Poverty (State-level Data)

Second Stage Results with State-level Data - Rate of People in Poverty

	Dependent variable:			
	log(Rate of People in Poverty)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.542***	0.014***	1.024***	-0.030
	(-0.042)	(0.001)	(-0.363)	(0.053)
log(Education)	-3.601***	-0.147***	-3.626***	-0.103***
	(-0.213)	(-0.003)	(-0.249)	(-0.036)
log(Population)	0.092***	1.626***	0.378*	1.394***
	(-0.030)	(0.002)	(-0.227)	(0.265)
Constant	14.462***		8.998*	
	(1.435)		(5.403)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	216	216
R^2	0.631	0.098	0.484	0.079
Adjusted R^2	0.625	-0.084	0.476	-0.106
Residual Std. Error (df = 212)	0.460		0.544	
F Statistic	120.622*** (df = 3; 212)	6.451*** (df = 3; 179)		18.627***
Note:	This table presents second-stage regression results on the relationship between corruption (log(CPIM)) and the rate of people in poverty. The OLS models show a consistent positive association between corruption and poverty, while the IV models highlight a significant positive relationship only in the absence of fixed effects. Once fixed effects are added, the association becomes statistically insignificant. Significance levels: * 90%, ** 95%, *** 99%.			

Source: Calculated by the author using R programming language.

The results for extreme poverty (Table 8) mirror many of the trends seen in Table 7 but with some notable differences. The OLS model (Column 1) shows a strong positive association, with a 1% increase in corruption linked to a 0.595% rise in extreme poverty rates. As with general poverty, this relationship weakens considerably in the fixed-effects OLS model (Column 2), where the estimated increase is only 0.035% for a 1% rise in corruption. The IV models indicate a stronger relationship between corruption and extreme poverty. In the IV model without fixed effects (Column 3), corruption is associated with a 0.842% increase in extreme poverty rates. Interestingly, when fixed effects are included (Column 4), the relationship turns negative but statistically insignificant.

Both OLS models (Column 1) show relatively high R-squared values (0.631 for poverty and 0.543 for extreme poverty), reflecting strong explanatory power. However, as with earlier results, including fixed effects reduces the R-squared significantly, reflecting the absorption of variation by unobserved state-level factors.

Table 8: Second-Stage Regression Results: CPIM and the Rate of People in Extreme Poverty (State-level Data)

Second Stage Results with State-level Data - Rate of People in Extreme Poverty

	Dependent variable:			
	log(Rate of People in Extreme Poverty)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.595^{***}	0.035^{***}	0.842^{**}	-0.017
	(-0.055)	(0.002)	(-0.331)	(0.129)
log(Education)	-3.395^{***}	0.315^{***}	-3.408^{***}	0.366^{**}
	(-0.285)	(-0.004)	(-0.301)	(-0.182)
log(Population)	0.191^{***}	2.567^{***}	0.338	2.301^{***}
	(-0.042)	(0.009)	(-0.214)	(0.676)
Constant	11.039^{***}		8.235	
	(1.966)		(5.344)	
Fixed effects	No	Yes	No	Yes
Observations	216	216	216	216
R^2	0.543	0.089	0.507	0.078
Adjusted R^2	0.537	-0.095	0.500	-0.108
Residual Std. Error (df = 212)	0.529		0.549	
F Statistic	84.064^{***} (df = 3; 212)	5.808^{***} (df = 3; 179)		16.255^{***}

Note: This table presents second-stage regression results on the relationship between corruption (log(CPIM)) and the rate of people in extreme poverty. Both OLS and IV models with state-level data are presented, including specifications with two-way fixed effects. The OLS models show a consistent positive association between corruption and extreme poverty, while the IV models highlight a significant positive relationship only in the absence of fixed effects. Once fixed effects are added, the association becomes statistically insignificant.

Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

Regional Analysis

The data explorations presented in the "Descriptive Statistics" section, as well as the existing literature, show that the socioeconomic indicators explored in this thesis have strong regional patterns. Though there are a variety of theories and no clear consensus on the origins of these inequalities, the fact that they exist is widely acknowledged (Da Silva, 2017). These inequalities indicate a need to run the previous models with variables aggregated at the region level in order to understand regional variations that could influence the strength and direction of these relationships. The table below outlines the aggregations made to the existing variables in the dataset prior to running the model. The resulting dataset has a total of 40 observations, with data for 8 years and five different regions (Central-West, North, Northeast, South, Southeast). Given the size of the sample, the results in this section should give orientation regarding the direction of relationships but must be interpreted with caution.

Table 9: Methods for aggregating data at the regional level

Variable	**Aggregation Method**
CPIM	Simple average
Gini Coefficient	Weighted average using the state's population as weight
% of people in poverty	Weighted average using the state's population as weight
% of people in extreme poverty	Weighted average using the state's population as weight
GDP per capita	Weighted average using the state's population as weight
IDH (M)	Weighted average using the state's population as weight
Education	Weighted average using the state's population as weight
Population	Sum all populations of states within the same region
Remuneration	Simple average

The following sections present the regional-level results of the second-stage models. As in the state-level specifications, each model incorporates controls for education and population and the results are displayed in tables that cover both OLS and IV specifications, including versions with fixed effects. The analysis for each dependent variable will further include a comparison with state-level results.

Gini Coefficient

Table 10 presents second-stage regression results exploring the relationship between corruption and inequality at the regional level, complementing the earlier state-level results. In the OLS model (Column 1), corruption is positively and significantly associated with inequality at the regional level, with a 1% increase in corruption linked to a 0.074% rise in the Gini coefficient. This is consistent with the state-level results, where the OLS model similarly found a positive association (0.061%). However, the inclusion of fixed effects in the regional OLS model (Column 2) alters the relationship, yielding a negative and significant coefficient of -0.029%. This shift mirrors the pattern observed in the state-level fixed-effects model (-0.008%), though the magnitude of the effect is stronger at the regional level.

In the IV models without fixed effects (Column 3), corruption is positively and significantly associated with inequality at the regional level (0.089%), a finding that aligns with the state-level IV model (0.085%). However, with the inclusion of fixed effects (Column 4), the relationship turns negative and remains statistically significant (-0.037%), similar to the state-level fixed-effects IV model (-0.044%).

The R-squared values for the regional-level OLS model (0.844) are slightly higher than those for the state-level model (0.236), reflecting the stronger explanatory power of corruption, education, and population in explaining regional-level inequality. However, as with the state-level results, the inclusion of fixed effects in both OLS and IV models leads to a substantial drop in R-squared.

The regional-level results are broadly consistent with the state-level findings but reveal stronger and more pronounced shifts in the relationship between corruption and inequality. The negative fixed-effects coefficients at the regional level are larger in magnitude than their state-level counterparts.

Table 10: Second-Stage Regression Results: CPIM and the Gini Coefficient (Regional-level Data)

Second Stage Results with Regional-level Data - Gini Coefficient

	Dependent variable:			
	log(Gini)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.074***	-0.029***	0.089***	-0.037***
	(-0.001)	(0.0001)	(-0.004)	(0.0002)
log(Education)	-0.056***	-0.319***	-0.001	-0.320***
	(-0.014)	(-0.0004)	(-0.022)	(0.001)
log(Population)	0.071***	-0.738***	0.080***	-0.791***
	(-0.001)	(0.0001)	(-0.003)	(0.0002)
Constant	-1.861***		-2.274***	
	(0.084)		(0.151)	
Fixed effects	No	Yes	No	Yes
Observations	40	40	40	40
R^2	0.844	0.333	0.825	0.327
Adjusted R^2	0.831	-0.041	0.811	-0.050
Residual Std. Error (df = 36)	0.024		0.026	
F Statistic	64.836*** (df = 3; 36)	4.156** (df = 3; 25)		11.004**
Note:	This table presents second-stage regression results exploring the relationship between corruption (log(CPIM)) and inequality (log(Gini)) at the regional level. The results indicate that corruption is positively and significantly associated with inequality in OLS and IV models without fixed effects, while the inclusion of fixed effects reveals a negative and significant relationship, suggesting that controlling for unobserved regional and temporal heterogeneity alters the direction of the association. Significance levels: * 90%, ** 95%, *** 99%.			

Source: Calculated by the author using R programming language.

GDP per Capita

Table 11 presents second-stage regression results with GDP per capita as the dependent variable. In the OLS model (Column 1), corruption exhibits a significant negative association with GDP per capita, with a 1% increase in corruption corresponding to a 0.278% decrease in GDP per capita. This result is comparable in direction to the state-level OLS finding (-0.198%), though the effect is stronger at the regional level. When fixed effects are added (Column 2), the relationship shifts direction and becomes positive, with a 1% increase in corruption associated with a 0.064% rise in GDP per capita. This positive shift is consistent with state-level findings, where corruption also exhibited a smaller positive coefficient (0.017%) in the fixed-effects OLS model.

The IV models provide additional insights. Without fixed effects (Column 3), the IV model estimates a strong negative relationship between corruption and GDP per capita, with a 1% increase in corruption linked to a 0.276% decrease in GDP per capita. This mirrors the state-level IV results, which showed a similarly strong negative relationship (-0.383%). However, with the inclusion of fixed effects (Column 4), the regional-level IV model yields a positive and significant relationship, with a 1% increase in corruption predicting a 0.148% increase in GDP per capita. This positive coefficient is larger than the corresponding state-level fixed-effects IV estimate (0.012%).

The R-squared values in the regional-level OLS model (0.835) are similar to those in the state-level results (0.712), indicating strong explanatory power. However, the inclusion of fixed effects at the regional level significantly reduces the R-squared (0.127 in Column 2), reflecting the absorption of variation by unobserved regional factors. Similarly, the IV model without fixed effects (Column 3) retains a high R-squared (0.835), comparable to its state-level counterpart (0.650). The IV model with fixed effects (Column 4) shows a much lower R-squared (0.115).

The regional-level results align closely with the state-level findings but highlight stronger shifts in the direction and magnitude of the corruption-GDP per capita relationship when fixed effects are included. The stronger positive coefficient in the regional fixed-effects IV model (0.148%) compared to the state-level model (0.012%) suggests that the impact of corruption on economic outcomes may be more pronounced when aggregated at the regional level.

Table 11: Second-Stage Regression Results: CPIM and GDP per Capita (Regional-level Data)

Second Stage Results with Regional-level Data - GDP per Capita

	Dependent variable:			
	log(GDP pc)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	-0.278^{***}	0.064^{***}	-0.276^{**}	0.148^{***}
	(-0.041)	(0.0005)	(-0.112)	(0.004)
log(Education)	2.562^{***}	0.875^{***}	2.570^{***}	0.877^{***}
	(-0.296)	(0.017)	(-0.548)	(0.038)
log(Population)	-0.267^{***}	0.828^{***}	-0.266^{***}	1.337^{***}
	(-0.047)	(0.006)	(-0.093)	(0.049)
Constant	-2.424		-2.488	
	(2.129)		(4.130)	
Fixed effects	No	Yes	No	Yes
Observations	40	40	40	40
R^2	0.835	0.127	0.835	0.115
Adjusted R^2	0.821	-0.362	0.821	-0.381
Residual Std. Error (df = 36)	0.171		0.171	
F Statistic	60.758^{***} (df = 3; 36)	1.212 (df = 3; 25)		5.788

Note: This table presents second-stage regression results examining the relationship between corruption (log(CPIM)) and GDP per capita at the regional level. The results show a significant negative association in OLS and IV models without fixed effects. The inclusion of fixed effects reveals a positive and significant relationship, suggesting that controlling for unobserved regional and temporal heterogeneity alters the observed association.
Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

HDI (M)

Table 12 presents second-stage regression with HDI as the dependent variable. In the OLS model (Column 1), corruption is negatively and significantly associated with HDI, with a 1% increase in corruption linked to a 0.039% decrease in HDI, mirroring state-level OLS findings (-0.025%). Including fixed effects (Column 2) reverses the direction of the relationship, with a small but significant positive coefficient of 0.0001%. This shift in direction does not happen on the state-level fixed-effects OLS model, though the effect was also substantially reduced.

The IV models provide additional insights. Without fixed effects (Column 3), the IV model estimates a negative relationship between corruption and HDI, with a 1% increase in corruption leading to a 0.054% decrease in HDI. This result is comparable to the state-level IV estimate (-0.082%). In the IV model with fixed effects (Column 4), the relationship again turns positive but remains small, with a 1% increase in corruption associated with a 0.004% increase in HDI.

As in the state-level models, the control variables show consistent and significant relationships with HDI. Education has a positive and significant association with HDI in all specifications, with more substantial effects in models without fixed effects. Population displays a negative relationship with HDI across all models, though the magnitude is larger in fixed-effects specifications.

The regional-level OLS model's R-squared values (0.960) are notably higher than those in the state-level OLS model (0.837), indicating greater explanatory power at the regional level. However, as with the state-level results, including fixed effects substantially reduces the R-squared (0.361 in Column 2 and 0.338 in Column 4), reflecting the absorption of unobserved heterogeneity. The IV model without fixed effects (Column 3) retains a high R-squared (0.943), consistent with the state-level IV result (0.557).

Table 12: Second-Stage Regression Results: CPIM and HDI (M) (Regional-level Data)

Second Stage Results with Regional-level Data - HDI (M)

	Dependent variable:			
	log(HDI)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	-0.039^{***}	0.0001^{***}	-0.054^{***}	0.004^{***}
	(-0.0002)	(0.00000)	(-0.001)	(0.00002)
log(Education)	0.462^{***}	0.264^{***}	0.405^{***}	0.264^{***}
	(-0.002)	(0.00003)	(-0.005)	(0.00001)
log(Population)	-0.005^{***}	-0.039^{***}	-0.014^{***}	-0.017^{***}
	(-0.0001)	(-0.00005)	(-0.001)	(-0.0001)
Constant	-2.022^{***}		-1.598^{***}	
	(0.009)		(0.035)	
Fixed effects	No	Yes	No	Yes
Observations	40	40	40	40
R^2	0.960	0.361	0.943	0.338
Adjusted R^2	0.957	0.003	0.938	-0.033
Residual Std. Error (df = 36)	0.013		0.015	
F Statistic	289.322^{***} (df = 3; 36)	4.701^{***} (df = 3; 25)		13.993^{***}

Note: This table presents second-stage regression results on the relationship between corruption (log(CPIM)) and the HDI at the regional level. Corruption is negatively associated with HDI in the OLS and IV models without fixed effects, with the coefficients suggesting small but significant declines in HDI as corruption increases. However, in the fixed-effects specifications (Columns 2 and 4), the relationship diminishes, with coefficients becoming positive but small, suggesting that fixed effects might capture much of the unobserved heterogeneity affecting this relationship.

Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

Poverty and Extreme Poverty Rates

Table 13 presents second-stage regression results examining the relationship between corruption and the rate of people in poverty at the regional level. The findings are generally consistent with state-level findings but show greater sensitivity at the regional level.

In the OLS model (Column 1), corruption is positively and significantly associated with poverty, with a 1% increase in corruption linked to a 0.805% increase in poverty rates. This effect is stronger than the corresponding state-level result (0.542%). When fixed effects are introduced (Column 2), the relationship becomes negative (-0.109%) but remains significant. The IV model without fixed effects (Column 3) estimates a strong positive relationship, with a 1% increase in corruption associated with a 0.902% rise in poverty rates. This is consistent with the state-level IV estimate (1.024%). However, when fixed effects are added (Column 4), the relationship turns negative but small (-0.073%), similar to the state-level IV with fixed-effects result (-0.030%).

Table 13: Second-Stage Regression Results: CPIM and the Rate of People in Poverty (Regional-level Data)

Second Stage Results with Regional-level Data - Rate of People in Poverty

	Dependent variable:			
	log(Rate of People in Poverty)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.805***	-0.109***	0.902***	-0.073***
	(-0.108)	(0.001)	(-0.282)	(0.001)
log(Education)	-3.216***	-2.239***	-2.847*	-2.239***
	(-0.904)	(0.023)	(-1.525)	(-0.025)
log(Population)	0.516***	-0.761***	0.574**	-0.544***
	(-0.114)	(0.012)	(-0.230)	(-0.004)
Constant	4.598		1.838	
	(5.952)		(10.955)	
Fixed effects	No	Yes	No	Yes
Observations	40	40	40	40
R^2	0.897	0.249	0.892	0.240
Adjusted R^2	0.889	-0.171	0.883	-0.186
Residual Std. Error (df = 36)	0.255		0.262	
F Statistic	104.718*** (df = 3; 36)	2.764* (df = 3; 25)		4.872

Note: This table presents second-stage regression results exploring the relationship between corruption (log(CPIM)) and the rate of people in poverty at the state level. Corruption is positively and significantly associated with poverty in the OLS and IV models without fixed effects (Columns 1 and 3). The introduction of fixed effects (Columns 2 and 4) diminishes the magnitude of the relationship, turning it insignificant in the IV model with fixed effects.

Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

Table 14 extends the analysis to extreme poverty, with patterns similar to those observed for general poverty rates but with notable differences in magnitude and significance. In the OLS model (Column 1), corruption is strongly and significantly associated with extreme poverty, with a 1% increase in corruption corresponding to a 0.843% rise in extreme poverty rates. This result is larger than the state-level OLS estimate (0.595%).

Introducing fixed effects in the OLS model (Column 2) reduces the relationship to near zero (0.002%) and becomes statistically insignificant. The fixed-effects IV model (Column 4) similarly shows a small negative but significant relationship (-0.030%), closely mirroring the state-level IV result for extreme poverty (-0.030%).

Table 14: Second-Stage Regression Results: CPIM and the Rate of People in Extreme Poverty (Regional-level Data)

Second Stage Results with Regional-level Data - Rate of People in Extreme Poverty

	Dependent variable:			
	log(Rate of People in Extreme Poverty)			
	OLS	OLS with FE	IV	IV with FE
	(1)	(2)	(3)	(4)
log(CPIM)	0.843^{***}	0.002	0.793^{**}	-0.030^{***}
	(-0.147)	(0.004)	(-0.385)	(0.007)
log(Education)	-2.551^{**}	-1.951^{***}	-2.740	-1.952^{***}
	(-1.162)	(0.098)	(-2.169)	(0.127)
log(Population)	0.635^{***}	1.147^{***}	0.606^{**}	0.955^{***}
	(-0.199)	(0.053)	(-0.303)	(0.066)
Constant	-1.220		0.193	
	(8.578)		(15.099)	
Fixed effects	No	Yes	No	Yes
Observations	40	40	40	40
R^2	0.849	0.120	0.848	0.114
Adjusted R^2	0.837	-0.372	0.835	-0.381
Residual Std. Error (df = 36)	0.310		0.311	
F Statistic	67.696^{***} (df = 3; 36)	1.139 (df = 3; 25)		3.457

Note: This table presents second-stage regression results examining the relationship between corruption (log(CPIM)) and the rate of people in extreme poverty at the regional level. OLS and IV models without fixed effects (Columns 1 and 3) show a positive and significant relationship, while Column 2 (OLS with fixed effects) shows no significant relationship. The IV model with fixed effects (Column 4) indicates a smaller, negative, but statistically significant coefficient.

Significance levels: * 90%, ** 95%, *** 99%.

Source: Calculated by the author using R programming language.

Across both tables, the control variables exhibit consistent trends with the state-level results. Education is negatively associated with poverty and extreme poverty in all models, though the magnitude decreases in fixed-effects specifications. Population shows mixed effects.

Robustness Tests

This section presents robustness checks to validate the findings of the primary analysis as described in the previous sections. The robustness tests are essential for ensuring that the results are not driven by specific model assumptions, variable definitions, or sample selections but rather reflect a genuine relationship between corruption and socioeconomic indicators. Firstly, the strategy to address heteroskedasticity across the sample will be outlined. Secondly, a series of placebo tests will be applied to the analysis to ensure that the findings are consistent.

Heteroskedasticity

To ensure the reliability of the regression results presented, it is a critical step to test for the presence of heteroskedasticity, which can undermine the validity of standard error estimates. The Breusch-Pagan (BP) test was employed for each model used in the state-level analysis to diagnose this issue. The BP test is designed to detect any linear form of heteroskedasticity by evaluating whether the squared residuals from the regression are correlated with the independent variables (Breusch et al., 1979). The results of the BP test for the state-level models, focusing on key socio-economic indicators, are presented in Table 10.

Table 15: Results of the BP test for each of the state-level models

	BP_Statistic	P_Value
Gini	5.2993	0.2579
GDP per capita	44.3856	0
HDI (M)	10.1664	0.0377
% of people in poverty	20.6749	0.0004
% of people in extreme poverty	17.2672	0.0017

Source: Calculated by the author using R programming language.

The BP tests presented in Table 15 indicate the presence of heteroskedasticity in several state-level models, particularly for the percentage of people in poverty and extreme poverty. Given these findings, each model presented throughout the thesis employed heteroskedasticity-consistent (HC) standard errors of type HC1 in the regression results to correct for the potential bias in the standard errors. The HC1 estimator adapts White's original estimator (HC0). It includes a degree of freedom adjustment, making it more appropriate for finite samples (Wooldridge, 2010). Applying robust standard errors makes the statistical inference drawn from the models more reliable.

It is essential to highlight that this analysis employs HC1 standard errors without clustering. This methodological choice was made based on the relatively small number of years and states included in the sample, which makes clustered standard errors potentially unreliable. While there is no universally agreed-upon threshold for the minimum number of clusters required for clustering to produce reliable estimates, the literature suggests thresholds ranging from fewer than 20 to fewer than 50 clusters in balanced panel data (Cameron et al., 2015). Given that this study is based on data for 27 states over an 8-year period, the number of clusters lies at the lower end of this range.

Model Selection

The primary analysis employs a fixed effects approach to control for unobserved heterogeneity across time and federative units. Given the strong regional and state-specific patterns described in previous sections, this approach is essential as it accounts for the likely correlation between unobserved, time-invariant factors specific to each federative unit and time period with the independent variables. A fixed effects model mitigates potential biases, leading to more precise and reliable estimates. The model's emphasis on within-unit variation aligns closely with the research question, which seeks to understand how changes within each unit influence the outcome variable over time.

To ensure the validity of this approach, a Wald test was conducted to assess whether the difference in coefficients between the fixed effects and random effects models is statistically significant enough to impact model specification. The results of the Wald test across all models suggest that the specifications are statistically equivalent, confirming that the fixed effects model does not compromise the robustness of the findings discussed in this thesis.

The results section also includes OLS models (with and without fixed effects) and an IV model without fixed effects to provide a broader perspective. While the findings differ across model specifications, these variations highlight the importance of accounting for endogeneity and unobserved heterogeneity. The fixed effects IV model is the most restrictive specification and is considered the most appropriate for addressing the research question, as it provides estimates that better control for potential confounding factors. Including alternative models allows for a more nuanced interpretation of the findings, helping to identify the conditions under which the relationship between corruption and development outcomes changes.

Placebo Test with Simulated Data

This section focuses on the most restricted specification (IV with fixed effects) for conducting placebo tests. This choice is motivated by its methodological rigour: the IV with fixed effects specification addresses both endogeneity and unobserved heterogeneity, making it the most robust model for assessing whether the foundational assumptions of the research design hold. As Eggers et al. (2024) discuss, spurious relationships could emerge if these assumptions are violated. The analysis minimises the risk of detecting false positives due to model weaknesses by concentrating on the most stringent specification.

Five artificial variables were generated to perform these tests, one for each model, which serve as placebo variables. These variables were created using the R programming language with a fixed seed of 123, ensuring reproducibility. Each placebo variable follows a log-normal distribution, with the mean and standard deviation set to match those of the independent variable it replaces in each respective model. This approach was selected to ensure that the placebo variables mimic the statistical properties of the original data, allowing a better assessment of the models' validity. By doing so, the test can detect potential biases or spurious effects resulting from model misspecification or violations of underlying assumptions.

The two tables presented below (Tables 16 and 17) show the results of the models with placebo variables as outcomes. The coefficient of CPIM in all models presented is statistically insignificant, with very small effect sizes. The R-squared values are also low across all models, indicating that the models do not explain much of the variance in the placebo outcomes.

The lack of significance and the small magnitude of coefficients indicate that the instrument has no spurious relationship with the placebo outcomes. These results collectively reinforce the robustness of the IV models by demonstrating that when the outcome variables are replaced with random placebo variables, the instrument does not produce significant effects. This suggests that the original findings regarding the CPIM and its impact on the outcome variables are likely not driven by spurious correlations or model misspecification.

Table 16: Results of the second stage OLS, with a placebo variable as outcome - Gini and GDP per capita

Second Stage Results with State-level Data - Placebo - Gini and GDP per Capita

	Dependent variable:	
	Log(Placebo Gini)	Log(Placebo GDP pc)
	(1)	(2)
log(CPIM)	-0.013	0.180
	(0.072)	(0.423)
log(Education)	-0.131	-1.029
	(0.319)	(1.882)
log(Population)	0.181	1.137
	(0.472)	(2.790)
Observations	216	216
R^2	0.00000	0.002
Adjusted R^2	-0.201	-0.198
F Statistic	1.217	0.455
Note:	This table presents placebo test results using the IV specification with fixed effects. The dependent variables are log-transformed placebo versions of Gini and GDP per capita. None of the coefficients for log(CPIM) are significant, indicating the robustness of the main analysis and suggesting no spurious relationships due to the specification. Significance levels: * 90%, ** 95%, *** 99%.	

Source: Calculated by the author using R programming language.

Table 17: Results of the second stage OLS, with a placebo variable as outcome - HDI (M), % of people in poverty, % of people in extreme poverty

Second Stage Results with State-level Data - Placebo - HDI, Poverty and Extreme Poverty Rates

	Dependent variable:		
	Log(Placebo HDI)	Log(Placebo Poverty)	Log(Placebo Extreme Poverty)
	(1)	(2)	(3)
log(CPIM)	0.059	0.383	0.216
	(0.065)	(0.472)	(0.528)
log(Education)	-0.455	-1.553	3.064
	(0.291)	(2.101)	(2.350)
log(Population)	0.422	2.428	0.252
	(0.431)	(3.115)	(3.484)
Observations	216	216	216
R^2	0.004	0.0001	0.011
Adjusted R^2	-0.196	-0.201	-0.188
F Statistic	3.429	1.136	2.813
Note:	This table presents placebo test results using the IV specification with fixed effects. The dependent variables are log-transformed placebo versions of HDI, poverty, and extreme poverty rates. The results show no significant associations with log(CPIM), further supporting the robustness of the main findings and mitigating concerns of specification bias. Significance levels: * 90%, ** 95%, *** 99%.		

Source: Calculated by the author using R programming language.

Discussion

This thesis empirically explored the relationship between corruption and socioeconomic indicators across Brazilian federative units and regions. The initial hypothesis was that corruption would either improve or hamper development across the country. The results offer nuanced insights into the complex relationship between these variables.

The findings indicate a mixed connection between corruption and broader socioeconomic indicators. While previous literature often frames corruption as uniformly detrimental (Mauro, 1995; Tanzi & Davoodi, 1997), the results of this study suggest it has both redistributive and obstructive effects. For instance, the negative association between corruption and HDI shows how systemic inefficiencies might compromise access to essential public services, such as healthcare and education (Bazie ad et al. (2024). At the same time, the potential redistributive effects suggested by the reduction in inequality and increase in GDP per capita may reflect the informal redistribution mechanisms, as indicated by Dreher et al. (2013) and Leff (1964).

The exploration of four model specifications in this thesis, with the IV fixed-effects model as the primary approach, seeks to address unobserved heterogeneity. The fixed-effects specification significantly altered the results in most models, revealing previously hidden within-unit variations and highlighting that localised governance structures likely influence corruption and development outcomes. For instance, corruption was negatively associated with GDP per capita and HDI in the IV model without fixed effects. However, including fixed effects revealed that corruption could exhibit a small but positive association with these variables at the state level. This aligns partially with findings from Lucio et al. (2020), who showcased that when spatial dependence is accounted for, the impact of corruption on GDP per capita at the state level in Brazil becomes nonsignificant.

As indicated by the statistically significant coefficient in the IV model with fixed effects, the positive association between corruption and GDP per capita challenges the conventional "sand the wheels" theory. This finding suggests that, in the Brazilian context, higher levels of corruption may be linked to increased economic activity, potentially supporting the "grease the wheels" hypothesis in regions where bureaucratic inefficiencies are prevalent. However, this interpretation needs careful consideration, as the positive correlation does not necessarily imply that corruption benefits long-term development. Given the short time frame of the dataset, it may reflect a short-term boost in economic output due to the circumvention of bureaucratic barriers, which could have detrimental effects on sustainable development. The results for GDP align with Spyromitros et al. (2022), who demonstrated that corruption does not necessarily hinder economic growth when other factors favour it, showing positive effects on growth in Latin American and MENA countries. Their study also highlights that corruption hinders development in African and Asian contexts and is likely contingent on regional institutional and economic conditions.

The study also reveals a significant negative relationship between corruption and income inequality, as measured by the Gini coefficient. This inverse correlation suggests income inequality decreases in regions with higher corruption levels. One possible explanation for this could be the redistribution of resources through informal or corrupt channels, which might reduce inequality in the short term. These findings show the opposite relationship from that identified by Camacho (2016), who found that a decrease in corruption decreases the Gini coefficient in OECD, Asian, African, and Latin American countries.

In contrast, the negative impact of corruption on the HDI indicates the potentially detrimental effects of corruption on social development. The findings suggest that higher levels of corruption are associated with lower levels of education, health, and income, which are the main components of the HDI indicator. This aligns with the broader literature that highlights the corrosive effects of corruption on public services and social welfare, where resources intended for education, healthcare, and infrastructure are often diverted, leading to poorer outcomes for the population. The misallocation assumption is aligned with Bazie ad et al. (2024) findings on how corrupt activity distorts the allocation of public resources.

Interestingly, the analysis shows a weak and statistically insignificant relationship between corruption and poverty rates, suggesting that corruption's impact on poverty may be less direct or more complex than its effects on other development indicators. This could be due to the multifaceted nature of poverty, which is influenced by a wide range of factors, including but not limited to economic policies, social safety nets, and external economic conditions. In the context of Brazil, this complexity is likely exacerbated by the presence of extensive social safety nets, such as *Bolsa Família*, which may alleviate any direct effects of corruption on poverty. Established in 2003, *Bolsa Família* is currently one of the world's most established Conditional Cash Transfer (CCT) programs (Fiszbein & Schady, 2009). In 2023 alone, *Bolsa Família* benefitted on average 21.3 million people per month. The program is estimated to reduce the percentage of people in poverty and extreme poverty by between 1.1 and 1.5 percentage points per year (Souza et al., 2019).

The regional analysis provides additional insights, indicating that the relationship between corruption and development varies significantly across different regions in Brazil. The stronger explanatory power of the models at the regional level suggests that aggregation may help capture more stable relationships between corruption and development indicators, possibly by reducing noise and focusing on broader patterns. Regional results were broadly consistent with state-level estimates, with few differences in the direction or significance of coefficients. This alignment reinforces the robustness of the primary findings and suggests that corruption's effects are not overly sensitive to the level of aggregation.

It is important to note that regional findings must be interpreted cautiously due to the small sample size (n=40). The limited number of observations may affect the precision of the estimates and increase the likelihood of sampling variability. As a result, while regional models provide valuable context for understanding broader trends, they should be viewed as complementary to the state-level analyses rather than as standalone evidence.

Limitations

One limitation of the analysis is the reliability of CPIM. The indicator strongly relies on the quality and completeness of the data from the TCU, and there's a possibility that not all instances of corruption are captured if they don't result in an irregular account judgment. Another point to be considered is that there is typically a time lag between the occurrence of corruption and its ruling as irregular. The delay means that the corruption might have occurred a few years before it was officially ruled as irregular, impacting the immediacy and responsiveness of the indicator in reflecting current corruption levels.

A second limitation lies in the challenge of establishing a causal relationship between corruption and development. One significant issue is the lack of comprehensive data at the state level over more extended periods of time, which limits the depth and robustness of the analysis. Without long-term data, capturing trends and making definitive conclusions about causality is challenging. Additionally, the unavailability of various relevant variables at the state level further compromises the statistical accuracy of the model. These variables, if accessible, could have provided a more nuanced and precise understanding of the interplay between corruption and development.

Conclusion

The findings of this thesis contribute to the existing literature by offering a more nuanced understanding of the complex relationship between corruption and socio-economic development outcomes in Brazilian federative units and regions. Contrary to the common belief that corruption uniformly hampers development, the results suggest a more varied impact, with some positive associations between corruption and economic indicators such as GDP per capita. The statistically significant positive correlation between the CPIM index and GDP per capita indicates that corruption may coincide with higher economic output in specific contexts, possibly by facilitating transactions in bureaucratically constrained environments. This does not necessarily imply that corruption is beneficial, but instead, it suggests that the effects of corruption on economic growth are multifaceted and context-dependent.

On the other hand, the significant negative relationship between corruption and human development indicators, such as the HDI, highlights the corrosive effects of corruption on broader social outcomes.

Likely through a distortion of the allocation of public resources, corruption appears to undermine critical aspects of development, such as education and healthcare, and income distribution, which are essential for improving the quality of life. Additionally, the analysis found a negative relationship between corruption and income inequality, suggesting that corruption may sometimes reduce inequality, potentially by redistributing resources in informal channels.

These findings have significant implications for policymakers aiming to promote development and combat corruption in Brazil. The mixed impact of corruption on economic and social indicators suggests that anti-corruption measures must be carefully designed to address specific contexts and objectives. Policies to reduce corruption should be integrated into broader development strategies, focusing on enhancing transparency, strengthening institutions, and improving governance. These policies are fundamental to mitigating the negative impacts of corruption, particularly on human development.

Targeted interventions in states and regions with higher levels of corruption could be particularly effective in addressing the negative aspects of corruption. By focusing on these areas, policymakers can work to ensure that the benefits of economic growth are more evenly distributed and that corrupt practices do not compromise social development. These interventions could include increasing public oversight, enhancing the capacity of local institutions to monitor and report corruption, and promoting civic engagement to hold public officials accountable.

Further academic research is needed to explore the relationship between corruption and development more comprehensively. One potential area of investigation is whether the relationship identified through the CPIM aligns with findings using subjective indicators, such as the CPI. The use of an objective indicator in this thesis is a significant contribution to the field, given the challenges associated with subjective measures of corruption, which are often influenced by perceptions rather than actual instances of corruption. The lack of long-term subjective data at the state or regional level limited the scope of this research, highlighting a need for future studies that can compare the outcomes using both indicators.

Given the complex findings, future research should also investigate the mediating role of institutional quality in shaping the relationship between corruption and development outcomes in Brazilian states. For example, examining how states with stronger judicial systems or greater public sector accountability mitigate the adverse effects of corruption can provide deeper insights. Drawing parallels with other federative systems like South Africa might also yield comparative insights into the interplay between governance structures and development trajectories.

References

Azevedo, E.B. et al. (2018) ‘Corrupção, governança e desenvolvimento: Uma análise seccional de dados para o Brasil', Revista Globalización, Competitividad y Gobernabilidad, 12(1), pp. 48–62. doi: 10.3232/GCG.2018.V12.N1.02.

Bazie, P., Thiombiano, N., & Maiga, E. W. H. (2024). ‘Allocating budget in developing countries, the need to fight corruption: evidence from Sub-Saharan African countries’. *Future Business Journal*, *10*(1), 2–12. doi: 10.1186/s43093-023-00289-y

Boll, J. L. S. (2010). ‘A corrupção governamental no Brasil: construção de indicadores e análise da sua incidência relativa nos estados brasileiros (Dissertação de mestrado)’. Pontifícia Universidade Católica do Paraná, Porto Alegre, RS, Brasil.

Breusch, T. S., & Pagan, A. R. A. (1979). ‘Simple test for heteroskedasticity and random coefficient variation’. Econometrica, 47(5), 987-1007. doi: 10.2307/1911963

Cameron, A.C. and Miller, D.L. (2015) ‘A Practitioner’s Guide to Cluster-Robust Inference’, The Journal of human resources, 50(2), pp. 317–372. doi: 10.3368/jhr.50.2.317.

Castor, B. V. J. (2002). ‘Brazil is not for amateurs: Patterns of governance in the land of "Jeitinho"’. Xlibris Corporation.

Corrêa, I., Camões, M., Meyer-Sahling, J., Mikkelsen, K. & Schuster, C. (2020) ‘Distorções de incentivo ao desempenho e redução de motivação no serviço público federal no Brasil’. Revista do Serviço Público , 71 (3) pp. 476-503. 10.21874/rsp.v71i3.3408.

Dreher, A. & Gassebner, M. (2013) 'Greasing the wheels? The impact of regulations and corruption on firm entry', Public Choice, 155(3/4), pp. 413–432. doi: 10.1007/s11127-011-9871-2.

Eggers, A.C., Tuñón, G. & Dafoe, A. (2024) ‘Placebo Tests for Causal Inference’, *American Journal of political science*, 68(3), pp. 1106–1121. doi: https://doi.org/10.1111/ajps.12818.

Fiszbein, A., & Schady, N. (2009). ‘Conditional cash transfers: Reducing present and future poverty’. A World Bank policy research report. Washington, DC: World Bank.

France, G. (2022) ‘Brazil: Setbacks in the legal and institutional anti-corruption frameworks’, Transparency International.

Garcia, G.P. (2022) 'Proposta de um indicador de corrupção no Brasil baseado em dados do Tribunal de Contas da União', Revista da CGU, 14(25), pp. 129–143. doi: 10.36428/revistadacgu.v14i25.401.

Gilman, S. C. (2018) ‘To Understand and to Misunderstand How Corruption is Measured: Academic Research and the Corruption Perception Index’, Public Integrity, S74-S88, doi: 10.1080/10999922.2018.1472974

Gyimah-Brempong, K., & Munoz de Camacho, S. (2006). ‘Corruption, Growth, and Income Distribution: Are there Regional Differences?’, Economics of Governance, 7(3), 245–269. doi: 10.1007/s10101-005-0008-2

Leff, N.H. (1964) 'Economic development through bureaucratic corruption', The American Behavioural Scientist (Beverly Hills), 8(3), pp. 8–14. doi: 10.1177/000276426400800303.

Li, H., Xu, L.C. & Zou, H. (2000) ‘Corruption, Income Distribution, and Growth’, Economics and Politics, 12(2), pp. 155–182. doi: 10.1111/1468-0343.00073.

Lucio, F.G.C., de Castro Pereira, R.A., & Gomes, J.W.F. (2020). 'Grease or Sand the Wheels? A State-Level Approach on Corruption in Brazil', Theoretical Economics Letters, 10(3), pp. 510-531. doi:10.4236/tel.2020.103032.

Mauro, P. (1995) 'Corruption and growth', The Quarterly Journal of Economics, 110(3), pp. 681–712. doi: 10.2307/2946696.

Melo, F. L. N. B. de, Sampaio, L. M. B. & Oliveira, R. L. de. (2015). ‘Corrupção Burocrática e Empreendedorismo: Uma Análise Empírica dos Estados Brasileiros’. Revista de Administração Contemporânea, 19(3), 374–397. doi: 10.1590/1982-7849rac20151611

Méon, P.-G. & Sekkat, K. (2005) 'Does corruption grease or sand the wheels of growth?', Public Choice, 122(1/2), pp. 69–97. doi: 10.1007/s11127-005-3988-0.

Nur-tegin, K. & Jakee, K. (2020) 'Does corruption grease or sand the wheels of development? New results based on disaggregated data', The Quarterly Review of Economics and Finance, 75, pp. 19–30. doi: 10.1016/j.qref.2019.02.001.

Onody, V. da S.M. et al. (2022) 'Corruption and FDI in Brazil: Contesting the “Sand” or “Grease” Hypotheses', Sustainability (Basel, Switzerland), 14(10), p. 6288. doi: 10.3390/su14106288.

Palotti, P. & Freire, A. (2015) ‘Perfil, composição e remuneração dos servidores públicos federais: trajetória recente e tendências observadas.’ Caderno Enap v. 42, p. 23-53, 2015

Da Rocha, M. L. (2019) ‘A compreensão das práticas sociais no Brasil a partir da percepção: o caso ‘jeitinho' e corrupção no Brasil’, 31 f. Trabalho de Conclusão de Curso (Bacharelado em Administração Pública) - Instituto Brasiliense de Direito Público, Brasília

Sallaberry, J. D., Quaesner, L. S., Costa, M. C., & Flach, L. (2020). ‘Measurement of damage from corruption in Brazil’. Journal of Financial Crime, 27(4), 1239–1239. doi: 10.1108/JFC-04-2020-0057

Da Silva, S. A (2017). ‘Regional Inequalities in Brazil: Divergent Readings on Their Origin and Public Policy Design’. EchoGéo [Online], 41. doi:10.4000/echogeo.15060

Tanzi, V. and Davoodi, H.R. (1997) 'Corruption, Public Investment, and Growth', IMF Working Paper, 97(139), p. 1. doi: 10.5089/9781451929515.

Souza, P. H. G. F., Osorio, R. G., Paiva, L. H., & Soares, S. (2019). 'Os Efeitos do Programa Bolsa Família sobre a Pobreza e a Desigualdade: um balanço dos primeiros quinze anos'. Working Paper, 2499. Rio de Janeiro: IPEA.

Spyromitros, E. and Panagiotidis, M. (2022) 'The impact of corruption on economic growth in developing countries and a comparative analysis of corruption measurement indicators', Cogent economics & finance, 10(1), pp. 1–30. doi: 10.1080/23322039.2022.2129368.

Trombini, M. E., Pohlmann, M., Valarini, E. & Oliveira, V. E. de. (2024). 'The Fight against Systemic Corruption : Lessons from Brazil (2013–2022)'. 1st ed. 2024. Wiesbaden: Springer Fachmedien Wiesbaden Imprint: Springer VS.

Wooldridge, J.M. (2010) 'Econometric analysis of cross section and panel data'. 2. ed. Cambridge, Mass. [u.a.]: MIT Press.