

Human Capital and Structural Change*

Tommaso Porzio[†]

University of California, San Diego

and

Gabriella Santangelo[‡]

University of Cambridge

[*Work in Progress*]

[*Click here for most current version*](#)

August 26, 2017

Abstract

What explains labor reallocation out of agriculture? We propose an accounting framework that leverages observable variation across birth cohorts to study the role of human capital accumulation. We model a dynamic overlapping generations economy where heterogeneous individuals choose whether to stay in or move out of agriculture, subject to mobility frictions. The model shows analytically that labor reallocation within- and across-cohorts pins down the relative role of human capital vs. sectoral prices/productivities in labor reallocation. We apply the framework to micro data from 52 countries. We document novel empirical patterns on labor reallocation by cohort and use them, through the lens of our model, to discipline the size of mobility frictions and show two results: (i) human capital explains one third of labor reallocation, on average; but (ii) it has a minor role in explaining why some countries have faster reallocation than others. Furthermore, we use years of schooling as a direct measure of human capital to validate our main approach and we exploit a large-scale school construction program in Indonesia as a natural experiment to study the effects of an exogenous increase in human capital. We show that the program led to labor reallocation out of agriculture.

**FIRST DRAFT*: 10th of August 2017.

We thank for helpful suggestions and criticisms Luigi Guiso, Hugo Hopenhyan, Joe Kaboski, David Lagakos, Tim Lee, Ben Moll, Guillermo Ordonez, Michael Peters, and conference and seminar participants at the 2017 IEA World Congress, the Conference on Structural Change at the University of Edinburgh, and the Einaudi Institute. Xiao Ma provided excellent research assistance.

[†]9500 Gilman Drive, La Jolla, CA 92093, email: tporzio@ucsd.edu.

[‡]email: gabriella.santangelo@econ.cam.ac.uk.

1 Introduction

Two features have robustly characterized the growth experiences of most countries of the world: reallocation of labor out of agriculture and increase in human capital, at least as measured by years of schooling. Take Brazil, for example. In 1960 more than half of its male population was employed in agriculture, and the average working age male spent a dismal two years in school. By 2010, those numbers changed significantly: less than fifteen percent of working age males were employed in agriculture, and they spent on average eight years in school.

In this paper, we ask whether the increase in human capital contributes to the reallocation of labor out of agriculture, and if so, by how much. To answer this question, we develop an accounting framework that leverages novel empirical evidence on labor reallocation within and across birth cohorts, which we document using micro-level data from 52 countries. We conclude that, for our set of countries, human capital accounts on average for approximately one third of total reallocation out of agriculture, but, at the same time, it has only a minor role in explaining why some countries had faster reallocation than others.

Our research question is motivated by a recent literature that documents – using micro level data from countries of all income level – that high skilled workers sort out of agriculture, and interprets this empirical pattern as evidence that non-agricultural production is more skill intensive.¹ It is then natural to conjecture that, as the average level of human capital grows over time, an increasing number of workers becomes endowed with skills that are valued more by the non-agricultural sector, thus triggering labor reallocation out of agriculture. The contribution of this paper is to bear this hypothesis to empirical scrutiny.

Our main challenge is to devise an empirical approach that allows us to disentangle the role of human capital relative to the traditional view on structural change, which treats labor as an homogenous input and argues that the decline in relative agricultural productivity is the driving force behind labor reallocation.²

The development accounting literature offers one possible path to directly measure the contribution of human capital. This literature, starting from the seminal work of Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999), has used Mincerian returns to school to create human capital stocks from educational attainments. While appealing, this approach entails a number of drawbacks that makes it unattractive in our context. In particular, it would require to use the cross-sectional relationship between years of schooling and individual probability of agricultural employment to forecast the effect, in the time-series, of an increase in average number years of schooling on the share of the population employed in agriculture. Unfortunately, through this methodology, we may either overestimate or underestimate the role of human capital. The cross-sectional estimates may provide an overestimate of the effect of an aggregate change in schooling

¹See Gollin et al. (2014), Young (2013), and Buera et al. (2015); for example. See also Caselli and Coleman II (2001), which was – to our knowledge – the first to argue that skills may be more useful out of agriculture. We review this literature and others below.

²The literature has traditionally distinguished between supply side (e.g. Baumol (1967)) or demand side (e.g. Gollin et al. (2002)) explanations for the decrease in agricultural revenue productivity. As we will explain, for our purpose the two are isomorphic.

since they capture omitted variables – such as unobserved skills – which are unlikely to vary over time. At the same time, they could also underestimate the role of human by restricting our focus only to changes in schooling years, thus omitting other sources of human capital increase over time, such as health improvements or better school quality.

In order to circumvent these concerns, in this paper we follow a new path. We develop a framework to indirectly back-out the role of human capital from individual choices, rather than from observed quantities of schooling and prices, as in the Mincer approach. We build on one simple insight: following a birth cohort of adult individuals over time keeps mostly constant the level of human capital,³ thus allowing to identify the effect of aggregate changes, such as those in relative agricultural prices/productivity; instead, comparing different cohorts at one point in time, which are exposed to the same aggregate conditions, can pin down the role of human capital.

As an example of our approach, consider Figure 1a. It plots the agricultural employment in Brazil, from 1960 to 2010, separately for different ten-years birth cohorts. Following a cohort of individuals as they age, we see that fewer and fewer of them work in agriculture: this suggests that the returns from agricultural production has decreased over time, thus pushing workers to reallocate. At the same time, if we compare across cohorts in any single year, we see that the younger ones have a smaller fraction of the workers in agriculture, suggesting that – due to their higher human capital – they have a stronger comparative advantage towards non-agriculture. Over time, higher human capital cohorts enter the labor market and replace lower human capital ones, thus contributing to an aggregate decrease in agricultural employment. Not all countries look like Brazil. As a comparison, in Figure 1b we plot a similar graph for India: in this case, within cohorts reallocation over time is mostly muted, while we still observe sizable across-cohort reallocation. In this paper, we systematically document this heterogeneity across countries, and exploit it to draw general conclusions on the role of human capital.

More in general, however, the simple insight on the map between reallocation by cohort and human capital might fail, since cohorts possibly differ for aspects other than their human capital. In particular, younger cohorts may face lower mobility frictions to change sector. A core contribution of this paper is to develop a simple model to analytically characterize how the reallocation within and across cohorts can be used to back out the role of human capital, taking into account both mobility frictions and general equilibrium interactions across-cohorts. Equipped with the model, we use micro-level data for 52 countries to systematically document new facts on reallocation by cohort, along the lines of what just described for Brazil and India. We then use data and theory together to back out the role of human capital and to show our two main results: (i) human capital explains, on average, approximately one third of labor reallocation; but (ii) it does not explain why some countries have faster reallocation than others. We also show that mobility frictions play a minor role, which is instrumental in using reallocation by cohort to derive the main results. Finally, we turn back to schooling, and compare our approach with a direct measurement of human capital stocks using schooling. The two approaches are complementary.

³Of course, this insight holds as long as most of the human capital is acquired through schooling, and in the early phases of life. Otherwise, we'll underestimate the role of human capital.

Overview. Next we provide a detailed overview. The paper is organized in four sections.

In Section 2 we present a dynamic overlapping generation model. The model provides an accounting framework to leverage labor reallocation by cohort to quantify the relative role of human capital in aggregate labor reallocation out of agriculture. The general features of the model are the following: time is discrete; a finite number of cohorts are alive at each point in time; each period a cohort of individuals is born and enters the labor market and one dies; individuals are heterogenous in their human capital both within and across cohorts; average human capital grows across cohorts at a constant rate; there are two sectors: agriculture and non-agriculture; agriculture uses land and labor to produce; non-agriculture uses human capital; agricultural relative price and productivity, which give the relative revenue productivity, are exogenous and decrease at a constant rate; individuals choose, in each period in which they are alive, in which sector to work subject to two mobility frictions: (i) a one time fixed cost to be paid to change sector, and (ii) an iceberg-type cost that reduces the monetary value of non-agricultural wage each period; markets are complete and competitive.

We analytically characterize the equilibrium, which displays sorting across sectors, both within and across cohorts, and labor reallocation out of agriculture. We provide three sets of theoretical results. First, we show that the rate of labor reallocation out of agriculture is constant, does not depend on either mobility friction, and is increasing in the growth rate of relative non-agricultural revenue productivity, and in the growth rate of human capital across cohorts. This result highlights the two core forces that lead to labor reallocation out of agriculture: (i) decrease relative agricultural price and productivity; (ii) increase in human capital. Second, we decompose the rate of labor reallocation in two components: a year effect, which captures the rate at which a given cohort reallocates out of agriculture; and a cohort effect, which captures the gap in agricultural employment across cohorts. And we show that, absent mobility frictions and ignoring general equilibrium, the year effect pins down the relative contribution of prices/productivity, while the cohort effect pins down the relative contribution of human capital accumulation. This special case corresponds to our simple insight on the role of reallocation within and across cohorts. However, in general, mobility frictions and general equilibrium complicate the analysis, by tying together year and cohort effects. The theory provides further useful guidance: we show that only fixed costs are relevant to determine labor reallocation by cohort, and that old workers are more likely to be constrained by a fixed cost, since they have fewer periods to depreciate it over. As a result, comparison of labor reallocation rates across age groups informs us on the size of the frictions. Third, in search of additional ways to discipline the size of the mobility frictions, we describe how they affect the agricultural wage gap. We show that the wage gap for movers out of agriculture can be used to identify iceberg-type frictions – such as amenity costs that have to be paid each periods. However, we also show that fixed-cost-type frictions, which are the more relevant ones for our purpose, since they affect the map between cohorts effects and human capital, cannot be inferred from wages. In fact, a small wage gap for movers out of agriculture is consistent with an arbitrarily large fixed cost.

In Section 3 we turn to the data. In this section we describe three novel empirical results,

leaving their interpretation through the lens of the model to Section 4. We use micro level data available from IPUMS international for 52 countries around the world. The data are either censuses or large sample labor force surveys representative of the population. For each country, we have at least two repeated cross-sections distant 10 years apart. On average, for each country there are 28 years from the oldest to the most recent cross-section. For some countries, such as Brazil, our data cover half a century of labor reallocation. The 52 countries cover roughly $\frac{2}{3}$ of the world population, and span five continents and the income distribution from Liberia to the United States.

For each country, we compute year and cohort effects as defined in the model. On average, year and cohort effects are of similar size, thus giving our *first empirical result*: the across-cohorts-reallocation accounts on average for approximately half of the overall labor reallocation out of agriculture. We then study the year and cohort effects across countries as a function of their rate of labor reallocation. The year effects are strongly correlated with the rate of labor reallocation, while the cohort effects are more similar across countries, and less strongly correlated with the overall rate of reallocation. Formally, we decompose the cross-country variance of the rate of labor reallocation and show that differences in the across-cohorts reallocation explains approximately one quarter of it, which is our *second empirical result*. Finally, we compute for each country, the year effect separately for individuals of different ages, and show our *third empirical result*: individuals of different ages have similar year effects.

Section 4 uses theory and data together to decompose, in an accounting sense, the relative roles of human capital and prices/productivity for labor reallocation out of agriculture. First, we show that, without taking a stand on the size of the frictions or the strength of general equilibrium, we are able to provide an upper bound to the relative contribution of human capital: the first empirical results above directly implies that human capital accounts for at most half of average labor reallocation. That is, absent human capital accumulation the average rate of labor reallocation out of agriculture could be as low as just half the observed one. Second, we use our theoretical results to infer a value for the mobility frictions, and thus be able to provide a point estimate for the role of human capital in partial equilibrium. To back out the size of the friction, we follow two different approaches. First, we use the prediction on reallocation rates by age: the third empirical result above is not consistent with sizable mobility frictions, which would imply that old individuals reallocate at slower rates. Second, we show that, under the assumption that the mobility friction is constant within a subset of countries, the second empirical result above is not consistent with sizable mobility frictions either: mobility frictions tie together the cohort and year effects, and thus would predict that countries with faster labor reallocation have *both* larger year and cohort effects. The data reject this hypothesis as well. Therefore, both approaches are not consistent with a large role for frictions. In fact, we show that, in partial equilibrium, human capital accumulation accounts for 37 – 56% of average labor reallocation, depending on the chosen estimate for the frictions. Third, an elementary calibration exercise suggests that the general equilibrium forces are unlikely to overturn the quantitative results: taking into account general equilibrium reduces the role of human capital accumulation to 19 – 52%. Using our favorite

estimates for the size of the friction and for the GE calibration, we obtain that human capital accounts for approximately one third of labor reallocation out of agriculture. Fourth and last, we focus, rather than on the average rate of labor reallocation, on its variance across countries. We show that, while human capital explains a sizable fraction of labor reallocation *on average*, it has at most a minor role in explaining why some countries have faster rate of labor reallocation than others.

Finally, in *Section 5* we turn back to the usual approach of the literature and exploit schooling as a direct measure of human capital. Using schooling is useful for two purposes. First, it allows us to validate the main empirical approach, by showing that our model-inferred human capital stocks align well with direct measurement through schooling, both in levels and in changes across cohorts. Second, using schooling enables us to provide a proof of concept on the possibility that policies designed to increase human capital can trigger labor reallocation out of agriculture. We follow closely Duflo (2001) and exploit the INPRES school construction program in Indonesia as an exogenous variation in schooling. We show that the exogenous increase in schooling decreased the agricultural employment of the affected cohorts.

Related Literature. We draw upon insights from a rich literature on related topics. We here discuss our contribution relative to the most closely related articles.

Our work builds on the seminal work of Caselli and Coleman II (2001) and Acemoglu and Guerrieri (2008). To our knowledge, Caselli and Coleman II (2001) first recognized the interaction between aggregate changes in human capital and structural change. It noticed that non-agriculture is more skill-intensive than agriculture, and, therefore, an aggregate increase in schooling raises the relative supply of non-agricultural workers. It focused on the effect of human capital increase on relative wages, and argued that taking it into account is necessary to match the path of relative agricultural wages. Acemoglu and Guerrieri (2008) formalized the general insight that changes in the relative prices of inputs may lead to structural transformation if sectors vary in the intensity with which they use inputs. In its analysis, Acemoglu and Guerrieri (2008) considered capital and labor as the two inputs of interest.

We owe to these two papers the broad notion that human capital accumulation may be relevant in explaining reallocation out of agriculture. Relative to their work, our contribution is to provide an accounting framework and to use reallocation by cohorts to separately account for the role of human capital (or supply of agricultural labor) relative to the role of relative agriculture prices/productivity (or demand for agricultural labor). As discussed in the introduction, the recent literature that studies the cross-sectional allocation of heterogeneous workers to sectors motivates us to interpret human capital accumulation as a change in the relative supply of agricultural labor.⁴ At the same time, in our work we bundle together the traditional views of structural change that focus on demand or supply of agricultural *goods*, since they both similarly affect the relative revenue productivity of agriculture, hence the demand for agricultural *labor*.⁵

⁴See Lagakos and Waugh (2013), Young (2013), Gollin et al. (2014), Herrendorf and Schoellman (2015), Buera et al. (2015), Hicks et al. (2017), and Porzio (2017).

⁵Some relevant examples. For supply side theories based on changes in relative productivity: Baumol (1967),

With respect to this aim of separate the role of demand for and supply of labor as drivers of sectoral reallocation, our work is, in fact, mostly related to Lee and Wolpin (2006). Lee and Wolpin (2006) devised and structurally estimated a rich model to study the process of labor reallocation from manufacturing to services in the United States. We see our work as complementary to it, to the extent that we are interested in a similar question, but we tackle it from a different perspective. Specifically, our approach aims to impose the minimal structure to interpret the data, closer in spirit to the accounting literature.

Relative to the three papers above, we also depart in extending our empirical analysis to as many countries as we could gather data for, rather than focusing only on the United States. Using multiple countries allows us to provide additional ways to identify the size of the mobility frictions, which is an important component of our analysis.

More broadly, our work is related to a rich literature that studied the contribution of human capital to growth and development. This literature showed that the level of human capital is significantly correlated with consequent growth (See Nelson and Phelps (1966), Barro (1991), Mankiw et al. (1992)). However, the effects of changes in human capital stocks have been much more elusive (See Benhabib and Spiegel (1994) and Pritchett (2001)). Pritchett (2001) in particular, in a famous article, asked: “Where has all the education gone?”. In this respect, our work provides some encouraging answers: we show that growth of human capital stocks matters for explaining reallocation out of agriculture. Methodologically, we are in debt to the approach developed by the growth and developing accounting literature (See Jorgenson and Griliches (1967), Barro (1999), Hsieh and Klenow (2010), and more recently Gennaioli et al. (2013)). Relative to this literature, we show that observable variation across birth-cohorts can be used in an accounting framework, and we introduced a way to measure the role of human capital without having to rely on prices.

In terms of the purely empirical contribution of this paper – that is, documenting novel cross-country patterns in reallocation out of agriculture by cohort – our work relates to Kim and Topel (1995), Lee and Wolpin (2006), and Perez (2017), which document sectorial reallocation by cohort but limit their focus to, respectively, South Korea and United States and Argentina; and especially to Hobijn et al. (2017), which, in ongoing work, is also using the IPUMS dataset to document patterns on reallocation across sectors by cohorts, and use them to motivate a model linking demographic forces to structural change.

Our model combines elements and insights already presents in Matsuyama (1992), Lucas (2004), and Herrendorf and Schoellman (2017). To the best of our knowledge, we are the first to provide a tractable framework to analytically characterize reallocation within and across-cohorts in a context with general mobility frictions. Hsieh et al. (2016) also exploits year and cohort effects to calibrate a model of allocation of talent. It uses them to discipline the relative role, for the aggregate efficiency of the allocation of talent, of changes in frictions that affect human capital investment and frictions that distort the labor market. Relative to this paper, we focus on a simpler framework that allows us to consider fixed-cost-type frictions, which turn out to be

Ngai and Pissarides (2007). For demand side theory based on non-homothetic utility: Gollin et al. (2002), Echevarria (1997), Kongsamut et al. (2001), and more recently Boppart (2014) and Comin et al. (2015).

crucial to correctly identify the role of human capital.

Finally, our work relates to a growing literature that uses longitudinal wage data to reconsider the agricultural productivity gaps and that shows that these gaps are more consistent with sorting across-sectors than with large mobility frictions; (See Alvarez (2017), Herrendorf and Schoellman (2017), and Hicks et al. (2017)). We contribute to this literature in two ways: we provide a model that highlights when wage data can be informative on frictions; and we show, without relying on wage data, additional evidence corroborating the sorting explanation and casting doubts on the presence of large mobility frictions.

2 Model

In this section we introduce an elementary overlapping generation model of labor reallocation out of agriculture. We build the model to sharply characterize how observable information on labor reallocation within- and across-cohorts can be used to decompose, in an accounting sense, the relative contribution of human capital vs. prices/productivity in aggregate labor reallocation. We add to the model only the minimal structure necessary to analytically derive a map between human capital accumulation, mobility frictions, and reallocation within and across cohorts, which can be brought to the data. In this dimension, we follow the tradition of the accounting literature, although we need to impose more structure due to the nature of our exercise. It is also worthwhile to emphasize that, while we consider *one* particular micro-structure – in our opinion the simplest one – that generates a mapping between reallocation by cohorts, human capital, and prices/productivity, our empirical application hinges on the decomposition shown below in Proposition 2, and not on its specific micro-foundation. In fact, one appealing feature of our approach is that any micro-foundation that yields the same reduced form representation of Proposition 2, even if generated by different structural parameters, would give the same empirical results.

2.1 Environment

We next describe the economic environment. Time is discrete, markets are complete and competitive, and individuals have perfect foresight.

2.1.1 Demographics and Preferences

Each period a cohort, indexed by c , is born. A cohort is composed by a continuum of mass one of individuals. Individuals of cohort c enter into the labor market at time c and die after $N + 1$ periods; therefore, they work each period in $\{c, \dots, c + N\}$. Individuals face an increasing and non-satiated utility for an agricultural and a non-agricultural good. They have no disutility of labor.

2.1.2 Human Capital

Each individual is characterized by a pair (c, ε) , where c is the year her cohort is born and $\varepsilon \sim F$ is her idiosyncratic ability relative to the other members of her cohort. The human capital

of an individual (c, ε) is constant over time and given by

$$h(c, \varepsilon) = \kappa h_c^\gamma \varepsilon^{1-\gamma},$$

where κ is a scale constant, h_c captures the cohort- c specific human capital and $\gamma \geq 0$ is the elasticity of human capital with respect to the cohort human capital, and $1 - \gamma$ with respect to individual ability. It is simple to notice that when either $h_c = h$ for all c , or $\gamma = 0$, all cohorts have identically distributed human capital, when instead $\frac{h_{c+1}}{h_c} > 1$ and $\gamma > 0$, then, the distribution of human capital of cohort $c + 1$ first order stochastically dominates the one of cohort c . In fact, this latter is the relevant case, since we assume that human capital strictly increases across cohorts.

Assumption 1. *Human capital grows across cohorts at a constant rate*

$$\frac{h_{c+1}}{h_c} = g_h > 1.$$

Individual ability relative to the cohort is normalized on the interval $[0, 1]$ and distributed according to a Beta with parameters $(v, 1)$; v is a free parameter that captures how concentrated ability is. Additionally, we assume that the scaling constant κ is sufficiently large so that at least some individuals are in non-agriculture at each point in time. As will become evident, these assumptions on F and κ guarantee a balanced growth path in which the share of agricultural workers decreases at a constant rate, and in which a positive mass of workers from each cohort works in non-agriculture, as we observe in the data over the relevant time period.

2.1.3 Production and the Problem of the Firm

There are two sectors in the economy, we call them agriculture, indexed by A , and non-agriculture, indexed by M . Production of agricultural good requires land X and labor input $L_{A,t}$, while production of non-agricultural good only requires labor $L_{M,t}$. We assume that land is owned collectively by all individuals, who share the profits, and use them to finance consumption. Productivity in agriculture, $Z_{A,t}$, may differ from productivity in non-agriculture, $Z_{M,t}$. The relative price of agricultural goods, which is exogenous, is p_t . Production functions are Cobb-Douglas in each sector. Summing up, the revenue functions of agriculture and non-agriculture are given by

$$\begin{aligned} Y_{A,t} &= p_t Z_{A,t} X^\alpha L_{A,t}^{1-\alpha} \\ Y_{M,t} &= Z_{M,t} L_{M,t}. \end{aligned}$$

Few comments are in order. First, on the role of land. We introduce land in the production function in order to guarantee an interior solution in which a positive share of the population is in agriculture even as $p_t z_t \rightarrow 0$; moreover, as we will show, the presence of land generates general equilibrium interactions across cohorts. Second, it is useful at this stage to connect our setting to the literature on structural change. As we discussed in the literature review, the

traditional view on structural change distinguishes between supply or demand forces that lead to labor reallocation. In our context, this distinction is not relevant, since from the perspective of workers that choose whether to move out of agriculture the two are isomorphic, since both affect the relative agricultural wages. In fact, these aggregate forces of labor reallocation are here jointly captured by the growth rate of relative agricultural revenue productivity, given by

$$g_{pz,t} \equiv \frac{p_{t+1}z_{t+1}}{p_t z_t},$$

where $z_t \equiv \frac{Z_{A,t}}{Z_{M,t}}$. In this paper, we are being agnostic about the debate on demand vs supply forces, and in fact we are simply assuming an exogenous process for both prices and productivity. Specifically, we assume a constant decay of relative agricultural revenue productivity.

Assumption 2. *The rate of total factor revenue productivity (TFPR) is constant in both agriculture and non-agriculture, and the non-agriculture's one is bounded by the real interest rate: for all t*

$$\begin{aligned} \frac{p_{t+1}Z_{A,t+1}}{p_t Z_{A,t}} &= g_{pA} \\ \frac{Z_{M,t+1}}{Z_{M,t}} &= g_M \\ g_M &\leq \frac{1}{\beta}. \end{aligned}$$

Assumption 3. *Agricultural TFPR decreases relative to non-agricultural one, and at a sufficiently fast rate relative to human capital:*

$$\frac{p_{t+1}z_{t+1}}{p_t z_t} = g_{pz} \leq g_h^{-\frac{\alpha\gamma v}{1-\gamma}} < 1.$$

Assumption 3, as will become evident in the equilibrium characterization, is necessary and sufficient to have a non-negative rate of within-cohort labor reallocation out of agriculture; a feature that we observe in almost all countries in our sample.

Last, but crucial, we need to discuss how labor inputs in agriculture and non-agriculture are determined. Firms hire workers in a competitive market. Individuals, as we will explain next, take wages as given, and choose in which sector to work. We let $\omega_t(c, \varepsilon)$ be the occupational choice function, that is equal to 1 if individual (c, ε) at time t works in agriculture, and 0 otherwise. The two sectors differ to the extent that human capital only increases individual productivity in non-agriculture.

Assumption 4. *All individuals, irrespective of their human capital, are equally productive in agriculture, while human capital increases non-agricultural productivity. Specifically, for given*

occupational choice function ω_t , aggregate labor in agriculture and non-agriculture are

$$\begin{aligned} L_{A,t} &= \sum_{c=t-N}^t \int \omega_t(c, \varepsilon) dF(\varepsilon) \\ L_{M,t} &= \sum_{c=t-N}^t \int h(c, \varepsilon) (1 - \omega_t(c, \varepsilon)) dF(\varepsilon). \end{aligned}$$

The higher skill-sensitivity of non-agricultural is the reason why human capital accumulation triggers labor reallocation, hence it deserves some discussion. This assumption is consistent with sorting of high skilled workers to non-agriculture, as widely documented in the data (e.g. Gollin et al. (2014), Young (2013) Porzio (2017)). Moreover, it is consistent with the documented larger returns to skills in non-agriculture (see Herrendorf and Schoellman (2017)), and with patterns of mobility across sectors (see Hicks et al. (2017)).

2.1.4 Occupational Choice and Mobility Frictions

Finally, we describe the occupational choice problem of workers. Since we assume that markets are complete, and that there is no disutility of labor, then each individual (c, ε) chooses her occupation each period – $\{\omega_t\}_{t=c}^{N+c}$ – to maximize the present discounted value of her future income stream, taking as given the path of wages in agriculture – $\{w_{A,t}\}_{t=c}^{N+c}$ – and non-agriculture – $\{w_{M,t}(c, \varepsilon)\}_{t=c}^{N+c}$; and taking into account the mobility friction to change sector – $C_t(\omega_{t-1}, \omega_t, w_{A,t}, w_{M,t}(c, \varepsilon))$. That is, each individual (c, ε) solves

$$\begin{aligned} \max_{\{\omega_t\}_{t=c}^{N+c}} \sum_{t=c}^{N+c} \beta^{t-c} & \left(\underbrace{\omega_t w_{A,t} + (1 - \omega_t) w_{M,t}(c, \varepsilon)}_{\text{Income in period } t} - \underbrace{C_t(\omega_{t-1}, \omega_t, w_{A,t}, w_{M,t}(c, \varepsilon))}_{\text{Mobility Friction}} \right) \\ \text{s.t.} & \quad \omega_{c-1} = 1; \end{aligned}$$

where we are assuming that all individuals are born in agriculture, hence the constraint $\omega_{c-1} = 1$.

The mobility friction takes the following form

$$\begin{aligned} C_t(\omega_{t-1}, \omega_t, w_{A,t}, w_{M,t}) &= \underbrace{\mathbb{I}(\omega_t = 1) (i_t w_{M,t})}_{\text{Iceberg Cost if in M}} + \underbrace{\mathbb{I}(\omega_t < \omega_{t-1}) f_t w_{A,t}}_{\text{Fixed Cost if A to M}} \\ &+ \underbrace{\mathbb{I}(\omega_t > \omega_{t-1}) f_t w_{M,t}}_{\text{Fixed Cost if M to A}}, \end{aligned}$$

with an iceberg cost that reduces the monetary value of non-agricultural wage in each period, and a fixed cost to be paid to change sectors. The iceberg cost can be interpreted as an amenity cost – as in Lagakos et al. (2017b) – or as any other flow cost from leaving the agricultural sector, for example, generated by the exclusion from risk-sharing community – as in Munshi and

Rosenzweig (2016) and Morten (2016). The fixed cost can be interpreted as a one time mobility cost, which might be driven by the actual moving expenses, if a move is necessary to change sector, or by any other associated costs, such as retraining, idle time in between jobs, or even one time emotional/distress costs. Finally, we assume that the mobility frictions are constant over time, and are bounded above, respectively by \bar{i} and \bar{f} , which are explicit functions of the parameters – included in the appendix – that guarantee that at least some workers reallocate out of agriculture; while restrictive, this assumption is necessary for a balanced growth path.

Assumption 5. *Mobility frictions are constant over time: for all t*

$$\begin{aligned} i_t &= i \in [0, \bar{i}] \\ f_t &= f \in [0, \bar{f}]. \end{aligned}$$

2.2 Equilibrium

Next, we provide a definition of the equilibrium, or more specifically of a balanced reallocation path, and provide an overview of some of its main properties. Then, the few following sections will provide a formal characterization of the main results that we bring to the data.

Definition: Balanced Reallocation Path *A balanced reallocation path is given by a series $\{L_{A,t}, w_{A,t}, w_{M,t}(c, \varepsilon), \omega_t(c, \varepsilon)\}$ for all $c \in [N - t, t]\}_{t=0}^\infty$, such that, given a path of relative prices and productivity, $\{p_t Z_{A,t}, Z_{M,t}\}_{t=0}^\infty$, firms maximize profits taking wages as given, individuals choose optimally their occupation at each point in time taking wages as given, labor market clears in both agriculture and non-agriculture, and the aggregate agricultural labor decreases at a constant rate, $g_{L_A} = \frac{L_{A,t+1}}{L_{A,t}} < 1$.*

Along the reallocation path, since markets are competitive, wages are given by the marginal product of labor in either sector. In fact, a worker (c, ε) would earn the following wages if he works in agriculture or non-agriculture

$$\begin{aligned} w_{A,t} &= (1 - \alpha) p_t Z_{A,t} X^\alpha L_{A,t}^{-\alpha} \\ w_{M,t}(c, \varepsilon) &= Z_{M,t} h(c, \varepsilon), \end{aligned}$$

where notice that only one is actually observed in the data.

In order to understand the structure of the equilibrium, it is useful to first consider the frictionless case – i.e., when $i = 0$ and $f = 0$. In this case, an individual (c, ε) would move out of agriculture if and only if she earns a higher wage in non-agriculture; that is

$$\omega_t(c, \varepsilon) = 1 \Leftrightarrow w_{M,t}(c, \varepsilon) \geq w_{A,t}.$$

Substituting the equilibrium wages, we see that an individual (c, ε) moves out of agriculture as

soon as her human capital is sufficiently high with respect to relative agricultural productivity

$$h(c, \varepsilon) \geq \hat{h}_t = (1 - \alpha) p_t z_t X^\alpha L_{A,t}^{-\alpha}. \quad (1)$$

In fact, since skills are more useful in non-agriculture, high-skilled people have a comparative advantage there. As a result, in equilibrium, high-skilled people sort out of agriculture. Moreover, using the expression for $h(c, \varepsilon)$, we can see that there is *sorting both within- and across-cohorts*. Within any given cohort, the most skilled individuals are the ones that move out of agriculture. Across cohorts, the younger ones, that have higher cohort specific human capital, have a larger share of individuals that find it worthwhile to move out of agriculture. Over time, as $p_t z_t$ decreases – Assumption 3 – and as human capital increases – Assumption 1 – more and more people move out of agriculture, thus generating aggregate labor reallocation.

The distributional assumption on the idiosyncratic ability ε plays an important role in generating a constant rate of labor reallocation. For a given cohort c , we can use equation (1) to find an expression for the ability cutoff – $\hat{\varepsilon}_t(c)$ – that defines the marginal individual that moves out of agriculture at time t

$$\hat{\varepsilon}_t(c) = \left[(1 - \alpha) \kappa^{-1} p_t z_t X^\alpha L_{A,t}^{-\alpha} h_c^{-\gamma} \right]^{\frac{1}{1-\gamma}}.$$

The mass of workers from cohort c – $l_{A,t}(c)$ – in agriculture is then equal to

$$l_{A,t}(c) = F(\hat{\varepsilon}_t) \propto \left[p_t z_t X^\alpha L_{A,t}^{-\alpha} h_c^{-\gamma} \right]^{\frac{v}{1-\gamma}}, \quad (2)$$

and the reallocation out of agriculture for a given cohort is

$$\frac{l_{A,t+1}(c)}{l_{A,t}(c)} = g_{pz}^{\frac{v}{1-\gamma}} g_{L_A}^{-\alpha \frac{v}{1-\gamma}},$$

which is constant – as long as the aggregate g_{L_A} is constant as well, as we will prove. In doing this derivation, we used the fact that the CDF of a Beta($v, 1$) – over the relevant domain – is homothetic. The Beta($v, 1$), is the only distribution which satisfies this property.

Last, we can use equation (2) to notice that the ratio between agricultural employment at time t of cohort c and $c + 1$ is given by

$$\frac{l_{A,t}(c+1)}{l_{A,t}(c)} = \left(\frac{h_{c+1}}{h_c} \right)^{-\frac{\gamma v}{1-\gamma}} = g_h^{-\frac{\gamma v}{1-\gamma}} :$$

the faster human capital grows across cohorts, the bigger the difference in their agricultural employment. As we will show formally, differences across cohorts are informative about the growth rate of human capital.

Summing up, the aggregate constant reallocation rate hides substantial heterogeneity within and across cohorts. The assumption made, guarantee tractability despite the rich heterogeneity along two dimensions: age and ability.

Next, we discuss the role of mobility frictions. The iceberg cost i simply introduces a wedge between agricultural and non-agricultural wages, as such, it does not affect reallocation rates, but only the level of agricultural employment at each point in time. Instead, the fixed cost f is more consequential, since it may constraint *some* workers, but not others, from reallocating even though they would earn a higher wage out of agriculture. In particular, as we will show in details, relatively old individuals are more likely to be constrained. Intuitively, if a relatively young worker finds it worthwhile to move out of agriculture, then the fixed cost is unlikely to bind, since it is discounted over her whole future life-cycle. Instead, if a worker is still in agriculture when old, thus having only few periods left to work, then even a small fixed cost may trap her into agriculture. As we will show, the fixed cost implies that must divide cohorts into two groups, the constrained and the unconstrained, based on their age.

Next, we turn to the four main results – each expressed in one proposition – that characterize the equilibrium and that we are going to use in Section 4 to make inference about the size of the frictions and ultimately the role of human capital. All the proofs of the paper are included in the online Appendix.⁶

2.3 Aggregate Labor Reallocation

First, we characterize the aggregate rate of labor reallocation.

Proposition 1: Aggregate Rate of Labor Reallocation

The aggregate rate of labor reallocation, $g_{LA} \equiv \frac{L_{A,t+1}}{L_{A,t}}$ is given by

$$\log g_{LA} = \underbrace{\Theta \log g_{pz}}_{\text{Prices/Productivity}} - \underbrace{\gamma \Theta \log g_h}_{\text{Human Capital}}$$

where $\Theta \equiv \frac{v}{1-\gamma+\alpha v}$.

Proof. See Appendix. \square

It is useful to emphasize two features of g_L . First, labor reallocation out of agriculture – i.e. $\log g_{LA} < 0$ – is triggered by two distinct forces: (i) decay in the agricultural relative price or productivity, the usual forces emphasized by the structural change literature; (ii) increase in aggregate human capital through the entrance of more skilled cohorts in the labor market, the novel channel that we quantify in this paper. Second, mobility frictions are irrelevant for the rate of labor reallocation.

The first feature is at the core of the economic mechanism we are considering. As human capital increases, a larger share of the population has a comparative advantage in non-agriculture, which is more skill-intensive, and this leads to reallocation on aggregate. The second feature may seem surprising at first, since we might expect that mobility frictions “trap” people in agriculture, slowing down reallocation. The first part of the intuition is true: for given relative agricultural

⁶The online appendix is available at <https://sites.google.com/a/yale.edu/tommaso-porzio/research>.

price and productivity, the larger the mobility frictions, the larger the share of the population in agriculture. However, the level of agricultural employment is affected, but the reallocation rate is not, since mobility frictions affect each period similarly. Also, notice that we are assuming that frictions are sufficiently small as to generate positive reallocation. Trivially, if $f \rightarrow \infty$ or $i \rightarrow \infty$, there would be no reallocation. Proposition 1 thus shows that, in general, the reallocation is either zero, or does not depend on f and i . Our parametric restriction excludes the case in which reallocation is zero.

2.4 Labor Reallocation by Cohorts

The aggregate reallocation rate is not informative on the relative role of human capital and prices/productivity. We next show that, for given mobility frictions, labor reallocation within and across cohorts can be used to pin down the relative role of human capital.

As an intermediate step, we define year and cohort effects, first for each cohort c and then on average across all of them. The year effect captures the rate at which cohorts reallocate out of agriculture over their life-cycle – i.e. the within cohort reallocation. The cohort effect capture the reallocation across cohorts, relative to the aggregate trend – i.e. the across-cohort reallocation. By definition, the product of average year and cohort effects is equal to the aggregate rate of labor reallocation.

Definition: Year Effect. *The year effect for cohort c at time t is given by*

$$\psi_t(c) \equiv \frac{l_{A,t+1}(c)}{l_{A,t}(c)}.$$

The average year effect at time t is given by

$$\bar{\psi}_t \equiv \frac{\sum_{c=N-t+1}^t l_{A,t+1}(c)}{\sum_{c=N-t+1}^t l_{A,t}(c)}.$$

Definition: Cohort Effect *The cohort effect for cohort c at time t is given by*

$$\chi_t(c) = \bar{\psi}_t^{-1} \frac{l_{A,t+1}(c+1)}{l_{A,t}(c)}.$$

The average cohort effect at time t is given by

$$\bar{\chi}_t = \bar{\psi}_t^{-1} \frac{\sum_{c=N-t}^t l_{A,t+1}(c+1)}{\sum_{c=N-t}^t l_{A,t}(c)}.$$

The next result represents the main theoretical insight, and shows that, for a given level of frictions, cohort and year effects are informative about the relative role of human capital vs prices/productivity.

Proposition 2: Decomposition of Labor Reallocation

Labor reallocation out of agriculture g_{LA} can be decomposed as

$$\log g_{LA} = \underbrace{\log \bar{\psi}}_{\text{Year Effect}} + \underbrace{\log \bar{\chi}}_{\text{Cohort Effect}}$$

where

$$\begin{aligned} \log \bar{\psi} &= \underbrace{\left(1 - \underbrace{\lambda(f)}_{\text{Friction} \in [0, 1]}\right)}_{\text{Friction} \in [0, 1]} \underbrace{\left(\underbrace{\Theta \log g_{pz}}_{\text{Prices/Productivity}} + \underbrace{\alpha \Theta \left(\frac{\gamma v}{1 - \gamma}\right) \log g_h}_{\text{General Equilibrium}}\right)} \\ \log \bar{\chi} &= \underbrace{-\left(\frac{\gamma v}{1 - \gamma}\right) \log g_h}_{\text{Human Capital (P.E.)}} + \underbrace{\frac{\lambda(f)}{1 - \lambda(f)} \log \bar{\psi}}_{\text{Year Effect through Friction}} \end{aligned}$$

and the friction parameter $\lambda(f)$ satisfies the following: (i) it is equal to zero when the fixed cost is zero, $\lambda(0) = 0$; (ii) it is equal to one when individuals are unconstrained only in the first period of their life, $\lambda(\bar{f}) = 1$; (iii) it is increasing in the size of the fixed cost, $\lambda'(f) > 0$; and (iv) it does not depend on the size of iceberg cost.

Proof. See Appendix. \square

It is convenient to describe the proposition starting from the frictionless case, when $\lambda(f) = 0$.

First, the year effect. The faster the rate of decay of agricultural relative revenue productivity – $\log g_{pz}$ – the faster the rate of within cohort labor reallocation, i.e. the average year effect. This result is intuitive, as agricultural production becomes less rewarding, the minimum skill level necessary to have a comparative advantage in non-agriculture decreases, and thus more and more individuals of a given cohort move out of agriculture. We need however, to consider a general equilibrium interaction across cohorts – the second term in the equation for $\log \bar{\psi}$. As a cohort ages, younger, and more skilled cohorts enter the labor market. Therefore, as a cohort ages, it develops, relatively to the younger ones, a comparative advantage towards non-agriculture. In fact, the faster the rate of human capital growth across cohorts, the more a cohort would tend to *move in* agriculture as it ages: a general equilibrium effect. Assumption 3 guarantees that this general equilibrium is sufficiently weak and that there is positive reallocation, as we observe in the data.

Next, the cohort effect. The faster the rate of human capital growth, the more any two cohorts differ in their human capital, and thus the more they will differ in the average agricultural employment, thus leading to a large cohort effect. The cohort effect captures the partial equilibrium effect of human capital on labor reallocation. The aggregate effect of human capital, which has

to take into account the counteracting general equilibrium force, is weaker:

$$\underbrace{-\gamma\Theta \log g_h}_{\text{Aggregate Effect}} = \underbrace{-\left(\frac{\gamma v}{1-\gamma}\right) \log g_h}_{\text{Direct P.E. Effect}} + \underbrace{\alpha\Theta \left(\frac{\gamma v}{1-\gamma}\right) \log g_h}_{\text{Counteracting G.E. Effect}}. \quad (3)$$

Finally, the role of $\lambda(f)$. When the fixed cost is positive, $\lambda(f)$ is positive, and thus a share of the aggregate changes in relative prices/productivity shows up as a cohort effect. Why does this happen? Intuitively, when the fixed cost is zero, the problem is essentially a static repeated one, since each individual only needs to compare her current wages in either agriculture and non-agriculture. When, instead, the fixed cost is positive, dynamic considerations emerge: individuals evaluate the future path of wages when deciding whether to move out of agriculture. Therefore, when the fixed cost is positive two cohorts of different ages differ not only due to their human capital, but also because they have a different number of years left to work. This latter difference is larger the faster the change in aggregate economic conditions.

Recall that our overarching goal is to use reallocation within- and across-cohorts to decompose aggregate reallocation into the contribution of human capital and prices/productivity. Proposition 2 shows that to accomplish this goal it is necessary to pin down the size of the mobility friction parameter $\lambda(f)$. In Section 4 we will show that – under further assumptions – Proposition 2 can be used to bound the size of λ by using variation across countries, or across regions within countries, in the year and cohort effects. Nonetheless, we further explore our theoretical structure in search for further empirical predictions that allow to pin down a value for $\lambda(f)$. Specifically, we show that the reallocation rates by ages depend on $\lambda(f)$: young cohorts always reallocate at the unconstrained rate, while depending on the size of $\lambda(f)$, the relatively old ones might be constrained and not reallocate at all. In Section 4 we will use reallocation rates by ages as one of our approaches to back out a value for $\lambda(f)$.

Definition: Unconstrained Year Effect *Let ψ^u be “unconstrained” year effect at which individuals would reallocate out of agriculture in the frictionless benchmark with $i = 0$ and $f = 0$; that is, let*

$$\log \psi^u \equiv \Theta \log g_{pz} + \alpha\Theta \left(\frac{\gamma v}{1-\gamma}\right) \log g_h.$$

Proposition 3: Year Effects by Age

If the fixed cost is equal to zero, $f = 0$, then individuals of all ages reallocate at the unconstrained rate; that is

$$\log \psi_t(c) = \log \psi^u$$

for all (t, c) . If, instead, $f > 0$, and thus $\lambda(f) > 0$, then there exists an age cutoff, $\hat{a}(\lambda(f))$, with $\hat{a}_\lambda > 0$, such that individuals younger than $\hat{a}(\lambda(f))$ reallocate at the unconstrained rate, while those older than $\hat{a}(\lambda(f))$ are constrained and thus do not reallocate at all; that is

$$\log \psi_t(c) = \log \psi^u$$

for all (t, c) such that $t - c < \hat{a}(\lambda(f))$ and

$$\log \psi_t(c) = 0$$

for all (t, c) such that $t - c \geq \hat{a}(\lambda(f))$.

Proof. See Appendix. \square

2.5 Wages and Frictions

Finally, we turn to wages and characterize the agricultural wage gaps. We follow this route for two reasons: (i) in search of additional ways to pin down a value for the size of the mobility friction; (ii) in order to provide further validation for the model by showing that it is consistent with recently documented empirical facts.

Proposition 4: Agricultural Wage Gaps.

Let $(\hat{c}_t, \hat{\varepsilon}_t)$ be a mover to M at time t and $\bar{w}_{M,t} \equiv \sum_{c=N-t}^t \int w_{M,t}(c, \varepsilon) dF$ be the average wage in M , then for all periods t

$$\underbrace{\log \bar{w}_{M,t} - \log w_{A,t}}_{\text{Cross-Sectional Wage Gap}} > \underbrace{\log w_{M,t}(\hat{c}_t, \hat{\varepsilon}_t) - \log w_{A,t}}_{\text{Wage Gap for Movers}},$$

and the wage gap for movers is given by

$$\log w_{M,t}(c, \varepsilon) - \log w_{A,t} = -\log(1 - i) + \log(1 + (1 - g_{pA}\beta)f).$$

where i is the iceberg cost, f is the fixed cost, and g_{pA} is the growth rate of $p_t Z_{A,t}$.

Proof. See Appendix. \square

Related to our first objective above, Proposition 4 gives a relatively negative result. It shows that even the wage gap of movers from agriculture to non-agriculture is not directly informative about the size of $\lambda(f)$. Specifically, Proposition 4 demonstrates that observing a low wage gap for movers, as has been shown in recent literature (see Hicks et al. (2017), Alvarez (2017), Herrendorf and Schoellman (2017)), is sufficient to exclude a large iceberg cost, but not to exclude a large fixed cost. In fact, conditional on an individual not being constrained, the fixed affects her moving decision on the margin only through discounting.⁷ Intuitively, there are two reasons

⁷It is worthwhile to stress that this result is driven by two features of our environment: (i) the decision to move

Unfortunately, Proposition 2 showed that the fixed cost is the most important component in determining the friction parameter $\lambda(f)$. It is also important to stress that even long panels are not sufficient to back out the size of the fixed cost from wages. In fact, we would need to observe the whole wage paths in agriculture *and* in non-agriculture for both movers and non-movers. Of course, such data is simply impossible to generate. For this reason, in Section 4 we will use a different empirical approach to back out $\lambda(f)$; one that does not rely on wages, but rather on variation in reallocation rates across countries and age groups.

Related to our second objective, instead, Proposition 4 is encouraging. The model is able to generate one robust recent empirical fact (see again, Hicks et al. (2017), for example): wage gaps between agriculture and non-agriculture in the cross-section are much larger than the wage gaps observed for movers out of agriculture. The model naturally matches this fact since movers are indifferent between agriculture and non-agriculture, and thus they are less skilled than the average non-agricultural worker, and more skilled than the average agricultural worker.

3 Labor Reallocation by Cohort: Data and Facts

We next describe how we use data to quantify the role of human capital in labor reallocation out of agriculture. The key component of our empirical approach is labor reallocation within and across cohorts. This is motivated by the results in Proposition 2, which provide an accounting framework to link within- and across-cohorts labor reallocation to the relative contribution of human capital vs. prices/productivity in reallocation out of agriculture. In this Section, we start by documenting a number of novel cross-country facts about reallocation by cohort using micro level data for a large set of countries. Most of the cross-country evidence available to date only covers aggregate rates of reallocation. Our paper is among the first to provide micro level evidence on the behavior of different cohorts of workers in the process of structural transformation⁸. We present the patterns descriptively to convey information on what the data say in a transparent format, focusing on the novelty of the findings rather than on their role in our approach. In Section 4 we will instead interpret the observed patterns through the lens of theory. There, we make explicit how Proposition 2 can be brought to the data to make inference on whether human capital matters for the movement of workers out of agriculture.

Below we introduce the data and measurement approach, and then discuss the novel cross-country findings on reallocation by cohort.

out of agriculture is a dynamic one, hence individuals can choose to postpone it, (ii) the relative wages change over time. As a result of these two features, the fixed cost mainly affects the timing of the movement out of agriculture, while it impacts the wage gap only marginally through discounting.

⁸Kim and Topel (1995), Lee and Wolpin (2006), and Perez (2017) document sectorial reallocation by cohort but limit their focus to, respectively, South Korea and United States and Argentina. Hobijn et al. (2017) in ongoing work are also using the IPUMS dataset to document patterns on reallocation by cohorts. In particular, they document results consistent to our Fact 1 below, but considering reallocation between three sectors.

3.1 Data

We use micro level data from the Integrated Public Use Microdata Series (IPUMS)⁹. The data are either censuses or large samples from labor force surveys that are representative of the entire population.

We include in our analysis all IPUMS countries for which we have available at least two ten-years apart repeated cross-sections with available information on age, gender, and working industry. This gives us a sample of fifty two countries covering about two thirds of the world population. For fifty one countries, the IPUMS data also include geographical information at the sub-national level (e.g. states or districts) which we use in our analysis as a source of additional variation.¹⁰ For twenty three countries, we observe four or more cross-sections, for seventeen we observe three or more. On average, we observe countries over a period of 28 years. For some countries, such as United States and Brazil, our data cover a long time span of half a century or more of labor reallocation. Table A.1 in the Appendix lists the countries in our sample, the income level of each country, in 2010, relative to the one of United States, the years of coverage, the agricultural employment shares, and the number of observed cross-sections. The countries in the sample comprise a wide range of income levels, from the United States to Liberia and El Salvador. Eight countries are high-income countries, twenty five are middle-income countries and the remaining nineteen are low-income¹¹. Our sample also spans a large geographical area, covering Asia and Oceania (nine countries), Africa (twelve countries), Central and South America (nineteen countries), and Europe and North America (twelve countries).

We focus on males and restrict our attention to those aged 25 to 59. This is meant to capture working age and identify the period after education investment is completed, which allows to consider human capital as constant. We exclude women from the current analysis given the large cross-country differences in female labor force participation.

3.2 Measurement

Cohort and year effects represent the key objects of the accounting framework developed in Proposition 2 and hence are at the core of our empirical approach. In order to compute them in the data, we follow closely the above definitions of average year and cohort effects. The only difference is that we need to take explicitly into account (through weighting) the demographic composition of different cohorts, while in the model we assumed that each cohorts is of identical size. As we show below, this does not make a quantitatively relevant difference.

Here is how we proceed. For each country and between any two cross-sections at times t and $t + k$, where $t + k$ is the subsequent cross-section available in the data, we compute the average

⁹Integrated Public Use Microdata Series, International: Version 6.5 [dataset], see(King et al. (2017)).

¹⁰Specifically, we use the IPUMS created variable geolev1.

¹¹By high-income (low-income) countries we mean those with income per capita greater (smaller) than 45% (10%) of the one of the United States at PPP, in 2010.

year effect at time t as

$$\bar{\psi}_{t,j} = \left(\frac{\sum_{c=N-t+k}^t \omega_t(c) l_{A,t+k}(c)}{\sum_{c=N-t+1}^t \omega_t(c) l_{A,t}(c)} \right)^{\frac{1}{k}},$$

where $\omega_t(c)$ is the share of population in cohort c at time t .¹² For most countries, we observe more than two repeated cross-sections. Let \mathbb{T}_j be the set of cross-sections that we observe for country j . We compute the overall average year effect for country j as

$$\log \bar{\psi}_j = \frac{1}{|\mathbb{T}_j|} \sum_{t \in \mathbb{T}_j} \log \bar{\psi}_{t,j}.$$

The average cohort effects are given by

$$\begin{aligned} \bar{\chi}_{t,j} &= \bar{\psi}_{t,j}^{-1} \left(\frac{\sum_{c=N-t}^t \omega_{t+k}(c+k) l_{A,t+k}(c+k)}{\sum_{c=N-t}^t \omega_t(c) l_{A,t}(c)} \right)^{\frac{1}{k}} \\ \log \bar{\chi}_j &= \frac{1}{|\mathbb{T}_j|} \sum_{t \in \mathbb{T}_j} \log \bar{\chi}_{t,j}. \end{aligned}$$

It is simple to verify that calculating year and cohort effects over k periods does not impact the mapping between cohort and year effects and the model structural parameters – that is, the results in Proposition 2 hold for any $k < N$.

Last, we compute the rate of labor reallocation out of agriculture as

$$\begin{aligned} L_{A,t} &= \sum_{c=N-t}^t \omega_t(c) l_{A,t}(c) \\ g_{L_{A,t},j} &= \left(\frac{L_{A,t+k}}{L_{A,t}} \right)^{\frac{1}{k}} \\ \log g_{L_{A,j}} &= \frac{1}{|\mathbb{T}_j|} \sum_{t \in \mathbb{T}_j} \log g_{L_{A,t},j}, \end{aligned}$$

where $L_{A,t}$ is the share of overall population employed in agriculture at time t .

It is worth noticing that, consistently with Proposition 1, the empirical rate of reallocation out of agriculture in country j between time t and time $t+k$ can be written as the product of average cohort and year effects:

$$g_{L_{A,t},j} = \frac{L_{A,t+k}}{L_{A,t}} = \bar{\psi}_{t,j} \bar{\chi}_{t,j}.$$

We obtain $\log g_{L_{A,j}}$, $\log \bar{\psi}_j$ and $\log \bar{\chi}_j$ for all countries in our sample. To fix ideas, it is helpful

¹²For few countries, notably India, we observe age-heaping. We adjust for it by smoothing out $\omega_t(c)$ with a quadratic equation. In Section 3.3.5, we show that for all countries, but India, the adjustment is mostly inconsequential.

to go back to the patterns described in the introduction for Brazil and India and think about which year and cohort effects our method delivers for these two countries. The year effect is -1.9% for Brazil and -0.2% for India. The cohort effect is -0.9% for Brazil and -0.4% for India. These numbers provide a quantitative statement for the qualitative evidence that emerges by comparing the figures: over time cohorts move out of much faster in Brazil than in India; on the other hand, for both Brazil and India, at any given point in time younger cohorts are employed in agriculture to a smaller extent. Table A.2 in the Appendix includes the rate of labor reallocation and the average year and cohort effects for each country.

3.3 Facts

We now introduce three novel empirical facts about reallocation within and across cohorts during the process of structural change out of agriculture. The three empirical facts are summarized in columns (1)-(6) of Table 1.

3.3.1 Fact 1: Decomposition of Average Reallocation Rate out of Agriculture

In Figures 2a, 2b and 2c, we plot the cross-cross-country distribution of, respectively, the rate of reallocation out of agriculture, the average year effect and average cohort effect. For most countries, the rate of reallocation out of agriculture is negative, suggesting, unsurprisingly, that most countries in our sample underwent reallocation from agriculture to non-agriculture. We can also observe that in most countries both the year and cohort effects are negative, indicating that they both positively contributed to structural change. The key pattern to notice is that the two distributions have similar means, namely, -0.9% and -1.1% . Remember that, by construction, the total rate of reallocation out of agriculture can be decomposed in the sum of year and cohort effects. That is, we can write

$$\mathbb{E}[\log g_{L_{A,j}}] = \underbrace{\mathbb{E}[\log \bar{\psi}_{t,j}]}_{\text{Within-Cohort}} + \underbrace{\mathbb{E}[\log \bar{\chi}_{t,j}]}_{\text{Across-Cohorts}}.$$

This implies that similar cross-country means for year and cohort effects suggest that, on average, within- and across-cohort reallocations contribute to a similar extent to reallocation out of agriculture. Specifically, across-cohorts reallocation accounts, on average for all countries, for 56% of overall labor reallocation out of agriculture. If we restrict the attention only to low-, middle-, or high-income countries we obtain that across-cohorts reallocation accounts for respectively 64%, 52%, and 57% of the overall labor reallocation.

3.3.2 Fact 2: Decomposition of Cross-Country Variance of Reallocation Rates

In Figures 3a and 3b, we plot, respectively, year and cohort effects as a function of the rate of labor reallocation out of agriculture. Two patterns emerge: (i) the year effects are strongly positively correlated with the rate of labor reallocation, while (ii) the cohort effects are more similar across countries and weakly correlated with the rate of labor reallocation. In other words, countries experiencing both slow and fast structural transformation have quite similar reallocation across cohorts, while countries undergoing fast reallocation have much larger reallocation within cohorts. This is consistent with what the Brazil vs. India case suggested: Brazil experienced fast

structural transformation and displays a very large year effect.

To make this discussion more formal, we decompose the cross-country variation in the rate of reallocation out of agriculture as follows

$$\text{Var} [\log g_{L_{A,j}}] = \underbrace{\text{Cov} [\log g_{L_{A,j}}, \log \bar{\psi}_{t,j}]}_{\text{Within-Cohort}} + \underbrace{\text{Cov} [\log g_{L_{A,j}}, \log \bar{\chi}_{t,j}]}_{\text{Across-Cohorts}}.$$

We obtain that the, on average across all countries, the across-cohort component accounts for 28.9% of the dispersion of reallocation rates. If we focus to only low-, middle-, or high-income countries, we obtain that the across-cohorts component contribution to total variance is respectively 20.6%, 18.5%, and 16.5%.¹³

3.3.3 Fact 3: Reallocation Rates by Age Groups

Our third piece of novel empirical evidence is related to differences in labor reallocation across age. We are interested in assessing whether, for given changes over time in the relative returns of agriculture and non-agriculture, individuals of different ages are equally likely to move across sectors. We may for instance expect that older individuals are less likely to change sector than younger individuals. Within our framework, we can test whether this is the case looking at the year effects of individuals of different ages. For each country in our sample, we compute the year effect for a given age group as follows

$$\psi_{t,j}(c) = \left(\frac{l_{A,t+k}(c)}{l_{A,t}(c)} \right)^{\frac{1}{k}},$$

and then

$$\log \bar{\psi}_{j,age} = \frac{1}{|\mathbb{T}_j|} \sum_{t \in \mathbb{T}_j} \sum_{c \in \mathbb{C}_{age}(t)} \log \psi_{t,j}(c) \quad (4)$$

where the set $\mathbb{C}_{age}(t)$ – for the age group 25-35, for example – is given by

$$\mathbb{C}_{25-35}(t) = \left\{ c \text{ s.t. } \left(t + \frac{k}{2} - c \right) \in [25, 35] \right\}.$$

Figures 4b, 4a, and 4c plot these year effects, respectively, for the age groups 25-35, 35-50 and 50-59 against the year effect computed on the entire population. The figures clearly show that the year effects are very similar across the age groups. That is, within-cohort reallocation occurs to the same extent at different ages. To summarize the information of the Figures, we compute the gaps between the average year effect and the ones calculated for individuals aged 25-35 and 50-59: they are respectively -0.14% and -0.20%.

¹³Notice that the average contribution of the within-cohort component does not need to be a weighted average of the contributions within each income group. In fact, the overall variance of reallocation rates takes into account also the differences across income groups.

3.3.4 Analysis at the Sub-National Level

We replicate our analysis using sub-national units – instead of countries. This provides useful additional variation that we exploit in Section 4. Figures 5, 6, and 7 replicate, respectively, Facts 1, 2 and 3 in this extended sample. Summary statistics for each fact are in the bottom half of Table 1. The evidence aligns well with that obtained using only cross-country variation.

3.3.5 Controlling for Demographic Composition

In the model, we assumed that each cohort has equal size and that size is constant over the life-cycle. In the data, however, we observe that cohorts have different sizes, and that the size of a given cohort changes over time due to mortality. We may be concerned that these demographic compositional effects are relevant in explaining the cross-country variation in cohort and year effects. We here perform a series of exercise to show that, reassuringly, demographic composition does not mechanically drive our estimates. First, we compute the year effect weighting each cohort by its share in the population at time $t + k$ rather than at time t . In Figure 8a, we show that this change does not make any difference. Second, we compute the year effect with the raw data – i.e. without smoothing the demographic distribution, which we did in the benchmark exercise to adjust for age-heaping. As is well know, Indian data suffer from extreme age-heaping. Consistent with this, we see in Figure 8b that only in India the year effect estimated with the raw data is different from the benchmark one. Third, we recompute the rate of labor reallocation out of agriculture keeping constant the demographic structure at time t – i.e. we compute $L_{A,t+k}$ weighting each cohort according to $\omega_t(c)$. Figure 8c, shows that the estimated rate of labor reallocation are almost identical to the benchmark ones. Finally – in Figure 8d – we perform the same exercise, but keeping constant the demographic structure at time $t + k$. Again, we conclude the demographic changes do not have relevant mechanical effects. In Figure A.3 in the appendix, we recompute the same exercise using sub-national units. We find similar results.

4 Human Capital and Labor Reallocation: Data and Theory Together

This section introduces our empirical approach and discusses how we make use of the data and patterns described above to quantify, in an accounting sense, the relative contribution of human capital and prices/productivity to labor reallocation out of agriculture. Our main starting point is Proposition 2, which tells us that we can do so by leveraging within- and across-cohorts reallocation. Proposition 2 provides a mapping between two observable objects, year and cohort effects, and our two main objects of interest, the contribution of human capital and of prices/productivity to labor reallocation. The mapping, however, is made challenging by the possibility that labor mobility frictions bind, and by general equilibrium. The spirit of our empirical exercise in this section is to exploit – through multiple approaches – observable variation in year and cohort effects both across- and within-countries to discipline the size of the frictions. We then use further micro level data to calibrate the strength of the general equilibrium.

We provide a range of estimates depending on parameter values, but the overarching conclusion, which we reach in Section 4.1, is that human capital explains approximately one third

of total reallocation out of agriculture. In Section 4.2, we show that, at the same time, human capital has at most a minor role in explaining why some countries have faster reallocation rates than others.

4.1 Human Capital and Average Cross-Country Reallocation

We are interested in answering the following question: On average, by how much does human capital contribute to the reallocation out of agriculture observed across countries? We start from Proposition 1, which decomposes the rate of reallocation into a human capital and a prices/productivity component. We can obtain the empirical counterpart of Proposition 1, which tells us that the average rate of labor reallocation across countries can be expressed as:

$$\mathbb{E}[\log g_{L_A,j}] = \underbrace{\Theta \mathbb{E}[\log g_{pz,j}]}_{\text{Prices/Productivity}} - \underbrace{\gamma \Theta \mathbb{E}[\log g_{h,j}]}_{\text{Human Capital}}.$$

This expression captures the essence of our accounting exercise. On its own, however, it does not provide a viable path to do accounting using cross-country data. The core of our approach is instead the empirical counterpart of Proposition 2, which states that cross-country variation in year and cohort effects can be exploited to infer the roles of human capital and prices/productivity in labor reallocation. In particular, Proposition 2 directly leads to Lemma 1 below, which shows that – for a given size of the frictional parameters $\lambda_j(f)$, and strength of the general equilibrium – the observable average cohorts and year effects can provide an estimate for the average role of human capital in labor reallocation; that is, for $-\gamma \Theta \mathbb{E}[\log g_{h,j}]$.

Lemma 1: Contribution of Human Capital to Average Reallocation

The relative contributions of human capital to average labor reallocation is given by:

$$\underbrace{-\gamma \Theta \mathbb{E}[\log g_{h,j}]}_{\text{Aggregate Effect}} = \underbrace{\left(\frac{\frac{1-\gamma}{v}}{\frac{1-\gamma}{v} + \alpha} \right)}_{\text{G.E. Wedge}} \underbrace{\left(\mathbb{E}[\log \bar{\chi}_j] - \mathbb{E}\left[\frac{\lambda_j(f)}{1 - \lambda_j(f)} \log \bar{\psi}_j \right] \right)}_{\text{Direct P.E. Effect}}.$$

Proof. See Appendix. \square

Unfortunately, neither frictions nor the strength of the general equilibrium are directly observable. Therefore we proceed in three steps, informed by Lemma 1 above. First, we show that remaining agnostic about the size of the friction and the strength of the general equilibrium we are able to derive an upper bound to the relative contribution of human capital to labor reallocation. Second, we directly tackle the issue of measuring $\lambda_j(f)$ to obtain point estimates for the partial equilibrium effect of human capital. Third, we calibrate the strength of the general equilibrium using additional micro level data to provide a point estimate for the overall, aggregate, effect of human capital.

4.1.1 Upper Bound to the Contribution of Human Capital

Recall that the model implies $\lambda_j(f) \in [0, 1]$ for all country j . Moreover, notice that reallocation out of agriculture implies that $E[\log \bar{\psi}_j]$ is negative – both overall, and within the three different income groups. Finally, it is immediate to see that, since $\gamma \in (0, 1)$, $v \geq 0$ and $\alpha \geq 0$, the general equilibrium wedge is smaller than one. As a result of these three considerations, the observable average cohort effect, $E[\log \bar{\chi}_j]$, provides an upper bound for the aggregate effect of human capital on labor reallocation, $-\gamma\Theta E[\log g_{h,j}]$. We can then easily compute, using the results from Section 3, such upper bound, both overall, and separately by income group.

Results on Upper Bound of Human Capital Contribution. All results, throughout Section 4.1.1, on $-\gamma\Theta E[\log g_{h,j}]$ are reported as a percentage of the rate of labor reallocation $E[\log g_{LA,j}]$. Fact 1 in Section 3, showed that the cohort effect accounts for 56% of the average reallocation observed across countries. Interpreted through Lemma 1, Fact 1 then suggests that human capital can account for at most 56% of labor reallocation out of agriculture. Further, still using the results of Fact 1 we can obtain similar upper bounds within each income classes, we get 64%, 52% and 57% respectively for low-, middle-, and high-income countries. These results are included in Table 5. In the same table, we show the results of the same exercise using sub-national units: they are similar.

4.1.2 Backing Up Frictions

Next, we provide estimates for the size of the friction through a series of different exercises. The results of our estimates are summarized in Table 2. Conceptually, we follow two main alternative methods, each tied to a source of variation that can be exploited to back out the size of the frictions.

Our first method builds on Proposition 3, which relates the size of labor mobility frictions to the reallocation rates of individuals of different ages. The fixed cost traps in agriculture individuals that would otherwise move to non-agriculture in a frictionless environment. This effect, however, is not symmetric across ages, and is instead stronger for older individuals: they benefit from future increases in non-agriculture wages for fewer years and hence, for given fixed cost, face a stronger constraint. This means that the presence of frictions causes old individuals – the constrained – to reallocate at a slower rate than young individuals – the unconstrained. Fact 3 in Section 3 showed that old and young individuals reallocate at similar rates, thus providing evidence against a sizable role for mobility frictions. The following Lemma shows that this intuition can be used to provide a direct estimate for the size of the frictions.

Lemma 2.a: Size of Frictions and Country-Specific Year Effects by Age

Let $\psi_j^u = \psi_{t,j}(c)$, where c is an unconstrained cohort at time t . Then,

$$\lambda_j(f) = 1 - \frac{\log \bar{\psi}_j}{\log \psi_j^u}.$$

Proof. See Appendix. \square

For each country j , we compute ψ_j^u as the year effect for young individuals aged 25 to 35 – $\bar{\psi}_{j,25-35}$, measured as shown in 4. Lemma 2.a tells us that we can obtain a country-specific measure of frictions comparing the year effect for young individuals (assumed unconstrained) to the year effect for the entire population. When implementing Lemma 2.a, and computing the ratio $\frac{\log \bar{\psi}_j}{\log \psi_j^u}$ for each country, there are two practical concerns: (i) for few countries, $\log \psi_j^u > 0$, which would violate the assumptions underlying Lemma 2.a; (ii) when dividing by a small number – $\log \psi_j^u$ – any source of noise is going to be amplified. In order to circumvent these concerns, rather than computing the $\frac{\log \bar{\psi}_j}{\log \psi_j^u}$ country by country, we compute $\frac{E[\log \bar{\psi}_j]}{E[\log \psi_j^u]}$ for different subsets of countries.¹⁴ Under the assumption that, within the chosen subsets, $\lambda_j(f)$ and $\log \psi_j^u$ are independent – which would trivially hold if $\lambda_j(f) = \lambda(f)$ – this method would recover $E[\lambda_j(f)]$. The first column of Table 2 summarizes the estimated size of the frictions. When pooling together all countries, we estimate the friction to be 14%. When pooling together all subnational units we also find 14%. However, this first method gives heterogeneous estimates across income levels: the friction is as large as 40.2% in low-income countries, while even negative in high-income ones. This method relies on the strong assumption that workers aged 25-35 correctly identify the unrestricted reallocation rate for the whole economy. We next consider a different approach, that relies on an alternative set of assumptions.

Our second method to estimate $\lambda(f)$ relies on Proposition 2 and exploits its implications for the co-movement of the rate of labor reallocation with cohort and year effects. Mobility frictions induce a correlation between cohort and year effects, and, as a consequence, cause the correlation between cohort effect and rate of labor reallocation to increase. That is, in presence of frictions, we should observe that countries experiencing fast labor reallocation display both large year and cohort effects. This means that we can use the strength of the correlation between the rate of labor reallocation and cohort vs. year effects to obtain a measure for $\lambda(f)$. More formally, consider the following two identifying assumptions (IA).

IA.1. *Mobility costs are identical for all countries $j \in \mathbb{J}$.*

IA.2. *In the set \mathbb{J} , countries with faster decline in relative agriculture revenue productivity also experience on average (weakly) faster increase in human capital:*

$$\text{Corr}[-\log g_{pz,j}, \log g_{h,j}] \geq 0.$$

Under these assumptions, we can derive the following result.

¹⁴Nonetheless, we still computed $\lambda_j(f)$ country by country for the countries for which $\log \psi_j^u < 0$. When constrained the $\lambda_j(f)$ within the admitted range $[0, 1]$, we obtain similar results than with the benchmark methodology. In particular, we obtain a value for $E[\lambda_j(f)]$ equal to 16%.

Lemma 2.b: Size of Frictions and Cross-Country Year vs. Cohort Effect

If IA.1 and IA.2 are satisfied for a set of countries \mathbb{J} , then

$$\lambda(f) \leq \frac{\text{Cov}[\log g_{L_A,j}, \log \bar{\chi}_j]}{\text{Cov}[\log g_{L_A,j}, \log \bar{\psi}_j] + \text{Cov}[\log g_{L_A,j}, \log \bar{\chi}_j]} \equiv \bar{\lambda}(\mathbb{J}).$$

Proof. See Appendix. \square

Lemma 2.b shows that the relative values of $\text{Cov}[\log g_{L_A,j}, \log \bar{\chi}_j]$ and $\text{Cov}[\log g_{L_A,j}, \log \bar{\psi}_j]$ can be used to provide an upper bound for the size of the mobility frictions in any set of countries or sub-national regions \mathbb{J} . Notice that Assumption IA.1 is necessary since we are exploiting cross-country variation. Assumption IA.2 excludes the case in which high values of the friction $\lambda(f)$ are consistent with low values of $\text{Cov}[\log g_{L_A,j}, \log \bar{\chi}_j]$ because of how human capital and prices/productivity co-move.

The second column of Table 2 reports the results. We first compute a bound $\bar{\lambda}(\mathbb{J})$ for the assumption that frictions are the same for all countries in our sample. This gives a value of 28.9%. We then compute a bound separately for low-income, middle-income and high-income countries. The values are, respectively, 20.6%, 18.5%, 16.5%. Therefore, also with this method, we do estimate smaller frictions for high-income countries, but the differences are quite small.¹⁵ Results computed using sub-national variation – shown in the bottom half of the same Table 2 – show a similar average magnitude, but they don’t support the relationship between income and frictions.

In addition, we can derive country-specific bounds using within-country variation across sub-national units. That is, under the assumption that frictions are the same for all regions within country j , we can obtain a bound for $\lambda_j(f)$. Since we need to use variation across regions within countries, we restrict our focus to the 37 countries – 11 low-income, 19 middle-income, and 7 high-income – for which we have information available for at least ten subnational units. This source of variation allows to compute one bound for each country, $\bar{\lambda}(j)$. In column (3) of Table 2 we report the averages of these bounds. Overall, we get similar results than when we exploit across countries variation. Specifically, $E[\bar{\lambda}(j)] = 28.4\%$, which is almost identical to the estimated $\bar{\lambda}(\mathbb{J}) = 28.9\%$. It is important to stress that there are no mechanical reasons for the two to be similar; in fact, they use different sources of variation: $E[\bar{\lambda}(j)]$ across regions within each country, and $\bar{\lambda}(\mathbb{J})$ across countries.

Overall, the evidence from this approach points towards a limited role of mobility frictions in labor reallocation, consistently with the results shown from Lemma 2.a. It is reassuring that the two approaches provide a similar answer, since they rely on different identifying assumptions and exploit different variation in the data.

Finally, we can add further evidence about the size of frictions using a third method that

¹⁵Notice that the average contribution of the within-cohort component does not need to be a weighted average of the contributions within each income group. In fact, the overall variance of reallocation rates takes into account also the differences across income groups.

combines features of the two main methods above. Under the assumption that mobility frictions are the same within a set of countries \mathbb{J} , we can use the cross-country variation in reallocation rates by age to pin down $\lambda(f)$ in \mathbb{J} . Formally, we derive the following Lemma.

Lemma 2.c: Size of Frictions and Cross-Country Year Effects by Age

If IA.1 is satisfied for a set of countries \mathbb{J} , then the regression across countries

$$\log \bar{\psi}_j = B_0 + B_1 \log \psi_j^u + \epsilon_j$$

recovers

$$\lambda(f) = 1 - \hat{B}_1.$$

Proof. *See Appendix.* \square

The advantage of this last approach is to provide a point estimate without having to rely on the assumption that the unconstrained rate of reallocation is identical across age groups. The fourth column of Table 2 reports the results of running the regression in Lemma 2.c using cross-country variation for all countries in our sample as well as separately for high-income and low-income countries. Overall, the results are consistent with the previous methods, and show a value for the $\lambda_j(f)$ around 20%. As for Lemma 2.b, we can further provide country-specific estimates of mobility frictions obtained under the assumptions that frictions are the same across all regions in country j . This last exercise gives quite higher values for $\lambda_j(f)$. This is likely due to large measurement error at this level of variation, which downward biases the estimated coefficients $\hat{B}_{1,j}$, and thus upward biases our estimates for $\lambda_j(f)$. For this reason, we don't consider the estimates from this last exercise, which are reported in the fifth column of Table 2, for our results below.

Results on the P.E. Effect of Human Capital. As discussed, each one of our four exercises to blackout a size for the mobility frictions – shown in columns (1)-(4) of Table 2 – exploits a different source of empirical variation, and holds under a different set of assumptions. All the assumptions needed are quite restrictive, nonetheless, we find reassuring the fact that the different methods provided quite consistent estimates. Since each method provides an independent estimate for $\lambda_j(f)$, in column (6) of Table 2 we average them. We can then use these estimates, together with the values for the cohort effects documented in Section 3, to provide a point estimate for the partial equilibrium effect of human capital. We obtain 46% on average for all countries; and 55%, 42%, and 52% for low-, middle-, and high-income countries. These results are reported in Table 4.

Furthermore, we can compute more conservative estimates for the partial equilibrium effect of human capital under the assumption that the friction is 30% in all countries, which, we argue, is a reasonable upper bound vis-a-vis the evidence of this section. These lower bounds are reported in Table 5, and they are equal to – in the usual order – 37%, 49%, 31%, and 49%. We conclude

that, even in the most conservative case, our evidence suggests that human capital has, in partial equilibrium, a quantitatively important effect on labor reallocation out of agriculture.

4.1.3 Calibrating the strength of General Equilibrium

Obtaining an estimate for the size of the mobility frictions allowed us to provide a point estimate for the partial equilibrium effect of human capital, which we show to be large. The P.E. effect of human is already a relevant object. It may even be the parameter of interest for some applications. For example, if we are interested in the distributional effects – in terms of agricultural employment – of human capital growth across cohorts. At the same time, as Lemma 1 showed, it is not sufficient to study the aggregate effect of human capital on labor reallocation out of agriculture. In order to quantify the aggregate effect of human capital, we need to provide values for α and $\frac{1-\gamma}{v}$ to quantify the general equilibrium wedge.

First we focus on α . It is the land income share in agriculture. Herrendorf et al. (2015) gives us an estimate for α for the United States: it finds $\alpha_{US} = 0.07$. Land, however, may have a high income share in lower income countries, where agricultural production is less capital-intensive. For this reason, we show results for different values of α . Notice that - by definition - α must be lower than one minus the labor share, which is usually assumed to be – on the aggregate – equal to $\frac{2}{3}$.¹⁶ At the same time, we recognize that labor might play a smaller role in agriculture. For these reasons, we provide estimates for the GE wedge for a range of α from 0.07 to 0.40, which we consider a reasonable upper bound for the land income share.¹⁷

Next, we focus on $\frac{1-\gamma}{v}$. We can estimate it using the model structure. Recall that the idiosyncratic ability ε is distributed as a Beta($v, 1$) and that the wage in the non-agricultural sector of an individual (c, ε) at time t is given by

$$\log w_{M,t}(c, \varepsilon) = Z_{M,t} + \log \kappa + \gamma \log h_c + (1 - \gamma) \log \varepsilon.$$

As a result, the cross-sectional wage variance within the non-agricultural sectors among individuals of cohort c is given by

$$\begin{aligned} \text{Var}_{\varepsilon} [\log w_{M,t}(c, \varepsilon)] &= (1 - \gamma)^2 \text{Var} [\log \varepsilon \mid \log \varepsilon \geq \log \hat{\varepsilon}_t(c)] \\ &= (1 - \gamma)^2 \text{Var} [-\log \varepsilon \mid -\log \varepsilon \leq -\log \hat{\varepsilon}_t(c)] \\ &\leq (1 - \gamma)^2 \text{Var} [-\log \varepsilon] \\ &= \left(\frac{1 - \gamma}{v} \right)^2, \end{aligned} \tag{5}$$

where the first equality comes from the fact that only individuals with ability above the threshold $\hat{\varepsilon}_t(c)$ select to non-agriculture; the second one from the fact that the support of ε is on $[0, 1]$ and thus the support of $\log \varepsilon$ is on $(-\infty, 0]$; the third and the fourth ones from the following

¹⁶A recent paper – Karabarbounis and Neiman (2013) – casts some doubts on this value for the labor income share, and in particular it argues that the labor share has been declining over time. Nonetheless, even in the more recent years, the estimates for the labor share are almost always above $\frac{1}{2}$, and in most cases above 60%.

¹⁷In ongoing work, Gollin and Udry (2017) estimates production function for Uganda and Ghana and find land shares in the range 0.40-0.50.

statistical properties of the Beta and the Exponential distributions: (i) if $\varepsilon \sim \text{Beta}(v, 1)$, then $-\log \varepsilon \sim \text{Exp}(v)$, and (ii) the variance of an exponential distribution truncated from above is smaller than the unrestricted variance, which is equal to v^{-2} . Equation (5) shows that we can use the, empirically observable, standard deviation of log wages in non-agriculture to provide a lower bound for the structural object of interest, $\frac{1-\gamma}{v}$. Moreover, notice that the general equilibrium wedge is itself increasing in $\frac{1-\gamma}{v}$, therefore, our empirical estimates provide also a lower bound to it, thus being conservative on the aggregate role of human capital.

Practically, consider a country j for which we observe a cross-section with individual level wage data. We compute the lower bound estimate for $\frac{1-\gamma}{v}$ as follows: (i) for each cohort, compute the standard deviation of log wages in non-agriculture – $\sigma_{W,j}(c)$; (ii) take the average across cohorts of the standard deviations computed at point (i) – $\bar{\sigma}_{W,j}$. The data discussed in Section 3, do not include wage data for most countries. However, Lagakos et al. (2017a) made to us available the statistic $\bar{\sigma}_{W,j}$ for each one of the eighteen countries in their sample, which span the income distribution from Bangladesh to the United States.¹⁸ On average, we obtain $E[\bar{\sigma}_{W,j}] = 0.67$, where the range across countries is from 0.38 for France to 0.94 for Brazil.

In Table 3, we calculate the general equilibrium wedges for different values of α and $\frac{1-\gamma}{v}$. Even for the most conservative estimates of $\alpha = 0.4$ and $\frac{1-\gamma}{v} = 0.4$, the general equilibrium wedge is at least 50%.

Results on Aggregate Effect of Human Capital. Combining the estimates for the general equilibrium wedge with the results on the P.E. effect of human capital allows us to quantify the aggregate contribution of human capital to labor reallocation. We argue that the most reasonable estimate for the general equilibrium wedge is the one with intermediate values for $\frac{1-\gamma}{v}$ and α , which give a wedge equal to 76%. That is, the aggregate effect is 76% of the partial equilibrium one. As a result, we obtain – as shown in the second column of Table 4 – that the aggregate contribution of human capital is on average 35%, or approximately one third. In other words, we find that, absent human capital accumulation, the cross-country average rate of labor reallocation out of agriculture would be approximately two thirds of that observed. This is our main result.

Further, we can compute reasonable intervals on the aggregate contribution of human capital, using the relative more or less conservative estimates. These results are shown in Table 5, and highlight that even in the most conservative case, human capital still has a sizable aggregate effect.

4.2 Human Capital and Cross-Country Differences in Reallocation

To conclude, we focus rather than on the average rate of labor reallocation, on its variance across countries. We are interested in determining to what extent differences in human capital accumulation can explain the differences in rate of labor reallocation out of agriculture that we observe across countries. Our starting point is again Proposition 1, whose empirical counterpart tells us that the cross-country variance of the rate of labor reallocation can be expressed as

¹⁸Refer to Lagakos et al. (2017a) for data description and details. Wages are constructed as earnings divided by total hours of work in the period of observation, which is either weekly, monthly, or yearly. We drop the top and bottom 1% of wages to check that the variance estimates are not driven by outliers. For each country, we keep the most recent available cross-section.

$$\text{Var} [\log g_{L_A,j}] = \underbrace{\Theta \text{Cov} [\log g_{L_A,j}, \log g_{p_z,j}]}_{\text{Prices/Productivity}} + \underbrace{\gamma \Theta \text{Cov} [\log g_{L_A,j}, -\log g_{h,j}]}_{\text{Human Capital}}.$$

We proceed through the same steps performed above to determine the average contribution of human capital to labor reallocation. First of all, we use again Proposition 2 to provide a counterpart to Lemma 1 for $\gamma \Theta \text{Cov} [\log g_{L_A,j}, -\log g_{h,j}]$, rather than for $-\gamma \Theta \text{E} [\log g_{h,j}]$. It is given by Lemma 3 below, and shows that – as expected – we need a value for the G.E. wedge and for the size of the friction in order to quantify the contribution of human capital to cross-country dispersion in rates of labor reallocation.

Lemma 3: Contribution of Human Capital to Reallocation Dispersion

The relative contributions of human capital to dispersion in labor reallocation is given by:

$$\underbrace{\gamma \Theta \text{Cov} [\log g_{L_A,j}, -\log g_{h,j}]}_{\text{Aggregate Effect}} = \underbrace{\left(\frac{\frac{1-\gamma}{v}}{\frac{1-\gamma}{v} + \alpha} \right)}_{\text{G.E. Wedge}} \times \underbrace{\left(\text{Cov} [\log g_{L_A,j}, -\log g_{h,j}] - \text{Cov} \left[\log g_{L_A,j}, \frac{\lambda_j(f)}{1-\lambda_j(f)} \log \bar{\psi}_j \right] \right)}_{\text{Direct P.E. Effect}}.$$

Proof. See Appendix. \square

Equipped with Lemma 3, we use the results on the observable objects $\text{Cov} [\log g_{L_A,j}, -\log \bar{\chi}_j]$ and $\text{Cov} [\log g_{L_A,j}, \log \bar{\psi}_j]$, together with the estimates for the size of the frictions and the general equilibrium wedge, allows us to estimate the unobserved object of interest $\gamma \Theta \text{Cov} [\log g_{L_A,j}, -\log g_{h,j}]$.

Results on Contribution of Human Capital to Reallocation Dispersion. We report results on $\gamma \Theta \text{Cov} [\log g_{L_A,j}, -\log g_{h,j}]$ as percentage of $\text{Var} [\log g_{L_A,j}]$. All the results are shown in columns (3) and (4) of Tables 4 and 5 and rely on estimates for the mobility frictions and the G.E. wedge previously calculated. Fact 2 shown in Section 3 directly provides upper bounds for the aggregate effect: they are equal to 29%, 21%, 19% and 17% respectively for the overall average across all countries, and for the low-, middle, and high-income countries. These upper bounds are quite sizable. However, they reduce significantly once we take into account the average estimates for the mobility frictions, shown in column (6) of Table 2. We now obtain 13%, 0.6%, 1.5%, and 8.5% respectively for the same four groups of countries. Taking into account general equilibrium further reduces by one fourth these estimates. We conclude that human capital has at most a minor role in explaining why some countries have faster rates of labor reallocation than others.

5 Schooling and Labor Reallocation

In the Section 4, we have shown that human capital accounts for one third of labor reallocation out of agriculture. In order to reach this conclusion, we have broadly defined human capital through two core characteristics – (i) it increases productivity out of agriculture; (ii) it improves across cohorts – and then use the model to indirectly identify the role of human capital through individual choices. As we discussed in the introduction, this approach has the advantage of not imposing any restriction on the determinants of human capital accumulation. However, this same feature necessarily leaves open two questions. First, it is unclear what exactly our definition of human capital captures, for example, it could in theory represents changes in preferences for non-agricultural work, which push young cohorts out of agriculture. Second, it is unclear whether it is possible to induce human capital growth, for example through government policy, and thus trigger reallocation out of agriculture.

In this final section, we use schooling, as a natural and broadly used direct measure of human capital, to address these two questions, and at the same time to provide validation for our main approach. Specifically, we follow the human capital accounting literature and use a functional form to map schooling years into human capital that generates a Mincer regression. In fact, we posit that human capital of a cohort c in country j with average years of school $s_{c,j}$ is given by

$$h_{c,j} = \exp \{ \rho_j s_{c,j} \} \quad (6)$$

where ρ_j is the return to school in terms of human capital.¹⁹

Equipped with a direct measurement of human capital, we tackle the previous two questions. In Section 5.1, we use the same data of Section 3, which provides individual level information on educational attainment, and show that cohort-level average schooling correlates with our measures of human capital both in levels and in changes across cohorts. This result validates our methodology vis-a-vis the literature and at the same time provides a natural interpretation of human capital as – at least partially – enclosing knowledge acquired through schooling. It also provides a simple explanation for the observed increase of human capital across cohorts: it is coherent with the global increase in schooling over the last half a century. In Section 5.2, we study a large-scale school construction program introduced in the 1970s in Indonesia. We follow Duflo (2001) in exploiting the quasi-random variation generated by the program, and we show that, at least in this context, a deliberate school construction policy caused an increased in human capital, as measured through the lens of our model, and led to reallocation out of agriculture.

5.1 Schooling and Human Capital: Validation of Our Approach

Recall that our model provides an invertible map between cohort-level agricultural employment and cohort-level human capital. To the extent that schooling is a proper measure of human capital, as argued in the literature, we should expect that cohorts with higher schooling have

¹⁹Our model is consistent with a Mincer wage regression in non-agriculture. In fact, the non-agricultural wage of an individual (c, ε) is given by $\log w_{M,t}(c, \varepsilon) = \log p_t z_t + \gamma \log h_c + (1 - \gamma) \log \varepsilon$. Using equation (6), this becomes $\log w_{M,t}(c, \varepsilon) = \log Z_{M,t} + \gamma \rho s_c + (1 - \gamma) \log \varepsilon$. The Mincer coefficient is given here by $\gamma \rho$.

lower agricultural employment. In fact, the model provides several structural relationships linking agricultural employment and human capital, which can be tested using schooling as a direct measure of human capital. We first look – in Section 5.1.1 – at predictions linking agricultural employment and human capital across cohorts within each country; we then study – in Section 5.1.2 – how average cohort effects varies across countries as a function of human capital growth.

5.1.1 Variation Across Cohorts within Countries

The benchmark model shows that the fraction of a cohort c that at time t is employed in agriculture is given by

$$\log l_{A,t}(c) = -\frac{\gamma v}{1-\gamma} \log h_c + \frac{v}{1-\gamma} \log \left((1-\alpha) \kappa^{-1} p_t z_t X^\alpha L_{A,t}^{-\alpha} \right). \quad (7)$$

We use the functional form for the human capital of a cohort c in country j – $h_{c,j} = \exp \{ \rho_j s_{c,j} \}$ – to estimate, separately for each country j , the following empirical counterpart of (7)

$$\log l_{A,t,j}(c) = \alpha_{t,j}^1 + \beta_j^1 s_{c,j} + \epsilon_{c,t,j}^1,$$

where the coefficients recover the structural parameters as follows

$$\begin{aligned} \alpha_{t,j}^1 &= \frac{v}{1-\gamma} \log \left((1-\alpha) \kappa^{-1} p_{t,j} z_{t,j} X_j^\alpha L_{A,t,j}^{-\alpha} \right) \\ \beta_j^1 &= -\frac{\gamma v \rho_j}{1-\gamma}, \end{aligned}$$

and $\epsilon_{c,t,j}^1$ is an error term. The superscript 1 on the coefficients highlights that this is the first testable relationship. We keep this notation throughout. The coefficient of interest is β_j^1 . The model predicts $\beta_j^1 < 0$. In figure 9a we plot the histograms of the point estimates for our set of 52 countries: for 51 countries the point estimates are negative and significant at 5%; for only 1 country, Vietnam, the point estimate is positive. The simple average of the point estimates is -0.16 , suggesting that, on average across countries, one extra year of school reduces agricultural employment by 16%.²⁰ This empirical relationship corroborates the model predictions. At the same time, we are aware that it may simply reflects omitted variable bias, such as a time trend that leads younger cohorts to have higher schooling and to work less in agriculture.

In order to alleviate this concern, we look for further model predictions which are less likely to be driven by omitted variables. In appendix B, we show that, once we enrich the model allowing for heterogeneous growth rates of human capital across cohorts, we obtain the following structural relationship between the cohort effects and the growth rates of human capital

$$\log \chi(c) = -\frac{\gamma v}{1-\gamma} \log g_h(c). \quad (8)$$

²⁰We also run a regression pooling together all countries, with country-year fixed effect, and constraining the coefficient β to be identical across countries. The point estimates of this second approach is -0.14 , significant at $< 0.1\%$.

Recall that we defined the cohort effect as $\chi_t(c) = \bar{\psi}_t^{-1} \frac{l_{A,t+1}(c+1)}{l_{A,t}(c)}$, and $g_h(c) \equiv \frac{h_{c+1}}{h_c}$ is the growth rate of human capital between cohorts c and $c+1$. The empirical counterpart of equation (8) is given by

$$\log l_{A,t+1,j}(c+1) - \log l_{A,t,j}(c) = \alpha_{t,j}^2 + \beta_j^2 (s_{c+1,j} - s_{c,j}) + \epsilon_{c,t,j}^2,$$

where

$$\begin{aligned} \alpha_{t,j}^2 &= \log \bar{\psi}_t \\ \beta_j^2 &= -\frac{\gamma^v \rho_j}{1 - \gamma} \end{aligned}$$

and $\epsilon_{c,t,j}^2$ is an error term. This second empirical regression is equivalent to the first one run in first differences. As such, it alleviates most of the concerns with that previous specification. In particular, any time trend for agricultural employment and education would not generate spurious correlation in this second specification. The model, again, predicts $\beta_j^2 < 0$, and, of course, it also implies that the point estimates for β_j^1 and β_j^2 should be identical. In Figure 9b we plot the histogram of the point estimates for our usual 52 countries: for 49 of them, the point estimates are negative; for 3 – Ireland, Mali, and Switzerland – they are positive. The simple average of the point estimates is -0.154 , very similar to the one for β_j^1 , which was -0.16 . However, in this second approach, the estimates for 16 out of the 49 countries are not statistically different from 0 at 5%.²¹ Also, and reassuringly, the three countries with positive point estimates are not statistically different from zero.²²

Finally, we show how β_j^1 and β_j^2 vary across countries. Notice that the model predicts that differences across countries in these coefficients should be driven by same underlying returns to school ρ_j . Therefore, we would expect that β_j^1 and β_j^2 are positively correlated across countries, and also that ρ_j is lower in low income countries as long as an additional year of schooling generates a larger human capital gain in rich countries, as argued in the literature. We show that both these results hold in the data. We restrict the analysis to the point estimates that are significantly different from zero. Both $\hat{\beta}_j^1$ and $\hat{\beta}_j^2$ are negatively correlated with GDP per capita. These correlation are significant at 5%. Moreover, in the appendix Figure A.4b we show that β_j^1 and β_j^2 are positively correlated across countries. The coefficient of a regression of β_j^1 on β_j^2 is marginally significant, with a p-value of 0.116.

5.1.2 Variation Across Countries in Average Cohort Effects

Proposition 2 showed that the average cohort effect is given by

$$\log \bar{\chi} = -\frac{\gamma^v}{1 - \gamma} \log g_h. \quad (9)$$

²¹In the appendix Figure A.4a we plot the histograms of the significant point estimates.

²²Also for this case, we run a regression pooling together all countries, with country-year fixed effect, and constraining the coefficient β to be identical across countries. The point estimates is -0.16 , significant at $< 0.1\%$.

Therefore, we would expect countries that have a faster growth rate of human capital to have a larger average cohort effect. Following the previous discussion, we measure a cohort human capital using its average schooling. For each country j , we approximate $g_{h,j}$ with the average human capital growth across cohorts: $E\left[\frac{h_{c+1,j}}{h_{c,j}}\right] = \rho_j E[s_{c+1,j} - s_{c,j}]$. We define $\Delta s_j \equiv E[s_{c+1,j} - s_{c,j}]$ for notational convenience. We run across countries the following specification, which is the empirical counterpart of equation (9)

$$\log \bar{\chi}_j = \alpha^3 + \beta^3 \Delta s_j + \varepsilon_j. \quad (10)$$

Under the assumption that all countries have the same returns to school equal to $\bar{\rho}$,

$$\begin{aligned} \alpha^3 &= 0 \\ \beta^3 &= -\frac{\gamma v \bar{\rho}}{1 - \gamma}. \end{aligned}$$

In practice, the returns to school likely differ across countries, as we just discussed, and therefore we are estimating a biased estimate of the structural parameter of interest. Specifically, let $\hat{\beta}^3$ be the estimate from regression (10), then

$$\hat{\beta}^3 \rightarrow -\frac{\gamma v \bar{\rho}}{1 - \gamma} + \frac{\text{Cov}[(\rho_j - \bar{\rho}) \Delta s_j, \Delta s_j]}{\text{Var}[s_j]}$$

where $\frac{\text{Cov}[(\rho_j - \bar{\rho}) \Delta s_j, \Delta s_j]}{\text{Var}[s_j]}$ is a bias term that goes to zero if ρ_j and Δs_j are uncorrelated.

We run regression (10) in our sample of 52 countries and find a negative and significant coefficient: $\hat{\beta}^3 = -0.059$ with standard error equal to 0.023. To visualize the fit of the correlation, in Figure 10a we plot $\log \bar{\chi}_j$ as a function of Δs_j . As predicted by the model, countries that have seen a faster increase in schooling, have also a larger cohort effect. However, possibly due to the bias just discussed, or to measurement error in our measure of growth rates of human capital, the estimated coefficient is smaller than the average of the ones estimated using across-cohorts variation.

For comparison, in Figure 10b, we plot the average year effects $-\log \bar{\psi}_j$ as a function of Δs_j : the point estimates is equal to -0.015 with standard error equal to 0.039. Faster increase in schooling does not seem to be correlated with aggregate changes in relative agricultural price or productivity. We find reassuring that human capital growth, as predicted, is positively correlated with average cohort effects, but not with average year effect.

Finally, we notice that the model also provides a structural relationship for the aggregate agricultural employment as a function of human capital. However, this last relationship depends not only on human capital, but also on relative agricultural prices/productivity, and as such it cannot be directly brought to the data.²³

²³Nonetheless, we did study how the log agricultural employment varies across countries in our dataset as a function of schooling. We show that, as expected due to the endogeneity of relative agricultural prices/productivity, which are likely to be lower in countries with higher average human capital, the coefficient on schooling is larger (it is equal to -0.22) than the ones estimated using variation across cohorts.

5.2 Policies and Human Capital: School Construction Program in Indonesia

In this section, we provide a proof of concept that it is possible to trigger reallocation out of agriculture through policies that successfully increase the educational attainment of the population. To do so, we identify the causal effect of schooling on labor reallocation out of agriculture in the context of a school construction program in Indonesia. Following the seminal work of Duflo (2001), we use the INPRES school construction program, which built 61,000 primary schools between 1974 and 1978, to provide quasi-experimental variation in schooling. While the intensity of the program, captured by the number of new schools per pupil, was not random, only some cohorts, those younger than 6 at the time the program started, were fully exposed to the program. Therefore, we can run a fairly standard difference-in-difference exercise: we compare cohorts fully exposed to the treatment to those not exposed to it, in districts with higher or lower treatment intensity. The data – the 1995 intercensal survey of Indonesia –, the identification strategy, and the specifications follow closely Duflo (2001). The reader should refer to that article for more details.

We first replicate Duflo (2001) using a dummy for agricultural employment, rather than log wages as in the original paper, and show that the program-induced exogenous increase in schooling led the affected cohorts to work less in agriculture. We then use a different specification to show that the magnitude of the effect of schooling on agricultural employment is in line with the estimates of Section 5.1.

5.2.1 Replicate Duflo (2001), but with Agricultural Employment as Outcome

The goal of this exercise is to exploit the INPRES school construction program as an exogenous source of variation in schooling attainment to study the causal effect of schooling on agricultural employment. We restrict the sample to males born between 1950 – 1977. Before showing the IV specification, we focus on the first stage and reduced form. Consider the following specification

$$y_{ijk} = \alpha_{ij} + \eta_{ik} + c_{1k} + \sum_{c=1951}^{1977} (T_{ik} \times \mathbb{I}_{ic}) \delta_c + \sum_{c=1951}^{1977} (c_{ik} \times \mathbb{I}_{ic}) \varphi_c + \epsilon_{ijk} \quad (11)$$

where (i, j, k) is an individual i , born in cohort j , and currently living in district k ; α_j is a cohort fixed effect, where the omitted cohort, η_k is a district fixed effect; T_k is treatment intensity, defined as number of school build per 1000 children; \mathbb{I}_c is a dummy that takes value equal to 1 if individual i is born in cohort c , where the control cohort is the one born in 1950; and last c_{ik} is the enrollment in 1972. The coefficients of interest are $\{\delta_c\}_{c=1951}^{1977}$, which be interpreted as an estimate of the effect of the program on a given cohort. We estimate specification (11) for three different left-hand sides: first with years of schooling, which is our first stage; second with a dummy equal to 1 for agricultural employment, which is our reduced form effect of the program; third with a dummy equal to 1 for non-agricultural employment, which is useful to control that the program does not simply lead workers to drop out of the labor force, but rather make them more likely to work in non-agriculture.

In Figures 11a, 11b, and 11c, we plot the estimates of the effect of the program by cohorts for

each one of the three left-hand side variables of interest: schooling years, agricultural employment, non-agricultural employment. The vertical dotted red line at birth cohort 1968 separates the treated cohorts from the untreated ones. Visual inspection shows a positive effect of the program on education; a negative one on agricultural employment, and a positive one on non-agricultural employment – as expected. The coefficients are normalized to average zero for the control cohorts, that should’ve been at most marginally affected by the treatment. The figures also build confidence on the exclusion restriction, to the extent that they suggest no differential trend prior to the program.²⁴ As in the original paper, coefficients are not significant.

In order to improve power, following again Duflo (2001), we focus on the comparison of two cohorts: a treatment cohort of individuals that were between 2 and 6 years old at the time the program was implemented, and a control cohort of individuals that were between 12 and 17 years old. The specification remains the same as in (11), but with only one treatment cohort, and thus one coefficient of interest, the interaction between program intensity and treatment cohort.

The first stage gives a 5%-significant point estimate equal to 0.137 (0.037): one extra school per 1000 children increases schooling by ~ 0.14 , just as in Duflo (2001). The reduced form gives a 5%-significant point estimates equal to -0.0086 (0.0043). In order to interpret the magnitude, we compute an IV where we instrument for years of schooling using the interaction between treatment intensity and treated cohort. One extra years of school significantly (at 5%) reduces agricultural employment by 6.27% (3.04%). This evidence suggests that it is possible to reduce agricultural employment through policies designed to increase educational attainment. However, the magnitude of this estimate are not comparable with those of Section 5.1, due to the different functional form. We next fill this gap.

5.2.2 Comparing Magnitudes

Recall that the model provides a structural relationship – equation (7) above – between the log agricultural employment of a cohort and its level of human capital, which we are currently measuring through its average schooling years. We estimated above, using variation across cohorts within countries, a coefficient of log agricultural employment on schooling years to be approximately -15% . We next estimate, in Indonesia, the same relationship when we instrument for years of schooling using exposure to the school construction program. Specifically, we compute for each cohort/district pair (j, k) , the average agricultural employment $l_{A,j,k}$, and then we regress,

$$\log l_{A,j,k} = \alpha_j + \eta_k + \beta^4 s_{j,k} + \epsilon_{c,t,j},$$

where years of schooling are instrumented with exposure to the program. We find that the coefficient on “exogenous” schooling years, β^4 , is equal to -0.104 (.094), but it is not precisely estimated. We notice, however, that we don’t expect to find an effect of schooling on agricultural participation for cohorts that are at a corner solution, hence that have either too little human capital for anyone to find it worthwhile to move out of agriculture, or so much human capital

²⁴When we omit the controls for children enrollment in 1972, schooling years show a slight pre-trend. For this reason, we keep the controls throughout our analysis.

that everyone prefers to work out of agriculture already. Consistently with this hypothesis, when winsorize the data at either the 1st and 99th or at the 2nd and 98th percentiles in terms of cohort agricultural employment, we obtain more precise, larger, and significant estimates for β^4 : respectively -0.181 (.105), and -0.219 (.104). We conclude that the data are mostly consistent with estimates for β^4 in the range between -0.10 and -0.20. As a comparison, the point estimates of Section 5.1 for Indonesia are $\beta_{IND}^1 = -0.127$ and $\beta_{IND}^2 = -0.157$. The estimates for the causal effect of schooling are similar in magnitude to the cross-sectional ones. We find this result reassuring, even though, admittedly, it is not obvious how the two point estimates should compare. Endogeneity would likely bias the cross-sectional estimates upward; but measurement error would bias them downward. It is also important to remark the specificity of the INPRES school construction program. It was a policy targeted at increasing primary school, and starting from a specific level of educational attainment. Moreover, the usual external validity concerns apply: we studied one observation, in one country, and in a specific historical period.

6 Conclusion

In this paper, we tackle an ambitious question; that is, how much does human capital accumulation contribute to labor reallocation out of agriculture? We develop a novel accounting framework that exploits labor reallocation by cohort to quantify the role of human capital; we then use micro level data for 52 countries to document novel empirical patterns on labor reallocation by cohorts; and, last, interpreting this evidence through the lens of the model, we conclude that human capital accounts for approximately one third of labor reallocation out of agriculture, therefore providing an answer to the question we pose. It is important, however, to qualify this answer. Our main exercise is in the spirit of the accounting literature: we have identified the proximate role of human capital for labor reallocation out of agriculture. Our results leave open the possibility that the role of human capital might be even larger, if human capital also contributed to explain changes in what we referred to as prices/productivity. At the same time, there is an even more ambitious question which is left mostly unanswered, that is: what explained human capital accumulation? In the last section of this paper, we have only scraped the surface of this topic, by using schooling and showing that a large school-construction program led to reallocation of workers out of agriculture.

7 Figures and Tables

Figure 1: Structural Change By Cohort

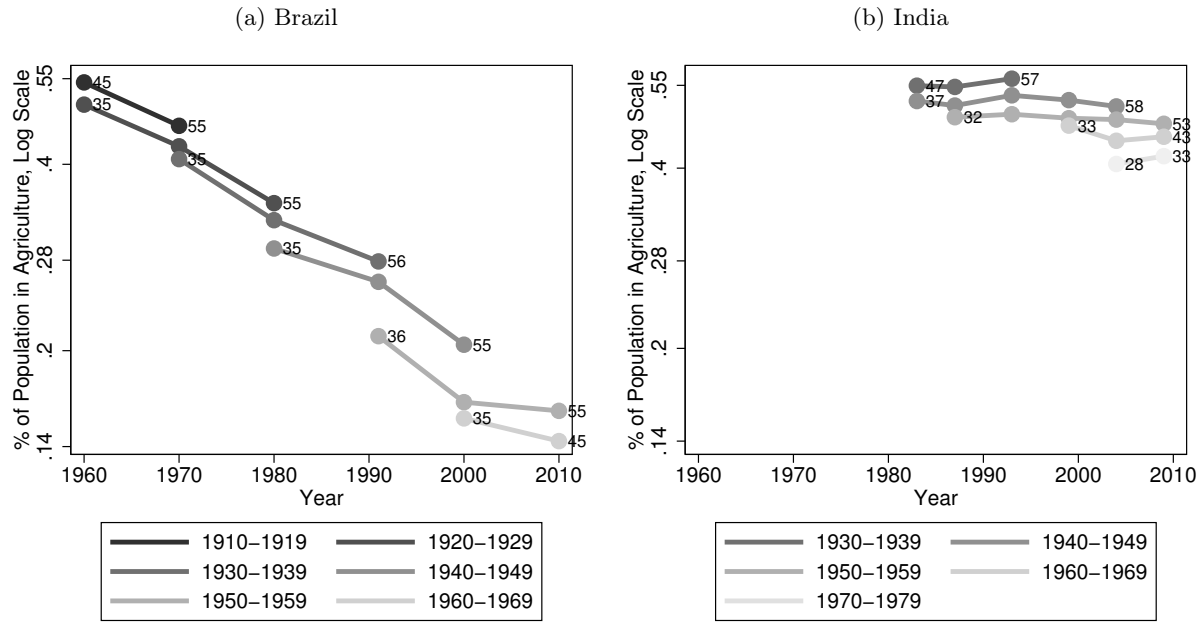


Figure 2: Distribution Across Countries (Fact 1)

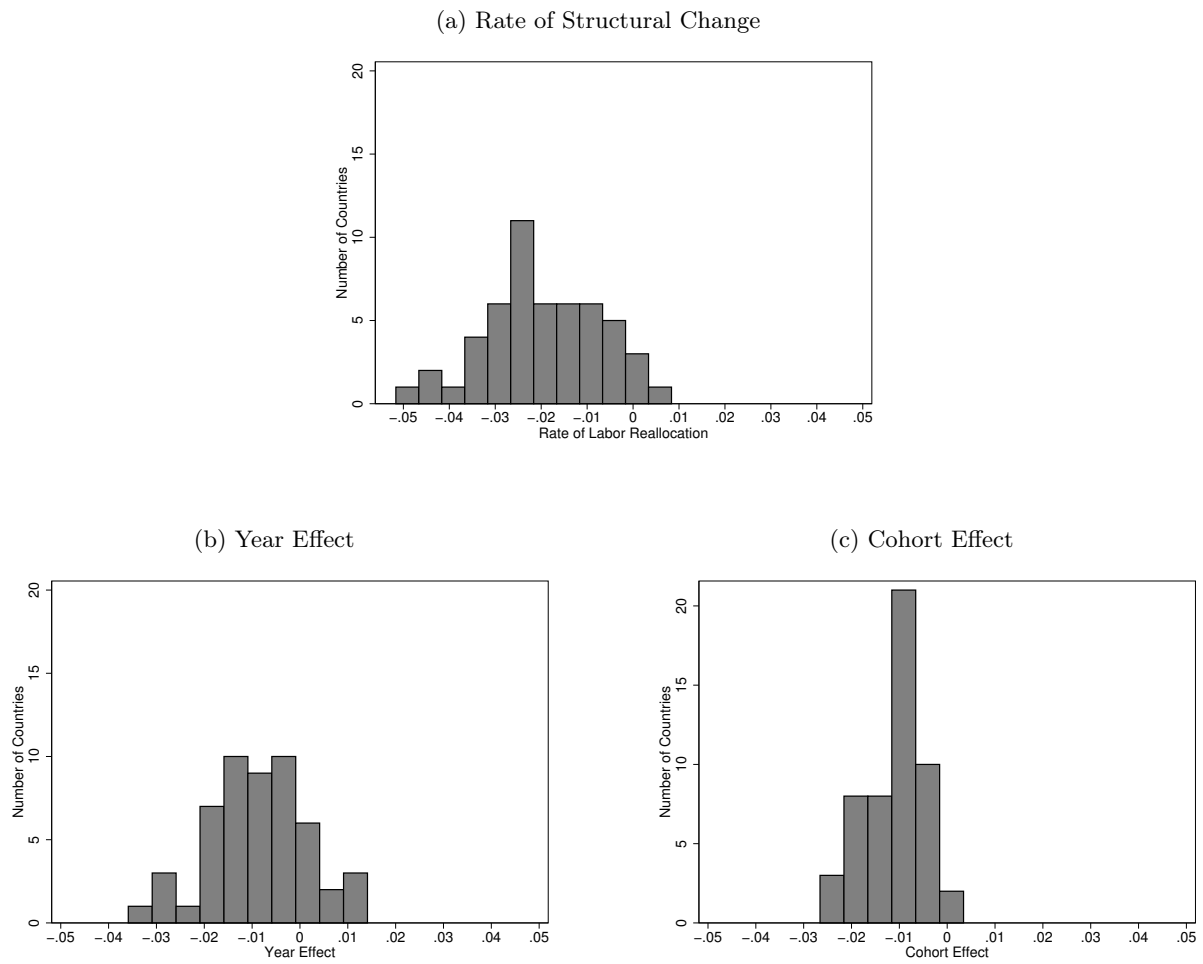


Figure 3: Variance Decomposition (Fact 2)

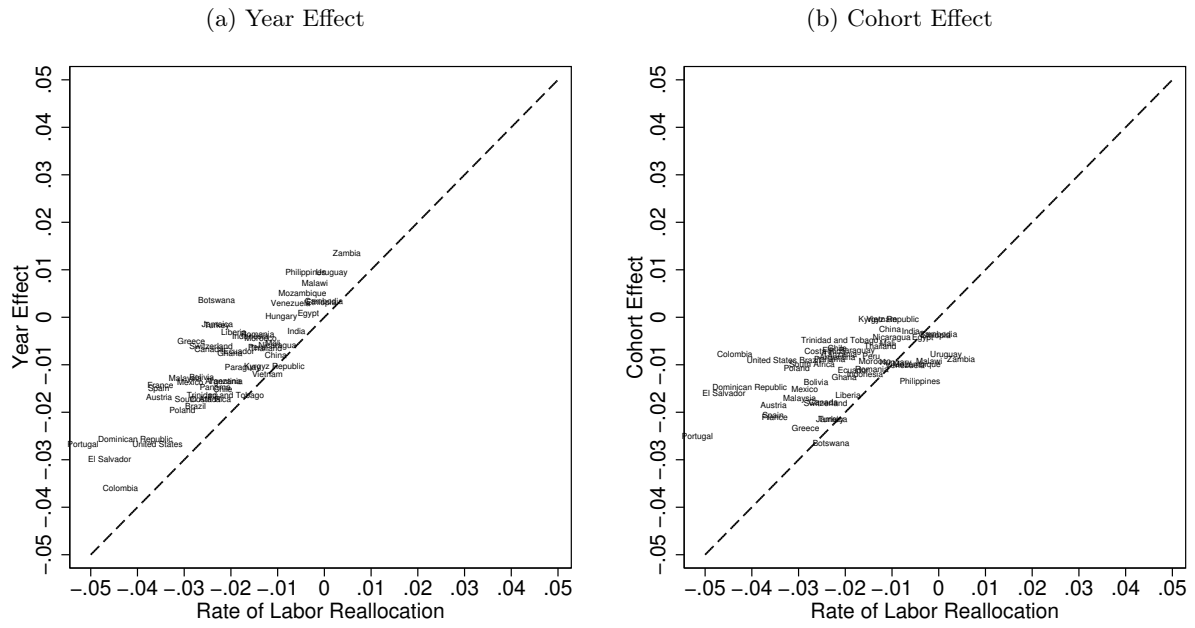


Figure 4: Comparing Year Effects Across Age Groups (Fact 3)

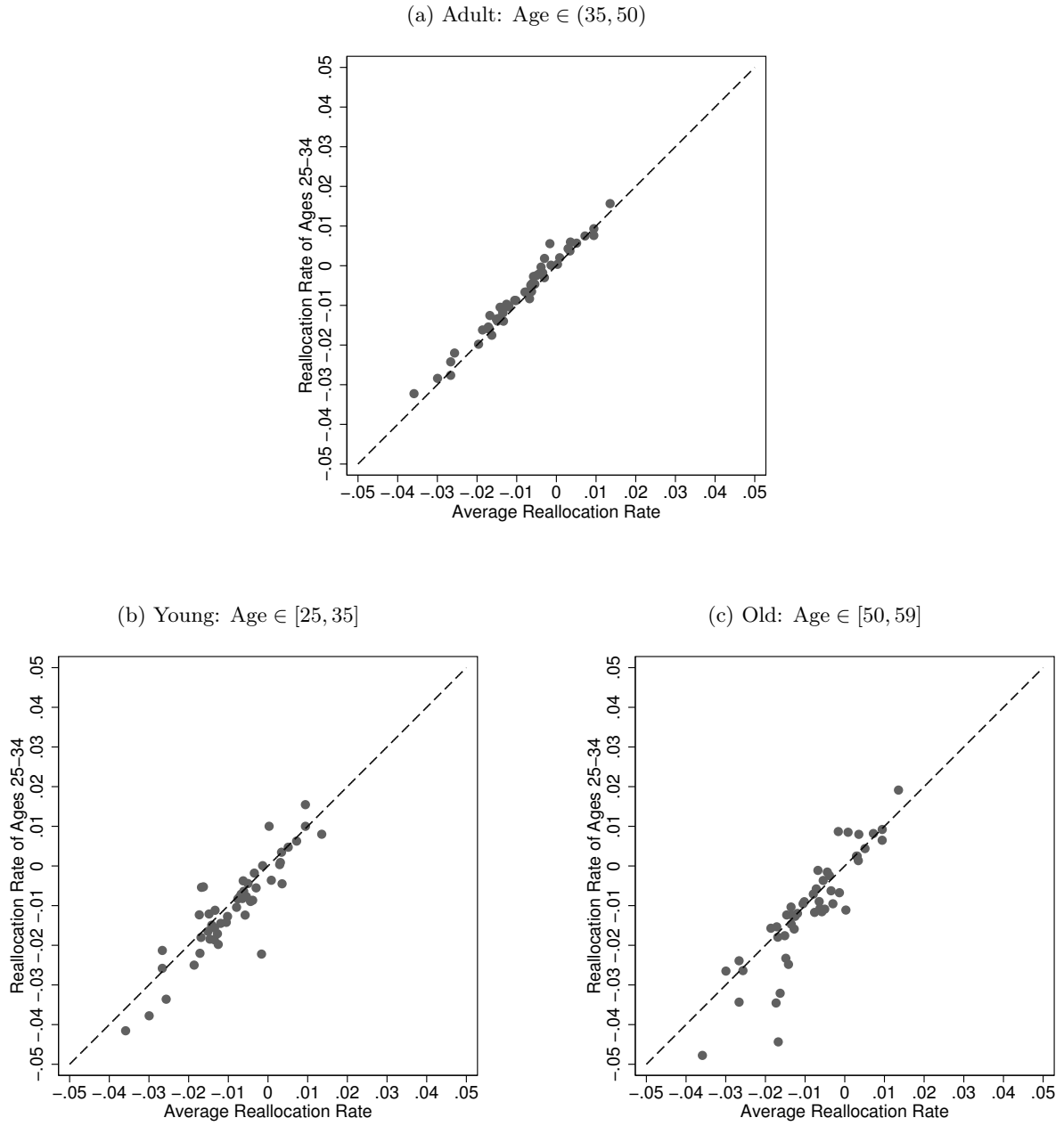


Figure 5: Distribution Across Regions

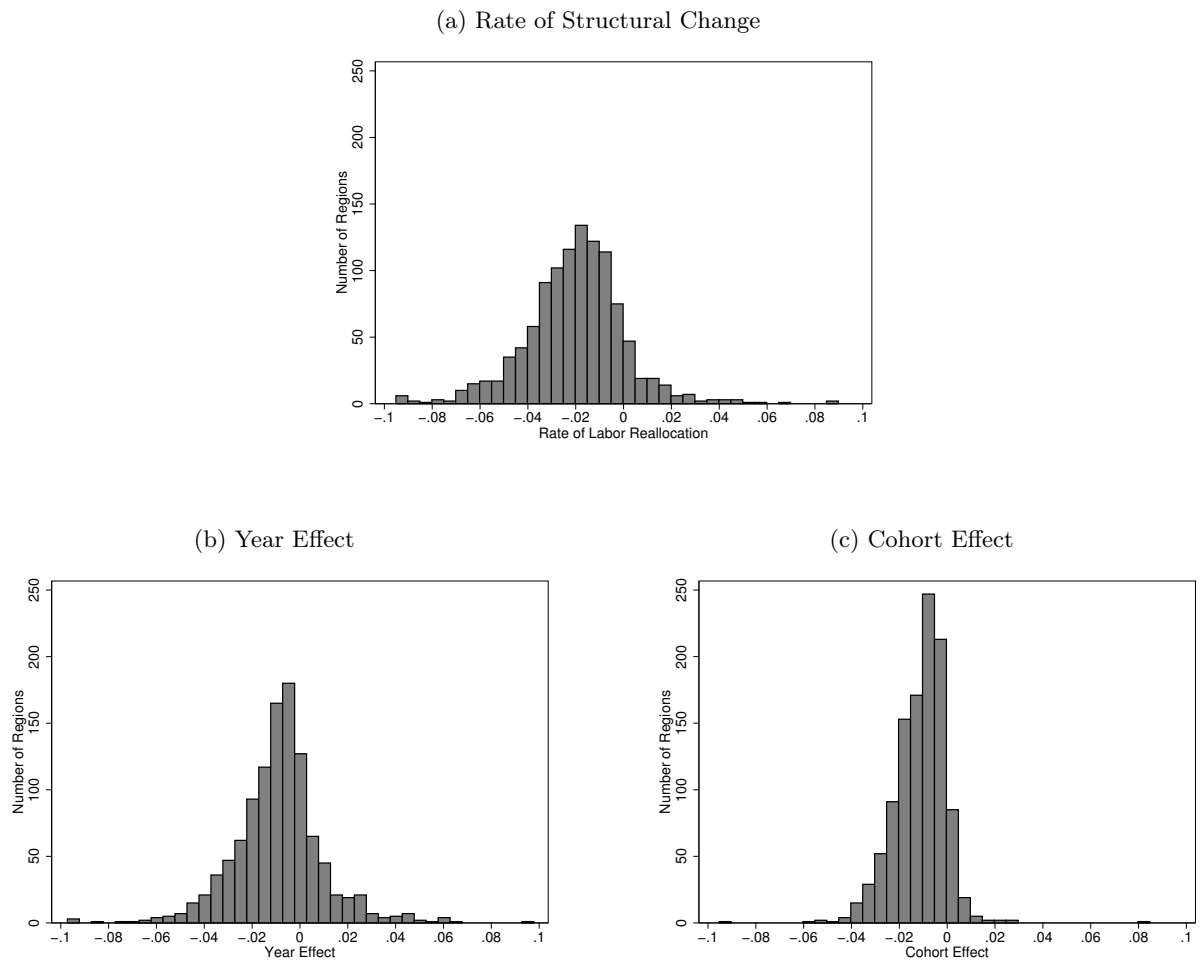


Figure 6: Variance Decomposition - Across Regions

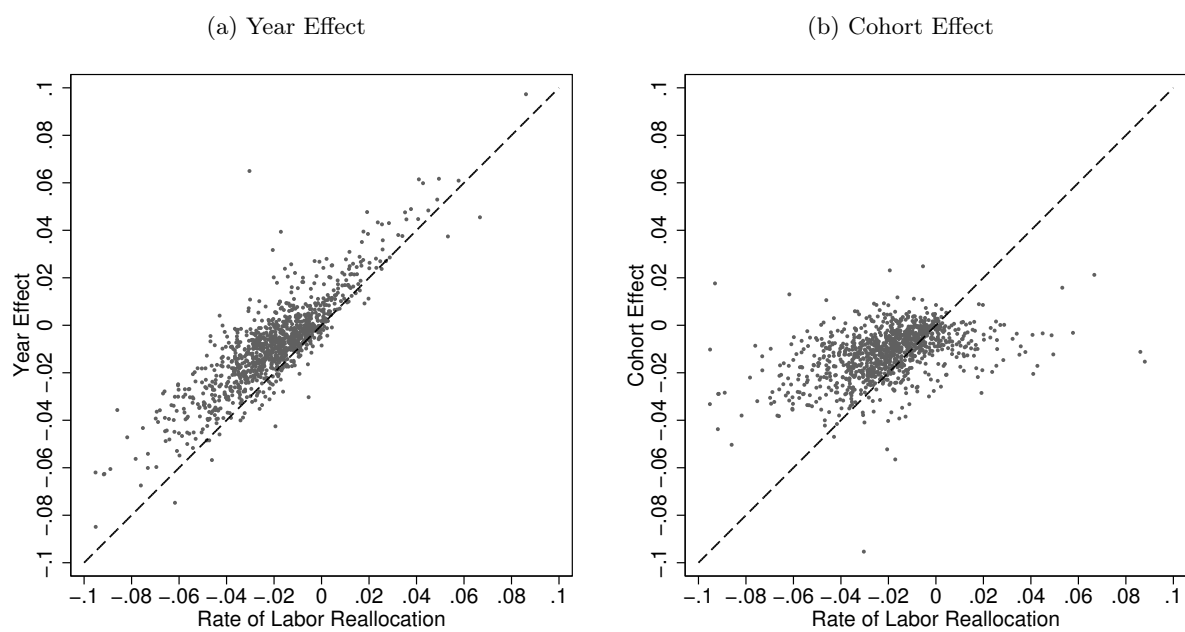


Figure 7: Comparing Year Effects Across Age Groups; by Regions

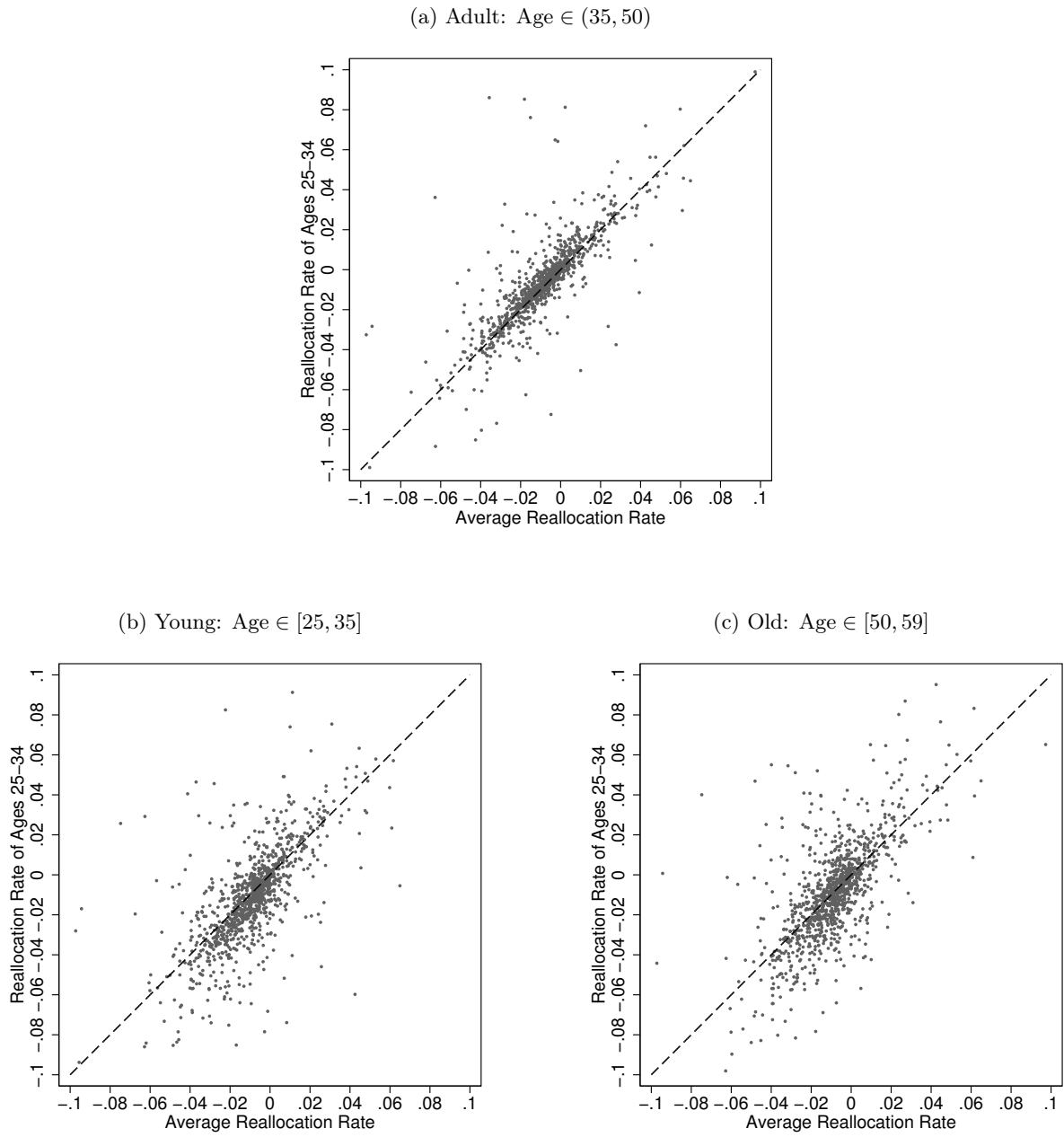
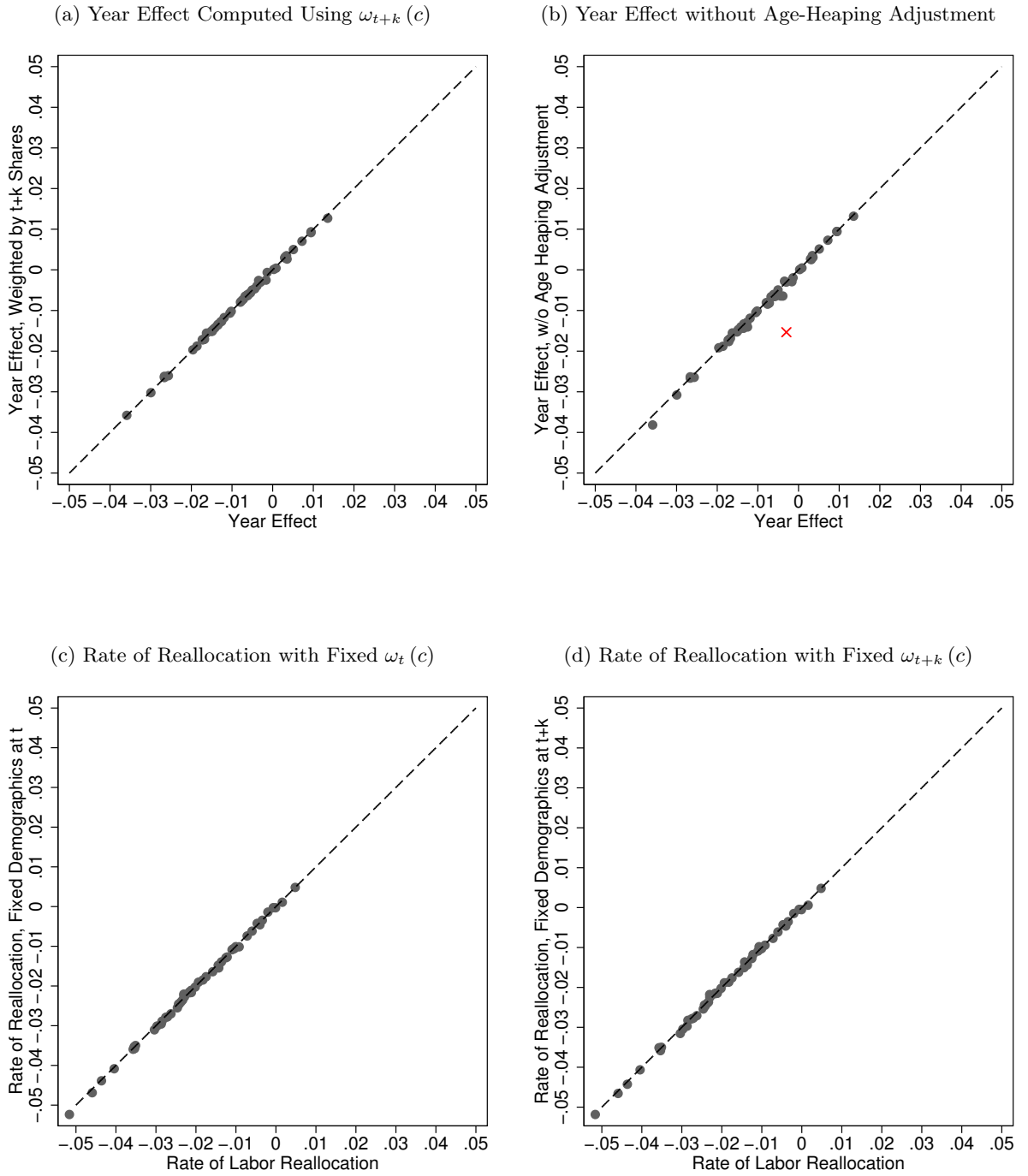


Figure 8: Robustness to Mechanical Demographics Effects



Notes: in panel (b), the red cross is India, which is an outlier due to the strong age-heaping present in its data.

Figure 9: Role of Schooling, Within Countries

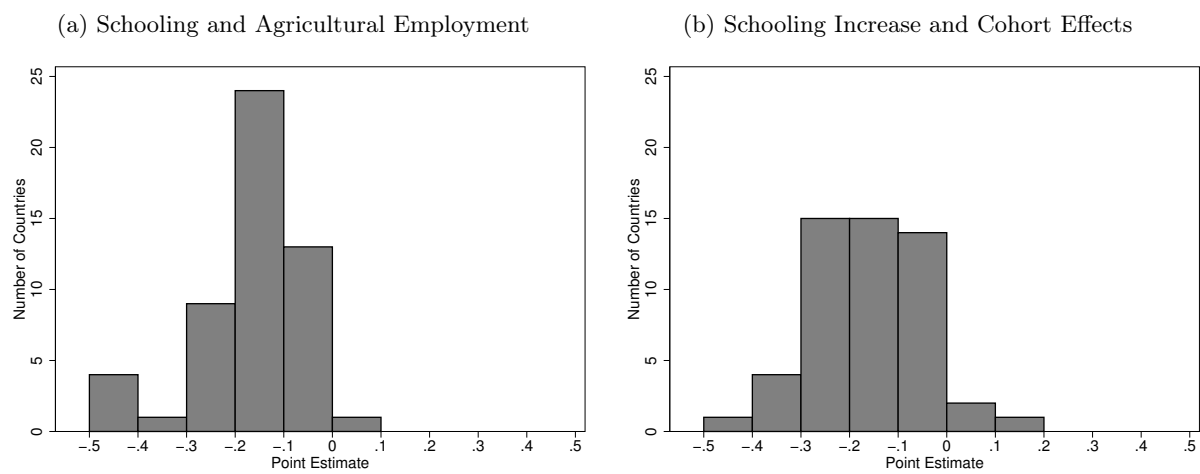
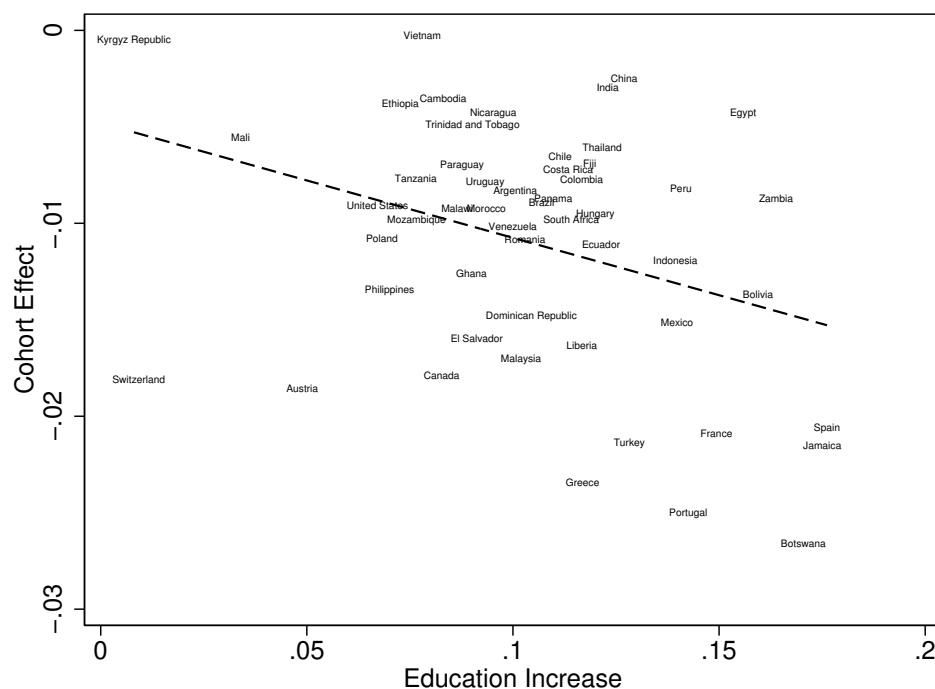


Figure 10: Role of Schooling, Across Countries

(a) Schooling Increase and Cohort Effects, Across Countries



(b) Schooling Increase and Year Effects, Across Countries

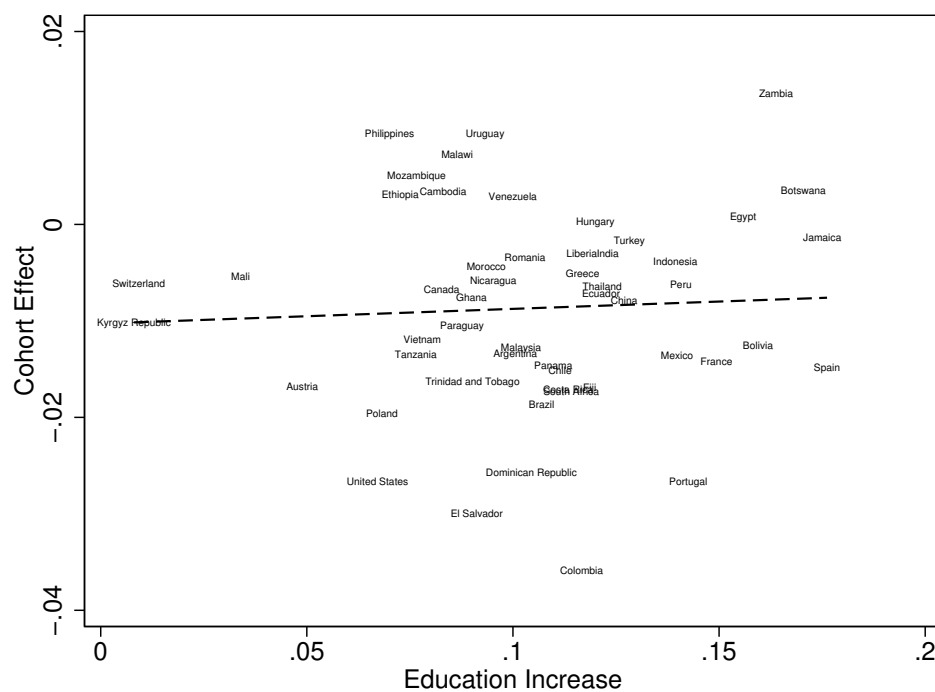
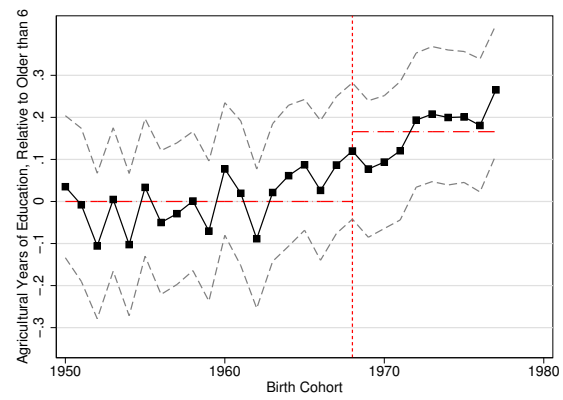
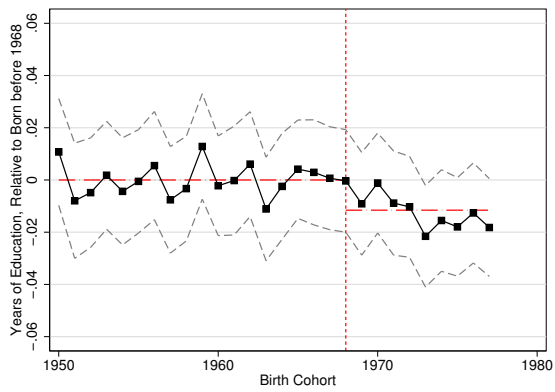


Figure 11: INPRES School Construction

(a) Point Estimates for Education



(b) Point Estimates for Agriculture



(c) Point Estimates for non-Agriculture

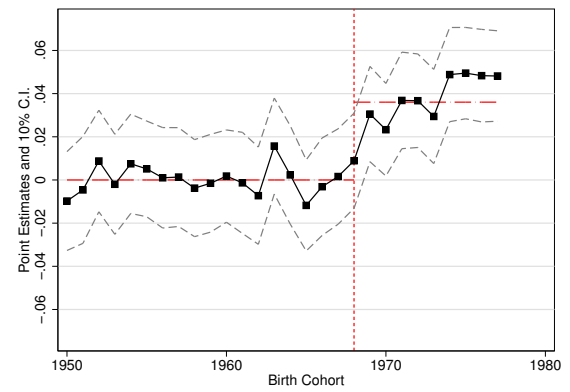


Table 1: Three Empirical Facts

	(1)	(2)	(3)	(4)	(5)	(6)
	Fact 1		Fact 2		Fact 3	
	$E[\log g_{L_A,j}]$	$\frac{E[\log \bar{\chi}_j]}{E[\log g_{L_A,j}]}$	$\text{Var}[\log g_{L_A,j}]$	$\frac{\text{Var}[\log \chi_j]}{\text{Var}[\log g_{L_A,j}]}$	$E\left[\log \frac{\psi_{j,young}}{\psi_j}\right]$	$E\left[\log \frac{\psi_{j,old}}{\psi_j}\right]$
All Countries	-1.97%	56%	0.016%	28.9%	-0.14%	-0.20%
Low Income	-1.27%	64%	0.014%	20.6%	-0.30%	-0.04%
Middle Income	-2.06%	52%	0.011%	18.5%	-0.14%	-0.14%
High Income	-3.38%	57%	0.008%	16.5%	0.24%	-0.71%
All Regions	-2.01%	56%	0.051%	17.3%	-0.16%	-0.14%
Low Income Reg.	-0.98%	81%	0.064%	5.6%	-0.24%	0.06%
Middle Income Reg.	-2.10%	52%	0.036%	21.3%	-0.24%	-0.14%
High Income Reg.	-3.34%	51%	0.037%	18.%	0.30%	-0.65%

Table 2: Estimates of the Size of the Frictions; $\lambda_j(i, f)$

	(1)	(2)	(3)	(4)	(5)	(6)
	Lemma 1.a	Lemma 1.b		Lemma 1.c		Overall Average
	$E[\lambda_j(f)]$	$\bar{\lambda}(\mathbb{J})$	$E[\bar{\lambda}(j)]$	$\lambda(f)$	$E[\lambda_j(f)]$	(w/o (5))
All Countries	14.4%	28.9%	28.4%	20.2%	50%	18.3%
Low Income	40.2%	20.6%	26.8%	12.8%	44%	20.1%
Middle Income	12.7%	18.5%	24.3%	31.2%	44%	17.3%
High Income	-20.0%	16.5%	42.1%	4.8%	74%	8.7%
All Regions	14.3%	17.3%	.	34.8%	.	22.2%
Low Income Reg.	27.1%	5.6%	.	31.9%	.	21.5%
Middle Income Reg.	19.6%	21.3%	.	30.9%	.	23.9%
High Income Reg.	-25.1%	18.0%	.	55.5%	.	16.1%

Table 3: Values of the General Equilibrium Wedge for Nine Parametrization

	$\alpha = 0.07$	$\alpha = 0.20$	$\alpha = 0.40$
$\frac{1-\gamma}{v} = 0.90$	93%	82%	69%
$\frac{1-\gamma}{v} = 0.65$	90%	76%	62%
$\frac{1-\gamma}{v} = 0.40$	85%	67%	50%

Table 4: Role of Human Capital in Labor Reallocation out of Agriculture

	(1)	(2)	(3)	(4)
	Average Rate		Dispersion of Rates	
	P.E.	G.E.	P.E.	G.E.
All Countries	46%	35%	13%	9.9%
Low Income	55%	42%	0.6%	0.4%
Middle Income	42%	32%	1.5%	1.1%
High Income	52%	40%	8.5%	6.5%

Table 5: Role of Human Capital in Labor Reallocation out of Agriculture [Intervals]

	(1)	(2)	(3)	(4)
	Average Rate		Dispersion of Rates	
	P.E.	G.E.	P.E.	G.E.
All Countries	37% – 56%	19% – 52%	-2% – 29%	-1% – 27%
Low Income	49% – 64%	24% – 60%	-13% – 21%	-7% – 20%
Middle Income	31% – 52%	16% – 48%	-16% – 19%	-8% – 18%
High Income	39% – 57%	19% – 53%	-19% – 17%	-10% – 16%

Notes: the lower bound are computed using a value $\lambda_j(f) = 30\%$ for all countries, and using the most conservative estimate for the GE wedge, 50%. The upper bound are computed using $\lambda_j(f) = 0$, and using the least conservative estimate for the GE wedge, 93%.

References

- Acemoglu, Daron and Veronica Guerrieri**, “Capital deepening and nonbalanced economic growth,” *Journal of political Economy*, 2008, *116* (3), 467–498.
- Alvarez, Jorge A**, “The Agricultural Wage Gap: Evidence from Brazilian Micro-data,” 2017.
- Barro, Robert J**, “Economic growth in a cross section of countries,” *The quarterly journal of economics*, 1991, *106* (2), 407–443.
- , “Notes on growth accounting,” *Journal of Economic Growth*, 1999, *4* (2), 119–137.
- Baumol, William J.**, “Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis,” *American Economic Review*, 1967, *57* (3).
- Benhabib, Jess and Mark M Spiegel**, “The role of human capital in economic development evidence from aggregate cross-country data,” *Journal of Monetary economics*, 1994, *34* (2), 143–173.
- Boppart, Timo**, “Structural Change and The Kaldor Facts in a Growth Model with Relative Price Effects and Non-Gorman Preferences,” *Econometrica*, 2014, pp. 2167–2196.
- Buera, Francisco J, Joseph P Kaboski, and Richard Rogerson**, “Skill Biased Structural Change,” Technical Report, National Bureau of Economic Research 2015.
- Caselli, Francesco and Wilbur John Coleman II**, “The US structural transformation and regional convergence: A reinterpretation,” *Journal of political Economy*, 2001, *109* (3), 584–616.
- Comin, Diego A, Danial Lashkari, and Martí Mestieri**, “Structural change with long-run income and price effects,” Technical Report, National Bureau of Economic Research 2015.
- Duflo, Esther**, “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, 2001, *91* (4), 795–813.
- Echevarria, Cristina**, “Changes in sectoral composition associated with economic growth,” *International economic review*, 1997, pp. 431–452.
- Gennaioli, Nicola, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer**, “Human Capital and Regional Development*,” *The Quarterly journal of economics*, 2013, *128* (1), 105–164.
- Gollin, Douglas and Christopher Udry**, “Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture,” 2017.
- , **David Lagakos, and Michael E. Waugh**, “The Agricultural Productivity Gap,” *Quarterly Journal of Economics*, 2014, *129* (2), 939–993.
- , **Stephen Parente, and Richard Rogerson**, “The role of agriculture in development,” *The American Economic Review*, 2002, *92* (2), 160–164.
- Hall, Robert E. and Charles I. Jones**, “Why Do Some Countries Produce So Much More Output Per Worker Than Others?,” *The Quarterly Journal of Economics*, February 1999, *114* (1), 83–116.

- Herrendorf, Berthold and Todd Schoellman**, “Why is Measured Productivity so Low in Agriculture?,” *Review of Economic Dynamics*, 2015.
- **and** –, “Wages, human capital, and Barriers to Structural Transformation,” *American Economic Journal: Macroeconomics*, 2017, (Forthcoming).
 - **, Christopher Herrington, and Akos Valentinyi**, “Sectoral technology and structural transformation,” *American Economic Journal: Macroeconomics*, 2015, 7 (4), 104–133.
- Hicks, Joan Hamory, Marieke Kleemans, Nicholas Y Li, and Edward Miguel**, “Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata,” Technical Report, National Bureau of Economic Research 2017.
- Hobijn, Bart, Todd Schoellman, and Alberto Vindas**, “Structural Transformation: A Race between Demographics and Technology,” Technical Report, Arizona State University 2017.
- Hsieh, Chang-Tai and Peter J. Klenow**, “Development Accounting,” *American Economic Journal: Macroeconomics*, January 2010, 2 (1), 207–23.
- **, Erik Hurst, Charles I. Jones, and Peter J. Klenow**, “The Allocation of Talent and U.S. Economic Growth,” Stanford Working Paper 2016.
- Jorgenson, DW and Z Griliches**, “The Explanation of Productivity Change,” *The Review of Economic Studies*, 1967, 34 (3), 249–283.
- Jr, Robert E Lucas**, “Life earnings and rural-urban migration,” *Journal of Political Economy*, 2004, 112 (S1), S29–S59.
- Karabarbounis, Loukas and Brent Neiman**, “The global decline of the labor share,” *The Quarterly Journal of Economics*, 2013, 129 (1), 61–103.
- Kim, Dae-II and Robert H Topel**, “Labor markets and economic growth: Lessons from Korea’s industrialization, 1970-1990,” in “Differences and changes in wage structures,” University of Chicago Press, 1995, pp. 227–264.
- King, Miriam, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick**, *Integrated Public Use Microdata Series, International: Version 6.5 [dataset]*, Minneapolis: University of Minnesota, 2017; <http://doi.org/10.18128/D020.V6.5>, 2017.
- Klenow, Pete and Andres Rodriguez-Clare**, “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?,” in “NBER Macroeconomics Annual 1997, Volume 12,” National Bureau of Economic Research, 1997, pp. 73–114.
- Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie**, “Beyond balanced growth,” *The Review of Economic Studies*, 2001, 68 (4), 869–882.
- Lagakos, David and Michael E Waugh**, “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, 2013, 103 (2), 948–80.
- **, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman**, “Life-Cycle Wage Growth Across Countries,” *Journal of Political Economy*, 2017, (Forthcoming).
 - **, Mushfiq Mobarak, and Michael E Waugh**, “The Welfare Effects of Encouraging Rural-Urban Migration,” 2017.

- Lee, Donghoon and Kenneth I Wolpin**, “Intersectoral labor mobility and the growth of the service sector,” *Econometrica*, 2006, 74 (1), 1–46.
- Mankiw, N Gregory, David Romer, and David N Weil**, “A Contribution to the Empirics of Economic Growth,” *The Quarterly Journal of Economics*, May 1992, 107 (2), 407–37.
- Matsuyama, Kiminori**, “A simple model of sectoral adjustment,” *The Review of Economic Studies*, 1992, 59 (2), 375–387.
- Morten, Melanie**, “Temporary migration and endogenous risk sharing in village india,” Technical Report, National Bureau of Economic Research 2016.
- Munshi, Kaivan and Mark Rosenzweig**, “Networks and misallocation: Insurance, migration, and the rural-urban wage gap,” *The American Economic Review*, 2016, 106 (1), 46–98.
- Nelson, Richard R and Edmund S Phelps**, “Investment in humans, technological diffusion, and economic growth,” *The American Economic Review*, 1966, pp. 69–75.
- Ngai, L Rachel and Christopher A Pissarides**, “Structural change in a multisector model of growth,” *The American Economic Review*, 2007, 97 (1), 429–443.
- Perez, Santiago**, “Railroads and the Rural to Urban Transition: Evidence from 19th-Century Argentina,” Technical Report, University of California, Davis 2017.
- Porzio, Tommaso**, “Cross-Country Differences in the Optimal Allocation of Talent and Technology,” February 2017.
- Pritchett, Lant**, “Where has all the education gone?,” *The world bank economic review*, 2001, 15 (3), 367–391.
- Young, Alwyn**, “Inequality, the Urban-Rural Gap and Migration*,” *The Quarterly Journal of Economics*, 2013, p. qjt025.