Distribution Agreement

The Effect of Expected Wage Differentials on the Informal Sector

by

Jonathan Kaminski

Sue Mialon
Adviser

Gordon Streeb
Adviser

Department of Economics

Sue Mialon

Adviser

Gordon Streeb

Adviser

Richard Doner

Committee Member

2018

The Effect of Expected Wage Differentials on the Informal Sector

By

Jonathan Kaminski

Sue Mialon and Gordon Streeb
Co-Advisers

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2018

Abstract

The Effect of Expected Wage Differentials on the Informal Sector
By Jonathan Kaminski

Following informal sector literature's infatuation with urbanization, this paper seeks to push authors to look under the hood of the urbanization phenomenon to explore the underlying engine as a means of understanding how to make better policy recommendations that allow governments to take the wheel with their informal sectors. Motivated by the Harris-Todaro Model for migration, I principally investigate the explanatory power urban-rural expected wage differentials have on the size of the informal sectors. As a result, I specify a Pooled OLS model including year dummies, and find a significant, positive effect suggesting governments can target reducing the urban-rural wage differential in order to limit and reduce the size of the informal sector. Second, I examine how government budget allocations on public healthcare and education affect the size of the informal sector, where I find slight evidence for increases in spending being associated with increases in the informal sector size. To this effect, the welfare spending result produces more questions than it does answers; there will need to be further research with attention to the urban-rural split of expenditure to make any concrete claims or recommendations. Concluding the study, I recommend governments subsidize the rural, agricultural sector with the help of urban, industrial tax revenues to subdue and shrink the informal sector. Additionally, governments should both evaluate current and future economic policy with the question of how a decision might affect the urban-rural expected wage differential to ensure it is not so urban-centric as to cause a burgeoning informal sector.

The Effect of Expected Wage Differentials on the Informal Sector

By

Jonathan Kaminski

Sue Mialon and Gordon Streeb
Co-Advisers

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2018

Acknowledgements

The author would like to thank the following individuals because without their support and assistance this study would not have been possible: Sue Mialon and Gordon Streeb; Rob O'Reilly, Christopher Curran, and Richard Doner; and friends and family.

Table of Contents

I. INTRODUCTION

The informal sector is a development trap. Yet, it was only in 2015 that the International Labour Organization (ILO) came to this conclusion, and formally recommended countries work to reduce the size of their informal economies to catalyze development (ILO 2015). There are many costs of informal sectors that governments and informal workers bear. For one, the informal nature of the work means lost tax revenues for the government, but there are also untold congestion costs on public goods, labor regulation, property rights, and human rights violations, and productivity losses from the low marginal productivity of informal labor. For these reasons, all governments, but especially those less along in the process of development, must look for ways to manage and reduce the size of their informal sectors.

The main purpose of this study is to build on past research on urbanization as a determinant of informal sector size through examining how well the Harris-Todaro Model of migration – which has been referenced, but never empirically incorporated into regression work – can explain variations in informal sector size. As a secondary objective, this study includes government spending on public healthcare and education[1], two large components of government welfare spending, to analyze how changes in allocation decisions impact the size of the informal sector. Existing findings and economic reasoning suggest high rates of urbanization drive increases in the size of the informal sector while higher welfare spending leads to a decrease in the informal sector (Elgin and Oyvat 2013).

In conducting this regression analysis, I use two samples of informal sector sizes, which are calculated with different approaches, for robustness. These two samples are referred to as Multiple Indicator, Multiple Cause (MIMIC) and Dynamic Stochastic General Equilibrium

[1] As a percentage of GDP.

(DSGE). Beyond the key variables of interest, expected wage differential and government expenditure on public healthcare and education, I include the urban share of population, two dummies for lower middle and upper middle income countries, Transparency International's Corruption Perceptions Index, and the Heritage Foundation's Business Freedom Index as regressors. The first set of regression techniques I apply are Pooled Ordinary Least Squares (OLS) and Pooled Two-Stage Least Squares Instrumental Variable (2SLS IV) for each of the samples. Next, to investigate a possible endogeneity problem between informal sector size and expected wage differential, I run a robust score chi-squared test and a robust regression F-test, reported in Table 7, where I fail to reject the null hypothesis that the model is exogenous. Following this, I run a further robustness check by examining time effects through Fixed Effects (FE) modelling. I run and report FE and FE IV for each sample in Table 8. From the FE modelling, it is clear there are significant time effects that should be accounted for in my Pooled OLS model. Subsequently, I add year dummies to my Pooled OLS and Pooled 2SLS IV models and report these in Table 9. Considering the results of the endogeneity tests, Pooled OLS with year dummies is the final model that I go on to discuss.

The major findings and contributions of this paper are two-fold:

1. The Harris-Todaro Model's expected wage differential is able to explain the changes in the size of the informal sector to a significant effect.
2. Government expenditure on welfare-type programs may increase the size of the informal sector, which is inconsistent with previous literature.

In the Pooled OLS with year dummies model, expected wage differentials are significant at the 1% level in the MIMIC sample and 5% for the DSGE sample with their impact being a 1% increase in relative informal sector size for every time a country's expected wage differential increases by 4.6%[2] (MIMIC) or 13.3% (DSGE) of its GNI per capita (PPP). Government welfare expenditure is significant at the 10% level in the MIMIC sample only, and the coefficient implies a 41.5% increase in relative informal sector size for every 1% increase in spending.

From these results, I conclude this study by making two major recommendations to policymakers. First, for countries wishing to reduce the size of the informal sector, economic policy must include significant subsidies to the rural sector with urban sector tax revenues in order to support rural wage growth, thereby closing the urban-rural wage differential and lessening the incentive for rural-to-urban migration. This action is necessary as continued spending in urban areas only exacerbates the wage differential, further increasing the sizeable costs of a large informal sector. Second, countries should examine the geographic distribution of healthcare and education spending to ensure it is equitable and not an additional migration incentive for rural citizens. Further research is needed to confirm this second recommendation.

II. THEORETICAL FRAMING

For decades, rural-urban migration, or urbanization, has been thoroughly studied from a theoretical perspective in order to understand the rationale for continued migration in the face of vast urban unemployment. There is a lengthy set of literature exploring this contradiction.

Lewis (1954) first attempts to identify the forces behind the urbanization phenomenon, giving rise to the Lewis Model. The Lewis Model presents a dual-sector model that suggests as the urban capitalist sector grows, labor is extracted from the rural subsistence sector due to both higher marginal productivity and higher wages being present in the "capitalist sector," or urban-formal, employment (Lewis 1954). However, the Lewis Model has been shown to empirically

[2] This percent is calculated by dividing 1 by 21.6 to get the ratio.

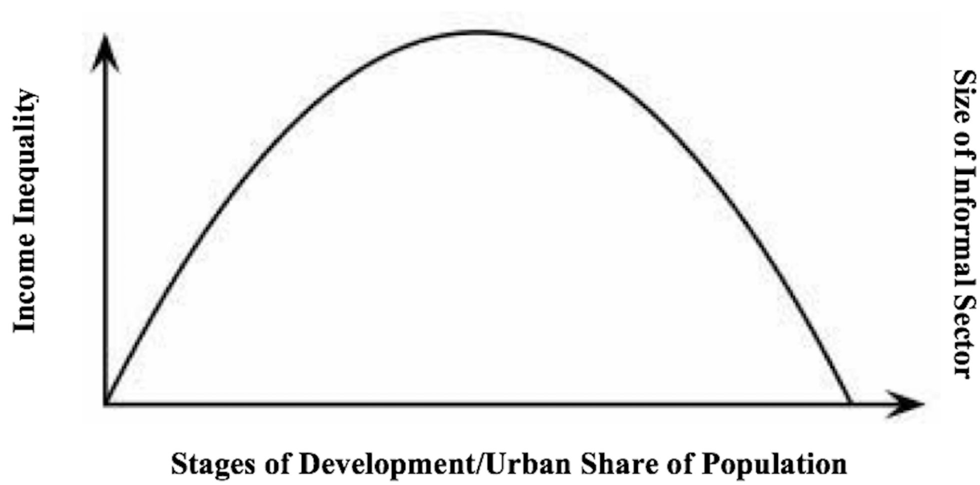
fail in accurately predicting migration flows and has suffered heavy criticism for its lack of addressing resultant changes to the rural sector and the existence of urban unemployment. The Harris-Todaro Model improves on the Lewis Model by framing the decision for a rural worker to migrate as not just a question of higher wages, but as a probabilistic model weighting the urban wage by the probability the migrant is able to find employment in the urban sector (Harris and Todaro 1970). Within the Harris-Todaro Model, a rural worker will only migrate to an urban area when their expected wage, that is the real urban wage multiplied by the labor force participation rate in the urban sector, is greater than their current, real wage in the rural sector. The Harris-Todaro Model redirects attention in the study of labor mobility away from previous models' emphasis on the real wage differential and toward the expected wage differential, thereby incorporating urban unemployment as a variable affecting the model's equilibrium. This is a marked improvement, but still, the Harris-Todaro Model has critics.

One of the most prominent criticisms of the Harris-Todaro model is that it does not adequately address the existence of an urban-informal subsector that employs a significant portion of urban residents and even draws migrants in its own right. Fields (1975) revises the Harris-Todaro model to include the existence of this "murky," or urban-informal, sector as a part of the model by including the urban-informal wage rate as a variable. As a result of Fields' revision, the model is able to more realistically predict urban unemployment rates instead of the overprediction experienced with the original Harris-Todaro Model due to its lumping of the underemployed with the unemployed. However, there is little to no reliable, global data on informal sector wages. Therefore, the Fields' Model is very difficult to empirically test and impossible to utilize in this study. Consequently, this empirical study employs the unsegmented, rural-urban expected wage differential from the Harris-Todaro Model as its main variable of interest.

Taking a step further, there is considerable literature detailing the relationship urbanization has with development. Notably, Zelinsky (1971), through the "Hypothesis of the Mobility Transition," first links the "mobility transition," of which urbanization is a part, to the "vital transition," or what is simply called development today. In the Zelinsky Model, both the vital and mobility transitions occur sequentially with "fatalistic inevitability" as seen when the rate of rural-to-urban migration increases as a country industrializes until it reaches an inflection point where the rate of rural-urban migration decreases due to rural, agricultural employment reaching an "optimum economic return" (Zelinsky 1971, 245, 249). The inflection point described by Zelinsky is the same point captured through the equilibrium point in the Harris-Todaro Model when the expected wage differential between urban and rural employment is equal to zero. According to Zelinsky, at this turning point, the rate of urbanization steadily decreases as the rural labor supply shock driven by the preceding rapid rural-to-urban migration causes an increase in per capita rural incomes. As rural incomes rise and urban unemployment proliferates, the wage differential between rural and urban employment shrinks, reducing the pressure for rural workers to migrate to urban areas (Todaro 1969, Harris-Todaro 1970, Fields 1975). Therefore, both Zelinsky and Harris-Todaro support an inverted-U relationship between the rate of urbanization and development. To bolster Zelinsky's qualitative approach and Todaro's theoretical approach, and building on the proposed inverted-U relationship, Ledent (1982) puts forward a quantitative analysis of urbanization and industrialization, creating a mathematical framework that further validates the inverted-U curve relationship by producing appropriate estimated results. Synthesizing these findings, both the Zelinsky Model and the Harris-Todaro Model corroborate an inverted-U curve relationship where within the development process the rate of urban migration first rises, reaches a maximum point, and then declines. This relationship between urbanization and development is vital to understanding the relationship between urbanization and the informal sector.

Rauch (1993) first proposes the inverted-U curve relationship between the level of urbanization and the size of the urban-informal subsector. Inspired by Kuznets (1955)'s hypothesis stating the inverted-U relationship between income inequality and development and rooted in Zelinksy (1971), Rauch (1993) transitively reasons that the same relationship must exist between income inequality and the share of urban population. Second, Rauch states that income inequality must be driven by increasing underemployment and excess labor supply as the underdeveloped urban-formal sector struggles to employ all urbanizing migrants, resulting in many incoming migrants taking low paying jobs in the informal sector to survive. Moreover, fueling still greater increases in both income inequality and the size of the urban-informal sector, increased urban-formal growth can further increase underemployment as the increased number of industrial jobs draws even more rural-urban migrants, as per the Harris-Todaro Model. Eventually, as mass industrialization catches up with the size of the urban population and formally employs those underemployed in the urban-informal sector, both income inequality and the size of the urban-informal sector decrease. Taken together, due to the interrelatedness of income inequality with the relative size of the informal sector, Rauch (1993) claims there must exist "another inverted-U [between] the share of the informal sector in the total labour force" and the share of urban population (Rauch 1993, 904). In sum, by overlaying conclusions from Kuznets (1955), Zelinsky (1971), and Rauch (1993) these relationships can be represented as Figure 1, below.

Figure 1. *Income Inequality and Size of the Informal Sector by Urban Share of Population*



Previously, as I describe above, most research has focused on establishing the theoretical foundations for relationships between development, income inequality, urbanization, and the informal sector. More recently, empirical work supporting and utilizing these relationships in regional or country-specific studies has become the focus of the field, especially the relationship between the size of the informal sector and urbanization (see Banerjee 1983 for India, Rauch 1993 for Latin America, Magazzino et. al. 2011 for Caribbean countries, Acosta-Gonzalez et. al. 2014 for OECD countries, and Bourhaba and Mama 2016 for Morocco, among others). These studies use either the rate of urbanization or the urban share of population as an explanatory variable to estimate the effect urbanization has in explaining variation in the informal sector. Buoyed by Elgin and Oyvat (2013), which has been the most comprehensive and robust empirical study exploring the informal sector's relationship with urbanization to date, the consensus in the literature is that urbanization is a significant regressor in explaining the informal sector.

Many of these studies reference the Harris-Todaro Model as a step in their theoretical motivation, yet none of them actually investigate whether the foundation of the Model, the expected wage differential, holds the same explanatory power as urbanization, itself. To the best of my knowledge, there are no empirical studies exploring this underlying variable that is supposed to act as the economic incentive engine for the increase in the share of urban population. Therefore, in this study, I conduct an analysis using country-level, urban and rural wage data to explore how well the foundation of the Harris-Todaro Model, the urban-rural expected wage differential, explains variation in the size of the informal sector due to its relationship with urbanization.

If the model is found to be significant, this study will provide novel, supporting evidence for urban-rural expected wage differential as a determinant of the size of the informal sector, supplement the claim that the majority of rural-to-urban migrants find work in the informal sector, and add further validity to the Harris-Todaro Model as a functional model for urbanization.

III. DATA AND EMPIRICAL STRATEGY

This study uses data from the Quality of Government dataset, the World Bank, Euromonitor's Passport, and data provided by Dr. Friedrich Schneider to conduct a global analysis into the effect expected urban-rural wage differentials have on the relative size of informal sectors. To provide more robust results, this study uses two sets of informal sector size estimates, one calculated using the popular MIMIC model approach and one using a newer DSGE model approach, as well as reports coefficient estimates from Pooled Ordinary Least Squares (Pooled OLS), Two-Stage Least Squares Instrumental Variable (2SLS IV), Fixed Effects (FE), and Fixed Effects Instrumental Variable (FE IV) regression techniques. The

MIMIC model set of informal sector estimates is a balanced panel provided by Hassan & Schneider (2016) that contains 2,297 observations for 157 countries from 1999 to 2013. Upon request, the Hassan & Schneider (2016) data was kindly provided by Dr. Schneider. The DSGE model set of estimates provided by Elgin and Öztunali (2012) and accessible as part of the Quality of Government (Standard, Time-Series, 2017) dataset is a 161-country unbalanced panel spanning from 1950 to 2009 that includes 7,396 observations, making it the largest dataset available on the size of the informal economy.

There are many methods to estimate the size of the informal sector, and there are upsides and downsides to each approach. I utilize the Hassan & Schneider (2016) dataset because it employs a popular and well-regarded approach, and I use the Elgin and Öztunali (2012) dataset as a robustness check because it both employs a novel approach designed to overcome problems with the MIMIC approach and it is the largest set of informal sector estimates available[3].

For the purposes of this study, I have restricted the original MIMIC and DSGE informal sector datasets to contain only capitalist economies that have both urban and rural populations[4]. Countries with socialist economies are excluded from these two samples because the structure and composition of both their formal and informal sectors are significantly different from those in capitalist economies. Completely urbanized countries without any rural population are also removed from the two samples as they do not have urban-rural wage differentials. Finally, as my final dataset is unbalanced, the samples are restricted by the data availability of my other regressors. After these restrictions, the MIMIC model sample contains 677 complete observations for informal sector size and the DSGE model sample contains 585.

---

[3] See Elgin and Öztunali (2012) for a comprehensive review and comparison of informal sector estimation approaches.

[4] Due to this, I have omitted informal sector data from China, Laos, Vietnam, North Korea, Singapore, Monaco, and Nauru.

To investigate the effect of the Harris-Todaro migration forces on the size of informal sectors, I combine annual and country-specific data on urban disposable income, rural disposable income, and labor force participation rates to calculate expected urban-rural wage differentials[5]. Both the urban and rural disposable income data is available from Euromonitor's Passport database, while the labor force participation data can be found within the World Bank Open Data database. Lastly, I divide each expected urban-rural wage differential by its respective GNI per capita (PPP) to reduce the disproportionate weight high income countries have in the model due to their significantly higher urban incomes and, consequently, dramatically larger differentials.

A. Outcomes

The main outcome of interest for this paper is to better understand the relationship rural-to-urban migration has on the size of countries' informal sectors. According to the Harris-Todaro model, a rural migrant's decision to migrate to an urban center is based on whether the differential between the expected urban wage and the current, real rural wage is positive. If it is positive, the individual has an incentive to migrate. If a country experiences an increase in the expected wage differential in favor of rural-to-urban migration, we would expect to see increased migration flows from rural areas to urban centers and therefore a respective, or even disproportionately large, increase in the relative size of the informal sector borne out in the proposed model. Although many migrants are first enticed by the wages of the formal sector, the majority of rural-to-urban migrants end up finding employment in the informal sector and do not continue to look for employment in the formal sector (Banerjee 1983). Due to this, the Harris-Todaro expected wage differential may actually have more explanatory power than a Fields Model with formal-informal segmented expected wage differentials.

[5] *Expected urban-rural wage differential$_{ct}$ = (urban disposable income$_{ct}$ × labor force participation rate$_{ct}$) − rural disposable incomect*

The secondary outcome of interest in this paper is the measure of welfare spending. The proposed model uses the aggregate of government spending on public healthcare and education to investigate the effect the size of a government's welfare spending budget has on the size of its informal economy. The government spending data on public healthcare and education is measured as a percentage of the country's GDP, and it is from the World Development Indicators database produced by the World Bank and accessible in the Quality of Government dataset (Standard, Time-Series 2017). If there is not a significant reduction in the size of a country's informal sector when the government is spending more on education and healthcare, governments will need to reevaluate these spending initiatives and perhaps take more direct approaches to reducing the size of their informal sectors.

B. Other Explanatory Variables

To parse out the specific effects of the expected urban-rural wage differential and government welfare spending, I include a few factors evidenced in informal sector literature as having an effect on the size of this sector. These factors are the percentage of the country's population living in urban areas, the country's income-level, the country's ease of business rating, and the country's level of corruption. For urban population data, this study uses the United Nations Urbanization Prospects (2014 revision), which contains a 233-country dataset starting in 1950 and forecasting urbanization trends to 2050. I use only data from 1950 to 2013 to provide data for the entire year ranges of both the MIMIC and DSGE datasets. For income-level data, I use GNI per capita (PPP) data from the World Bank and accessible both in the World Bank's DataBank and in the Quality of Government (Standard, Time-Series 2017) dataset in conjunction with the World Bank's 2017 Income Level Classifications[6] to create a set of dummy

[6] 2017 World Bank Income Level Classifications based on GNI per capita: Low-income < $1,005, Lower-middle income $1,006 - $3,955, Upper-middle income $3,956 - $12,235, High-income > $12,235.

variables for High Income, Upper Middle Income, and Low/Lower Middle Income countries. For country ease of business data, I use the Heritage Foundation's Business Freedom Index (BFI), which is based on the World Bank's Doing Business study. The BFI is on a 100-point scale where 100 represents maximum business freedom. For corruption, I use unbalanced panel data from Transparency International's Corruption Perceptions Index (CPI) that contains data from 1995 to 2015 on 176 countries. The CPI is also on a 100-point scale where 100 represents a complete lack of corruption.

C. Empirical Strategy

I run a combination of Pooled OLS and 2SLS IV regression methods to estimate coefficients for the following model:

$$infsize_{ct} = \beta_0 + \beta_1 exp\ wage\ diff\ (GNI)_{ct} + \beta_2 popurban_{ct} + \beta_3 LLM\ dummy_{ct}$$
$$+ \beta_4 UM\ dummy_{ct} + \beta_5 corrpercep_{ct} + \beta_6 businessfree_{ct} + \beta_7 welfarespend_{ct} + \varepsilon$$

Where *infsize* is the size of a given country's informal sector during a given year, and is measured as a percentage relative to the size of the country's formal sector GDP in the respective year, *exp wage diff (GNI)* is the expected wage differential between urban and rural areas weighted by GNI per capita (PPP), *popurban* is the percentage of a country's population living in urban areas, *LLM dummy*[7] and *UM dummy* are a set of dummies for income levels, *corrpercep* is the Corruptions Perceptions Index, *businessfree* is the BFI, and *welfarespend* is government spending on public health and education as a percentage of GDP.

[7] LLM Dummy combines Lower and Lower Middle income countries.

Initially, I run Pooled OLS regressions for the MIMIC and DSGE samples to estimate the coefficients of the above equation. I first use the OLS estimated equations as benchmarks to empirically check the relationships between the dependent and explanatory variables. Next, due to a potential endogeneity problem from simultaneity bias between the dependent variable, *infsize*, and my principal regressor, *exp wage diff (GNI)*, I conduct a set of 2SLS IV regressions in order to test the extent of endogeneity and determine whether Pooled OLS regression or Pooled 2SLS IV regression produces the more efficient estimates. I draw on government military expenditure as a % of GDP, *expmil*, as an IV for *exp wage diff (GNI)*, and then I run a robust score test and a robust regression F-test to better understand the possible endogeneity bias in the OLS model. To conduct a further robustness check, I run a set of regressions investigating whether there is any omitted variable bias due to time not being included in either of the OLS and 2SLS models. This set of regressions includes running FE and FE IV regressions, using time (i.e., year) as my panel identifier, and then Pooled OLS and Pooled 2SLS, both with year dummies, for each sample.

D. Limitations

Due to the nature of studying informality, most studies suffer limitations from the availability of data, and this study is no different. There are no exact measures of the size of informal sectors, and the estimated, indirectly measured measurements are the only option available to study the size of informal sectors. Through analyzing samples from two differently calculated sets of informal sector size estimates, I attempt to provide stronger and more trustworthy results to account for this unavoidable uncertainty. Another limitation is that I must drop the square of urban population, which I was including to capture the inverted-U relationship proposed by Rauch (1993), because of near perfect collinearity with urban population. Not being able to account for this relationship will introduce a small amount of omitted variable bias. Additionally, as this is a global study, and I wish to apply my findings to as many countries as possible, it is important to have a balanced sample of high, middle, and low income countries of varying levels of development, but most global data suffers from selection bias in its availability as less developed nations are less likely to have data. In Table 3, I explore the makeup of my samples to ensure there are adequate observations from all income levels to provide enough of a random sample. Further limitations may arise from simultaneity bias due to the aforementioned endogeneity problem between *infsize* and *exp wage diff (GNI)*. I aim to address this bias by utilizing and reporting IV regression techniques as well as by testing for endogeneity post-regression.

E. Summary Statistics

In Tables 1 and 2 below, I present summary statistics for the variables of interest for both the MIMIC and DSGE samples.

Table 1. *Summary Statistics of MIMIC Sample*

Variable	Mean	Std. Dev.	Min	Max
Informal sector size (as a % of GDP)	0.2903436	0.1447423	0.0843	0.812
Expected wage differential, weighted by GNI	-0.044981	0.1255886	-0.60939	0.3404961
Urban population (as a % of total population)	0.6824699	0.1598269	0.1955	0.98261
Dummy for Low and Lower Middle income countries	0.0723781	0.2593046	0	1
Dummy for Upper Middle income countries	0.3146233	0.464709	0	1
Corruption Perceptions Index	52.55965	23.52731	15	100
Business Freedom Index	63.61093	15.99873	18	100
Government spending on public healthcare and education	0.0938749	0.0315062	0.0227915	0.1824409

N = 677

Table 2. *Summary Statistics of DSGE Sample*

Variable	Mean	Std. Dev.	Min	Max
Informal sector size (as a % of GDP)	0.2645449	0.1333866	0.0807	0.7542
Expected wage differential, weighted by GNI	-0.049771	0.1315161	-0.60939	0.3404961
Urban population (as a % of total population)	0.6818042	0.1618492	0.1955	0.98261
Dummy for Low and Lower Middle income countries	0.0888889	0.2848268	0	1
Dummy for Upper Middle income countries	0.3094017	0.4626426	0	1
Corruption Perceptions Index	54.51607	24.30018	10	100
Business Freedom Index	64.27658	14.66095	20	100
Government spending on public healthcare and education	0.0933045	0.0311766	0.0174893	0.1824409

N = 585

Overall, as seen above, there is variation within both samples. The MIMIC sample has an average informal sector size of 29% while the DSGE sample has a mean of 26.5%. Because both samples have roughly the same breakdown across income levels, see Table 3 below and the coefficients of income dummies in Tables 1 and 2, this suggests the DSGE estimation method may be more conservative in its estimates than the MIMIC approach.

Table 3. *Sample Makeup Analysis*

Type of Observation	MIMIC	DSGE
Lower Middle Income	49	52
Upper Middle Income	213	181
High Income Income	415	352
Total Number of Observations	677	585
Number of Countries	69	66

Next, in Tables 4 and 5 below, I report correlation matrices for both samples. The only moderately high correlation is between *corrpercep* and *welfarespend*, where it is 0.76 for both samples. This likely inflates my standard errors, but I decide to keep this explanatory variable because of its important effect on *infsize* I wish to analyze.

Table 4. *Correlation Matrix for MIMIC Sample*

	Inf. Size (MIMIC)	Exp. Wage Diff.	Urban Pop.	LLM Dummy	UM Dummy	CPI	BFI	Gov. wel. spend
Inf. Size (MIMIC)	1.0000	-	-	-	-	-	-	-
Exp. Wage Diff.	0.4616	1.0000	-	-	-	-	-	-
Urban Pop.	-0.3643	-0.3713	1.0000	-	-	-	-	-
LLM Dummy	0.1706	0.2380	-0.4473	1.0000	-	-	-	-
UM Dummy	0.5214	0.4158	-0.2749	-0.1893	1.0000	-	-	-
CPI	-0.6061	-0.4467	0.6047	-0.3421	-0.4913	1.0000	-	-
BFI	-0.0257	-0.2466	0.0575	-0.0680	-0.2250	0.1428	1.0000	-
Gov. wel. spend	-0.4585	-0.4402	0.5057	-0.3263	-0.4563	0.7611	0.1542	1.0000

Table 5. *Correlation Matrix for DSGE Sample*

	Inf. Size (DSGE)	Exp. Wage Diff.	Urban Pop.	LLM Dummy	UM Dummy	CPI	BFI	Gov. wel. spend
Inf. Size (DSGE)	1.0000	-	-	-	-	-	-	-
Exp. Wage Diff.	0.4171	1.0000	-	-	-	-	-	-
Urban Pop	-0.4206	-0.3948	1.0000	-	-	-	-	-
LLM Dummy	0.3355	0.2609	-0.4920	1.0000	-	-	-	-
UM Dummy	0.5136	0.4147	-0.2361	-0.2091	1.0000	-	-	-
CPI	-0.6943	-0.4263	0.6022	-0.3905	-0.4968	1.0000	-	-
BFI	0.0323	-0.2021	0.0346	-0.0447	-0.2247	0.1371	1.0000	-
Gov. wel. spend	-0.5592	-0.4161	0.4961	-0.3579	-0.4638	0.7640	0.1247	1.0000

IV. Results

A. Pooled Ordinary Least Squares Estimation

The computed OLS coefficients for both the MIMIC sample and the DSGE sample can be found in Table 6 below:

Table 6. *Pooled OLS and 2SLS IV Coefficient Estimates*

	OLS		2SLS IV	
	(1)	(2)	(3)	(4)
Informal sector estimate approach	MIMIC	DSGE	MIMIC	DSGE
Expected wage differential, weighted by GNI	.2144362***	.0600604*	.1813512*	.0454274
	(.0395433)	(.0328556)	(.1085939)	(.1002591)
Urban population (as a % of total population)	.056076	.0844091**	.0515584	.0802714**
	(.0412302)	(.0366751)	(.0418902)	(.0406671)
Lower and Lower Middle Income Dummy	.0564608**	.1348654***	.0611662**	.1389613***
	(.0246732)	(.0243201)	(.0265829)	(.0241981)
Upper Middle Income Dummy	.0991322***	.1110191***	.1075348***	.1182379***
	(.0124511)	(.01051)	(.0155698)	(.0150993)
Corruption Perceptions Index	-.0028432***	-.0026162***	-.0028622***	-.0026276***
	(.0003312)	(.0002429)	(.0003296)	(.0002436)
Business Freedom Index	.0013686***	.0018949***	.0013109***	.0018475***
	(.0002824)	(.000274)	(.0002736)	(.0002725)
Welfare Total	.4482559**	.1336673	.4916795**	.2049153
	(.2405885)	(.2001507)	(.2508013)	(.2025142)

Note. Robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

With OLS estimation, most regressors are to some extent significant in both samples except for urban population, *popurban*, in the MIMIC sample and government spending on public healthcare and education, *welfarespend*, in the DSGE sample.

The effect of *exp wage diff (GNI)* on *infsize* is positive and significant at the 1% level in the MIMIC sample and at the 10% level in the DSGE sample. Because of the weighting by GNI per capita (PPP), the interpretation of the *exp wage diff (GNI)* coefficient is difficult to understand and readily apply, but essentially it can be understood as, ceteris paribus, the informal sector grows 21.4% (MIMIC) or 6% (DSGE) when a country's expected wage differential increases by an amount equivalent to the country's GNI per capita (PPP). Alternatively, a country's *infsize* will increase by 1% if the urban-rural expected wage differential increases by 4.6% (MIMIC) or 16.7% (DSGE) of the country's GNI per capita (PPP).

As mentioned, the effect of *popurban* is not significant in the MIMIC sample, but it is at the 5% level in the DSGE sample. In the DSGE sample, when all other variables are fixed, a 1% increase in the share of urban population sees a 8.4% increase in the relative size of the informal sector. This adds further support to the relationship between urbanization and the informal sector first theoretically presented by Rauch (1993).

For the income level dummies, the *UM dummy* is significant at the 1% level across both MIMIC and DSGE samples while the *LLM dummy* is significant at the 5% and 1% levels, respectively. All of the income dummy coefficients are positive, which indicates that lower middle and upper middle income countries have larger informal sectors than the base group, high income countries. For lower middle income countries, the relative informal sector size ranges from 5.6% (MIMIC) to 13.5% (DSGE) larger than higher income countries, and for upper middle income countries, the range is 9.9% to 11.1% larger.

The CPI, *corrpercp*, coefficients are negative and significant at the 1% level for both samples, with a 10-point increase in CPI resulting in a 2.7% decrease in the relative size of a country's informal sector. Because the CPI is on a counterintuitive scale where 100 represents a complete lack of corruption, I synthesize: the lower the CPI, the more corrupt a country, the greater the relative size of its informal sector. This effect is of the same direction and magnitude of previous literature. For example, Johnson, Kaufman, and Zoido-Lobatón (1998) found a 10-

point[8] increase in a country's CPI resulted in a 5.1% decrease in *infsize*. This is due to reasons such as a corrupt country being more likely to have corrupt officials willing to take bribes from informal companies so they can stay informal, increased levels of criminal-informal activity, and an overall lack of citizen trust of the government with the tax revenues, therefore citizens prefer to work in the informal sector to avoid paying taxes.

The BFI, *businessfree*, is positive and significant at the 1% level in both samples, suggesting increased business freedom (i.e., less regulation) increases *infsize* by 1.3% to 1.9% for every 10-point increase in a country's BFI. This is an interesting finding as previous literature suggests the relationship between business freedom and *infsize* is inverse as more business regulation would incentivise an individual or business to keep to or move to the informal sector (Johnson, Kaufman, Andrei Shleifer 1997). This could be indicative of measurement error in the data, as countries are likely to underreport the complexity or difficulty of doing business in their countries. This effect could also be entirely due to correlation over the years from governments with large *infsize*, typically less developed, introducing certain basic business regulation for the first time while countries with small *infsize*, often more developed, prune their existing regulations to encourage growth and business formation.

Finally, *welfarespend* is positive and significant at the 5% level in the MIMIC sample while insignificant in the DSGE sample. The result in the MIMIC sample suggests for every 1% increase in welfare spending, *infsize* increases by 45%. This is a very large effect and the exact opposite of what is expected. This result suggests that increased spending on health and education is actually correlated with increases in *infsize*. The discrepancy in significance across the two samples for *welfarespend* suggests further research needed to confirm this finding.

[8] Johnson, Kaufmann, and Zoido-Lobatón used the CPI when it was scaled to 10, I have adjusted their result to reflect its interpretation with the CPI now on a 100-point scale.

B. Pooled Two-Stage Least Squares Estimation

The computed 2SLS IV coefficients for both the MIMIC sample and the DSGE sample can be seen in Table 6 above, to the right of the OLS estimates. There exist a lot of similarities in the coefficients between the two models. Specifically, *urbanpop*, *LLM dummy*, *UM dummy*, *corrpercep*, and *businessfree* all have estimated coefficients nearly identical from the Pooled OLS model, so I do not provide in depth interpretation as I did with the OLS coefficients.

After implementing the IV *expmil* in the 2SLS regression[9], I see notable changes in the coefficients *exp wage diff (GNI)* and *welfarespend*. For *exp wage diff (GNI)*, the MIMIC sample is again the only significant coefficient, this time only at the 10% level, and the increase in *infsize* associated with the increase of the expected wage differential by the country's GNI per capita (PPP) drops to 18.1%. For *welfarespend*, the MIMIC sample is again, also, the only significant coefficient, at the 5% level once again, and the effect of 1% additional government expenditure on healthcare and education rises to a 49.2% increase in *infsize*.

C. Testing for Endogeneity

Next, I present test results exploring the potential endogeneity between *exp wage diff (GNI)* and *infsize* that is accounted for in the 2SLS IV regression in order to determine whether it performs better than the Pooled OLS model. I report the results of robust score chi-squared and robust regression F-tests in Table 7, below.

---

[9] See Appendix for First Stage Summary Statistics and 2SLS Wald Test for Instrument Strength.

Table 7. *Endogeneity Test Results*

Null: Model is exogenous	MIMIC	DSGE
Robust score chi-squared	.266355	.093446
	(0.6058)	(0.7598)
Robust regression F(1, 668) and F(1,576), respectively	.262641	.091685
	(0.6085)	(0.7622)

Note. P-values for test reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Both tests have null hypotheses of the model being exogenous, and consequently, a rejection of the null is in favor of the 2SLS IV regression model. However, as seen in Table 7, the p-values of both tests across both samples do not allow for the rejection of the null at any critical value. With this result, I conclude that the Pooled OLS model is the more efficient model.

D. Robustness Check: Fixed Effects

Further still, I conduct and report FE and FE IV regressions in Table 8 below to provide a robustness check on whether time effects should be accounted for in the model.

Table 8. *Fixed Effects Coefficient Estimates*

	FE		FE IV	
	(1)	(2)	(3)	(4)
Informal sector estimate approach	MIMIC	DSGE	MIMIC	DSGE
Expected wage differential, weighted by GNI	.2157779***	.0747212***	.1460001*	.0430683
	(.0253126)	(.0188985)	(.0839224)	(.0809443)
Urban population (as a % of total population)	.057481**	.0814943***	.0521553	.0773812**
	(.0241325)	(.0199144)	(.0342775)	(.0316318)
Lower and Lower Middle Income Dummy	.0749662**	.1411487***	.0818981***	.1449112***
	(.0263508)	(.0243413)	(.0222174)	(.020105)
Upper Middle Income Dummy	.1120259***	.11739***	.1181489***	.120983***
	(.0132309)	(.0090587)	(.0139725)	(.0143447)
Corruption Perceptions Index	-.0026352***	-.0025654***	-.0026498***	-.0025579***
	(.0002419)	(.0001508)	(.0003046)	(.0002693)
Business Freedom Index	.0013866***	.0019209***	.001316*	.0018942***
	(.0002926)	(.0003238)	(.000274)	(.00026)
Welfare Total	.4147529***	.2350567*	.3784028*	.2230768
	(.0776611)	(.119085)	(.2107629)	(.1892903)

Note. Robust standard errors are reported in parentheses. Panel ID is year. *** p<0.01, ** p<0.05, * p<0.1

Comparing the FE and FE IV estimated coefficients to Pooled OLS and Pooled 2SLS IV, there appear to be only small differences. The noteworthy changes are a large increase in the *welfarespend* coefficient in the FE DSGE sample model and increased regressor significance[10] due to the general efficiency gains of FE's lower standard errors. These efficiency gains suggest that a Pooled OLS model including year dummies is likely to perform better than an OLS model without a time component.

E. Pooled OLS and Pooled 2SLS IV with Time Dummies

Next, I compute and report both Pooled OLS and Pooled 2SLS IV with time dummies in Table 9, below, to capture the efficiency gains noted in the FE models.

[10] Regressor significance increases are as follows: DSGE: *exp wage diff (GNI)*, MIMIC & DSGE: *popurban*, and DSGE: *welfarespend*.

Table 9. *OLS and 2SLS IV Coefficient Estimates with Year Dummies*

	OLS with Year Dummies		2SLS IV with Year Dummies	
	(1)	(2)	(3)	(4)
Informal sector estimate approach	MIMIC	DSGE	MIMIC	DSGE
Expected wage differential, weighted by GNI	.2157779***	.0747212**	.1460001	.0430683
	(.0416139)	(.0341875)	(.1053984)	(.0996616)
Urban population (as a % of total population)	.057481	.0814943**	.0521553	.0773812*
	(.0413882)	(.037298)	(.0416641)	(.0407794)
Lower and Lower Middle Income Dummy	.0749662***	.1411487***	.0818981***	.1449112***
	(.0244931)	(.0246264)	(.0269774)	(.024658)
Upper Middle Income Dummy	.1120259***	.11739***	.1181489***	.120983***
	(.0126848)	(.0109556)	(.0157446)	(.0156836)
Corruption Perceptions Index	-.0026352***	-.0025654***	-.0026498***	-.0025579***
	(.0003458)	(.0002589)	(.0003376)	(.0002555)
Business Freedom Index	.0013866***	.0019209***	.001316***	.0018942***
	(.0002746)	(.0002786)	(.0002628)	(.0002751)
Welfare Total	.4147529*	.2350567	.3784028	.2230768
	(.2487756)	(.2107385)	(.2535034)	(.2081429)
1999 Dummy	-.0689757*	ø	-.0736461*	ø
	(.040326)		(.0410402)	
2000 Dummy	-.0745374*	ø	-.0782238*	ø
	(.0404574)		(.0409865)	
2001 Dummy	ø	ø	-.0672687*	ø
			(.0403466)	
2002 Dummy	ø	ø	-.0675652*	.0310533*
			(.0407365)	(.0168302)
2003 Dummy	-.072011*	ø	-.0765213*	ø
	(.0395465)		(.040277)	
2004 Dummy	-.0755732*	ø	-.0804765**	ø
	(.0391773)		(.0400006)	
2005 Dummy	-.0694916*	ø	-.0738968*	ø
	(.0392793)		(.0398908)	
2006 Dummy	ø	.0565524*	ø	.0355924**
		(.0326788)		(.0152659)
2007 Dummy	ø	ø	ø	.0294312**
				(.0143651)
2008 Dummy	ø	ø	ø	.0277782*
				(.0154429)

Note. Robust standard errors are reported in parentheses. Only year dummies that are significant at the 10% level or lower are reported, insignificant year dummies are denoted by a "ø". *** p<0.01, ** p<0.05, * p<0.1

For the OLS and 2SLS regressions on the MIMIC sample, the year dummies reported are negative and significant at the 10% level. These negative signs suggest that during these years, roughly 1999 to 2005, there was a significant trend of informal sector reduction in size. These years lineup with years of global economic expansion, which suggests that the global economic growth experienced during this time may have led to decreases in the size of informal sectors. Contradictory, for the DSGE sample, all year dummies are positive and significant at either the 5% or 10% level. This is most likely due to the unbalanced nature of the DSGE sample as the time effect is weakened when there is not balanced, sequential observations for each country within the dataset. The more balanced the dataset, the more informative the year dummies are. Therefore, I contend the coefficients of the year dummies from the MIMIC sample are more rigorous and accurate.

V. DISCUSSION

This Discussion section seeks to consolidate and distill my findings as well as provide analysis of significant effects related to the outcomes of interest, including economic reasoning and policy recommendation. Following the regression analysis reported above, Pooled OLS with year dummies is evidently the best specified model. As previously mentioned, after conducting endogeneity tests on the the 2SLS model, the results show the model is safely exogenous, with p-values around 0.6 and 0.75 for both tests of both samples. These results clearly show that OLS is more appropriate than 2SLS. Further, after conducting FE modeling to better understand potential time effects and finding significance in a few year dummies, I conclude that the addition of time effect dummies provides enough of an improvement in the model to warrant their inclusion. Therefore, the Pooled OLS with year dummies regression is the model I focus this discussion on.

Foremost, examining the coefficient of *exp wage diff (GNI)*, there is positive significance at the 1% level for MIMIC and 5% for DSGE where we see an effect ranging from a 7.5% (DSGE) to 21.6% (MIMIC) increase in *infsize* for every increase in the expected wage differential equivalent to the country's GNI per capita (PPP). By instead using a ratio out of the magnitude of the effect, we can create a standardization that better articulates the impact of the expected wage differential. For example, for the MIMIC sample, if a country's expected wage differential increases by 4.6%[11] of its GNI per capita (PPP), the *infsize* can be expected to increase 1% in relative size. Although there is a discrepancy between the two samples about how strong the effect is, *exp wage diff (GNI)* remains one of the strongest, most significant effects in the model. This finding supports the theoretical framework and motivation of this study, which is in using the driver of the Harris-Todaro migration model as a predictor, and lever, for the size of the informal sector. This result adds further evidence to the Harris-Todaro model as a realistic model for migration, presents expected wage differential as a new urbanization-related explanatory variable to consider in the field, and refocuses governments on controlling urbanization through expected wage differentials as a solution space for managing the size of their informal sectors.

As for policy recommendations, if governments wish to reduce the size of their country's informal sectors, as the International Labor Organization (ILO) recommends for development, this result implies they ought to focus on reducing the expected wage differential (ILO 2015). This can be accomplished through urban wages decreasing, labor participation decreasing, rural wages growing, or a combination of the aforementioned. However, no country would rightly

[11] This percent is calculated by simply dividing 1 by 21.6.

consider decreasing wages or labor participation a justifiable policy, therefore countries must focus on expanding rural opportunity. This finding provides a powerful argument for cross-subsidization from the urban sector to the rural sector. High-value, urban-industrial work, and the tax revenue thereof, would be better leveraged in subsidies to lower-value agricultural work in a country's rural area than in congested urban centers. Further expenditure in urban centers only widens the expected wage differential and causes further erosion in the marginal utility of government spending. Urban-to-rural cross-subsidization as a part of economic policy allows countries to optimally use resources while also reducing informal sector size via its reduction of the urban-rural expected wage differential.

Nonetheless, there are limitations to urban-to-rural cross-subsidization as an effective solution for all countries. The dynamics of expected wage differentials are likely different for high-income countries that have reached a steady-state in the urbanization phenomenon as compared to lower and upper middle income countries. More developed, higher income nations are more likely to be highly urbanized, and therefore they lack a sizeable enough rural population being drawn to the urban sector to make an impact on the size of the informal sector. Because of these differences, cross-subsidization may be a less effective strategy in further decreasing the size of informal sectors for high-income countries.

There are additional complexities if premature deindustrialization is considered. Premature deindustrialization, the transition of a developing country's economy from heavily manufacturing to serivce-oriented prior to it attaining full development, causes a loss of low-skill, urban manufacturing jobs, and therefore lessens the expected wage differential between urban and rural areas (Rodrik 2016). If a country is experiencing premature deindustrialization, the informal sector may grow even faster as previously formal sector employees are forced into underemployment. If this is the case, the fostering of rural, non-industrial jobs for low-skill laborers and spending on education to increase human capital levels becomes of utmost importance in order to open new channels to development as manufacturing dries up.

Beyond this guidance, it is vital that countries evaluate current and future economic development efforts to maintain balanced urban-rural, or even pro-rural, agendas. If these findings are not taken into account, disproportionate urban wage growth can cause massive informal sector growth through the vessel of urbanization. Instead, by focusing on decreasing the wage differential through efforts to increase opportunity and wages in rural areas, governments can limit the incentive to migrate to urban centers and accordingly reduce or limit the size of the informal sector. Understandably, this is a difficult balance as countries may simply wish to generate economic opportunity for their citizens, regardless of the location and as quick as possible. Still, in order to avoid the many costs of large informal sectors such as lost tax revenues, congestion costs, lower productivity, and human rights violations, to name a few, government leaders should consider the effect development initiatives have on urban-rural expected wage differentials and how that drives the size of the informal sector.

Welfare spending, the secondary outcome of interest in this study, has a much more puzzling effect. The *welfarespend* coefficient is only significant in the MIMIC sample and has a sizeable effect – a 41% increase in *infsize* given a 1% increase in aggregate government spending on public healthcare and education, as a percentage of GDP. Most interesting is that the effect is positive, which contradicts existing notions that increased welfare spending decreases the size of the informal sector as it provides resources that help diminish barriers between informal sector workers and formal sector employment, such as fostering additional human capital and lessening injury or sick time. There may be an explanation hidden in the geographic breakdown of welfare spending. It is likely that a disproportionate amount of public education and healthcare spending is concentrated in urban areas, and any increases in spending only further increases the incentives for rural-to-urban migration, thus fueling informal sector growth. Assuming this hypothesis is true, similar to policy recommendations affecting expected wage differential, governments must focus on being more even handed in the distribution of resources for education and healthcare. Unfortunately, this study's model does not allow for a fully causal interpretation, and, due to this uncertainty, I recommend further research examines *welfarespend*'s effect on the size of the informal sector with a focus on *welfarespend* by urban-rural split. To further qualify, this effect is only significant at the 10% level and for only one sample, so it may still be that this interpretation is Type I error.

Looking to the future, this study is the first exploration of expected wage differentials as a driver of informal sector shifts, and therefore, there is need for further, more robust research to confirm the findings presented in this paper. This study only examines the foundation of the Harris-Todaro probabilistic model of urbanization, which is one of the simpler probabilistic models. Due to data availability, it is difficult to empirically study other models, but that does not mean that it should not be pursued when the necessary granularity of data is available. This study also provides preliminary evidence supporting the theoretical model of how urbanization interacts with the informal sector. Still, the informal sector is a diverse beast, and there are countless additional variables that have influence on its size. There are likely different outcomes derived from expected wage differentials based on a country's stage of development, economic strengths, and culture. This being said, further research is absolutely warranted to explore the intricacies of these unaccounted for factors.

Appendix

Table 10. *First Stage Summary Statistics*

R-squared	0.4714	0.4364
R-squared, Adjusted	0.4659	0.4296
R-squared, Partial	0.2206	0.1648
Robust F(1,576)	56.7991***	40.8633***
	(0.0000)	(0.0000)

Note. P-values for test reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11. *2SLS Wald Test for Instrument Strength*

Null: Instrument is weak		
Critical Value at 10%	16.38	16.38
Minimum eigenvalue statistic	189.395	113.856

Note. There are no critical values producable for lower percentages. Given the minimum eigenvalues are much higher than the 10% critical value, we can reject the null in favor of the instruments being sufficiently strong.

References

Acosta-Gonzalez, Eduardo, Fernando Fernandez-Rodriguez, and Simon Sosvilla-Rivero. 2014. "An Empirical Examination of the Determinants of the Shadow Economy." Applied Economics Letters 21, no. 4-6: 304-307.

Bhattacharya, Prabir C. 1998. "Migration, Employment and Development: A Three-Sector Analysis." Journal Of International Development 10(7): 899-921.

Banerjee, Biswajit. 1983. "The role of the informal sector in the migration process: A test of probabilistic migration models and labour market segmentation for India." Oxford Economic Papers 35(3): 399–422

Bourhaba, Othmane, Mama, Hamimida. 2016. "An Estimation of the Informal Economy in Morocco." International Journal of Economics and Finance, 8(9): 140-147.

Elgin, Ceyhun, Öztunali, Oguz. 2012. "Shadow Economies around the World: Model Based Estimates." No 2012/05, Working Papers, Bogazici University, Department of Economics.

Elgin, Ceyhun, Oyvat, Cem. 2013. "Lurking in the cities: Urbanization and the informal economy." Structural Change and Economics Dynamics, 27(C): 36-47.

Euromonitor. 2018. Economies and Consumers Annual Data. Disposable Income by Urban/Rural Split. Retrieved from Passport.

Fields, Gary S. 2004. "Dualism in the labor market: A perspective on the Lewis model after half a century." The Manchester School, 72(6): 724-735.

Hassan, Mai & Schneider, Friedrich, 2016. "Size and Development of the Shadow Economies of 157 Countries Worldwide: Updated and New Measures from 1999 to 2013." IZA Discussion Papers 10281, Institute for the Study of Labor (IZA).

ICTD/UNU-WIDER, 'Government Revenue Dataset', June 2016, https://www.wider.unu.edu/project/government-revenue-dataset'.

Johnson, Simon, Daniel Kaufmann, and Andrei Shleifer. 1997. "The Unofficial Economy in Transition." Brookings Papers on Economic Activity, 1997 (2): 159-239.

Johnson, Simon, Kaufmann, Daniel, Zoido-Lobaton, Pablo. 1998. "Regulatory Discretion and the Unofficial Economy." Discussion paper, The World Bank, Washington, DC.

ILO. 2015. R204 - Transition from the Informal to the Formal Economy Recommendation. (No. 204). http://www.ilo.org/employment/units/emp-invest/informal-economy/WCMS_443501/lang--en/index.htm.

Ledent, Jacques. 1982. "Third World Migration and Urbanization: A Symposium." Economic Development and Cultural Change 30(3): 507-538.

Lewis, W.A. 1954. "Economic Development With Unlimited Supplies of Labor." Manchester School of Economic and Social Studies 22: 139-91.

Magazzino, Cosimo, Buscemi, Antonino, and Yallwe, Alem H.. 2011. "The Underground Economy in the Caribbean Countries." International Journal Of Economics And Business Modeling 2, no. 2: 124-132.

Rauch, J. E. 1993. "Economic Development, Urban Underemployment, and Income Inequality." Canadian Journal of Economics, 26 (4), 901-18

Rodrik, Dani. 2016. "Premature deindustrialization," Journal of Economic Growth 21(1): 1-33.

Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko & Richard Svensson. 2017. The Quality of Government Standard Dataset, version Jan17. University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se doi:10.18157/QoGStdJan17

United Nations, Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision.

World Bank. 2017. World Development Indicators 2017. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/26447 License: CC BY 3.0 IGO."

Zelinsky, W.. 1971. "The Hypothesis of the Mobility Transition." Geographical Review, 61: 219-49.