“Productivity dispersion and dual labour markets: evidence from 14 Latin American Countries”

TESIS PARA OPTAR AL GRADO DE

Magister en Economía

Alumno: Alejandro González

Profesor Guía: Esteban Puentes

Santiago, Noviembre de 2019

Productivity dispersion and dual labour markets: evidence from 14 latin American Countries

Alejandro González*

November 12, 2019

Abstract: A key feature of developing economies is the coexistence of large pools of subsistence workers alongside workers employed in the modern sector of the economy. Traditional theories of dualism predict that increases in the productivity of the modern sector will eventually absorb the subsistence sector. I test this hypothesis by estimating the reduced-form correlation between productivity and the probability of belonging to the subsistence sector by merging measures of industry wide productivity from urban registered firm with labor market outcomes from 14 Latin American countries. My empirical strategy uses a finite mixture model which estimates jointly the wages corresponding to each sector and the segment to which the worker belongs, which is treated as a latent variable. Surprisingly, I find that my measure of industry wide productivity has a negligible influence on the probability that workers switch from the subsistence to the modern sector. Nevertheless, I find strong evidence of dualism and substantial hourly wage differentials across sectors; the average treatment effect of switching to the modern sector is a 14% wage gain.

Key words: Dualism, Productivity, Latin America
JEL codes: O12, O14, J31

*University of Washington. Mail: agonc@uw.edu. The author would like to thank Esteban Puentes for his guidance, patience and multiple discussions of previous drafts of this paper. He would also like to thank Tomas Rau and Mauricio Tejada for helpful discussions regarding this paper, as well as seminar participants at the University of Chile for comments on previous versions of this paper. Finally, he thanks deeply ECLAC and Xavier Mancero in particular for providing the harmonized data on social surveys used in this paper.

1 Introduction

Models of economic dualism continue to be a popular framework through which we can understand growth, technical change and income distribution in developing economies. In Lewis (1954) seminal contribution, the dual economy is composed of a modern, capitalist sector, and a large subsistence pool of low-productivity workers. Furthermore, the supply of labor is perfectly elastic at the wage rate determined by the average product of the traditional sector, and the process of capital accumulation in the modern sector is the driver of long-run growth: As capital accumulates in the modern sector, workers from the subsistence sector are reallocated to the modern sector, and the process of development is simply the shrinkage of a traditional mode of production in favor of the modern capitalist sector.

The Lewis model makes a set of rich empirical predictions; one of them concerns the asymmetrical impact productivity growth has on wages and the share of subsistence workers. If productivity increases in the modern sector, then labor is reallocated from the subsistence to the modern sector at a given wage rate, resulting in a diminishing share of subsistence workers. If productivity increases in the traditional sector, however, then the wage rate increases as well, diminishing profits in the modern sector and slowing down capital accumulation; this leaves the share of subsistence workers unaffected. Modern formulations of the Lewis model retain this prediction, albeit the mechanisms are somewhat different (Temple, 2005; Vollrath 2009a).

In light of these theoretical predictions, **it's perhaps surprising that the vast empirical literature on dualism has ignored the potential impact of productivity on the allocation of worker's between sectors.** While there is a vast literature testing whether workers sort themselves between sectors or are excluded from the modern sector (Maloney, 1999; Radchenko, 2017; Contreras, Gillmore and Puentes, 2018) and how worker's characteristics influence either the queuing or sorting process, none of these papers investigate how firm or industry level characteristics impact the share of subsistence workers. This paper contributes to filling this empirical gap.

In order to test the hypothesis that increases in the productivity of the modern sector decrease the share of subsistence workers, we merge measures of industry-wide productivity from urban registered firms in the World Enterprise Surveys (WES), which are assumed to represent the modern sector, with data on labor market outcomes from social surveys previously harmonized by ECLAC in 14 Latin American countries. Then, we use a Finite Mixture Model (FMM) to estimate jointly the wage distributions in both sectors and the allocation of workers among them. FMM models have the crucial advantage of being a data-dependent method to classify which workers belong to the subsistence sector and to the modern sector; intuitively, the method maximizes the distance between the two conditional mean functions which describe wages in each sector; in turn, these conditional mean functions are simply modeled as standard earnings regressions. This contrasts heavily with other approaches which treat the subsistence sector as an observable variable, which is either modeled as the self-employed, as informal worker or as agricultural workers.

Surprisingly, while we find substantial evidence that two regimes describe the data better than one regime, and thus, of dualism, productivity has a negligible role in explaining how the share of subsistence workers varies between industries: Moving a worker from the 25th percentile to the 75th percentile of the productivity distribution increases the probability that a workers switches to the modern sector by roughly 0.42%. The level of schooling seems to be the most important determinant of the share of subsistence workers that a given industry employs; and given the strong correlation between an industry average schooling and its productivity, the strong correlation between productivity and the share of informal workers found in other studies might reflect, **in a large degree, the correlation between schooling and productivity**. Another prediction of dual economy models is that increases in the productivity of the modern sector should progressively shrink the wage gap between the modern and the subsistence sector during the development process. **I also empirically examine this proposition by computing** the correlation between the degree of overlap in wage distributions for each sector and the productivity of the modern sector at the industry level; this prediction is also not borne by the data: the productivity of the modern sector is uncorrelated with the degree of overlap of the wage distributions of both sectors.

The rest of the paper proceeds as follows: Section 2 discusses how the productivity of the modern sector might impact the share of subsistence workers. Section 3 presents our data, both from social surveys and from the WES database, and discuss our measures of productivity, along with the characteristics of our sample. Section 4 briefly presents FMM and rationalizes our modeling strategies. It then discuses how our estimated subsistence sector might correspond or not to the informal sector. Section 5 presents our main results, which include our wage and selection equations, the distribution of marginal effects along the productivity distribution, a characterization of the subsistence sector, average treatment effects, and our estimated wage densities vs the actual density. Finally, section 5 concludes and discuss some policy advice.

2 Dual Labor Markets and Productivity Dispersion

2.1 A Benchmark Model

The literature on dual economies is vast; and the emphasizes of theorists usually depend on whether they focus on growth or labor markets. To motivate our reduced form specification and focus the subsequent discussion, we briefly present a model due to Vollrath (2009a), which can be read as a modern interpretation of the Lewis model. **It should be clarified from the outset that we will not directly estimate this model, rather, we'll use it's implications to organize our empirical findings.** In this, model, the subsistence sector is identified by agriculture and the modern sector with manufacturing, nevertheless, we index both sectors with s for subsistence and m for modern, since our data exclude agricultural firms and workers. Total Labor is L, of which a fraction $a_t \in (0,1)$ work in the subsistence sector, and $1-a_t$ work in the modern sector. Each worker has a unit of time $s \in (0,1)$, which can be allocated either to wage work or to a non-market activity, which could be home production or child rising. Two assumptions are crucial to the analysis: first, workers in the subsistence sector face a non separability between market time and domestic production, despite having the same utility function as modern workers; second, workers consume a fixed bundle of subsistence goods, $\bar{b}$.[1]

Production in the subsistence sector is given by a constant returns to scale function of labor effort and some fixed resource (land in Vollrath's formulation). Capital is omited. Total subsistence production is given by:

$$Y_t^S = A_t^S F(R, E_t^S) \tag{1}$$

[1]The non-separability from subsistence workers comes from the fact that the subsistence sector does not have a labor market. Absent a labor market, subsistence workers will internalize the diminishing returns to their marginal product, and chose a lower amount of work effort that modern workers. See the original paper for more details.

Where A_t^S is total factor productivity in the agricultural sector, R is the amount of a fixed resource in this sector (land in Vollrath's formulation), and $E_t^S = s_t a_t L_t$ is total labor expended. As usual, the function is concave in both arguments. Each worker in the subsistence sector maximizes their net income, which is equal to:

$$I_t^S = p_t^S A_t^S F(R, E_t^S) - \rho_t r_t \tag{2}$$

Where the p_t^S is the relative price of the subsistence good, ρ_t is the rental price of the fixed resource, and r_t is the quantity of fixed resource employed (which is exogenously given.) Maximization of this income program subject to the time constraints of the subsistence worker, and assuming constant returns to scale, gives:

$$I_t^S = s_t p_t^S A_t^A F_E \tag{3}$$

Where F_E denotes the derivative of the subsistence technology with respect to labor. This equation simply states that total income in the subsistence sector is equal to the monetary value of the marginal product of labor effort ($p_t^S A_t^A F_E$) multiplied by the total amount of time spent working s_t. Note that this departs from Lewis' treatment of the subsistence sector wage as equal to the average product of land, which is a common assumption in dual economy models. Another crucial observation about the previous equation is that illustrates why the subsistence sector is productively inefficient but maximizes overall welfare: Due to the non-separability of non-market work, workers in the subsistence sector internalize the choice of s_t on their income. Thus, expending more hours in market work lowers the marginal product of their labor; this lowers the relative time cost of non-market activities, and therefore subsistence workers optimally choose to work less. Absent this non-separability, the dual economy would not exist.

To complete our exposition, we need an equation governing the wages of the modern sector. Here Vollrath assumes that the wage paid is proportional to total factor productivity in the modern sector and that there are diminishing returns to labor in the modern sector as well:

$$I_t^M = w_t^M s_t = A_t^M g(a_t) s_t \tag{4}$$

Thus, total income depends on productivity A_t^M, total work time s_t and a function which is increasing on the share of workers employed in the subsistence sector $g(a_t)$.[2] We'll skip a detailed description of the utility functions and worker optimization to get to the main results. Evidently, an equilibrium requires that workers in both sectors chose their work time s_t to maximize their utility, and that there is no arbitrage in terms of the utility of workers - that is, that they earn the same utility in both sectors. Additionally, since the consumption of subsistence goods is given, it is required that total demand over subsistence goods is equal to its supply: $\bar{b}L = A_t^S F(R, E_t^S)$. Unfortunately, there is no closed-form solution for the above model (including the equilibrium share of subsistence workers), but the author is able to derive some key analytic results regarding the equilibrium of this economy, **which we now discuss**.

The first contention is that the marginal product of a worker in the modern sector will be higher than the marginal product of a worker in the subsistence sector, or in other words:

$$w_t^M s_t^{*M} > s_t^{*S} p_t^{*S} A_t^S F_E \tag{5}$$

[2] Absent population growth, a_t tracks the total quantity of labor employed in the labor sector. Thus, this is equivalent to saying that there are diminishing returns to labor in the modern sector.

Where an asterisk over a variable denotes its optimal value. As mentioned before, the intuition behind this result is market time spent in the modern sector s_t^{*M} is higher than in the subsistence sector s_t^{*S}, due to the aforementioned non-separability. Thus, subsistence workers optimally choose to work less and enjoy more non-market time. A corollary of the above result is that output gains per worker are possible if workers are transferred from the subsistence to the modern sector; furthermore, they also imply that wage gains are possible by moving workers from the subsistence to the modern sector.

The second contention is that increases on the productivity of the modern sector (A_t^M) decrease the share of labor in the subsistence sector (a_t), while also increasing the market time devoted to workers in both sectors. Increases on the productivity of the subsistence sector also decrease its share of labor, but they lower market work in both sectors. A corollary of these results is that increases in the productivity of the modern sector decrease the difference in work effort in both sectors, which decreases the gap between marginal productivities in both sectors and the wage gap between both sectors.

2.2 Discussing deviations and extensions

Not all growth models which employ dual economies predict these asymmetric effects. There are two papers where technical progress in the modern sector does not lead to a diminishing share of subsistence workers with clearly delineated mechanisms are worth emphasizing: First, Eswaran and Kotal (1993) build a fairly standard Lewis model where consumer have hierarchical needs: Agents only consume subsistence goods (Y_t^S in the previous notation) until a certain threshold of income, say $\bar{G}$ has been met. This implies that the degree of substitution between the good produced in the modern sector and the subsistence sector is close to zero from the consumption side. Thus, whenever there is an increase in the productivity of the modern sector, this increases the real wage of modern workers and lowers the relative price of the modern good, but labor is not reallocated from the subsistence to the modern sector, since the demand for the modern good is insensitive to its relative price, as long as incomes are below the $\bar{G}$ threshold. Second, Proto (2007) builds a Lewis model where jobs in the modern sector need a certain level of schooling, while jobs in the subsistence sector require no schooling. Workers in the subsistence sector own no assets, and given imperfect capital markets (i.e, workers cannot borrow against promises of future labor income from schooling) they are wealth-constrained. This implies that if productivity in the modern sector increases, the share of subsistence workers does not need to decrease, since subsistence workers do not have enough schooling and cannot obtain such schooling (due to the mentioned wealth constraints) to be hired in the modern sector. Only if the increase in productivity is accompanied by policies which increase education (such as an education subsidies or a redistribution of wealth, in the form of an agrarian reform) will the share of subsistence workers decrease.

While growth theorists usually use the manufacture/agriculture divide to operationalize the subsistence/modern distinction employed by Lewis, most models of economic dualism used by labor economists identify the subsistence sector with the informal sector, and the modern sector with the formal sector. While an exact definition of the informal sector is a controversial aspect of the literature, most theorists posit that the informal sector avoids costly labor and tax regulations, but has a probability of being audited by the government and being charged a fine. The opposite occurs in the formal sector. In these models, each firm draws a productivity realization from a common distribution: firms which draw low productivity choose to sort themselves into the informal sector, while firms with high productivity sort themselves into the formal sector. Typically, the probability of being caught is made a function of the amount of labor employed, by arguing that bigger firms are easier to audit by the government. Thus, if the productivity distribution shift to the right, a higher fraction of firms will choose to formalize, given that on the margin, the increase in the probability of getting caught offset the costs from evading labor and tax regulations (Bosch and Esteban-Prestel (2012), Albrecht, Navarro and Vroman (2012), Meghir, Narita and Robin (2015)).

Since these models assume a common productivity distribution for both the informal and formal sector, unlike the growth models we reviewed, there is nothing special about the productivity of the modern sector vis-a-vis the the subsistence sector. However, the recent contribution by Ulyssea (2018) highlights the possibility of asymmetrical effects: in his model, firms employed in the formal sector are allowed to hire both formal workers and workers "off the books", by which they avoid labor regulations. In contrast, unregistered firms hire all of their labor informally. If productivity increases for unregistered firms, they will have incentives to formalize themselves, but will still hire all of their labor off the books, which implies the share of informal labor will increase. However, if firms in the formal sector face a productivity increase, they will hire a larger share of their workers formally, since employing more labor will make them more visible to government audits, and the increased costs of (expected) fines offsets the cost of paying labor regulations. Thus, the aforesaid paper is central in establishing how the asymmetrical effects of productivity increases in the modern and subsistence sector, but it also emits a note of caution in establishing a one-to-one relationship between subsistence sector *firms* and subsistence sector *workers*, which is a tacit assumption in the Vollrath model examined above. This distinction will play an important role in the subsequent empirical analysis.

For the purposes of analyzing the effect of modern sector productivity on the share of subsistence workers, it is irrelevant whether workers are excluded from the modern sector and queuing for a job in this sector, or if they sort themselves into each sector according to comparative advantage, as in the Roy (1954) model or the Vollrath model presented above. While there is a large literature devoted to test these competing hypotheses (Magnac, 1991; Maloney, 1999; Radchenko, 2017; Contreras, Gillmore and Puentes, 2018), in both self-selection and exclusion models an increase in the productivity of the modern sector will decrease the share of subsistence workers: If jobs in the modern sector are rationed and there are queues to get one, then increases in productivity will expand employment, and diminish the size of the queue, which in turns shrinks the share of subsistence workers. Likewise, if workers sort themselves between sectors, and the expected wage of participating in the modern sector increases (due to the increase in modern productivity), then, *ceteris paribus*, a larger share of workers will choose to reallocate themselves into this sector. Thus, the share of subsistence workers always falls after the modern sector increases its productivity.

To summarize, a common prediction of dual economy models is that increases in the productivity of the modern sector will decrease the share of subsistence workers. Additionally, we would expect these expansions in productivity to diminish the gap in marginal products - and thus, wages - between the modern and the subsistence sector, and that shifts from the subsistence to the modern sector entail wage gains for workers. To the best of our knowledge, this is the first empirical work to try to contrast these hypotheses using detailed data on labor market outcomes and measures of productivity at the industry level for a large number of developing countries; however, in a recent paper, Allen, Nataraj and Schipper (2018) examine many of these predictions at the *firm* level for a cross-section of manufacturing industries in India. They equal the subsistence sector with the informal sector, and the modern sector with the formal sector, and confirm that industries with higher average productivity contain a smaller share of informal firms, and that industries with higher average productivity have a smaller productivity gap between formal and informal firms. However, their analysis relies solely and descriptive statistics and fails to account for the confounding effect of other covariates, such as average worker schooling, an issue which is considered in our analysis.

3 Data

In order to test our hypotheses, we need data on the productivity of the modern sector, a method to identify which workers belong to the modern sector and which workers belong to the subsistence sector, and wages of the workers in both sector. To construct a measure of the productivity of the modern sector, we draw on firm-level data from the World Enterprise Surveys, which contains data on productivity for registered, urban firms around the developing world. We then aggregate these data to the 2-digit industry level to merge it with individual worker data drawn from social surveys previously harmonized by ECLAC.

An ideal dataset to test these hypotheses exploiting cross-sectional variation would need wages from workers, a measure of productivity at the firm level, and a method to classify firms and workers into the modern or subsistence sector. Our dataset faces two main limitations: First, it's not possible to match firm-level data with worker-level data. Thus, we opt to aggregate firm level data to the 2-digit industry level and merge this measure of productivity with data on workers. Second, measuring the modern sector and subsistence sector is not a trivial task: as mentioned above, some researchers use the agricultural sector as a measure of the subsistence sector; while others use the informal sector for the same purposes. In this respect, our firm-level dataset only contains firms located in urban sectors, and which are registered in the tax system of their respective countries. Thus, we make the assumption that our sample of firms is an accurate representation of the modern sector. However, we still have the task of measuring the share of subsistence workers; in this respect, our dataset does not contain information on the formality status of workers - we do not know who has a written labor contract or who receives some form of social security. To match our sample of firms, we restrict the sample of workers to urban areas; however, it seems overly restrictive to equate all urban workers with the share of modern workers. Thus, we opt to treat the share of subsistence workers as a latent variable and use an estimation procedure which will allow us estimate whether a worker belongs to the subsistence or modern sector. This will be discussed in more detail below. We now turn to a discussion of some relevant features of our firm and worker level data.

3.1 Industry-Level Data

As mentioned above, our Industry-Level data comes from the World Enterprise Surveys (WES) conducted every four years by the World Bank. The WES database covers non-agriculture firms in the major cities of virtually all south American countries, as well as many Caribbean countries. The database contains modules on finance, R&D, competition, corruption, pricing decisions and productivity, among others. In order to be eligible for the survey, all firms have to be registered in their respective countries, which means that they comply with costly tax and labor regulations.

We posses data of sales per worker disaggregated at the 2 digit level, which we'll use as our measure of productivity. Even though Value Added or Total Factor Productivity (TFP) are usually employed as measures of productivity in various studies, we do not employ Value Added since measures of intermediate inputs are only available to manufacturing industries, and TFP estimates in the absence of panel data carry along with them a plethora of problems which make them unreliable, such as measurement error in capital stocks, simultaneity and self-selection issues (Blundell and Bond, 2000; Levinsohn and Petrin, 2003). Furthermore, if the technology available to firms is of the Leontief-Sraffa-Von-Nueman type, then, there is a close link between productivity and sales per worker. Let sales per worker be:

$$\frac{Sales}{Worker}_j = \frac{p_j Y_j}{L_j} = p * b \tag{6}$$

Where p_j is the product price in industry j, Y_j is a measure of output, and L employment. b is simply the fixed average labor productivity or labor coefficient in a production function of the type $Y = \min\{bL, vK\}$. Thus, our simple measure captures the dispersion of average labor productivity and the dispersion of prices, and its assumed that the dispersion of physical productivity across industries is more dominant than the variation in relative prices.[3] Therefore, we will refer to sales per worker subsequently as Y/L, average labor productivity. This measure is larger than Value Added per worker and TFP, and thus, we would mechanically estimate a smaller coefficient using sales per worker.[4]

Our measure of productivity has a reasonably good amount of coverage in terms of sectors, since it includes manufacturing, construction, transport, hotels, restaurants and wholesale and retail trade. This makes our data more representative of the average worker in Latin American countries than traditional manufacturing census employed in other studies which measure productivity in Latin America (e.g, Busso, Madrigal and Pagés, 2013), since on average only 15.2% urban workers are employed in the manufacturing sector.[5] However, the downside of this is that the sample is significantly smaller than the more commonly employed manufacturing census: On average, sample sizes go from 475 for small countries (Bolivia, Costa Rica, Honduras, Nicaragua and Paraguay) to 900-950 for big countries (Argentina, Brazil, Colombia and Mexico).

In many cases, this implies that for sector j in country k we will only count with a handful of firms, going from a hundred to merely 4 o 5. In order to increase sample size, we pool together two distinct years of the WES surveys, which are available for all countries except for Brasil. The first year usually corresponds to the period 2004 - 2007, and the second to 2009 - 2010. Despite having a panel dimension, we do not exploit it in order to increase sample sizes.

In order to get a reasonable number of observations to estimate industry-level average productivity, we decide to use a cutoff rule of 25 observations to compute industry-wide productivity. Table 1 shows the number of observations for each industry in each country, and highlights in bold character those industries which satisfy our cutoff.

[3] Note that the inability to separate the dispersion in relative prices from physical productivities plagues all the literature on the subject; that's why some researchers prefer to refer TFP as TRFP - total *revenue* factor productivity.

[4] Card et. al (2018) show in an unrelated context that regression of Sales per Worker on wages usually give smaller coefficient estimates than regressions of Value Added or TPF on wages, relying on a literature review that covers two dozens of papers for different countries and datasets.

[5] Detailed data on the distribution of employment and the representativeness of our sample can be found on the appendix.

Table 1: Number of observations per sector and cutoff rule

	Argentina	Bolivia	Brasil	Colombia	Costa Rica	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Nicaragua	Paraguay	Peru	Uruguay
Manufacturing of Food, Beverages and Tobacco	**292**	**126**	**132**	**262**	**81**	**118**	**176**	**189**	**147**	**296**	**127**	**87**	**104**	**197**
Manufacturing of Textiles	**156**	15	**137**	**179**	8	**30**	**68**	**100**	**26**	**156**	19	5	11	**66**
Manufacturing of Wearing Apparel	**197**	**95**	**141**	**253**	4	**31**	**120**	**92**	**38**	**250**	**29**	16	**43**	**81**
Manufacturing of Leather Items	**31**	4	**119**	**29**	2	9	3	9	3	24	17	4	7	14
Sawmill and Planning of Wood	4	5	15	4	12	7	2	**36**	**49**	7	**36**	4	18	4
Manufacturing of Paper and Paper Products	13	3	7	2	4	6	1	9	6	6	4	11	10	6
Publishing, Printing and Reproduction fo Recorded Media	18	9	18	19	13	19	8	10	10	15	20	24	**31**	12
Manufacturing of Coke and Chemicals	**157**	**52**	**111**	**246**	17	**73**	**58**	**42**	**43**	**288**	**45**	11	**71**	**105**
Manufacturing of Rubber and Plastic Products	**48**	23	**33**	**61**	16	**33**	24	**28**	13	**136**	19	14	**28**	**48**
Manufacturing of Mineral Products	16	**25**	18	4	**27**	7	**70**	**35**	**36**	**171**	**69**	16	**28**	9
Manufacturing of Basic Metals	6	2	12	15	7	4	18	21	8	14	10	1	11	4
Manufacturing of Fabricated Metal Products	**109**	**25**	**91**	**115**	**26**	**36**	**62**	**37**	**33**	**153**	**44**	24	20	10
Manufacturing of Machine and Equipment	**216**	0	**137**	22	17	8	9	8	5	**245**	6	9	5	2
Manufacturing of other Machinery	15	6	**41**	6	8	6	3	2	6	**141**	3	1	15	4
Manufacturing of Transport Equipment	9	1	**98**	8	2	8	3	6	1	7	4	3	3	4
Manufacturing of Furniture	7	1	**150**	13	10	21	14	**28**	**47**	**115**	**50**	19	15	2
Recycling	1	0	3	1	1	0	0	0	0	5	0	0	2	1
Construction	23	10	**29**	19	7	12	16	7	5	23	12	4	18	10
Sales of Motor Vehicles	14	12	**34**	6	5	15	12	16	17	21	13	5	13	12
Wholesale Trade	**37**	**27**	**34**	**40**	21	**71**	**42**	**31**	**27**	**32**	**47**	24	**37**	**34**
Retail Trade	**92**	**31**	**68**	**105**	**76**	**94**	**67**	**81**	**46**	**102**	**51**	**66**	**35**	**71**
Hotels and Restaurants	13	16	15	6	**40**	5	13	15	15	2	24	5	6	9
Transport	11	9	4	10	11	18	13	12	8	22	10	9	**39**	**27**
Computer and Related Activites	**41**	4	**103**	**35**	4	5	3	3	3	**67**	6	1	17	15

As it can be readily appreciated, most countries satisfy their cutoff rule for both wholesale and retail trading, and light manufacturing. In general, heavy manufacturing, related to metals, wood, or machinery - capital goods industries - tends to have far fewer observations, and the only cells which satisfy our cutoff rule consistently belong to Brazil and Mexico. The only industries which never satisfy our cutoff rule are manufacturing of paper products, manufacturing of basic metals and recycling, while construction, sales of motor vehicles and transport only satisfy our rule for one or two countries. Overall, after pooling across years, we count with 126 observations for median productivity per industry for our whole sample.[6]

One of ideas implicit in the Lewis model is that the differences in productivity between the modern sector and the subsistence sector are as large as the differences in average productivity in a cross-section of countries: Indeed, the evidence reported in Vollrath (2009b) and Temple and Woessmann (2006), among others, is consistent with this. Much of the development thought in Latin America after Lewis embraced this idea and added that productivity differentials even across the modern sector could be as large as that between countries (Anibal Pinto, 1970). Table 2 presents evidence consistent with this view: at the industry level, differences in productivity within countries can be as large as differences between countries. A quick calculation reveals that the standard deviation of (average) country productivity is 14,610; while the standard deviation of productivity within countries is higher than this in 11 out of the 14 countries; thus, in many cases, the primary source of variation is within-country industry productivity as opposed to cross-country productivity differences.

3.2 Data from Workers

We obtain our worker level-data from social surveys previously harmonized by the Economic Commission of Latin America (ECLAC). These surveys contain information on labor market outcomes as well as demographic variables. We restrict our sample to urban workers who are wage earners and are aged form 25 to 65 years old. Only urban areas are selected to match our industry database, which only samples urban firms; the restrictions on age are to avoid life-cycle decisions

[6] Our choice of median instead of mean productivity is guided by the fact that the productivity distribution is highly skewed to the right and leptokurtic, which implies that the mean is a poor measure of the central tendency of the productivity distribution.

Table 2: Summary Statistic of Labor Productivity at the Industry Level

Country	Mean	Std. Dev.	$N_{Industries}$
Argentina	54,753	23,607	11
Bolivia	19,758	14,828	7
Brazil	46,088	45,244	16
Colombia	37,070	27,119	10
Costa Rica	29,818	20,709	5
Ecuador	53,305	28,906	8
El Salvador	24,801	12,970	8
Guatemala	23,576	23,747	11
Honduras	13,861	10,249	10
Mexico	22,995	8,485	13
Nicaragua	16,140	17,529	9
Paraguay	18,036	15,490	2
Peru	24,940	17,948	9
Uruguay	49,262	50,085	8

such as schooling and retirement, and the restrictions to wage earners is due to the fact that all of our industries are constructed with firms that employ more than one worker - excluding the productivity of the self-employed from our firm database. The decision to exclude the self employed makes our hypothesis harder to test, since it's commonly assumed in the models cited above that increases in the productivity of the modern sector increase the scale of production, increasing wages and drive part of the self-employed out of the subsistence sector to become wage workers. Thus, we are only observing the changes in the share of subsistence workers come from wage earners who switch from the subsistence to the modern sector, which are likely to be a smaller fraction than the overall share of the workforce who switches to the modern sector.[7].

All of the surveys contain data on the industry where the wage-earners are employed; we homogenize these data to ISIC Rev 3.1, and aggregate them at the 2-digit industry level in order to merge this data with our industry level data. We transform all wages from local currency units to U.S real dollars with 2009 as the base year, adjusted by Purchasing Power Parity. These changes to wage data make them fully comparable to our data on productivity, which are also valued at 2009 U.S real dollar adjusted by PPP. Table 3 shows some descriptive statistic regarding our sample of workers, which, as mentioned above, are only urban wage earners: as expected, there is a great deal of heterogeneity in schooling, age, the share of women in urban salaried work, the share of salaried workers, and inequality measured by the Gini coefficient. Additionally, these country present wide variation in other institutional features, with highly unionized and regulated labor markets in countries as Argentina and Uruguay, and other with low unionization levels and relatively deregulated labor markets, such as Colombia. For the sake of comparison, log labor productivity along with their mean and standard deviation are computed for all the economies, measured at the firm level (as opposed to the industry level, which was shown in table 2).

[7] To the best of my knowledge, there is not a single model á la Lewis which suggest that increases in the productivity of the modern sector might drive some wage-earners to switch into self-employment. On the contrary, the fact that urban wage-earners returned to agricultural-self employment in the light of increases in the productivity of the agricultural sector is one of the stylized facts that led to the development of the Harris-Todaro model.

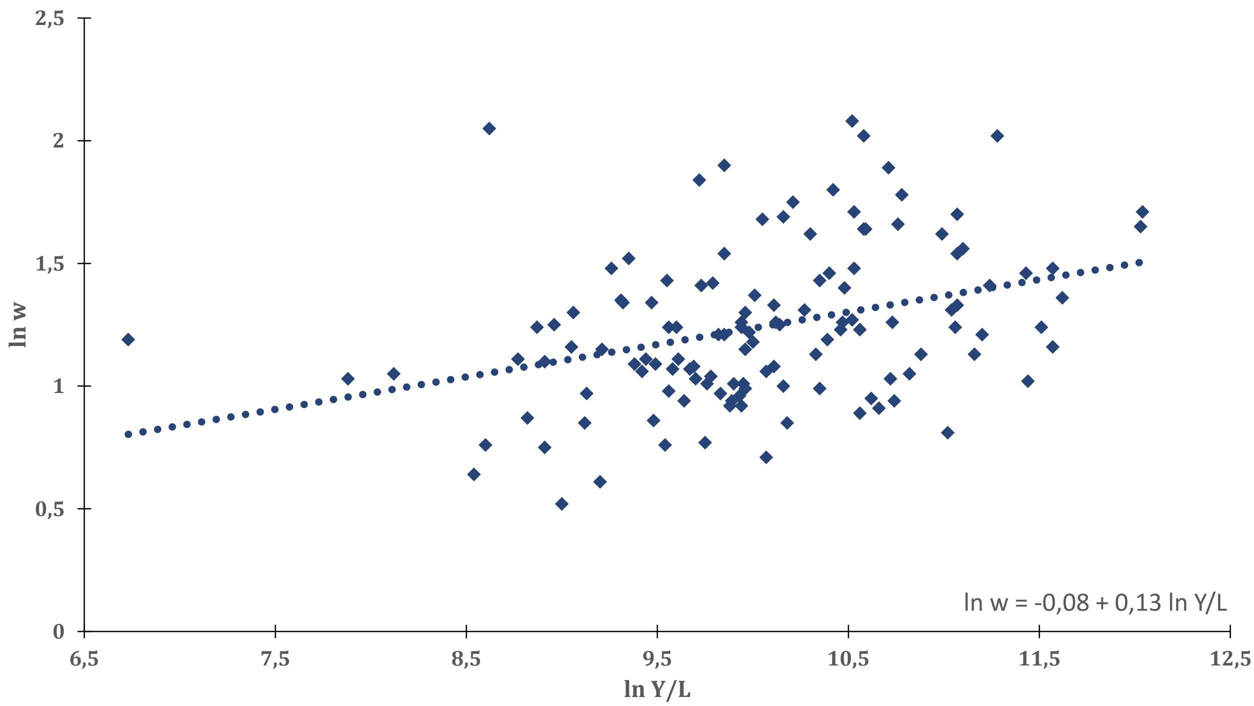
Table 3: Descriptive Statistics

Country	Schooling Mean	Schooling Sd	Age Mean	Age Sd	Woman	Salaried Workers	Wages Mean	Wages Gini	Productivity Mean	Productivity Std. Dev.	$N_{Workers}$
Argentina	12.06	3.99	37.62	12.57	0.38	0.69	6.90	0.40	10.90	1.04	117,614
Bolivia	11.86	4.43	34.29	12.21	0.33	0.49	5.31	0.45	9.76	1.41	4,445
Brazil	9.90	3.85	34.61	11.97	0.40	0.63	5.76	0.49	10.02	2.28	100,316
Colombia	11.04	4.25	35.03	11.53	0.43	0.44	4.74	0.46	10.41	1.27	230,274
Costa Rica	10.32	4.27	35.87	12.09	0.38	0.70	6.38	0.41	10.34	1.35	5,044
Ecuador	11.30	4.72	36.30	13.50	0.34	0.55	4.51	0.40	10.79	1.12	20,850
El Salvador	9.59	4.74	34.15	12.06	0.36	0.55	4.09	0.40	10.08	1.43	17,275
Guatemala	7.75	4.80	32.06	13.00	0.31	0.52	4.85	0.45	9.58	1.19	6,072
Honduras	9.02	4.54	32.59	12.40	0.39	0.55	4.41	0.46	9.55	1.46	14,293
Mexico	10.51	4.44	35.51	12.71	0.36	0.71	4.50	0.48	9.86	1.26	37,459
Nicaragua	9.17	4.54	33.07	12.09	0.35	0.53	3.03	0.42	9.26	1.63	8,337
Paraguay	10.88	4.07	33.48	12.54	0.33	0.54	5.16	0.43	10.49	2.05	5,141
Peru	12.27	4.22	35.50	12.71	0.35	0.47	4.44	0.47	10.13	1.30	23,948
Uruguay	10.63	3.81	38.56	13.21	0.43	0.65	6.43	0.42	10.52	1.09	51,978

Our descriptive statistics show that the south cone countries - Argentina and Uruguay - have the oldest workforce in the region, the highest share of salaried workers, the lowest inequality along with the highest mean wages, and also count with some of the highest schooling. These economies also feature extensive union coverage and labor legislation. The Andes economies - Ecuador, Peru, Bolivia and Colombia - follow closely the level of schooling set by the south cone economies, have a slightly younger workforce, a substantially lower share of salaried workers and average wage, along with a higher Gini - except for Ecuador, which has a Gini comparable to that of Argentina and Uruguay. These economies also feature much lower union coverage and more deregulated labor markets. Finally, the central American economies, such as El Salvador, Guatemala, Honduras and Nicaragua, feature the lowest schooling and the youngest workforce in the region, while showing a higher share of salaried workers than the Andes countries, some of the lowest average wages in our sample, and heterogeneous levels of inequality - El Salvador and Nicaragua are as egalitarian as the south cone countries, while Guatemala and Honduras are as unequal as the Andes economies. These countries show the lowest level of labor market regulation and unionization. Mexico, Brazil, Paraguay and Costa Rica do not fit neatly into these patterns; perhaps the simplest way to describe Costa Rica is as a South Cone country in the middle of Central America. It should also be noted that the share of woman participating in the salaried labor force also does not follow a clear pattern with the clusters of countries we defined: The highest share of women appear in both Colombia and Uruguay, while the lowest are in Guatemala, Bolivia and Paraguay.

One implication of the theories discussed above was that industries with higher productivity would contain both smaller share of subsistence workers, and given the wage gap between the modern sector and the subsistence sector, higher wages. Figure 1 plots median log productivity per industry against average log wages per industry, and, as expected, there is a strong positive correlation between productivity and wages at the industry level; a back-of-the-envelope calculation suggest a elasticity of 0.13. In the developed world, most explanations on the link between higher productivity and higher wages rely on monopolistic labor markets, some degree of bargaining power from workers, or some degree of imperfect monitoring of worker effort (see Card, Cardoso, Heining and Kline, 2018, for a discussion in the developed world). However, the model of the previous section, which features perfect competition, perfect observability of effort, and no bargaining power on the side of workers, can perfectly rationalize this stylized fact in the developing world. The econometric methods to which we now turn to explain are a powerful tool to discriminate between these alternative explanations and dual economy effects.

Figure 1: Average log wages against average log productivity at the industry level



Source: Authors own calculation based on data obtained from WES and ECLAC.

4 Estimation Strategy

4.1 Finite Mixture Models

As it was made clear in the discussion above, to operationalize the concepts of modern and subsistence sector is not a straightforward task. Some authors identify the modern sector with manufacturing and the subsistence sector with agriculture; other authors use the formal/informal divide, which has become more popular in recent time; however, the definition of informality is a highly controversial topic: some definitions stress the absence of social security, others of a written employment contract which guarantees compliance with labor regulations, and other define all small firms as informal. Given this plethora of definitions, we adopt an agnostic approach: we treat both sectors as latent variables and define them in terms of the outcomes they generate for workers, in particular by their wage determination mechanism. In other words, we allow the conditional mean function governing each sector to differ in terms of their constant and slope. This approach has a long tradition, dating back to Dickens and Lang (1985, 1987).

Besides having the advantage of letting the data speak for themselves on the absence or existence of dual labor markets, an additional restriction that motivates this approach is the absence of data on both agricultural firms and the formality status of workers in different countries. Moreover, given that labor regulations and what is considered formal employment varies substantially from country to country, to impose an ad-hoc common criteria seems even more problematic.

Thus, to obtain a data-dependent measure of the subsistence and modern sector, a Finite Mixture Model (FMM) will be used to estimate jointly the allocation of workers across sectors and the wage equations in each sector. A FMM model is a flexible parametric model composed of n equations which describe the continuous variable of interest, and a multinomial logit equation which assigns agents to each regime. If both sectors posses distinct wage determination mechanisms, a FMM will recognize this through different earning equations with different parameters for the same set of covariates; if their mechanisms are not distinct, then the heavy overparametrization implies by the FMM model will not compensate the additional goodness of fit. A two regime FMM model is a more general case of earlier "switching" regression models used by Dickens and Lang (1985), and of the "essential heterogeneity" model used by Heckman, Urzua and Vytlacil (2006). Formally, the conditional distribution of wages is modeled by FMM as:

$$p(\ln w|x) = \sum_{j=1}^{m} \alpha_j^m(x)\phi(y, \mu_j^m(x), \sigma_j^m) \tag{7}$$

Where m is the number of regimes, α_j^m are the mixing probabilities, and $\phi(y, \mu_j^m(x), \sigma_j^m)$ is the Gaussian distribution j with mean μ and variance σ^2. x denote de vector of covariates, which enter in both the conditional mean of each regime, and in the mixing probabilities or the selection equation. The set of covariates can differ across equations.

Finite mixture models have a number of attractive properties; in particular, finite mixtures of Gaussian distributions can approximate a wide class of nonparametric functions under reasonably weak regularity conditions (Norets, 2010); and can be estimated in a relatively straightforward fashion, at least when compared to other nonlinear models.

For our purposes, $m = 2$ and the mixing probabilities are modelled as a logit; thus, the model is described by the following set of equations:

$$\ln w_{1ijkt} = X_{1ijkt}\beta_1 + \alpha_{1k} + \alpha_{1t} + \varepsilon_{1ijkt} \tag{8}$$

$$\ln w_{2ijkt} = X_{2ijkt}\beta_2 + \alpha_{2k} + \alpha_{2t} + \varepsilon_{2ijkt} \tag{9}$$

$$z^*_{ijkt} = Z_{ijkt}\Gamma + \gamma(Y/L)_{jk} + \alpha_{2k} + \alpha_{2t} + u_{ijkt} \tag{10}$$

Where $\ln w_{1ijkt}$ denotes the log wage of worker i in industry k, country j and time t for regime 1; X_{1ijkt} is a vector of covariates that varies at the individual level, including a constant, α_{1k}, α_{1t} are fixed effects across countries and time, and ε_{1ijkt} corresponds to the error term. Likewise, z^*_{ijkt} is a latent variable that assigns workers across sectors; the only regressor of interest which varies across industries and countries is $(Y/L)_{jk}$, our measure of productivity, which enters only in the selection equation. The first two equations describe the conditional mean functions governing wages in each sector; the third equation describes how workers self-select themselves into each sector. This equation is crucial in our framework, since it will allow to test us whether productivity (γ) influences how workers allocate themselves between sectors. Absent this equation, we do not have a way of testing the predictions of the Vollrath model, nor do we have a way of explicitly measuring how workers sort themselves into different sectors based on their observables or unobservables.[8]

Despite being a flexible and agnostic way in which to both describe the wage distribution and test the existence of dual labor markets, our empirical strategy is not devoid of problems. First, despite the fact that we model the allocation of workers into each sector, we do not model participation decisions into the labor market, which might be potentially important (Günther and Launov, 2012). However, the likelihood of the model is rendered nearly intractable by the inclusion of this additional dimension, which we choose to exclude from the analysis. Secondly, suppose workers sort themselves into the modern or subsistence sector according to their ability, which is unobserved to the econometrician. If workers with higher ability sort themselves into the high productivity, modern sector, then the error term will be positively correlated with average productivity, producing inconsistent and biasing upward estimates of the effect of productivity on the share of subsistence workers. Nevertheless, since our measure of productivity is constructed at the industry level, only industry-level sorting is present, which might be less severe than firm-level sorting. While one could potentially solve this endogeneity problem with instrumental variables, the joint treatment of non-linearity and endogenity is still on its theoretical infancy; thus, we choose to treat our estimates as the upper bound of the true causal effect.

Given there is no closed form solution for our estimator, numerical techniques must be employed to estimate the model. Appendix 1 details the computational details of the estimation process. Here it suffices to say that our estimation procedure is recursive: First, we concentrate out the share of subsistence workers by treating them as an exogenous parameter, obtaining estimates for the wage equations for both sectors. Then, we take the previous estimates of the wage equations as initial values and estimate jointly the wage equations and the selection equation.

5 Results

5.1 Number of Regimes

A first and elemental assumption of the dual economy models reviewed earlier is that labor markets should be composed of two distinct sectors; if the data was generated by a competitive labor market without dualism, then there isn't any point in trying to test the predictions discussed above. A simple way to test this hypothesis is to compute the Bayesian Information Criteria (BIC) of an FMM model with only one regime, which is equal to running a simple Mincer-type regression; and then compute the BIC of a model with two regimes. If a dual economy model generated the data, then we would expect the goodness of fit to heavily outweigh the additional number of parameters. An alternative approach could be to simply test a Likelihood Ratio test of the two regime model against a single regime model, however, the standard regularity conditions needed for conventional

[8] Strictly speaking, it's not possible to distinguish whether workers self-select or are excluded from the modern sector with our specification; if the reader prefers, the selection equation can be interpreted as describing what sort of characteristics are correlated with being selected from a queue of workers which attempt to get a job in the modern sector. For a discussion of different ways to test the self-selection vs the exclusion hypothesis, see the literature cited at the introduction.

asymptotics break down in the case of FMM models; thus, the Likelihood ratio statistic converges to a non-standard distribution, which requires burdensome numerical methods to compute the test statistic. In contrast, computing the BIC is numerically simple and straightforward, since it only requires knowing the number of parameters and the goodness of fit of the model.[9]

Table 4 shows the result of this exercise. The first column (M1) shows a baseline Mincer specification where log wages are regressed against schooling, age and age squared, gender, and a set of country and time fixed effects. The second column (M2) shows my preferred specification, which differentiates schooling according to its tertiary years and primary/secondary years[10], and allows the return to schooling to vary across gender. This preferred specification is the product of a specification search where models with different interactions across the baseline specification were estimated and the BIC computed for each one of them; among them, this was the model with the lowest BIC.[11]

Table 4: Testing Integrated versus Dual Labor Markets

Regimes	M1	M2
1	238,023	234,377
2	213,574	211,919

The results clearly show a dramatic improvement in terms of the BIC when a two regime model is estimated. Despite containing more than the double number of parameters, the increased in goodness of fit decreases the BIC dramatically. Furthermore, as it can be appreciated in the table, the decrease in the BIC is substantially more pronounced by allowing two regimes than by allowing more covariates or specifying a a functional form with more interactions. Thus, it seems that a model generated by two distinction wage determination mechanisms, an essential ingredient of many dual economy models, finds ample support in the data. In what follows, we use our preferred specification (M2 with two regimes) to examine the theoretical predictions of the models outlined above.

5.2 Wage and Selection Equations

Table 5 shows the results for our preferred specification. The first column shows a simple OLS regression of log hourly wages on our measure of productivity and the relevant set of covariates, while the three next columns shows the results for our finite mixture model, which includes the two wage equations for each sector and the selection equation. We specify the selection equation in a somewhat simpler way than the wage equations, without adding interactions or differentiating between tertiary schooling years from other schooling years. Adding a more complicated selection equation does not alter our results; however, it makes estimation substantially longer and takes up more computing power, which is chosen to report results from this simpler specification.

[9] The breakdown of the conventional regularity conditions is due to the fact that a parameter is not identified under the null hypothesis; see Hansen (1999) for a general discussion of non-linear models and bootstrap methods to solve this issue. A discussion fo the advantages of the BIC criteria can be found in Steel & Raftery (2010).

[10] We use CINE, the classification system of UNESCO to homogenize tertiary schooling across countries.

[11] The specifications included models where gender interacted with country dummies and where schooling (without differentiating between tertiary and others) was allowed to have a different slope for each country, among others. Further results are available upon request.

Table 5: Regression Results

	OLS	FMM			
Variables		Regime 1	Regime 2	Logit	Logit 2
$\ln Y/L$	-0.021			-0.003	-0.028
	(0.022)			(0.017)	(0.040)
Basic Schooling	0.058***	0.029***	0.087***		
	(0.002)	(0.003)	(0.005)		
Basic Schooling * Women	0.001	-0.001	0.024***		
	(0.003)	(0.003)	(0.001)		
Tertiary Schooling	0.141***	0.072***	0.148***		
	(0.008)	(0.015)	(0.010)		
Tertiary Schooling * Women	-0.003	0.001	0.026***		
	(0.005)	(0.004)	(0.007)		
Schooling				-0.034***	-0.060
				(0.004)	(0.042)
Schooling * $\ln Y/L$					0.003
					(0.004)
Age	0.040***	0.017***	0.064***	-0.008**	-0.008**
	(0.003)	(0.002)	(0.005)	(0.003)	(0.003)
Age Squared	-0.001***	-0.001***	-0.001***	-0.001*	
	(0.001)	(0.001)	(0.001)	(0.001)	
Women	-0.224***	-0.076*	-0.640***	0.077***	0.076***
	(0.038)	(0.041)	(0.075)	(0.006)	(0.0203)
Constant	0.227***	0.988***	-0.954***		
	(0.133)	(0.085)	(0.184)		
σ		0.30	0.74		
Obs	133,093				

Clustered Standard Errors in Parenthesis
Includes a country and year fixed effect
The base category on the selection equation is Regime 1.
The selection equation shows average marginal effects.

A comparison between OLS and FMM estimates reveals the immense heterogeneity present in the data. Regime 1 could be classified as the subsistence sector, since it exhibits low returns to both basic and tertiary schooling, experience, and is relatively homogeneous, in the sense that wage dispersion is less than half of the dispersion in the modern sector. Additionally, women and men share essentially the same returns in this sector, and the wage penalty for women is roughly 10% of the wage penalty present in the modern sector.

The modern sector exhibits the exact opposite characteristics. Besides having rates of return to schooling and experience which exceed greatly those found in the subsistence sector, perhaps the most striking feature of this sector is its wage dispersion, which is reflected both in its standard deviation and on the huge wage penalty which women experience, despite having higher returns to schooling than men. The selection equation shows, unsurprisingly, that higher schooling and age increase the probability of being in the modern sector, while being a women decreases such probability. While the subsistence sector does not exhibit flat age or education profiles, as it has been sometimes found in previous studies (Dickens and Lang, 1985; 1993), it's safe to say these characteristics correspond tightly to the usual description of the subsistence sector.

The central regressors of interest, median industry productivity, shows an economically small and statistically insignificant negative effect on the probability of switching to the subsistence sector. Two central candidates which could explain this small effect of productivity on the share of subsistence workers **within the literature we have reviewed above** are the following: first, consumers could have hierarchical needs over goods produced in the subsistence sector, which is equivalent to assuming that they have near-zero elasticity of substitution between them (Eswaran and Kotwal, 1993); second, jobs in the modern sector require formal education in order to access to them (Proto, 2007). The first hypothesis seems unlikely at the disaggregated level implied by our data: it requires that consumers have a very low elasticity of substitution over beer produced by a modern-sector firm and wine produced by a subsistence-sector firm. While there is some evidence that, even between narrowly defined industries, goods produced in the modern sector have higher quality (Porta and Shleifer, 2008), which can be interpreted as evidence of imperfect substitution, it seems implausible to assume that their substitution elasticity is zero. The second hypothesis seems to have more promise; after all, the selection equation in column 4 does shows that schooling is a powerful predictor of belonging to the modern sector. A natural way to test this directly is by adding an interaction term between education and median productivity in the selection equation: if the effect of productivity on switching to the modern sector is mediated by the schooling of a worker, we should expect this term to be positive and statistically significant. Column 5 shows, surprisingly, that the point estimates on this interaction term indicate the opposite: when a worker has higher schooling, the impact of productivity on the probability to switch into the modern sector is *lower*, not higher. Thus, explanations based on the schooling requirement of the modern sector also seem unsatisfactory.

To check the robustness of this result, two additional exercises were carried out. First, the same model is estimated while adding industry fixed effects on both the selection and wage equations. These fixed effects should absorb all other time-invariant industry characteristics which affect the allocation of workers across sectors; for example, there could be idiosyncratic probabilities of being audited which vary across sectors, which play a key role explaining the size of the subsistence sector in formal/informal models of dual economies (Ulyssea, 2018; Meghir et. al, 2015). As discussed above, if the probability of getting audited increases in a given industry, more registered firms hiring informal workers will get caught evading labor regulations, which will decrease the share of subsistence workers and rise average industry productivity. Thus, these omitted fixed effects could be positively correlated to industry productivity, biasing upwards the true effect of average productivity on the share of subsistence workers. Table 6 shows that adding fixed effects drops the average marginal effect to a negative value, however, the confidence intervals are so wide that a formal hypothesis of equality between the two would not reject the null.

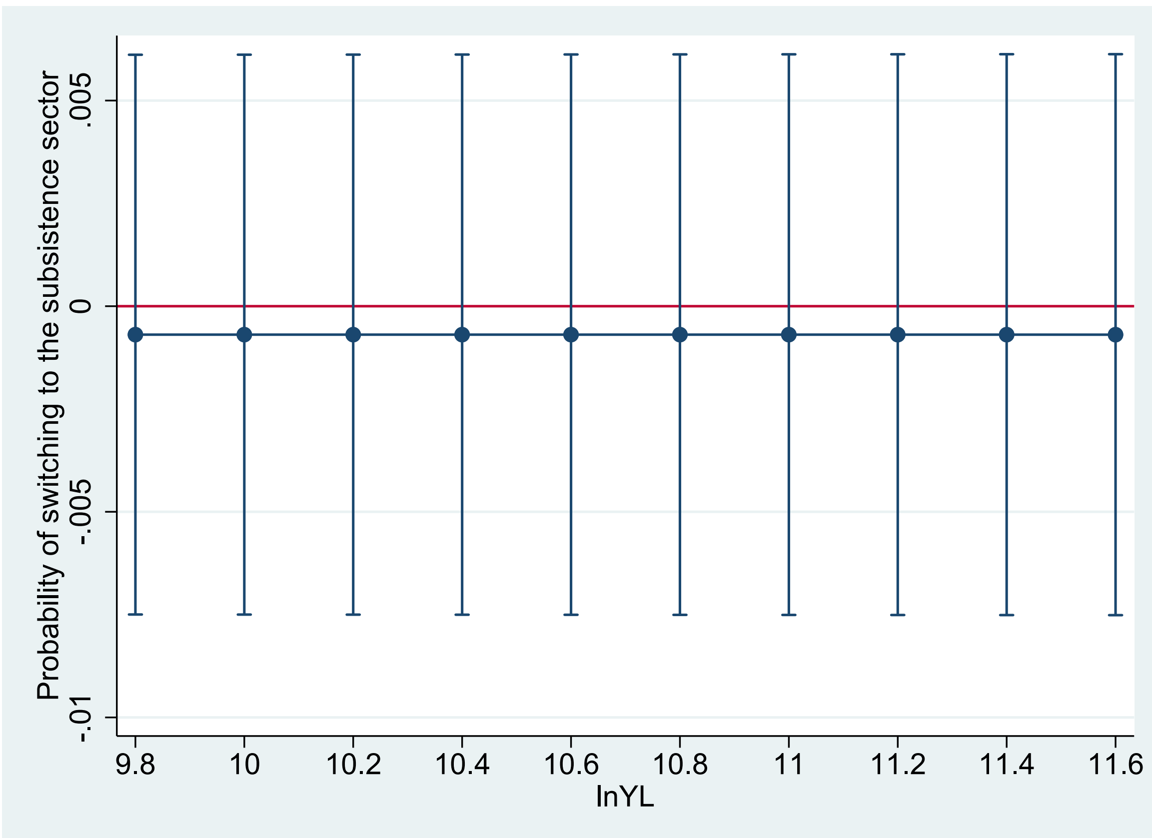
The second robustness exercise we carry out is the following: Given that in logit models, marginal effects are non-linear, we plot in figure 2 the marginal effect of productivity on the probability of switching to the subsistence over a wide range of the support of the productivity distribution, while holding the values of other covariates constant. This exercise shows that the marginal effect, surprisingly, is linear. This implies that a, at face value, taking a worker from the 25 percentile of the productivity distribution (10) and locating him at the 75 percentile of the distribution (11.2), decreases the probability that he belongs to the traditional sector by a mere 0.42%, which is negligi-

Table 6: Average Marginal Effect of median productivity on the probability of switching to the subsistence sector

Measure of Productivity	Baseline Model	Baseline Model + Fixed Effects per Industry
Log Sales per Worker	-0.0035	0.0058
	(0.0036)	(0.0153)

ble in economics terms, specially when compared to the effects of other covariates such as schooling.

Figure 2: Marginal effect of productivity on the probability of switching to the subsistence sector



Overall, while these results show that there is ample evidence to support the idea that Latin American countries seem to be characterized by a dual labor market, the proposition that productivity differences across the economy should influence the allocation of workers between sectors seem to be at best poorly supported by the data; in particular, increases in the productivity of the modern sector do not decrease the share of subsistence workers. Thus, differential growth rates of productivity between the modern and the subsistence sector are not interesting candidates to explain the size of the subsistence sector in a cross-section of industries in a wide set of developing countries.

5.3 The share of subsistence workers

Given the parameter estimates, an attractive feature of FMM models is that they allow the researcher to construct measures of the share of observations which belong to each regime. This can be done in two ways: for each observation, the posterior probability of belonging to each of the regimes can be computed, or if the researcher prefers so, the predicted regime to which the observation belongs can also be computed. In this section, posterior probabilities are computed for a number of observables which are of interest, such as year, country, schooling level, and different stages of the life cycle. This will allow to characterize richly the subsistence sector in a wide part of the developed world.

A natural way to asses the goodness of fit of our measure of the subsistence sector is to compare its size to the size of the informal labor force provided by the International Labor Organization. Given that most researchers commonly operationalize the subsistence sector as the informal sector, we would expect the measure of the subsistence sector estimated by the FMM model to be closely correlated to the informal sector. Table 7 shows the result of this exercise: Column 1 shows the fraction of wage earners which as a percentage of the total labor force, for our subsample of interest. Column 3 shows what fraction of this labor force is estimated to belong to the subsistence sector, by computing the posterior probability of the subsistence sector for each country. Column 3 shows the fraction of the labor force which works as self-employed, which most analysts regard as part of the informal sector. Finally, column 4 shows a measure of the subsistence sector as the sum of the self employed and the fraction of wage earners which belong to the subsistence sector. For a large subse of countries, although not all of them, this measure compares reasonably with the measure of informality employed by the ILO, which is shown on the final column. Additionally, it should be noted that the measure constructed using the FMM model is readily comparable across countries, while the ILO itself recognizes that the definition of the informal sector is kept deliberately vague in order to allow national agencies to compute their preferred measure of informality; thus, their measure is hardly comparable across countries.[12]

A few interesting facts stand out when comparing the size of the subsistence sector with the measure of informality provided by the ILO. First, the measure of the subsistence sector reveals that countries which have a small fraction of wage earners in the subsistence sector, such as Bolivia and Peru, have very large informal sectors. This discrepancy is accounted entirely by the percentage of self employed workers, which is among the highest in this countries. Second, in some cases, the size of the subsistence sector is substantially larger than that of the informal sector measured by the ILO, particularly in the case of Brazil, Costa Rica, Ecuador and El Salvador. A possible interpretation of this discrepancy, is that many *de jure* formal wage earners act as *de facto* informal wage earners. There is some evidence in the literature to support this interpretation; for example, Meghir et. al (2012) show that up to 8% of formal wage earners in Brazil are not paid the minimum wage, which constitutes one of the key labor regulations that formal firms most comply with; thus, even though these workers appear formally as part of the modern sector, their wage determination mechanism is more akin to the subsistence sector.

An interesting implication of the most dual economy models we have reviewed is that the share of the subsistence sector is counter cyclical. In Vollrath's model, productivity shocks to either the modern or the subsistence sector expand output and contract the share of subsistence workers. Table 8 shows the size of the subsistence and modern sector before and after the crisis; given that

[12] https://www.ilo.org/ilostat-files/Documents/description_IFL_EN.pdf

Table 7: Comparing the subsistence and the informal sector

Country	Wage Earners (2)	Subsistence Wage Earners (3)	Self Employed (4)	(2)*(3) + (4)	Informal ILO
Argentina	60,7	42,0	30,0	55,5	-
Bolivia*	27,1	35,4	57,1	66,5	71,8
Brazil**	58,6	54,2	32,0	63,6	36,9
Colombia	60,9	65,3	34,6	74,2	-
Costa Rica	34,7	61,9	53,9	75,4	43,2
Ecuador	63,2	62,5	29,8	69,0	31,5
El Salvador	35,5	71,6	51,1	76,7	21,6
Guatemala	42,3	36,4	46,5	61,8	66,8
Honduras**	37,0	61,9	50,7	73,6	73,4
Mexico**	62,3	35,6	19,5	41,9	53,9
Nicaragua	42,9	58,8	52,4	77,7	75,0
Paraguay**	34,4	58,6	52,4	72,7	64,4
Peru**	30,0	28,6	57,1	65,8	68,8
Uruguay**	66,5	26,7	23,7	41,6	33,2

Source: Author's own calculations based on ECLAC, WES and ILOSTAT databases. ILO measure taken in 2015, unless otherwise noted. *=2009, **=2013

pre-crisis years are usually 2006-2007, this was the peak of an expansion for most of the countries in the sample. For 7 out of the 11 countries for which we have data on pre and post crisis years, the share of subsistence workers expanded significantly; thus, broadly speaking, it seems to subsistence sector behaves in a counter cyclical fashion. This findings are consistent with previous work by Bosch and Maloney (2008) and Bosch and Esteban-Prestel (2012) for Brazil and Mexico.

Another interesting way in which we can use our posterior probabilities is to characterize the age and schooling profiles in each sector. Table 9 and Figure 3 show the respective probability distributions. Two features stand out: First, the percentage of workers in the subsistence sector goes monotonically down with age, with a prominent 60% at the beginning of the life cycle and only 30% at the end of the life cycle. Second, Figure 3 reveals that schooling is a crucial predictor of the share of subsistence workers, with the share of subsistence workers going from 70% for workers with no schooling to 20% for workers who have some degree of tertiary schooling. Its also interesting to note that workers who have over 20 years of schooling experience a spike in the probability of belonging to the subsistence sector, by given the sparsity of the data for these workers, this increase is very imprecisely estimated. These two pieces of evidence indicate that workers belonging to the modern sector count with substantially more human capital, measured as work experience or schooling, a stylized fact absent from the models we have presented in the previous sections.

Finally, I average posterior probabilities by industry, which sheds light on which productive sectors concentrate more subsistence workers and which concentrate higher shares of modern workers. These results are shown in table 10. The results are remarkably similar with those obtained by Dickens and Lang (1985, 1993) for the U.S: Light manufacturing, mainly apparel and leather, along with the service sector, with the exception of wholesale trade, concentrate a higher percentage of subsistence sector; in contrast, heavy manufacturing, specially publishing prints, chemicals, mineral

Table 8: Probability distribution before and after the Financial Crisis

	Pre Crisis		Post Crisis	
Country	Subsistence	Modern	Subsistence	Modern
Argentina	0.29	0.71	0.42	0.58
Bolivia	0.28	0.72	0.35	0.65
Brazil	-	-	0.54	0.46
Colombia	0.66	0.34	0.65	0.35
Costa Rica	-	-	0.62	0.38
Ecuador	0.46	0.54	0.62	0.38
El Salvador	0.62	0.38	0.72	0.28
Guatemala	0.36	0.64	-	-
Honduras	0.48	0.52	0.62	0.38
Mexico	0.27	0.73	0.36	0.64
Nicaragua	0.43	0.57	0.59	0.41
Paraguay	0.62	0.38	0.59	0.41
Peru	0.31	0.69	0.29	0.71
Uruguay	0.25	0.75	0.27	0.73

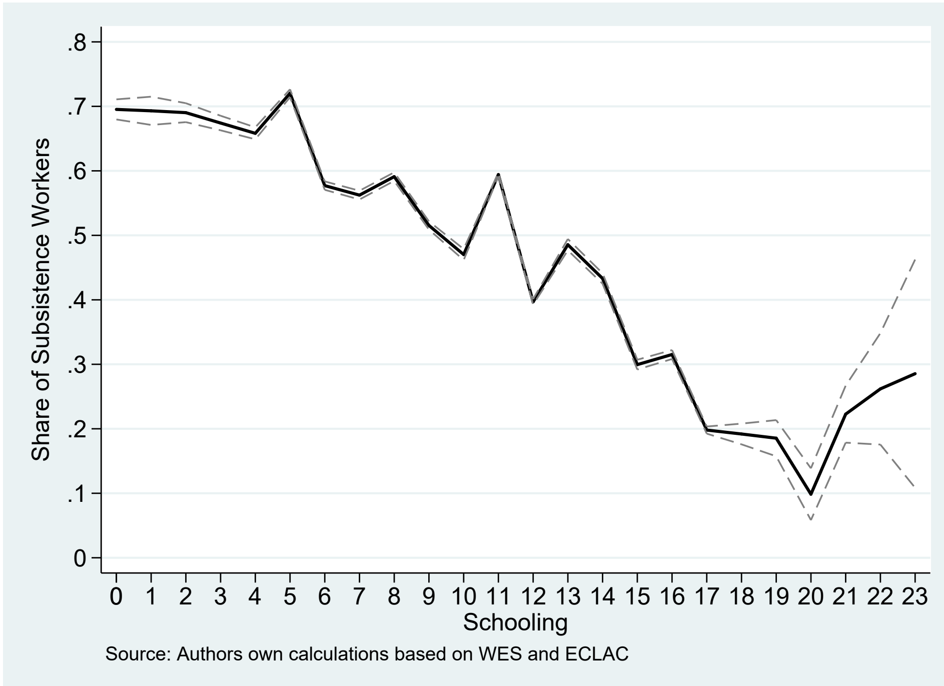
Table 9: Share of subsistence workers during the life cycle

Age	Subsistence	Modern
25 a 30	0.59	0.41
31 a 35	0.55	0.45
36 a 40	0.52	0.48
41 a 45	0.47	0.53
46 a 50	0.42	0.58
51 a 55	0.36	0.64
56 a 60	0.31	0.69
Mean	**0.51**	**0.49**

products and manufacturing of machinery and equipment - which are prototypical capital goods industries - concentrate a large share of modern sector workers[13]. Given the small magnitude of our productivity parameters in the selection equation, and the fact that this is our covariate which varies by industry, the size of the subsistence sector likely reflects the age and schooling composition of the labor force in each industry.

[13]Transport shows such an anomalous behavior because we aggregated all forms of transport - land, sea and air - in a single industry in order to have enough data to reliable estimate the productivity of this sector. As such, it's highly heterogeneous.

Figure 3: The share of subsistence workers against schooling



5.4 Average Treatment Effect

In the original Lewis model, there is a uniform wage rate set by subsistence workers in both sectors.[14] The Vollrath model presented in the previous section allows wages to differ on a per-worker basis, given that workers in the modern sector expend a higher share of their time in wage work, but the hourly wage is also equal across sectors. However, there is a large tradition in dual economy models starting with Harris and Todaro (1970) which posits that the wage in the modern sector is higher than the wage in the subsistence sector. A natural way to measure the wage differences between sectors is to conceptualize an assignment to the modern sector as a treatment and compute the Average Treatment Effect of switching from the subsistence to the modern sector, since a Finite Mixture Model with two regimes is equivalent to the essential heterogeneity model of Heckman, Urzua and Vytlacil (2006). In this context, the ATE is computed as follows:

$$E(y^1 - y^0) = \bar{X}(\beta^1 - \beta^0) \tag{11}$$

Where, in this case, the treatment is switching a worker from the subsistence sector to the modern sector, and the outcome of interest $(y^1 - y^0)$ is the log wage difference. As it can be seen, the ATE is obtained by subtracting the vector-valued coefficients of the modern regime with those from the subsistence regime $((\beta^1 - \beta^0))$, and multiplying them by matrix-valued average values of relevant covariates $\bar{X}$. We compute this ATE for each country and for the whole sample; whenever we compute these ATE's at the country level, we average the values of covariates at the country level. The results are presented in table 10.

[14]This is a crucial assumption in the model; if wages in the modern sector were to rise, the profit share would be squeezed along the steady growth path, which would then prevent the concomitant increase in the savings rate, capital accumulation, and growth

Table 10: Probability Distribution by Industry

Industry	Subsistence	Modern
Manufacturing of Food, Beverage and Tobbaco*	0.52	0.48
Manufacturing of Textile	0.54	0.46
Manufacturing of Wearing Apparel	0.64	0.36
Manufacturing of Leather Items	0.66	0.34
Sawmill and Planning of Woord	0.50	0.50
Publishing, Printing and Reproduction of Recorded Media	0.26	0.74
Manufacturing of Coke and Chemica	0.39	0.61
Manufacturing of Rubber and Plastic Productos	0.53	0.47
Manufacturing of Mineral Products	0.41	0.59
Manufacturing of Fabricated Metal Products	0.54	0.46
Manufacturing of Machinery and Equipment	0.36	0.64
Manufacturing of other Equipment**	0.42	0.58
Manufacturing of Transport Equipment	0.35	0.65
Manufacturing of Furniture	0.49	0.51
Construction	0.57	0.43
Sales of Motor Vehicles	0.52	0.48
Wholesale Trade	0.47	0.43
Retail Trade	0.57	0.43
Hotels and Restaurants	0.64	0.36
Transport	0.20	0.80
Computer and related activities	0.34	0.66

Notes: * Uses together ISIC industry codes 15 and 16. ** Uses together ISIC industry codes 30, 31, 32 and 33, which correspond to manufacturing of accounting, electrical, telecommunication and medical machinery

As it can be seen, the ATE for the whole sample, whether we weight it or not by the respective sample weights, is a substantial 7.1%-8.7%. Nevertheless, the ATE's for individual countries are very heterogeneous: The upper bound is composed by Brazil, Uruguay and Peru, which have ATE's of 34.6%, 21.6% and 19.0%, and the lower bounds are negative, which correspond to Argentina, Bolivia, Honduras and Mexico. This somewhat surprising result is driven by the large negative penalty that women experience in the modern sector; in all countries, ATE's are positive for men.[15] As such, the large wage penalties for women outweigh the small wage gains for men. Nevertheless, the ATE is positive, and for 6 out of 14 countries the wage gains associated with switching to the modern sector exceed 10%. Thus, the evidence seems mostly consistent with large wage differentials between sectors which are causally explained by the effect of the modern sector.

5.5 Wage Densities

One of the interesting implication of the Vollrath model reviewed above regards the wage gap between sectors; namely, that the wage gap should diminish as the productivity of the modern sector increases. There are several ways to try to test this

[15]Results available upon request.

Table 11: Average Treatment Effect by Country

	ATE
Argentina	-21.0%
Bolivia	-5.9%
Brazil	34.6%
Colombia	4.2%
Costa Rica	16.9%
Ecuador	10.5%
El Salvador	10.9%
Guatemala	9.5%
Honduras	-1.8%
Mexico	-3.7%
Nicaragua	4.0%
Paraguay	1.1%
Peru	19.0%
Uruguay	21.6%
Average	7.1%
W Average	8.7%

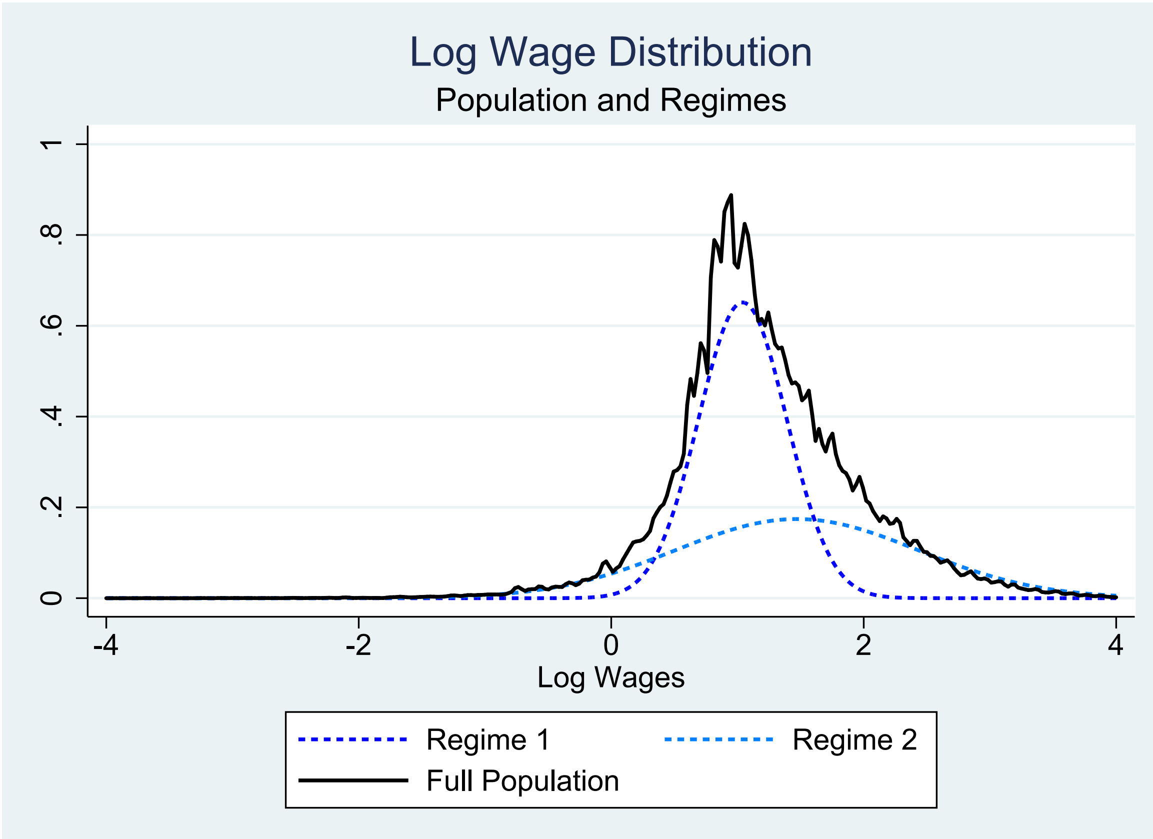
prediction; the first one could be to measure the correlation between the median productivity and the ATE of each sector. While certainly of interest,the recent literature on dualism has established that while the modern sector pays a positive differential relative to the subsistence sector, the wage distributions of both sectors overlap substantially (Meghir et. al, 2015; Ulyssea, 2018; Allen et. al, 2018). This is evident in our data, too: as Figure 4 shows, the distribution of the subsistence sector (regime 1) has a lower mean wage than the modern sector, but the overlap between both distributions is substantial. Even more, in contrast with Ullysea (2018) for Brazil and Bobba et. al (2018) for Mexico, the modern sector tends to dominate the left end of the support, which is a surprising result.[16]

In light of this observation, we can follow in the spirit of Allen et. al (2009) and report the maximum Kolmogorov-Smirnoff (KS) statistic, which formally tests whether two distributions of data are identical, but which can be used as a device to measure the degree of convergence between the wage distributions of the subsistence and the modern sector for each industry. Intuitively, as the KS statistic gets smaller, the two distributions become more similar; in the limit, both distributions become equal and the null hypothesis of equality cannot be rejected at any conventional statistical level. **Thus, the gradual convergence in wage distributions across sectors could be examined by measuring the correlation between the productivity of the modern sector and the value of the KS statistics; a negative correlation would provide support to the model, while a positive correlation would reject one of its key implications.**

Table 12 shows that the convergence predictions of the Vollrath model, at the industry level, are

[16] Recall that the mean observed wage in regime i is equal to $E[y^i|D_i = 1] = X\beta^i + E[u_i|D_i]$, that is, it includes both the observed components of the wage equations, and the unobserved components which arise from the selection equation.

Figure 4: Wage densities for each regime and the whole population



Notes: The density for each regime is calculated using a Gaussian distribution. The density for the whole population is estimated non-parametrically using an epanechnikov kernel and a bandwidth of 0.02

rejected by the data. The point estimates on the correlation coefficient imply that industries with higher median productivity also show a higher divergence between the wage distribution of the subsistence and the modern sector; however, this weak positive correlation is not statistically significant at any conventional level. Quite surprisingly, the same occurs with the level of schooling: industries with higher average schooling experience higher divergence between the wage distributions of both sectors. These findings are inconsistent with those of Allen et. al (2018), which find that the opposite holds for a narrow set of manufacturing industries in India. We do confirm, however, the finding that industries with higher shares of subsistence workers increase the divergence between the wage distributions of both sectors.

Table 12: Correlation between KS statistic and selected variables

Variable	Correlation	P-value
Share of Subsistence Workers	0.16	0.07
$\ln Y/L$	0.10	0.29
Average Schooling	0.09	0.31

6 Conclusion

Dual economy models make a rich set of empirical predictions regarding the impact of productivity growth on the existence and size of the subsistence sector in developing economies. In particular, they suggest that productivity growth in the modern sector tends to shrink the subsistence sector, as well as to decrease the wage gap between the modern and the subsistence sector. This paper attempts to take such predictions to the data by using a set of 14 Latin American countries which count with data on labor market outcomes and industry-wide productivity on registered, urban firms. Overall, our results show that productivity in the modern sector has a negligible economic and statistical effect on the share of subsistence workers, and increases in productivity of the modern sector are not correlated with measures of convergence in the wage distributions of the modern and subsistence workers. Furthermore, some explanations in the theoretical literature as to why increases in productivity in the modern sector might not have any impact on the share of subsistence workers - such as hierarchical preferences or low-educated wealth-constrained individuals in the subsistence sector - are hard to rationalize in the light of our results.

However, the paper does find that a dual-labor market model substantially improves the fit of the data with regards to a single labor market; the share of subsistence workers is counter-cyclical, as the model predicts; the degree of convergence in the wage distribution is smaller whenever the share of subsistence workers is larger, worker in the modern sector earn a substantial wage premium, and the share of subsistence workers are characterized by young and low skilled workers. Overall, these features of the data suggest that our estimation strategy does pick up the salient features of a dual economy, as opposed to some other mechanism such as compensatory differences or heterogeneity in cognitive ability. Furthermore, given the large wage gains associated with switching to the modern sector and the positive correlation between the wage gap among sectors and the share of subsistence workers, reducing the share of subsistence workers remains an indispensable objective for development countries. **However, if productivity growth does not play the role envisioned by Lewis and others in progressively reducing the subsistence sector until its disappearance, then the crucial question is what drives the reduction in the subsistence sector in developing countries. Future research should be devoted to answering this question.**

Bibliography

Albrecht, J., Navarro, L., & Vroman, S. (2009). The effects of labour market policies in an economy with an informal sector. The Economic Journal, 119(539), 1105-1129.

Allen, J., Nataraj, S., & Schipper, T. C. (2018). Strict duality and overlapping productivity distributions between formal and informal firms.

Almeida, R., & Carneiro, P. (2012). Enforcement of labor regulation and informality. American Economic Journal: Applied Economics, 4(3), 64-89.

Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: an application to production functions. Econometric reviews, 19(3), 321-340.

Bosch, M., & Esteban-Pretel, J. (2012). Job creation and job destruction in the presence of informal markets. Journal of Development Economics, 98(2), 270-286.

Bosch, M., & Maloney, W. (2008). Cyclical movements in unemployment and informality in developing countries. The World Bank.

Busso, M., Madrigal, L., & Pagés, C. (2013). Productivity and resource misallocation in Latin America. The BE Journal of Macroeconomics, 13(1), 903-932.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. Journal of Labor Economics, 36(S1), S13-S70.

Contreras, D., Gillmore, R., & Puentes, E. (2017). Self-Employment and Queues for Wage Work: Evidence from Chile. Journal of International Development, 29(4), 473-499.

Eswaran, M., & Kotwal, A. (1993). A theory of real wage growth in LDCs. Journal of Development Economics, 42(2), 243-269.

Fields, G. S. (2009) Segmented Labor Market Models in Developing Countries. In The Oxford Handbook of Philosophy of Economics.

Günther, I., & Launov, A. (2012). Informal employment in developing countries: Opportunity or last resort?. Journal of development economics, 97(1), 88-98.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. The American economic review, 60(1), 126-142.

Heckman, J. J., Urzua, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. The Review of Economics and Statistics, 88(3), 389-432.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. The Review of Economic Studies, 70(2), 317-341.

Magnac, T. (1991). Segmented or competitive labor markets. Econometrica: journal of the Econometric Society, 165-187.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. The World Bank Economic Review, 13(2), 275-302.

Meghir, C., Narita, R., & Robin, J. M. (2015). Wages and informality in developing countries. American Economic Review, 105(4), 1509-46.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. The review of economic studies, 61(3), 397-415.

Norets, A. (2010). Approximation of conditional densities by smooth mixtures of regressions. The Annals of statistics, 38(3), 1733-1766.

La Porta, R., & Shleifer, A. (2014). Informality and development. Journal of Economic Perspectives, 28(3), 109-26.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. The manchester school, 22(2), 139-191.

Pinto, A. (1970). Naturaleza e implicaciones de la" heterogeneidad estructural" de la América Latina. El trimestre económico, 37(145 (1), 83-100.

Porta, R. L., & Shleifer, A. (2008). The unofficial economy and economic development (No. w14520). National Bureau of Economic Research.

Radchenko, N. (2017). Informal employment in developing economies: Multiple heterogeneity. The Journal of Development Studies, 53(4), 495-513.

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. Journal of monetary economics, 55(2), 234-250.

Steele, R. J., & Raftery, A. E. (2010). Performance of Bayesian model selection criteria for Gaussian mixture models. Frontiers of statistical decision making and bayesian analysis, 2, 113-130.

Temple, J. R. (2005). Growth and wage inequality in a dual economy. Bulletin of Economic Research, 57(2), 145-169.

Temple, J., & Wößmann, L. (2006). Dualism and cross-country growth regressions. Journal of Economic growth, 11(3), 187-228.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. American Economic Review, 108(8), 2015-47.

Vollrath, D. (2009a). The dual economy in long-run development. Journal of Economic Growth, 14(4), 287.

Vollrath, D. (2009b). How important are dual economy effects for aggregate productivity?. Journal of development economics, 88(2), 325-334.

Appendix 1: Estimating the model

FMM estimators do not have closed form solutions, which means we must employ numerical techniques to obtain the estimator of interest. The model is estimated sequentially: In the first step, we treat the share of observations in each regime, $(\alpha, 1-\alpha)$ as exogenous parameters, i.e, the selection equation is treated only as a constant. We obtain initial estimates of the share of observations using the expectations maximization (EM) algorithm. Given the model is linear conditioning on $(\alpha, 1-\alpha)$, each wage equation is simply estimated by OLS in this step. After the (EM) algorithm converges, we count with initial estimates of both the share of observations in each regime and the parameters of the wage equations.

The second step involves using as initial values the estimated parameters of the previous step, and jointly estimating the selection and wage equations using full-information maximum likelihood. A standard Newton-Raphson algorithm is employed, and convergence was achieved quickly. In practice, the parameters from the wage equations changed little from step 1 to step 2, so the second stage essentially pins down the selection equation only. Convergence was relatively fast overall, and despite randomizing the initial values on the first stage convergence on the same parameters was always achieved, which makes us confident that we are looking at a global and not a local maximum.

Appendix 2: Figures and Tables

Table 13: Distribution of GDP by Sector - Year 2009. Selected Latin American Countries

Country	Manufacturing	Construction	Transport	Wholesale and Retail Trade; Hotel	Total
Argentina	15.7	4.4	6	14.1	40.2
Bolivia	11.4	2.6	8.7	9.9	32.5
Brasil	12.5	5.4	7.4	10.4	35.7
Chile	11.4	6.0	8.0	9.1	34.6
Colombia	13.0	7.5	6.2	11.5	38.2
Costa Rica	14.3	6.2	6.6	12.2	39.3
Ecuador	13.5	9.4	7.6	12.3	42.8
El Salvador	18.7	3.9	8.0	19.9	50.6
Guatemala	18.5	4.7	7.9	19.3	50.4
Honduras	16.4	6.4	6.5	16.5	45.7
Mexico	15.0	8.2	8.3	17.0	48.5
Nicaragua	13.7	4.4	5.0	13.7	36.9
Paraguay	11.9	6.2	6.1	17.1	41.3
Peru	15.3	5.8	7.7	13.6	42.4
Uruguay	14.2	7.8	6.6	13.2	41.8

Source: CEPALSTAT.

Table 14: Employement Distribution by sector - Latin America, 2009

Country	Manufacturing	Construction	Transport	Wholesale and Retail Trade; Hotels	Total
Argentina	13.3	9.0	8.1	22.9	53.3
Bolivia	15.1	10.2	9.4	29.1	63.8
Brasil	15.4	8.3	5.5	24.7	53.9
Chile	14.2	9.8	8.2	21.1	53.3
Colombia	14.8	6.3	9.9	30.4	61.4
Costa Rica	12.8	6.4	8.5	27.0	54.7
Ecuador	13.6	7.3	7.7	32.6	61.2
El Salvador	17.6	5.1	5.3	33.9	61.9
Guatemala*	19.9	7.9	4.1	29.4	61.3
Honduras	18.0	8.6	5.6	31.2	63.4
Mexico	16.9	8.2	6.2	21.6	52.9
Nicaragua	15.9	5.8	6.3	30.9	58.9
Paraguay	13.7	7.5	6.3	32.5	60.0
Peru	13.0	6.5	9.3	32.0	60.8
Uruguay	13.6	7.2	5.9	22.9	49.6

Notes: *=2006. Source: Cepalstat.