Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps across Sectors in Indonesia*

José Pulido[†] and Tomasz Święcki[‡]

This version: November 2019

Abstract

The existence of large income gaps between agricultural and non-agricultural workers in developing countries is well known. However, the source of these gaps is still being debated and the two main hypotheses - barriers to labor mobility and sorting of workers based on unobserved productivity - have opposing implications for aggregate efficiency. We use a panel of Indonesian workers to move beyond the cross-sectional gaps and document that workers moving out of agriculture see income gains of over 20% while those moving into agriculture see similar income losses, with large flows of workers in both directions. To interpret these findings, we structurally estimate a model featuring both sorting and barriers to sectoral mobility. Our estimates indicate that while self-selection is important, there is more misallocation than is suggested in the recent literature. Removing mobility barriers would lead one third of workers to reallocate and would increase aggregate output by as much as 21%.

JEL Numbers: E24, J24, J31, J62, O13, O15, O53.

Keywords: mobility barriers, sorting, income gaps, agriculture, Indonesia.

*We would like to thank Paul Beaudry, Matilde Bombardini, Todd Schoellman, Michael Waugh, as well as audiences at the 2015 Workshop on Structural Transformation at St. Andrews, 2015 WIEM, 2018 CEA, 2018 SITE Summer Workshop, 2018 LACEA, 2019 SED, NUS, and Toronto for helpful comments and suggestions. We acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.

[†]Banco de la República - Colombia. E-mail: jpulidpe@banrep.gov.co

[‡]Corresponding author, Vancouver School of Economics, University of British Columbia. E-mail: Tomasz.Swiecki@ubc.ca.

1 Introduction

The relative plight of agricultural workers compared to non-agricultural workers in developing countries is well documented. Even factoring out differences in educational attainment and other observable characteristics, non-agricultural workers on average earn substantially higher incomes (Herrendorf and Schoellman, 2018), enjoy higher consumption (Young, 2013), and have higher labor productivity (Gollin, Lagakos and Waugh, 2014) than agricultural workers. What exactly accounts for these residual gaps is still being debated, however. One long-standing view is that these gaps exist because workers cannot arbitrage them away due to broadly understood barriers to labor mobility across sectors. To the extent that these barriers are at least partially induced by policies, this view implies that labor is misallocated. Given the magnitude of the gaps, there are potentially large aggregate efficiency gains from mitigating these frictions. In this spirit, Restuccia, Yang and Zhu (2008) calculate that frictions in the allocation of labor between agriculture and non-agriculture play an important role in explaining cross-country income differences. An alternative explanation for the residual gaps that has recently gained influence is that these gaps are the result of worker sorting across sectors based on characteristics that are known to them but that are not observed by researchers. In this vein, Lagakos and Waugh (2013) adapt the classic model of Roy (1951) to show how agricultural incomes can be lower, on average, as a result of lower-ability workers sorting into agriculture. Young (2013) builds on their framework to argue that sorting can account for rural-urban disparities. Importantly, in this recent view, residual gaps can exist despite the allocation of labor being efficient.

The goal of this paper is to evaluate what the observed income gaps, along with other moments in the data, can really tell us about the presence of labor mobility barriers and worker sorting and to quantify the aggregate losses from any uncovered worker misallocation. The key to separating the two mechanisms is observing workers who actually switch sectors. We thus exploit the panel dimension of the Indonesia Family Life Survey (IFLS; Strauss, Witoelar and Sikoki, 2016). As a large-scale longitudinal survey spanning more than 20 years in the fourth most populous country in the world, the IFLS is uniquely well fitted to our goals.

We begin our analysis by documenting three robust features of the Indonesian data. First, as in other developing countries, the ILFS data shows the existence of a large cross-sectional income gap across sectors in Indonesia. Controlling for observable worker characteristics, non-agricultural workers earn 78% more than agricultural workers. Our second finding is more surprising. Despite the large difference in average incomes across sectors, we see almost as many workers transitioning from non-agriculture to agriculture as workers moving in the opposite direction during the same period. The gross flows across sectors are significantly higher than the net flows even within geographically narrow labor markets. The third finding is equally stark. Switching sectors is associated with substantial changes in individual workers' incomes. The same worker tends to earn 39% more in non-agriculture than in agriculture.

Importantly, the non-agriculture premia reported above already condition on the rural versus urban location. Much of the literature tends to associate rural employment with agriculture and urban employment with non-agriculture. This implicit isomorphism might lead to an intuition that a non-agricultural premium is to be expected even in the absence of sorting or frictions because it compensates workers for the real cost of their rural-to-urban migration. This logic would be misguided in the case of Indonesia.[1] Nearly half of the country's rural workforce has primary employment outside of agriculture and 1 in 10 urban workers is employed primarily in agriculture, so in our study we can meaningfully separate the non-agricultural and urban premia. We emphasize the sectoral dimension because most of the rural-urban residual income gap can be accounted for by differences in the sectoral compositions of rural and urban areas, combined with the large non-agriculture premium.[2] Using detailed migration data, we can go even further and isolate the spatial mobility entirely. The within-worker non-agriculture premium is surprisingly strong even within villages. Workers who move out of agriculture see an income gain of 22% while those who move into agriculture see an income loss of 23%, even if they stay in the same village.

Taken together, the three reduced-form facts - the large cross-sectional non-agriculture premium, the smaller but still significant within-worker non-agriculture premium, and the prevalence of two-way worker flows between sectors - are difficult to reconcile with the existing workhorse models of labor markets in developing countries. In particular, the two-way flows and income cuts experienced by workers switching away from non-agriculture are inconsistent with the canonical Roy model, in which worker's productivity is fixed over time. In order to account for the patterns observed in the data, we thus extend the canonical sorting model in three dimensions by introducing: (i) sector-specific time-varying idiosyncratic productivity shocks, (ii) compensating differentials, and (iii) barriers to sectoral labor mobility. The idiosyncratic sector-specific shocks (capturing, e.g., local weather shocks affecting field-level yields in agriculture) generate two-way worker flows and help explain the variability in individual incomes over time. The compensating differentials allow workers to systematically attach different values to the non-pecuniary aspects of working in the two sectors and, thus, voluntarily switch sectors despite taking an income cut. The barriers to sectoral mobility take the form of frictions preventing some individuals from working in their preferred sector and are intended to represent, e.g., exogenous job separations and life events forcing workers to switch their sector of employment even if it involves an income cut.

Armed with the model, we can revisit our motivating reduced-form findings and ask what we can infer from them. As it turns out, we can learn surprisingly little from these findings when they are taken in isolation. This is because the same pattern of cross-sectional non-agriculture premium, within-worker non-agriculture premium, and sectoral transitions can be rationalized by different combinations of idiosyncratic shocks and barriers to mobility. In particular, it is possible

[1] See Bryan and Morten (2018) for a study of costly migration in Indonesia in the presence of spatial sorting.

[2] The direct urban premia estimated at 23% in the cross-section of workers and 9% identified from rural-urban switchers are substantially smaller than the corresponding 78% and 39% non-agriculture premia.

to generate two-way flows and large within-worker premium in the absence of any barriers (or compensating differentials). When workers switch sectors in response to the sector-specific shocks, the shocks with a larger dispersion have a higher chance of taking extreme values, resulting in larger average increase in income for workers shifting to the sector with the larger variance. Thus, the within-worker premium commanded by a sector can, in principle, simply reflect this sector's higher variance of shocks. Similarly, it is possible to have large barriers to mobility despite observing zero non-agricultural premium for switching workers. The premia *by themselves* have limited empirical content in that they cannot tell us whether there is any worker misallocation or not. This result provides a caveat to recent studies downplaying the importance of intersectoral mobility frictions based on finding modest within-worker sector premia (Hicks et al., 2017; Alvarez, 2018; Herrendorf and Schoellman, 2018).

In order to separate the role of barriers to mobility and sorting, we need to impose some parametric assumptions and augment the three motivating facts with additional moments of the joint sector-income distribution over time. We use this richer set of data in the structural estimation of our model by indirect inference.

Our estimation results show that while the sorting on productivity emphasized in the recent literature clearly occurs, it is not sufficient to quantitatively capture all of the salient features of the panel data. To fit the data well, the model ultimately needs a force explaining why workers sometimes switch sectors despite taking an income cut. If we insist that the choice of the sector is always voluntary, then the model relies on an extreme preference for agriculture, through the estimated compensating differential, to rationalize why so many workers make the income-reducing move to that sector. However, we find no evidence of such a preference for agricultural work in the self-reported job satisfaction level in the IFLS.

The alternative is to recognize that not all sectoral transitions are voluntary, as we do in the model allowing for barriers to mobility. In fact, our central estimate implies that more than half of the observed transitions from non-agriculture to agriculture (but not the other way around) happen randomly rather than in response to productivity shocks. Once a worker lands in their sub-optimal sector, moving to their preferred sector is difficult. The model with such barriers to mobility not only performs best in our estimation, but its underlying mechanism is also consistent with direct evidence on job separations. For a subset of the sample, we can classify job transitions based on the self-reported reason for separation as either voluntary (e.g., looking for a higher wage) or forced (e.g., employer closed down). The data shows that forced separations are quite common (20% of all separations being a conservative estimate) and are associated with large income cuts relative to voluntary transitions. In line with the model's predictions, forced transitions are particularly prevalent among switches from non-agriculture to agriculture, while voluntary transitions are particularly prevalent among switches in the opposite direction.

The barriers to mobility are quantitatively important. To make this point, we conduct a counterfactual exercise in which the frictions are entirely removed. This thought experiment is standard in the misallocation literature, even though it is extreme because we do not know how the frictions could be completely eliminated in practice.[3] With that caveat in mind, removing all of the barriers to intersectoral mobility would result in a large reallocation of workers. Overall, 35% of the workforce would be employed in a different sector than in the baseline equilibrium. Since the initially misallocated workers reap large income gains from the reallocation, the adjustment has a sizable effect on aggregate output, increasing it by 21.5%. Agricultural employment contracts by 8 percentage points, but output and productivity increase by double digits in percentage terms in both sectors.

Related Literature

Our main contribution is to evaluate the role of barriers to mobility and self-selection in explaining the observed patterns of income differences and worker transitions across sectors. While our application is to agriculture and non-agriculture in Indonesia, our approach can be applied to other settings in which workers sort over time across sectors/occupations/firms based on their comparative advantage but subject to mobility frictions. In particular, our results show that the reduced-form premia estimated from the within-worker variation by themselves have limited empirical content in the presence of time-varying selection.

Among the extensive literature on the income gap between agriculture and non-agriculture, the most closely related work consists of a handful of studies that also exploit individual-level panel information for developing countries. Beegle, De Weerdt and Dercon (2011) offer early evidence of large individual gains in Kenya, but their focus is on consumption gains from migration rather than on the more puzzling income gains from sector switching conditional on not migrating. Alvarez (2018) using a calibrated version of the Lagakos and Waugh (2013) model finds that sorting explains most of the income gap among formal workers in Brazil. Our data on poorer Indonesian workers shows distinct patterns: agriculture accounts for a much larger share of employment, the gaps are larger, and workers are moving between sectors in both directions in large numbers. We need our richer model to explain these patterns. In concurrent work, Hicks et al. (2017) also use data from the IFLS and find smaller within-individual non-agricultural premium than we do. We explain the divergence in our reduced-form findings in Appendix B. More importantly, we argue that the non-agricultural premium by itself is not necessarily an informative statistic, and we build and estimate a structural model that allows us to quantitatively evaluate the importance of barriers to sectoral mobility.

While our results are non-experimental and the magnitudes we report depend on the structural assumptions we make, we believe our key finding of the existence of barriers to mobility is also broadly consistent with the limited existing experimental evidence. In a randomized small-scale setting, Bryan, Chowdhury and Mobarak (2014) find substantial gains from inducing workers in

[3]Job transitions involving income cuts are common even in well developed and flexible labor markets like the US, cf. Grigsby, Hurst and Yildirmaz (2019).

Bangladesh to work outside their village, though again the focus is on consumption gains from migration making direct comparison difficult. More closely, Sarvimaki, Uusitalo and Jantti (2018) using a natural experiment in Finland find large income gains for workers who abandoned farming as a result of forced migration.

Finally, our approach shares the spirit of studies using detailed individual-level panel data to interpret wage differences across sectors (Krueger and Summers, 1988) and, more recently, across firms (Abowd, Kramarz and Margolis, 1999; Taber and Vejlin, 2016; Card et al., 2018) in developed countries. However, this literature often focuses on rent-sharing arrangements between firms and their employees (such as efficiency wages and collective bargaining) or monopsony power of employers, considerations less relevant in a developing country where the predominant form of employment is working on a family farm or for a family firm.

2 Data

In this section we describe the data, but only highlighting the features of the dataset most relevant for our analysis. Comprehensive details about the design and implementation of the IFLS are reported in Strauss, Witoelar and Sikoki (2016).

Our primary data source is the Indonesia Family Life Survey. The first IFLS was conducted in 1993, with subsequent survey waves in 1997, 2000, 2007, and 2014. From the outset, the IFLS was designed as a long-term panel survey, which allows us to compare the life trajectories of individuals making different occupational choices. Further, the IFLS puts considerable effort into tracking individuals over time. This feature is rare among longitudinal household surveys in developing countries, where it is typical to lose respondents who have moved out of the original survey area. As a measure of tracking success, Thomas et al. (2012) report that the 2007 IFLS managed to interview 87% of individuals who were selected to be tracked. Tracking movers is important in a country undergoing the process of urbanization when the decision to migrate is not random.

The IFLS is a large-scale survey conducted in 13 of the 27 Indonesian provinces. As the excluded provinces are mostly the peripheral ones, the sample is representative of 83% of the Indonesian population. The first survey wave interviewed 22,019 individuals and by the fifth wave the number of respondents has grown to 58,337. In our analysis, we restrict our attention to adults (those 15 years or older) who are employed and therefore answer the questions on the survey's detailed work module. The definition of employed is expansive and comprises all persons who answered affirmatively to any of the following criteria: i) their primary activity during the past week was working, trying to work or helping to earn income; ii) had worked for pay for at least 1 hour during the past week ; iii) had a job or business, but were temporarily not working during the past week; iv) had worked at a family-owned (farm or non-farm) business during the past week.

For those individuals, the dataset we construct records their annual income, the sector where they worked according to the job that consumed the most of their time, years of schooling, work experience by sector, and standard demographic characteristics such as age and gender. In addition, we use information on the household location in each survey wave and the movements recorded in the migration module of the survey to construct individual location histories at various levels of administrative detail.

Our main outcome variable of interest is annual income. The annual income can be derived from wages, from net profits of a business (such as a farm), or from other sources such as government transfers. We believe that total income is the appropriate measure in a setting where working on a family farm is pervasive and where half of the workforce does not report any wage income.

Following a standard distinction for developing countries, we split the sectors into agriculture and non-agriculture (comprising all other sectors).[4] Locations are classified according to whether they are rural or urban, as determined by the Indonesian Central Bureau of Statistics (BPS) based on multiple criteria.

Table 1 reports the descriptive statistics for the constructed dataset. In our analysis below we focus on the 22,829 individuals we observe in at least two waves of the survey, for a total of 70,586 observations.[5]

3 Income Gaps and Transitions across Sectors

3.1 Baseline Results

In this section we present the key patterns of the income gaps and worker transitions across sectors in Indonesia. The gaps are estimated by using Mincerian regressions with the following general form

$$\ln y_{islt} = X_{it}\beta + D_N + D_U + D_i + \varepsilon_{islt}, \tag{1}$$

where y_{islt} denotes the income of individual i working in sector s (agriculture or non-agriculture) and living in location type l (rural or urban) in year t. X_{it} represents the standard individual covariates such as gender, years of education, work experience and work experience squared, as well as the year and province dummies. D_N and D_U capture the non-agriculture and urban premia of interest, while D_i captures the time-invariant component of individual heterogeneity.

The baseline specification is a reduced-form relationship between income and certain observable and unobservable worker characteristics. On the one hand, if workers randomly switch between sectors, then the D_N premium has a simple interpretation of an average gain that a worker can obtain by moving from agriculture to non-agriculture. On the other hand, if workers sort across

[4]This two-sector partition is common in the macro-development literature and is sufficient to illustrate the puzzle of low agricultural incomes. We also divided non-agriculture further into manufacturing and services. The income gaps between manufacturing and services are small relative to the gaps between these two sectors and agriculture.

[5]Depending on the specification the effective sample size can be smaller as we do not observe all variables for all individuals.

sectors based on their unobserved comparative advantage as according to Roy (1951), then the premia estimated using equation (1) need not have a simple interpretation and an exhaustive analysis requires a structural model. While our argument in this paper is that sorting is indeed important and we, therefore, estimate a structural model later in the paper, we begin by discussing the reduced-form OLS estimates as there is a long tradition of their use (e.g., Katz and Summers, 1989) and we use them as auxiliary models in our structural estimation.

As a starting point, we estimate equation (1) without any controls except for the sector dummies.[6] This specification simply compares average incomes across sectors and, as can be seen in the first column of Table 2, these incomes vary greatly. Compared to agriculture, incomes in non-agriculture are on average 84 log points [lp] (or 131%) higher.[7] The second column of Table 2 compares urban and rural incomes. The urban premium stands at a similarly dramatic 65 lp (or 91%). A natural question is whether the urban and sectoral premia capture the same variation in the data.

Many studies take a dichotomous view of economic activity in developing countries. A classical divide in the development literature goes along the rural versus urban dimension. Macroeconomists tend to work with sectoral data and hence use the agriculture versus non-agriculture split.[8] But both literatures often implicitly consider both partitions as interchangeable; for example, by associating structural transformation (a decline in the share of agricultural employment) with urbanization (an increase in the urban share). The joint distribution of workers across sectors and locations, shown in Table 1, suggests that such interchangeability is too crude in Indonesia. In 2000 (around the middle of our sample period), the share of rural workers at 59% was quite a bit higher than the 37% share of agricultural workers. Among rural workers, 45% had primary employment outside of agriculture, while 11% of the urban workforce was employed in agriculture.

A natural question is whether the raw urban premium can be explained by the different compositions of the sectors in rural and urban locations or whether urban workers are paid more in the same sectors. Column 3 of Table 2 jointly estimates the urban and sectoral premia. Controlling for sectors significantly reduces the urban premium, bringing it down to 41 lp. Controlling for the type of location has a smaller impact on the sectoral premium, still at 69 lp. These numbers are the first indication that the sector of employment might have a stronger and more direct effect on the income level than the place of residence (urban or rural).

This point is further strengthened by controlling for individual worker characteristics in the

[6] In all specifications we control for the year and province fixed effects. The results are robust to controlling for survey month effects to account for possible seasonality and controlling for alternative levels of geographical aggregation. Observations are weighted by their longitudinal survey weights and the standard errors are clustered at the level of the primary sampling units of the survey.

[7] Because the coefficients of interest are often large in magnitude we report them directly in log points and only occasionally translate them to exact percentage differences. All the reported coefficients are statistically significant at 5% level or lower.

[8] We assign workers to a sector based on their main job in a given year. In less than 10% of cases workers have a secondary job in a sector different from their main job. The secondary sector accounts for a small share of their total income, as discussed further in Appendix A.4.

Mincer regression. Column 4 shows an urban premium of 21 lp and a much larger non-agriculture premium of 57 lp. Controlling for the observables reduces the urban premium by half, while the sectoral premium again changes by much less. These residual (controlling for observables) income gaps are also about as much as what can be calculated with cross-sectional data. They therefore correspond most directly to the gaps calculated in most other studies. We summarize this discussion of cross-sectional sectoral income gaps as the following (well-known) fact.

Fact 1. *Workers in non-agriculture earn significantly more than observationally similar workers in agriculture.*

Using the panel structure of our data we are in a position to begin addressing the issue of sorting based on the unobservables. The specification in column 5 adds worker fixed effects to the set of controls. If workers sort themselves into their initial careers based on their unobservable skills but subsequently switch sectors randomly, then the cross-sectional non-agriculture premium would suffer from a selection bias. However, the worker fixed-effects premium could still be interpreted as an average gain from switching to agriculture. Using only the within-worker variation to identify the gaps reduces the urban premium by more than half, to 8 lp. While not trivial, a 9% additional income gain associated with moving from a rural to an urban location while keeping the same sector of employment is not shocking either. In contrast, the non-agriculture premium is still surprisingly large. The same worker switching from agriculture to non-agriculture without changing their rural-urban status sees, on average, an additional income gain of 33 lp (or 39%). Column 6 paints a similar picture while using a slightly more flexible specification with a full set of interactions between the sector and the urban dummies. Staying in a rural area and switching away from agriculture results in an income boost of 33 lp.

Because the large non-agricultural premium estimated for switchers is perhaps the most novel and surprising reduced-form finding of this paper, we now carefully explore the mobility pattern in our data. Panel A of Table 3 presents the count of wave-to-wave transitions between the sectors and panel C shows the associated transition matrix.[9] About 20% of workers in agriculture transition to non-agriculture between survey waves, and 12% of workers in non-agriculture switch to agriculture. Overall, 24% of workers in our sample change the sector of employment at least once. Because non-agriculture is the larger sector, these transition probabilities mean that there are almost as many cases of workers moving into agriculture as cases of moving out. This fact might seem puzzling in light of the large negative premium associated with working in agriculture. To illustrate the prevalence of these two-way flows, panel E of Table 3 reports the ratio of gross flows to net flows across sectors at various levels of spatial aggregation. At the country level, gross flows between agriculture and non-agriculture are 10 times larger than the net flows; at the province level, they are 6 times larger; even within districts - relatively small labor markets - they are over 3 times

[9] These are transitions between two consecutive observations for each worker rather than year-to-year transitions. The time between the waves of the survey varies form two to seven years.

larger.[10] We summarize these transition patterns as the following fact.

Fact 2. *Gross flows between agriculture and non-agriculture are significantly larger than net flows.*

Table 3 also records analogous transitions along the rural-urban dimension. There is less mobility between rural and urban areas, with a change in location status in 9% of cases. About 17% of workers move between rural and urban locations at least once while in our sample. As expected in a developing country, there are more than twice as many transitions from rural to urban areas than in the opposite direction, resulting in net migration to urban areas.

The sectoral and urban premia reported so far show an average effect of moving in and out of the sector or location. Given that flows in both directions are the norm, we now reevaluate the income gaps while taking into account the direction of transitions. The estimating equation now takes the form

$$\Delta \ln y_{islt} = \Delta X_{it}\beta + \Delta D_{s's} + \Delta D_{l'l} + \Delta \varepsilon_{islt}, \tag{2}$$

where $\Delta D_{s's}$ and $\Delta D_{l'l}$ capture the direction of the sectoral and locational transitions. The results are reported in the first column of Table 4. Along the locational dimension, workers who move from rural to urban areas see an income increase of 9 lp relative to those who stay in rural areas. Workers who move into rural areas incur an income shortfall of 16 lp $(= -20 - (-4))$ relative to those who stay in urban areas. Once again, the results for the non-agriculture premium are much stronger. Relative to workers who remain in agriculture, workers who switch out of agriculture see an additional income growth of 22 lp. Workers who switch from non-agriculture to agriculture see an income loss of 33 lp $(= -39 - (-7))$ relative to workers who remain in non-agriculture.

So far, we have established the existence of a significant income premium for working in non-agriculture while controlling for movements between rural and urban locations. But it is still conceivable that movements within rural and urban locations, if correlated with sector switching and having an independent effect on income, might bias the estimates of the sectoral premium. Now we completely isolate geographic mobility, using the detailed migration information provided by our dataset. We interact the direction of the sectoral transition variable $\Delta D_{ss'}$ with an indicator for whether a worker migrated across a village boundary (or correspondingly narrowly defined location for cities). The second column of Table 4 shows the results, with the reference category being workers who stay in agriculture and who stay within a village. Workers who migrate and also move out of agriculture experience the largest income gains; workers who migrate and move into agriculture suffer the largest relative income losses. But perhaps the most striking results are for workers who do not migrate: those who switch out of agriculture gain an additional 20 lp in income, relative to those who remain in agriculture in the same village. Those who switch into agriculture see an income loss of 26 lp $(= -38 - (-12))$, relative to non-movers who remain employed in non-agriculture. That such large non-agricultural premium can be identified from within-worker sector

[10] There are 217 districts in our sample, out of over 400 in Indonesia.

switches within very narrow geographical areas might seem truly surprising. We summarize these within-worker findings as the following fact.

Fact 3. *Workers switching from agriculture to non-agriculture see significant income increases, while workers switching in the opposite direction see significant income cuts.*

Taken together, the three empirical facts we document are not easily reconciled with the workhorse models of labor markets in developing countries. If workers are sorting across sectors according to a comparative advantage that is fixed over time and if switching is frictionless and occurs because of changes in sectoral prices (as in the canonical Roy model), then we could explain Fact 1 (as is done in Lagakos and Waugh, 2013). But we should not expect to see a large premium for switchers, and we should expect flows to be only in one direction.[11] If switching is costly and occurs only if the income gain justifies incurring the mobility cost, then we should see a positive premium regardless of the direction of the voluntary switch. In contrast, we see workers switching to agriculture taking systematic cuts to their incomes that are of similar magnitude as the gains for workers switching in the opposite direction. Thus, it seems that there is a pure premium associated with working in non-agriculture. There are several possible rationalizations for why the premium along with the two-way flows might exist in an equilibrium. First, allowing for shocks to comparative advantage over time can explain the two-way flows. More subtly, in principle, it can also explain the non-agricultural premium, as we demonstrate later. Second, workers might simply attach higher non-monetary value to working on a farm than to other jobs. Finally, frictions can prevent some workers from being employed in their desired sector. In order to quantify the relevance of these explanations, in the next two sections we develop and estimate a structural model capturing all these forces.

3.2 Robustness

But before we get to the model, we illustrate the robustness of our motivating reduced-form findings. In this subsection we only briefly summarize a number of robustness checks regarding the existence and interpretation of the non-agricultural premia, with details relegated to Appendix A. There we show that: (i) premia exist for both self-employed (the largest group in Indonesia) and wage workers; (ii) premia exist if we look at the wage income and consumption instead of the total income; (iii) premia are robust to allowing for heterogeneity in the Mincerian returns across sectors; (iv) premia are robust to taking secondary jobs and home production into account; (v) premia are robust to controlling for the number of hours worked; (vi) switching sectors affects income even in the long run. In Appendices A.5 and B we also compare our results to those in the concurrent work of Hicks et al. (2017), also using the IFLS data.

[11] Two-way rural-urban flows could in principle be explained by workers moving from their birthplace to a location where they can realize their comparative advantage when they enter the workforce as in Young (2013). This mechanism would not be able to account for the observed two-way sector flows within narrow geographical locations or for workers switching sectors later (or multiple times) in their working lives.

4 Model of Sorting with Barriers to Sectoral Mobility

In the previous section we establish that income differences between the agricultural and non-agricultural sectors in Indonesia are evident not only in the cross-section of workers, but also among workers who switch sectors. Furthermore, such sectoral switches occur in both directions. One of the main goals of this paper is to investigate what these reduced-form findings tell us about the extent of the self-selection and the barriers to sectoral mobility. To this end, in this section we introduce a simple discrete-time model of the labor supply in which heterogeneous workers self-select into sectors in each period based on the value of their human capital. Workers switch across sectors due to exogenous variation in the prices of human capital over time, and due to idiosyncratic time-varying sector-specific productivity shocks. In addition, there are two potential reasons why workers find themselves in a sector that does not maximize their income: they might have a preference for one sector over the other, or there might be frictions preventing them from working in their desired sector.

4.1 Basic Frictionless Economy

In the basic frictionless version of the model, workers choose the sector at each time t to maximize their contemporaneous utility.[12] Let Ω_{it} be a vector of the state variables for an individual i at time t. The income an individual receives in sector $s \in \{A, N\}$ is a product of the exogenous price of human capital in sector s at time t, R_t^s, and of the amount of human capital h^s the worker can supply to that sector:

$$y_t^s\left(\Omega_{it}\right) = R_t^s h^s\left(\Omega_{it}\right).$$

The supply of human capital depends on both the observable and unobservable components. The former are gathered in a vector of covariates X_{it}, which in our estimation includes gender, the urban-rural location, years of schooling, years of work experience and the square of the work experience. Notice that since we emphasize the sectoral dimension of the residual income premia, we abstract from the choice of location and treat the urban-rural choice just as another covariate. Since the sectoral premia are robust to the heterogeneous Mincerian returns across sectors, for simplicity we assume homogenous returns to the covariates and, hence, we focus our attention on self-selection based on the unobservable components. The set of unobservables includes a time-invariant component θ_i^s, representing the permanent productivity of worker i in sector s, and an idiosyncratic time-varying term ε_{it}^s, representing a transitory productivity shock that affects the

[12] Given our empirical findings on long-run income growth in Appendix A.6 we abstract from modeling the intertemporal choices. Turning the model into a dynamic one would greatly complicate the analysis without necessarily affecting the insights from our simpler framework.

comparative advantage of the same worker i in sector s at time t:[13]

$$h^s(\Omega_{it}) = \exp\left(X_{it}'\beta + \theta_i^s + \varepsilon_{it}^s\right).$$

As in standard selection models, the functional form assumptions on the distribution of the components of comparative advantage are key for identification. We assume that the permanent component θ_i^s is i.i.d. across individuals, drawn from a normal distribution $N(\mu_\theta, \Sigma_\theta)$. The productivity shocks are also normal i.i.d. across individuals and time, $\varepsilon_{it}^s \sim N(\mu_\varepsilon, \Sigma_\varepsilon)$ and, for identification purposes, orthogonal across sectors (i.e., Σ_ε is diagonal). We impose the normalization $\mu_\theta = \mu_\varepsilon = \mathbf{0}$ in order to identify the evolution of the prices of human capital over time.

Formally, the worker's problem is to choose where to work at the beginning of period t. At the time of the decision, they know the value of the comparative advantage components and the prices of human capital. Their problem is:

$$V(\Omega_{it}) = \max_s \{V^s(\Omega_{it})\}, \tag{3}$$

where the value of working in sector s in the frictionless economy case V_{fe}^s is simply the logarithm of the income, $V^s(\Omega_{it}) = V_{fe}^s(\Omega_{it}) = \ln y_t^s(\Omega_{it})$.

Finally, we assume the researcher observes individual income $\hat{y}_{it}^s$ subject to a pure idiosyncratic measurement error ν_{it}:

$$\ln \hat{y}_{it}^s = \ln y_t^s(\Omega_{it}) + \nu_{it}.$$

We assume the measurement error has a mean of zero and is normal i.i.d. across individuals and time, $\nu_{it} \sim N(\mathbf{0}, \sigma_\nu^2)$.[14]

For ease of reference, the set of all structural parameters, denoted as Θ, is listed and described in column 1 of Table 5.

4.2 Generalization I: Compensating Differential

In the first generalization of the basic model, workers care not only about the monetary reward from working in a given sector but they also directly value the sector's amenities (cf. Rosen, 1986; Taber and Vejlin, 2016). The value of working in sector s becomes:

$$V_{cd}^s(\Omega_{it}) = \ln y_t^s(\Omega_{it}) + \ln C^s,$$

[13]Permanent comparative advantage arises because work in the two sectors requires different fundamental skills. Comparing the self-reported job attributes in the IFLS, we find that work in agriculture relies more on physical effort while non-agriculture jobs require more concentration and people skills. Idiosyncratic productivity shocks capture events such as localized weather shocks in agriculture and shocks to profitability of a village store in non-agriculture.

[14]An alternative interpretation of this error is as an ex-post productivity shock that affects the worker's observable income, but not their sectoral choice.

where

$$C^s = \begin{cases} cd & \text{if } s = A \\ 1 & \text{if } s = N \end{cases}$$

and, thus, the compensating differential cd measures the additional utility obtained by working in agriculture (relative to working in non-agriculture) that is common to all workers. Notice that the differential acts as if it is proportionally scaling income up or down, so it can be interpreted in terms of annual earnings. This specification adds one parameter (cd) to the set of structural parameters Θ.

4.3 Generalization II: Barriers to Sectoral Mobility

In the second generalization of the model, we introduce barriers to sectoral mobility. We want to capture an idea that workers do not always get to work in the sector they would like, even if they have a strong comparative advantage in that sector. These frictions represent instances such as forced job separations (due to, e.g., the employer going out of business) or life events forcing individuals to switch to a sector where their earnings are lower. In a richer setting, such frictions could be rationalized by, e.g., on-the-job search frictions (Gautier, Teulings and Van Vuuren, 2010; Gautier and Teulings, 2015).

In our simple specification, we assume that at the beginning of each period an individual gets a random draw, such that they will be able to choose the sector they desire with probability $1-p\left(\Omega_{it}\right)$ and they will be forced to work in the other sector with probability $p\left(\Omega_{it}\right)$. The probability p of being forced to accept a job in a sector other than desired can depend on the worker's state. In particular, we want to allow for the possibility that it might be more difficult to switch sector than to keep working in the same sector, by letting the probability differ between those who desire to switch and those who desire to stay:

$$p^{s_{t-1}s_t}\left(\Omega_{it}\right) = p^{s's} = \begin{cases} p^T & \text{if } s \neq s' \\ p^S & \text{if } s = s' \end{cases}.$$

This specification adds two parameters (p^T,p^S) to Θ.

In Appendix F we present an alternative formulation of the barriers to sectoral mobility, i.e., the utility costs of switching sectors. As we explain there, this alternative formulation turns out to be less useful for explaining the patterns we see in the data.

5 Structural Estimation

In this section we describe the estimation procedure and the identification of the parameters of the structural model. This section is more technical than the rest of the paper; readers primarily interested in our substantive results might wish to just skim it before moving on to the next section.

5.1 Estimation Procedure

We use indirect inference (Gourieroux, Monfort and Renault, 1993) to estimate the model. The first step in the estimation procedure is to choose a set of auxiliary regression models that summarize the main features of the data we want to capture: the sectoral premia, the moments of the joint distribution of income, and the workers' sectoral decisions over time. These auxiliary models are used to compute the indirect inference loss function to be minimized and, hence, they must be simple to estimate multiple times. For identification, this method does not require that those auxiliary regressions are well specified (i.e., that they are exact reduced form equations derived from the structural model). However, the selected models need to provide enough information about the moments in the data to allow us to identify the set of structural parameters Θ.

We first reduce the dimensionality of Θ by estimating the Mincerian returns β in a linear regression of the log-income on the observables, while controlling for the interaction between the sectoral choice and the year. With the estimated $\hat{\beta}$, we construct the residual income net of the observables that is used for estimating the rest of the model. We can estimate β separately from the rest of Θ because the observables do not affect the self-selection of workers given our assumption of homogeneity in the Mincerian returns across sectors.[15]

To estimate the rest of the structural parameters in Θ, we select the following seven auxiliary models: i) a log-residual income linear regression on the sector choice; ii) a log-residual income linear regression on the sector choice, controlling for the individual fixed effects; iii) a log-residual income linear regression on the direction of the sector switching between the survey waves; iv) a log-residual income linear regression in first differences on the direction of the sector switching between the waves; v) a log-residual income linear regression on the interaction between the sectoral choice and the year; vi) a sectoral choice linear probability model on time dummy variables; and vii) a sectoral choice linear probability model on the previous sectoral choice. The role of each of these models in identifying the structural parameters is explained in the next subsection.

For efficiency reasons, we only use the coefficients of interest of the selected auxiliary regressions in the indirect inference loss function. Hence, we use the following 29 coefficients of the seven auxiliary models: 1-2) the non-agriculture premia in models i (cross-sectional premium) and ii (within-worker premium); 3-6) the sector-specific premia for switching workers from models iii (income of switchers relative to workers in the destination sector) and iv (income gain from switching relative to staying); 7-23) the full set of estimated coefficients from models v (evolution of the cross-sectional premium over time), vi (sectoral employment shares over time), and vii (sectoral transition

[15] A conceptually identical but numerically less efficient estimation could be performed by including β and using the log-income to compute all of the auxiliary regressions, while controlling for the observables. The homogeneity assumption is also the reason why we can safely drop the effects of the observables from our identification proof in Appendix D.

probabilities); 24-25) the sector-specific residual variance in model v; and 26-29) the sector-specific residual variance for non-switching and switching workers in model iv. Table 6 summarizes the auxiliary models as well as the selected coefficients.

Let the vector $\hat{\hat{\delta}}$ contain the values of the selected coefficients estimated on the observed data. The elements of vector $\hat{\hat{\delta}}$ remain fixed during the estimation procedure and are displayed in column 2 of Table 7. The indirect inference loss function is computed as the weighted sum of the squared differences between the values in $\hat{\hat{\delta}}$ and the values for the same set of coefficients obtained from the simulations of the structural model. For the weights, we use the factors that represent the importance of the estimated coefficient in the identification of the structural parameters of the model, assigned after extensive experimentation. Appendix C describes their magnitudes and presents the technical aspects of the estimation procedure in more detail.

In the specifications with the barriers to sectoral mobility there is an issue of endogeneity of observing the workers' initial sector allocation in the panel. We address this by introducing a pre-sample period we refer to as zero, during which a probability of being forced to work in the undesired sector is independent of the worker's state and equal to p^S. We use pre-sample information on the covariates, when available, to construct the distribution of the initial conditions. Thus, although the auxiliary regressions are computed only for the five years available in the sample, the model's data generating process produces draws also for period zero.

Our model abstracts from the possibility that workers drop out of the labor market, to focus attention on the role of sorting and the barriers to sectoral mobility in explaining the non-agriculture premia among active workers. For this reason, in the structural estimation we use the balanced panel of workers with income recorded in all of the five available waves of the IFLS.

5.2 Identification

We now discuss how the information obtained from the auxiliary regressions allows us to identify the structural parameters in Θ. We relegate the lengthy derivations to Appendix D, where we demonstrate how Θ is identified in a simplified version of the model with two periods. In the proof, we exploit the functional form assumptions to extend the standard results on the identification of a canonical log-normal Roy model from cross-sectional data to our richer model by utilizing the panel data. We fully expect our reasoning to generalize to the same setting with more than two periods, which we verify through multiple simulations.

Let us summarize the main insights from the demonstration in Appendix D. In the basic frictionless economy, sectoral decisions do not depend on workers' histories, so the model in the cross section in each period t looks much like the standard log-normal Roy model with productivity draws $\theta_i^s + \varepsilon_{it}^s$. In this case, we can use the standard arguments of Heckman and Honoré (1990) to identify the prices of human capital R_t^s and the variance matrix $\Sigma_\theta + \Sigma_\varepsilon$ (up to the variance of the measurement error σ_ν^2) from the following cross-sectional moments in each period: a) the share of workers in each sector and b) mean, c) variance, d) skewness of income observed in each sector. However, only with the panel data can we separately identify the variances of the permanent and transitory components of productivity and the variance of the measurement error. This decomposition is possible with the information from: e) the transition probabilities across sectors between periods, f) mean and g) variance of income growth for workers switching sectors. We verify that by adding this information from the switchers to the standard cross-sectional moments we can set up a system of equations with a unique solution for all of the parameters in Θ.

Generalizing the model to include a sectoral compensating differential requires only a minor adjustment to the derivations. In contrast, in the model with barriers to mobility, sectoral choices depend on workers' histories and this complicates the calculation of the relevant moments. Nevertheless, we are still able to derive closed-form solutions for all the moments (a-g), expressed in terms of moments of upper truncated multivariate normal distributions. The dimensionality of these distributions grows with the number of time periods in the panel, with the probabilities of forced switches acting as shifters in the distributions. Once again, we verify that the resulting system of equations has a unique solution for the structural parameters in Θ.

We now discuss in detail how the selected coefficients of the auxiliary regressions in our indirect inference loss function capture the set of moments (a-g) required for identification. First, the linear probability model vi describes the distribution of the sectoral choices in each cross-section (moment a). The linear probability model vii describes the average transition probabilities between waves. Combined, these models characterize the evolution of the joint distribution of sectoral choice over time and, hence, they deliver the probabilities of the sectoral transitions across all survey waves (moment e).

Second, for the cross-sectional moments, the non-agriculture premium from model i along with the sector-year interactions from model v inform us about the conditional mean income for each sector-year combination (moment b). To account for the cross-sectional variance (moment c), we collect the residual variances for the pool of workers in each sector from model v and the residual variances for non-switching workers in each sector from model iv. We do not estimate the skewness (moment d) directly, since it can be very sensitive to outliers. Instead, we substitute it with model iii, which compares the average performance of a switching worker to the peer group of stayers in the destination sector. This comparison provides us with information regarding the nature of the sorting in the data. In particular, it is informative about whether there is positive or negative hierarchical sorting into each sector. This is the same type of information as provided by the skeweness of the sectoral income distribution (Heckman and Honoré, 1990).

Finally, for the mean income growth of switchers (moment f), we use both the within-individual premium from model ii and the premia for switching workers relative to stayers in the model in first differences iv, with the latter model providing more granular information. For the variance of income growth (moment g), we use the residual variances for workers switching to each sector from model iv.

6 Results

In this section we present the structural estimation results and use them to quantify the importance of barriers to mobility and of self-selection. We begin with a discussion of the frictionless model and show that it fails to explain some salient features of the data. The model with compensating differential provides a better fit to the data but it relies on workers having an extreme preference for agriculture, which is at odds with the direct evidence from the survey. The model with barriers to mobility fits the data the best and it implies a large extent of labor misallocation in Indonesia. We end by discussing the empirical content of the reduced-form non-agriculture premia when viewed through the lens of the model.

6.1 Estimation Results

Column 2 of Table 5 shows the values of the indirect inference point estimates and the standard errors for the 16 structural parameters in the basic frictionless economy presented in section 4.1. Given these estimated parameters, the model generates the values for the 29 coefficients of the auxiliary regressions displayed in column 4 of Table 7. The last row of this table shows the value of the loss function, indicating the overall fit of the model (with smaller values indicating a better fit).

Perhaps surprisingly, the model without any frictions or differences in preferences is not only able to replicate the cross-sectional non-agriculture premium but also to generate a sizable within-individual premium. In the estimated basic frictionless economy, workers who switch from non-agriculture to agriculture see their incomes decline by 21 lp on average. This striking result can be explained by a selection effect generated by the transitory productivity shocks. As a result of this mechanism, the fixed effect premium is largely shaped by the variance of the transitory shocks across sectors. We formally state this result for a simplified version of the model in the following proposition.

Proposition 1. *Consider the basic frictionless model with two periods and human capital prices equal across sectors and over time. Then the average growth of the log income of workers switching from agriculture to non-agriculture is positive if and only if* $\sigma^2_{\varepsilon N} > \sigma^2_{\varepsilon A}$*. Further, the average growth of the log income of workers switching from non-agriculture to agriculture has the same magnitude but the opposite sign.*

Proof. See Appendix E. □

Since with two periods the fixed effects premium is simply equal to the average growth of the log income of the switchers (taken with the appropriate signs), we immediately have the following implication.

Corollary 1. *Under the same conditions as in Proposition 1, the non-agriculture premium identified from a regression with worker fixed effects is positive if and only if* $\sigma^2_{\varepsilon N} > \sigma^2_{\varepsilon A}$.

To understand these results, observe that after workers sort themselves into sectors in the first period, the only reason a worker would switch to a different sector in the next period would be due to a change in the balance of the productivity shocks, $\varepsilon^N_{it} - \varepsilon^A_{it}$. With equal variances of shocks across sectors, the average growth in income is the same for switchers in both directions, so the within-individual premium is null. But in the case of asymmetric variances, the shocks with a larger dispersion have a higher chance of taking extreme values, resulting in larger average increase in income for workers shifting to the sector with the larger variance. Thus, the sign of the non-agriculture premium after controlling for the worker fixed effects depends only on the relative size of the variance of the productivity shocks: it is positive when the variance is larger in non-agriculture and negative otherwise. This reasoning carries over quantitatively to the estimated general model with multiple periods and evolving human capital prices.

The main message from this discussion is that finding a large non-agricultural income premium after controlling for the worker fixed effects, as we find for Indonesia, *by itself* does not indicate that workers face any frictions or show intrinsic preferences when choosing their sector of employment. In principle, the premium can be explained simply by larger dispersion of productivity shocks faced by non-agricultural workers. But the patterns of the variances have observable implications for moments other than the sectoral premia. In particular, the frictionless model struggles to simultaneously account for the non-agricultural premia and the pattern of the residual variances of workers' earnings in the data (the variance is larger in agriculture). To generate the cross-sectional and fixed-effects non-agriculture premia, the frictionless model forces the relative magnitudes of the variances for both the permanent and transitory components of the comparative advantage to be opposite to the pattern observed in the residual variances. This enables it to display a relatively good fit for the premia (0.56 lp and 0.21 lp in the model versus 0.57 lp and 0.40 lp in the data for the cross-sectional and the fixed-effects premia, respectively). But this also comes at the expense of generating residual variances that are completely reversed relative to the data (compare coefficients δ_{24}, δ_{25} and δ_{26}, δ_{27} in columns 2 and 4 of Table 7). To explain jointly the premia and the patterns of the residual variances, we need to enrich the basic model.

First, we allow the two sectors to provide different non-monetary values as in section 4.2. Column 3 of Table 5 shows the estimates for the model with the compensating differential, and column 5 of Table 7 shows the corresponding coefficients of the auxiliary models. The key result in this specification is that workers show a very strong preference for agriculture. The interpretation of the estimated $\ln cd$ coefficient is that the amenity value that an individual derives from working in agriculture is equivalent to increasing her agricultural income by 61 lp (or 89%).

Because of this strong compensating differential, the model now has an easier time justifying why workers switch to agriculture. It can rationalize these workers' income cuts in terms of their preference for agricultural jobs, so it does not need to rely on the counterfactual pattern of the residual income variances. The generalized model can therefore generate both a within-individual premium that is close to the one observed in the data (0.35 lp in the model versus 0.40 lp in the data) and it can also deliver the correct qualitative patterns for the residual variances (larger variances in agriculture, see coefficients δ_{24} to δ_{27} in Table 7). In summary, the overall fit of the model with the compensating differential is noticeably better than in the case of the basic frictionless model (last row of Table 7).

A key question is whether such a strong systematic preference for working in agriculture over non-agriculture is plausible.[16] Ultimately, in a model built on revealed preferences (i.e., voluntary choices), the compensating differential is a residual force that allows the model to rationalize decisions that are otherwise difficult to explain. Fortunately, the IFLS contains information allowing us to directly test the compensating differential explanation. In the two most recent waves of the survey, workers were asked to report their job satisfaction. The results in the first two columns in Table 8 show that job satisfaction is higher on average in non-agriculture, both in a cross-section of workers and when reported by the same worker switching sectors. This result is perhaps not surprising, since non-agricultural incomes tend to be higher and income is very strongly correlated with satisfaction. But more importantly, once we control for income in columns 3 and 4, the sector of employment has no discernible effect on job satisfaction. This finding goes directly against the estimated compensating differential model, according to which workers in agriculture should be significantly more satisfied than workers in non-agriculture once we condition on their income. In the absence of such a relationship in the data, we cannot treat as satisfactory the explanation based on a utility preference for agriculture.[17]

In the second generalization of the basic model, we therefore explore an alternative conceptual approach to modeling sectoral mobility. Instead of treating all of the observed sectoral transitions as being the result of voluntary choices, the alternative is to recognize that sometimes workers switch sectors for reasons independent of their productivity. Column 4 of Table 5 reports the structural parameters of the model with barriers to sectoral mobility from section 4.3. The key parameters are the probabilities of being allocated to a different sector than the worker would desire. The probabilities can depend on whether the worker wants to switch or to remain in the same sector as in the previous period, thus, capturing the notion that switching might be more difficult than staying put. This is indeed the case: the estimated probability that a worker who wants to remain in a sector has to switch anyway is $p^S = 0.11$, whereas a worker who wanta to switch will most likely not get the chance to do so ($p^T = 0.81$). These numbers imply that 63% of the observed transitions from

[16]The role of compensating differentials in explaining pay differences and transitions across jobs has been long debated. See Rosen (1986) for a classic treatment and Sorkin (2018) for a recent application to wage differences across firms.

[17]In a recent panel study of casual rural laborers in India, Baysan et al. (2019) suggest that non-agricultural wages can be higher within a village because laborers find these jobs (concentrated in construction, brick laying and manufacturing, and mining) to be harder. In contrast, in the IFLS workers report the agricultural jobs to be more physically demanding than non-agricultural jobs.

non-agriculture to agriculture are driven by chance rather than being in response to productivity shocks. This effect is not symmetric, in that only 32% of the switches to non-agriculture are forced by randomness.

The explanation offered by this model for the prevalence of income-reducing transitions to agriculture is, thus, that these transitions are largely random events. Furthermore, once a worker finds themselves in a non-desired sector, they can become "trapped" there for a while, because it is difficult to transition to the other sector. The model with these features provides a considerably better fit to the data than the alternatives presented above, as can be seen from column 6 in Table 7. In particular, this model can closely match not only the qualitative pattern of non-agriculture premia and residual variances but also their magnitudes. This is also the only specification that can replicate the asymmetry in the magnitude of the income growth of switchers to agriculture and switchers to non-agriculture (coefficients δ_5 and δ_6) that is observed in the estimation sample.

While the model allowing for involuntary sector allocation can match the complex data patterns well, the remaining question is whether the underlying mechanism is compelling. We now present evidence suggesting that it is. For a small subset of the sample (wage workers in the last two survey waves who were either fired or had quit a job during the previous 5 years) the IFLS reports the reasons for job separation. These responses are informative about the prevalence and consequences of involuntary switches. In Table 9, we classify the separations into four groups: those clearly *voluntary* from the worker's perspective (*Wage/salary was too low, Not conducive working environment*), clearly *forced* (*Fired by the company because business was closed down/relocated/restructured, Fired for other reason, Refused being relocated*), those happening for *family/health* reasons (*Marriage, Childbirth, Other family reason, Prolonged sickness*), and the uninformative *other* category. To demonstrate that these categories have meaningful content, panel A of Table 9 relates them to the associated change in income. Relative to *voluntary* transitions, *forced* job switches result in a large 39 lp decline in wages.[18] Transitions for *family/health* reasons look much like *forced* transitions in terms of their impact on wages.[19] The *other* category lies half way between *voluntary* and *forced*, suggesting that at least some of these job transitions are also involuntary. Overall, these results support the notion that involuntary job changes are associated with income cuts.

Panel B of Table 9 reports the distribution of the reasons for job separation. Three points are noteworthy. First, involuntary separations are common. Even with the narrowest definition of involuntary separations (*forced* only), they account for one in five of the observed separations. Second, involuntary separations are particularly prevalent in transitions from non-agriculture to agriculture, while being the least prevalent in transitions in the opposite direction. Third, there is

[18] Large losses for forcibly displaced workers are well documented in other contexts, see e.g. Jacobson, LaLonde and Sullivan (1993) for the US.

[19] Whether these job switches should be interpreted as voluntary or involuntary is debatable. If a person has to accept a lower-paying job due to marriage then it could be seen as a random shock resulting in an involuntary allocation in our model of individual income maximization. It could be a voluntary decision in a richer model in which household formation and occupational choice are jointly determined. In any case, the patterns we discuss here would be qualitatively unchanged if we lumped the *family/health* group with either the *voluntary* or *forced* transitions.

an opposite pattern for voluntary separations in search of better pay. These patterns are consistent with the predictions of our model with involuntary sectoral switches.

In summary, the model with barriers to sectoral mobility driven by a chance of misalignment with the worker's desired sector fits the data well and its underlying premise is corroborated by the survey. This model is therefore our preferred specification and the basis for further analysis.[20]

6.2 Counterfactual Exercises

We now proceed to quantify the importance of mobility barriers across sectors by computing the counterfactual equilibrium in which the barriers are removed. While this counterfactual is intended to illustrate how the labor supply responds to the removal of such frictions, it is worth pointing out that this exercise lacks general equilibrium adjustments of the factor prices. Such adjustments can dampen the reallocation of workers, so our results should be regarded as an upper limit of the full impact.

We simulate the counterfactual data setting $p^S = p^T = 0$ while keeping the remaining elements of $\hat{\Theta}$ and the values of the covariates as in our baseline model. We first discuss the implications of eliminating the frictions for aggregate income and then present the sectoral outcomes. Denoting the total number of individuals in the panel as N, we compute the number of individuals reallocated after the barriers are removed, equal to M, and the fraction of the population that is reallocated, $m = \frac{M}{N}$. To decompose the impact of workers' misallocation on total income Y into its different margins, denote by Y_m the sum of the earnings of the misallocated individuals. Further, denote by ψ_m the ratio of the average income of the misallocated individuals to the average income of the population, $\psi_m \equiv \frac{NY_m}{MY}$. Then the percentage growth rate of the total income after removing the mobility frictions can be expressed as the product of three terms:

$$\Delta\%Y = m\psi_m \Delta\%Y_m.$$

Panel A of Table 10 presents the results of this calculation. The main finding is that removing barriers to workers' mobility across sectors leads to a significant reallocation of workers across sectors (35% of the total labor force) and to a large increase in the income of the misallocated workers (which doubles, on average). As a result, removing barriers produces a sizable impact in aggregate terms: a 21.5% increase in total income (pooled across all years). This effect would be even larger if the misallocated workers were average earners. However, in our estimated model the representative misallocated worker earns just over half of the average income ($\psi_m = 0.57$), largely because the misallocated workers cannot realize their full earning potential when they are employed

[20]Extending the model with barriers to mobility to also include the compensating differential has very little effect on the estimated structural parameters (cf. columns 4 and 5) in Table 5 and model fit (cf. columns 6 and 7 in Table 7). The estimated compensating differential is also small in magnitude. For these reasons we use the more parsimonious specification as our baseline. In Appendix F, we show that a model with barriers to sectoral mobility in the form of the utility costs of switching sectors significantly underperforms our preferred specification.

in the wrong sector. This fact moderates the effect of the reallocation of those workers on the adjustment in aggregate income. It is also worth noting that in our baseline specification income is the only determinant of utility, so increases in income result in identical increases in welfare.

Panel B of Table 10 breaks down the results by sector. Removing barriers to mobility would result in agricultural employment shrinking by 8.1 p.p. as a share of the total workforce. While this net change is not small, it is significantly smaller than the 35 p.p. gross flows of workers between sectors. The gross flows exceed the net flows because some workers are wrongly allocated in both sectors. Further, because the misallocated workers have on average lower productivity than the average worker in their sector, removing the misallocation increases (labor) productivity in both sectors, by 10.1% in non-agriculture and a whooping 44.4% in agriculture.[21] Consequently, output increases in both sectors. In particular, it increases by 14.2% in agriculture, despite the sector contracting in terms of employment.

In summary, our results indicate that labor is misallocated to a significant degree in Indonesia because of barriers to labor mobility across sectors. Eliminating such barriers would potentially lead to large aggregate productivity gains. Our work does not offer a practical guide to how the barriers can be eliminated in practice, but it highlights that policies that aim to ease the frictions workers face in making sectoral choices could have a large positive impact on the economy.

6.3 Industry Premia Revisited

With the structural model at our disposal, we now use it to shed more light on the empirical content of the reduced-form sectoral premia of the kind we estimated in section 3.

There is a strand in the literature arguing that if substantial cross-sectional non-agriculture premium largely disappears after controlling for worker fixed effects, then the data can be explained by efficient sorting of workers (e.g., Hicks et al., 2017; Alvarez, 2018; Herrendorf and Schoellman, 2018). In section 6.1, we explain that frictionless sorting does not imply that there should be a zero premium identified from within-worker variation. The flip side of this argument is that once we allow for barriers to sectoral mobility, the absence of the within-worker premium does not imply that the allocation is efficient. There can be many combinations of processes for the permanent and transitory components of comparative advantage draws and barriers to mobility that result in the same cross-sectional and (possibly zero) within-worker premia. To separately identify the role of frictions and sorting, we have to look beyond the industry premia at a rich set of moments observable in a panel of workers.

To illustrate this discussion, column 2 in Table 11 reports the cross-sectional and within-worker non-agriculture premia obtained from data in a counterfactual simulation removing frictions (discussed in the previous subsection). Even though the allocation is perfectly efficient in this case, the non-agriculture premium from a regression with worker fixed effects is not zero but, in fact, is

[21] The estimated processes of the permanent and transitory components of comparative advantage draws imply that both sectors are "standard" in the Roy model terminology of Heckman and Honoré (1990).

strongly negative at -31 lp. The negative premium is a natural consequence of a larger variance of the productivity shocks workers face in agriculture.

The level of the fixed-effects premium by itself, therefore, does not have clear implications for the strength of the barriers to mobility if sorting is also present. But the difference between the fixed-effects and cross-sectional premia does indeed indicate the presence of sorting. To illustrate this point, we consider an alternative counterfactual scenario in which self-selection is eliminated. Specifically, we set $\sigma^2_{\theta^A}$, $\sigma^2_{\theta^N}$, $\sigma^2_{\varepsilon^A}$, $\sigma^2_{\varepsilon^N}$ all to zero. In this case all workers are identical and would prefer non-agricultural employment as it offers higher prices for human capital. There is no sorting and both sectors employ workers because of the frictions restricting workers from selecting their preferred sector. As column 3 in Table 11 confirms, when transitions between sectors are purely random the fixed-effects premium effectively takes the same value as the cross-sectional premium.

To summarize, comparing the cross-sectional and within-worker sector premia can be a useful diagnostic for detecting self-selection. But detecting barriers to sectoral mobility in the observational data requires imposing a sufficient structure and using data that goes beyond the sectoral premia.

7 Conclusions

We present extensive reduced-form evidence of a substantial premium for working outside of agriculture in Indonesia. Not only are non-agricultural incomes around 80% higher on average but the same individual switching from agriculture to non-agriculture sees an income gain of about 20-30% while an individual switching in the opposite direction faces an income loss of a similar magnitude. Despite this non-agricultural premium, we systematically observe worker flows in both directions. We argue that in order to generate simultaneously those premia, the pattern of transitions across sectors, and the main moments of the joint distribution of income, we need to extend the models that attribute income gaps across sectors only to the sorting of workers by including barriers to sectoral mobility that misallocate workers across sectors.

We model the barriers to mobility as shocks that restrict the ability of workers to work in their desired sectors. Such frictions misallocate a large fraction of workers across sectors (35% in our baseline specification) and imply large income gains (of around 100%) for the misallocated workers when they are reallocated. As a result, output in Indonesia could increase by as much as 21% if barriers to mobility across sectors were to be removed.

In this paper we are agnostic about the root causes of the barriers to sectoral mobility. A fruitful avenue for future research would be to investigate what constitutes such barriers, why they persist, and what policies could be used to reduce their impact. There are two particular research directions we find promising. The first is to investigate the extent to which involuntary transitions capture the fact that, in the absence of effective unemployment insurance, the family farm is a fallback option for many workers in developing countries. The second is to extend our individual-level analysis to a household level. Sectoral choices that are puzzling from an individual standpoint might be easier to explain when employment decisions are jointly determined at a household level due to the prevailing social norms or missing markets.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Alvarez, Jorge. 2018. "The agricultural wage gap: Evidence from Brazilian micro-data." *American Economic Journal: Macroeconomics*, forthcoming.

Artuc, Erhan, Daniel Lederman, and Guido Porto. 2015. "A mapping of labor mobility costs in the developing world." *Journal of International Economics*, 95(1): 28 – 41.

Artuc, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *The American Economic Review*, 100(3): 1008–1045.

Baysan, Ceren, Manzoor H. Dar, Kyle Emerick, and Elisabeth Sadoulet. 2019. "The agricultural wage gap within rural villages." mimeo.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2011. "Migration and economic mobility in Tanzania: Evidence from a tracking survey." *Review of Economics and Statistics*, 93(3): 1010–1033.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." In . Vol. 5 of *Handbook of Econometrics*, , ed. James J. Heckman and Edward Leamer, 3705 – 3843. Elsevier.

Bruins, Marianne, James A. Duffy, Michael P. Keane, and Anthony A. Smith Jr. 2018. "Generalized indirect inference for discrete choice models." *Journal of Econometrics*, 205(1): 177–203.

Bryan, Gharad, and Melanie Morten. 2018. "The aggregate productivity effects of internal migration: Evidence from Indonesia." *Journal of Political Economy*, forthcoming.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh." *Econometrica*, 82(5): 1671–1748.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics*, 36(S1): S13–S70.

de Nicola, Francesca, and Xavier Gine. 2014. "How accurate are recall data? Evidence from coastal India." *Journal of Development Economics*, 106: 52 – 65.

Dix-Carneiro, Rafael. 2014. "Trade liberalization and labor market dynamics." *Econometrica*, 82(3): 825–885.

Dixit, Avinash, and Rafael Rob. 1994. "Switching Costs and Sectoral Adjustments in General Equilibrium with Uninsured Risk." *Journal of Economic Theory*, 62(1): 48 – 69.

Duncan, Greg J., and Daniel H. Hill. 1985. "An investigation of the extent and consequences of measurement error in labor-economic survey data." *Journal of Labor Economics*, 3(4): 508–532.

Gautier, Pieter A., and Coen N. Teulings. 2015. "Sorting and the output loss due to search frictions." *Journal of the European Economic Association*, 13(6): 1136–1166.

Gautier, Pieter A., Coen N. Teulings, and Aico Van Vuuren. 2010. "On-the-job search, mismatch and efficiency." *The Review of Economic Studies*, 77(1): 245–272.

Gibson, John, and Bonggeun Kim. 2010. "Non-classical measurement error in long-term retrospective recall surveys." *Oxford Bulletin of Economics and Statistics*, 72(5): 687–695.

Godlonton, Susan, Manuel A. Hernandez, and Michael Murphy. 2016. "Anchoring bias in recall data: Evidence from Central America." International Food Policy Research Institute (IFPRI) IFPRI discussion papers 1534.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2014. "The agricultural productivity gap." *The Quarterly Journal of Economics*, 129(2): 939–993.

Gourieroux, C., A. Monfort, and E. Renault. 1993. "Indirect inference." *Journal of Applied Econometrics*, 8: S85–S118.

Grigsby, John, Erik Hurst, and Ahu Yildirmaz. 2019. "Aggregate Nominal Wage Adjustments: New Evidence from Administrative Payroll Data." mimeo.

Heckman, James J., and Bo E. Honoré. 1990. "The empirical content of the Roy model." *Econometrica*, 58(5): 1121–1149.

Herrendorf, Berthold, and Todd Schoellman. 2018. "Wages, human capital and barriers to structural transformation." *American Economic Journal: Macroeconomics*, 10(2): 1–23.

Hicks, Joan Hamory, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel. 2017. "Reevaluating agricultural productivity gaps with longitudinal microdata." National Bureau of Economic Research Working Paper 23253.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *The American Economic Review*, 83(4): 685–709.

Kan, Raymond, and Cesare Robotti. 2017. "On moments of folded and truncated multivariate normal distributions." *Journal of Computational and Graphical Statistics*, 26(4): 930–934.

Katz, Lawrence F., and Lawrence H. Summers. 1989. "Industry rents: Evidence and implications." *Brookings Papers on Economic Activity*, 209–290.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, 56(2): 259–293.

Lagakos, David, and Michael E. Waugh. 2013. "Selection, agriculture, and cross-country productivity differences." *The American Economic Review*, 103(2): 948–980.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. 2008. "Agriculture and aggregate productivity: A quantitative cross-country analysis." *Journal of Monetary Economics*, 55(2): 234 – 250.

Rosen, Sherwin. 1986. "The theory of equalizing differences." In . Vol. 1 of *Handbook of Labor Economics*, 641 – 692. Elsevier.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford Economic Papers*, 3(2): 135–146.

Sarvimaki, Matti, Roope Uusitalo, and Markus Jantti. 2018. "Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations." mimeo.

Sauer, Robert M., and Christopher R. Taber. 2017. "Indirect inference with importance sampling: An application to women's wage growth." IZA Discussion Papers 11004.

Sorkin, Isaac. 2018. "Ranking Firms Using Revealed Preference." *The Quarterly Journal of Economics*, 133(3): 1331–1393.

Strauss, J., F. Witoelar, and B. Sikoki. 2016. "The fifth wave of the Indonesia Family Life Survey (IFLS5): Overview and field report." RAND working papers WR-1143/1-NIA/NICHD.

Taber, Christopher, and Rune Vejlin. 2016. "Estimation of a Roy/search/compensating differential model of the labor market." National Bureau of Economic Research Working Paper 22439.

Tallis, G. M. 1961. "The moment generating function of the truncated multi-normal distribution." *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1): 223–229.

Thomas, Duncan, Firman Witoelar, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, Cecep Sumantri, and Wayan Suriastini. 2012. “Cutting the costs of attrition: Results from the Indonesia Family Life Survey.” *Journal of Development Economics*, 98(1): 108 – 123.

Young, Alwyn. 2013. “Inequality, the urban-rural gap, and migration.” *The Quarterly Journal of Economics*, 128(4): 1727–1785.

Tables

Table 1: Descriptive Statistics

	IFLS 1: 1993	IFLS 2: 1997	IFLS 3: 2000	IFLS 4: 2007	IFLS 5: 2014
Share of male	0.60	0.62	0.59	0.58	0.57
Mean age	41.4	38.1	39.0	40.7	41.2
Mean years of schooling	5.4	6.1	7.1	7.8	8.7
Joint distribution over sectors and locations					
Total Agriculture	0.45	0.35	0.36	0.36	0.29
Rural Agriculture	0.42	0.31	0.32	0.31	0.24
Urban Agriculture	0.03	0.03	0.04	0.05	0.05
Total Non-Agriculture	0.55	0.65	0.64	0.64	0.71
Rural Non-Agriculture	0.27	0.30	0.27	0.25	0.27
Urban Non-Agriculture	0.28	0.35	0.37	0.39	0.44
Total Rural	0.69	0.62	0.59	0.56	0.50
Total Urban	0.31	0.38	0.41	0.44	0.50
No. observations	9714	12875	17931	20874	24475
Main sample: panel of workers with 2+ observations					
No. observations			70586		
No. individuals			22829		

Table 2: Sectoral and Urban Income Premia

	(1) Log Income	(2) Log Income	(3) Log Income	(4) Log Income	(5) Log Income	(6) Log Income
Non-Agriculture	0.839***		0.686***	0.574***	0.332***	
	(0.041)		(0.040)	(0.036)	(0.033)	
Urban		0.647***	0.405***	0.207***	0.084**	
		(0.045)	(0.042)	(0.036)	(0.032)	
Agr.×Urban						0.062
						(0.055)
Non-Agr.×Urban						0.416***
						(0.046)
Non-Agr.×Rural						0.326***
						(0.039)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.				Yes	Yes	Yes
Individual FE					Yes	Yes
Observations	48299	48308	48299	44494	44497	44497
R^2	0.412	0.394	0.424	0.503	0.518	0.518

Notes: Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Transitions across Sectors and Locations

(A) Distribution of Transitions across Sectors

Sector transitions	No. of cases	Share of total
AA	13214	27.68
AN	3886	8.14
NA	3546	7.43
NN	27098	56.76
Total	47744	100.00
Indiv. who switch at least once		23.89

(B) Distribution of Transitions across Locations

Location transitions	No. of cases	Share of total
RR	23299	48.79
RU	3171	6.64
UR	1166	2.44
UU	20121	42.13
Total	47757	100.00
Indiv. who switch at least once		16.91

(C) Transition Probabilities across Sectors

		Sector in T+1	
		Agricult.	Non-Agr.
Sector in T	Agricult.	0.78	0.22
	Non-Agr.	0.12	0.88

(D) Transition Probabilities across Locations

		Location in T+1	
		Rural	Urban
Location in T	Rural	0.90	0.10
	Urban	0.05	0.95

(E) Gross/Net Flows across Sectors

Spatial Unit	Ratio Gross/Net Flows
Country	9.65
Province	5.97
District	3.24

(F) Gross/Net Flows across Locations

Spatial Unit	Ratio Gross/Net Flows
Country	2.12
Province	1.76
District	1.26

Notes: XY indicates a transition from sector (or location type) X to Y between two consecutive observations for an individual. A - Agriculture, N - Non-agriculture, R - Rural, U - Urban.

Table 4: Premia for Switchers and Stayers

	(1)	(2)
	Δ Log Income	Δ Log Income
Sector transitions		
AN	0.220***	
	(0.050)	
NA	-0.392***	
	(0.049)	
NN	-0.066***	
	(0.023)	
Location transitions		
RU	0.091*	
	(0.047)	
UR	-0.199***	
	(0.058)	
UU	-0.040*	
	(0.023)	
Sector trans. × Migration		
AA × Migrate		-0.108
		(0.092)
AN × Stay		0.196***
		(0.053)
AN × Migrate		0.275**
		(0.108)
NA × Stay		-0.379***
		(0.054)
NA × Migrate		-0.472***
		(0.110)
NN × Stay		-0.117***
		(0.021)
NN × Migrate		-0.008
		(0.039)
Δ Year FE	Yes	Yes
Δ Province FE	Yes	Yes
Δ Indiv. cont.	Yes	Yes
Observations	27697	24858
R^2	0.075	0.075

Notes: XY indicates a transition from sector (or location type) X to Y between two consecutive observations for an individual. A - Agriculture, N - Non-Agriculture, R - Rural, U - Urban. Migrate indicates movement outside of the village boundary. Omitted categories: staying in agriculture (AA) and staying in the rural area (RR) in column 1; staying in agriculture within the same village (AA×Stay) in column 2. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 5: Parameter Estimates

(1) Parameter	(2) Basic frictionless	(3) Compensating differential	(4) Barriers to mobility	(5) Barriers to mobility + compensating differential
Variance of permanent comparative advantage in sector s ($\sigma^2_{\theta^s}$) and covariance ($\sigma_{\theta^{AN}}$)				
$\sigma^2_{\theta^A}$	0.29	0.52	0.41	0.40
	(0.03)	(0.05)	(0.02)	(0.02)
$\sigma^2_{\theta^N}$	0.63	0.48	0.64	0.61
	(0.04)	(0.04)	(0.03)	(0.02)
$\sigma_{\theta^{AN}}$	0.26	0.18	0.26	0.25
	(0.04)	(0.05)	(0.02)	(0.02)
Variance of transitory productivity shocks in sector s ($\sigma^2_{\varepsilon^s}$)				
$\sigma^2_{\varepsilon^A}$	0.00	0.12	0.25	0.25
	(0.00)	(0.03)	(0.02)	(0.02)
$\sigma^2_{\varepsilon^N}$	0.06	0.01	0.03	0.00
	(0.01)	(0.01)	(0.02)	(0.00)
Variance of measurement error (σ^2_ν)				
σ^2_ν	0.73	0.71	0.47	0.50
	(0.01)	(0.01)	(0.02)	(0.01)
Price of human capital in sector s at time t (R^s_t)				
R^A_1	0.80	0.47	0.77	0.77
R^A_2	1.29	0.75	1.15	1.20
R^A_3	1.18	0.62	1.10	1.10
R^A_4	1.41	0.88	1.51	1.60
R^A_5	1.74	1.12	2.00	1.94
R^N_1	1.08	1.31	1.48	1.56
R^N_2	1.74	1.94	2.20	2.18
R^N_3	1.36	1.66	1.79	1.86
R^N_4	1.77	2.16	2.15	2.09
R^N_5	2.16	2.50	2.52	2.66
Compensating differential				
$\ln cd$	–	0.61	–	0.11
		(0.04)		(0.04)
Probabilities of involuntary choices				
p^S	–	–	0.11	0.11
			(0.01)	(0.01)
p^T	–	–	0.81	0.81
			(0.02)	(0.02)

Notes: Standard errors from 100 bootstraps in parentheses. All standard errors for R^s_t are smaller than 0.1 times the point estimate value.

Table 6: Auxiliary Models and Selected Coefficients

Auxiliary model	Selected coefficients	Coefficient description
i) Log-residual income linear regression on the sector choice: $\ln \tilde{y}_{its} = c + 1\{d_{it} = N\}\, \delta_1 + D_t + \varepsilon_{ist}$	δ_1	Non-agriculture premium (cross-sectional)
ii) Log-residual income linear regression on the sector choice: $\ln \tilde{y}_{its} = c + 1\{d_{it} = N\}\, \delta_2 + D_t + D_i + \varepsilon_{ist}$	δ_2	Non-agriculture premium (within-individual)
iii) Log-residual income linear regression on the direction of sector switching: $\ln \tilde{y}_{its} = c + 1\{d_{it-1} = s, d_{it} = s'\}\, \gamma_{ss'} + D_t + \varepsilon_{ist}$	$\delta_3 = \gamma_{NA}$ $\delta_4 = \gamma_{AN} - \gamma_{NN}$	Premia for switchers to each sector relative to their peers post-switch
iv) Log-residual income linear regression in first differences on the direction of sector switching: $\Delta \ln \tilde{y}_{its} = 1\{d_{it-1} = s, d_{it} = s'\}\, \gamma_{ss'} + \Delta D_t + \varepsilon_{ist}$	$\delta_5 = \delta_{AN}$ $\delta_6 = \delta_{NA} - \delta_{NN}$	Premia for switchers to each sector relative to non-switching workers
v) Log-residual income linear regression on the interaction between sector choice and year: $\ln \tilde{y}_{its} = \delta_7 + \{1\{d_{it} = N\} \times 1\{d_{it} = t\}\}\, \gamma_{s\times t} + \varepsilon_{ist}$	δ_7 $\delta_8 = \gamma_{A\times 2} \ldots \delta_{16} = \gamma_{N\times 5}$	Constant Interactions of sector and year
vi) LPM of sector choice on time dummy variables: $1\{d_{it} = N\} = \delta_{22} + 1\{d_{it} = t\}\, \gamma_t + \varepsilon_{ist}$	δ_{17} $\delta_{18} = \gamma_2 \ldots \delta_{21} = \gamma_5$	Constant Year dummies
vii) LPM of sector choice on previous sector choice: $1\{d_{it} = N\} = \delta_{27} + 1\{d_{it-1} = N\}\, \delta_{28} + \varepsilon_{ist}$	δ_{22},δ_{23}	Constant and lagged sector choice
viii) Residual variances:	δ_{24}, δ_{25}	For workers in each sector from model v
	δ_{26}, δ_{27}	For non-switching workers in each sector from model iv
	δ_{28}, δ_{29}	For switching workers to each sector from model iv

Notes: LPM stands for linear probability model. $\tilde{y}_{its}$ is the residual income of individual i in time t working in sector s, calculated as $\ln \tilde{y}_{its} = \ln y_{its} - X'_{it}\hat{\beta}$, where y_{its} is the observed income, X'_{it} is the set of observables that includes gender, urban-rural location, years of schooling, years of work experience and the square of years of work experience, and $\hat{\beta}$ is the vector of the estimated coefficients on the observables in the log-income linear regression on the interaction between sector choice and year conditional on observables: $\ln y_{its} = \delta + X'_{it}\beta + \{1\{d_{it} = N\} \times 1\{d_{it} = t\}\}\, \gamma_{s\times t} + \varepsilon_{ist}$. D_t corresponds to the year fixed-effect (t=1,...,5) and D_i to the individual fixed-effect. Δx is the first difference of variable x. $1\{d_{it} = N\}$ is a dummy indicating whether individual i works in non-agriculture in period t, $1\{d_{it-1} = s, d_{it} = s'\}$ is a set of dummies indicating whether individual i in period $t-1$ worked in sector s and in period $t$worked in sector s', and $1\{d_{it} = t\}$ is a set of dummies indicating whether the observation of worker i corresponds to period t. The omitted category in models iii and iv is AA, in model v is $A \times 1$ and in model vi is $t = 1$.

Table 7: Coefficients of Auxiliary Regression Models

(1) Coefficient δ_i (weight Ω_i)	(2) Data ($\hat{\delta}_i$)	(3) Standard error in the data	(4) Basic frictionless	(5) Compensating differential	(6) Barriers to mobility	(7) Barriers to mobility + compensating differential
Non-agriculture premia: cross-sectional (δ_1) and within-individual (δ_2)						
δ_1 (1)	0.57	(0.03)	0.56	0.60	0.48	0.49
δ_2 (1)	0.40	(0.05)	0.21	0.35	0.40	0.41
Premia for switchers to agriculture (δ_3, δ_6) and non-agriculture. (δ_4, δ_5). The first element in (a, b) is relative to peers post-switch; the second to non-switching workers						
δ_3 (5)	-0.05	(0.06)	-0.05	-0.10	-0.04	-0.05
δ_4 (5)	-0.31	(0.05)	-0.41	-0.37	-0.24	-0.25
δ_5 (5)	0.15	(0.07)	0.21	0.31	0.24	0.24
δ_6 (5)	-0.42	(0.06)	-0.21	-0.33	-0.40	-0.40
Constant (δ_7) and coefficients on interaction sector and year ($\delta_8 : A \times 2,\ \delta_9 : A \times 3, \ldots \delta_{16} : N \times 5$)						
δ_7 (5)	-0.17	(0.10)	-0.18	-0.18	-0.18	-0.16
δ_8 (1)	0.38	(0.07)	0.47	0.45	0.41	0.43
δ_9 (1)	0.34	(0.07)	0.38	0.27	0.38	0.35
δ_{10} (1)	0.63	(0.07)	0.56	0.55	0.67	0.72
δ_{11} (1)	0.85	(0.08)	0.78	0.78	0.94	0.89
δ_{12} (5)	0.76	(0.06)	0.60	0.64	0.70	0.74
δ_{13} (1)	1.10	(0.06)	1.06	1.03	1.07	1.04
δ_{14} (1)	0.89	(0.06)	0.91	0.88	0.85	0.88
δ_{15} (1)	1.05	(0.06)	1.12	1.16	1.03	0.97
δ_{16} (1)	1.27	(0.07)	1.33	1.33	1.19	1.23
Constant (δ_{17}) and coefficients on year dummies ($\delta_{18} : t = 2,\ \delta_{19} : t = 3...$)						
δ_{17} (10)	0.70	(0.01)	0.67	0.68	0.67	0.66
δ_{18} (10)	0.01	(0.02)	0.00	-0.03	-0.03	-0.03
δ_{19} (10)	-0.02	(0.02)	-0.09	-0.02	-0.05	-0.05
δ_{20} (10)	-0.03	(0.02)	-0.04	-0.05	-0.07	-0.08
δ_{21} (10)	-0.04	(0.02)	-0.05	-0.09	-0.09	-0.09
Constant (δ_{22}) and lagged sector choice (δ_{23})						
δ_{22} (10)	0.21	(0.01)	0.20	0.22	0.16	0.15
δ_{23} (10)	0.68	(0.01)	0.66	0.62	0.71	0.72
Residual variance of workers in agriculture (δ_{24}) and non-agriculture (δ_{25})						
δ_{24} (3)	1.24	(0.04)	1.01	1.14	1.13	1.14
δ_{25} (3)	0.95	(0.03)	1.19	1.12	1.09	1.06
Residual variance of non-switching workers in agriculture (δ_{26}) and non-agriculture (δ_{27}), switching to non-agriculture (δ_{28}) and to agriculture (δ_{29})						
δ_{26} (3)	1.43	(0.06)	1.44	1.57	1.44	1.47
δ_{27} (3)	1.08	(0.04)	1.56	1.44	1.01	1.01
δ_{28} (3)	1.73	(0.14)	1.58	1.54	1.80	1.80
δ_{29} (3)	1.86	(0.14)	1.51	1.51	1.83	1.81
Overall fit (loss function)			2.013	1.462	0.414	0.380

Notes: A description of the auxiliary regressions is available in Table 6. Ω_i refers to the $i-$th element of the diagonal of matrix Ω. The overall fit is the value of the loss function being minimized by the estimation procedure.

Table 8: Self-Reported Job Satisfaction

	(1) Satisfied	(2) Satisfied	(3) Satisfied	(4) Satisfied
Non-Agriculture	0.019**	0.034**	-0.009	0.026
	(0.009)	(0.016)	(0.009)	(0.021)
Log Income			0.045***	0.028***
			(0.003)	(0.005)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes
Observations	23275	23279	19695	19698
R^2	0.026	0.015	0.043	0.021

Notes: The dependent variable is equal to one if the worker reports being Very Satisfied or Satisfied with the job and zero if Unsatisfied or Very Unsatisfied. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 9: Reason for Job Separation

(A) Wage Growth for Job Transitions by Reason

	Reason for separation				
Dep. variable	Voluntary	Forced	Family/Health	Other	Observations
Δ Log Wage	-	-0.393***	-0.447***	-0.241***	1410
	-	(0.071)	(0.072)	(0.057)	

(B) Job Transitions by Reason and Sector

Job transitions	Reason for separation (share of total)				
	Voluntary	Forced	Family/Health	Other	No. of cases
AA	22.90	17.56	23.66	35.88	131
AN	37.18	10.26	23.08	29.49	78
NA	20.86	22.46	28.34	28.34	187
NN	30.62	19.41	20.07	29.90	1669
Total	29.49	19.23	21.16	30.12	2065

Notes: Data for wage workers in IFLS wave 4 and 5 who were fired or quit in the preceding 5 years. The reported reason for separation from the previous job: voluntary: *Wage/salary was too low, Not conducive working environment;* forced: *Fired by the company because business was closed down/relocated/restructured, Fired for other reason, Refused being relocated;* family/health*: Marriage, Childbirth, Other family reason, Prolonged sickness;* other: *Other.* Panel A: Dependent variable is change in log wage between the last job and current job. *Voluntary* transitions are the omitted category. Controls: Year FE for current and last job, Province FE, Urban dummy, dummy for migrating outside of the village boundary. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Panel B: Fraction of job transitions occurring within and across sectors, broken down by reason for separation.

Table 10: Counterfactual: Aggregate Income

(A) Effect on Aggregate Income

Variable		Notation	Counterfactual
Growth rate (%) in total income: (1) ∗ (2) ∗ (3)		$\Delta\%Y_i$	21.5
			(2.3)
(1)	Fraction of the population reallocated	m	0.35
			(0.02)
(2)	Ratio of average income of reallocated workers to average income	ψ_m	0.57
			(0.02)
(3)	Growth rate (%) in total income of reallocated workers	$\Delta\%Y_m$	106.5
			(8.5)

(B) Effect on Sectoral Allocation and Productivity

Variable	Agriculture	Non-Agriculture
Baseline employment share	0.39	0.61
Counterfactual employment share	0.30	0.70
Counterfactual employment growth (%)	-21.0	13.1
Counterfactual output growth (%)	14.2	24.6
Counterfactual productivity growth (%)	44.4	10.1

Notes: The results correspond to the counterfactual exercise of eliminating involuntary switches. Panel A: Standard errors from 100 bootstraps in parentheses.

Table 11: Sectoral Premia in Counterfactuals

Coef.	(1) Baseline model	(2) No barriers	(3) No sorting
Non-agriculture premia: cross-sectional (δ_1) and within-individual (δ_2)			
δ_1	0.48	0.18	0.46
δ_2	0.40	-0.31	0.44

Notes: The baseline model is from column 6 of Table 7. No barriers imposes $p^T = p^S = 0$. No sorting imposes $\sigma^2_{\theta^A}$, $\sigma^2_{\theta^N}$, $\sigma^2_{\varepsilon^A}$, $\sigma^2_{\varepsilon^N}$ all equal to zero.

Appendix

A Robustness of Reduced-From Findings

In this Appendix we show that the existence of a non-agriculture premium documented in section 3 is robust to a number of concerns about measurement, interpretation and estimation. Our baseline point of reference is the 57 lp cross-sectional premium and 33 lp within-worker premium reported in column 4 and 5 of Table 2.

A.1 Job Type

The first exercise incorporates information on a type of job workers engage in as this helps to illuminate the nature of labor markets in Indonesia. Workers in IFLS can be consistently classified into 4 categories: self-employed, private workers, government workers and unpaid family workers. As panel A of Table A.1 reports, self-employment is the most common work status, accounting for almost half of employment. Private sector workers earning wages and salaries - a category that would usually be the focus in studies based on developed countries - constitutes less that a third of the workforce. Almost 15% of workers who typically help in household work or in a family business or farm are classified as unpaid family workers. These workers nevertheless can report income and are included in the analysis, but our results are robust to dropping this category altogether. Panel B of Table A.1 also reports the 10 most common occupations. The point of this table is to show what non-agriculture typically means in Indonesia. It is more about being a self-employed street vendor rather than having a formal factory job in manufacturing.

Controlling for the job type has a small impact on the non-agriculture premium, e.g., reducing it from 33 lp to 29 lp in the worker fixed effects regression. More interestingly, Table A.2 reports the results of interacting job type with a direction of switch. For the two main categories, self-employed and private workers, there is about 25 lp premium for switching to non-agriculture relative to staying in agriculture. Workers switching away from non-agriculture suffer a loss of similar magnitude relative to workers remaining in non-agriculture. The similarity of results for self-employed and wage workers can come as a surprise. The non-agriculture premium for wage workers could be in principle rationalized along similar lines as intersectoral or even inter-firm wage differentials documented for developed countries. There might be good non-agricultural jobs that pay more than bad agricultural jobs because employers in non-agriculture for some reason share rents with their employees. But such rent-sharing explanation would be silent as to why we see a similar premium for self-employed workers switching sectors since they are the residual claimants of their effort. The sectoral premium for the self-employed is thus our particularly surprising finding.

A.2 Wages and Consumption

While our preferred outcome variable is annual income, there can be concerns about the quality of that self-reported measure. The problem could be particularly stark for self-employed who often have to allocate family business income to individuals. As a robustness check we now restrict attention to annual wage income that is less likely to suffer from measurement problems. Doing so comes at the expense of restricting the sample by more then half to individuals who work for wages in the private or government sector. Table A.3 illustrates that the same pattern of premia can be observed using data for wages as for total income, though the magnitudes are a little smaller.

Controlling for worker fixed effects, the non-agriculture premium is 23 lp, while the urban premium 11 lp. Despite the sample size being significantly reduced the premia are still precisely estimated.

Since the IFLS records consumption expenditure, it offers an additional way of verifying that working in non-agriculture allows a higher standard of living. One drawback of consumption data in the present context is that it is recorded at a household level, whereas the focus of the paper is on individual decisions. This requires some adjustments to make the results comparable. The first column of Table A.4 reports results of a household-level cross-sectional regression of log per capita expenditure (Log PCE) on a continuous variable measuring the share of household income derived from non-agriculture and an urban dummy. Column 4 reports a corresponding calculation for per capita household income. Households that derive higher share of income from non-agriculture have a higher per capita consumption, though the elasticity is not as large as for income. The rest of Table A.4 reverts to individual level regressions, but with dependent variables still at the household level. Column 3 results indicate that if a member of a household moves from agriculture to non-agriculture than the average consumption in the household increases by over 7 lp. This might appear as a modest number compared to the baseline income premium so two comments are in order. First, since a survey worker typically accounts for less than 60% of income in his household, the coefficient should be scaled by the inverse of that share to be interpretable as an increase in consumption associated with all household workers switching to non-agriculture. This transformation would increase the non-agriculture consumption premium to about 13 lp. To illustrate that this transformation is reasonable column 6 performs it on per capita income variable. The transformed coefficient of 35 lp is very close to the baseline non-agricultural premium. Second, in similar specifications the consumption premium is still only 1/3-1/2 as large as income premium. In light of permanent income logic perhaps it should not be surprising that an income shock associated with switching sectors has only partial pass-through to consumption.

A.3 Heterogeneity in Mincerian Returns

The baseline regressions control for standard Mincerian determinants of income such as education and experience. The coefficients on these determinants do not vary between sectors and rural/urban locations, however. A recent paper by Herrendorf and Schoellman (2018) argues that this might lead to an overstatement of the residual income gaps, if, e.g, non-agriculture offers higher returns to education and experience. To address this concern, we now allow the Mincerian returns to vary be sector and location. Table A.5 reports the associated premia, calculated as the average marginal effects of switching for the population.[22] While some underlying returns do indeed differ by sector, this has no significant effect on the estimated premia of interest.

A.4 Additional Jobs and Home Production

Workers are assigned to a sector according to whether their main job is in agriculture or non-agriculture. Correspondingly, the annual income is constructed using the income from the main job. Some workers, however, have more than one job. If having a secondary job is more common for agricultural and rural workers then we might overestimate the non-agriculture and urban premia. Columns 3 and 4 of Table A.6 show the premia estimated when instead we take into account income from worker's both primary and secondary jobs. This adjustment reduces the premia by about a

[22] Results are similar if we calculate the average marginal effects for switchers instead.

fifth.[23]

A related concern is whether it is meaningful to assign a worker to a specific sector within a given year. Such an assignment could be problematic for workers frequently switching between farm and non-farm work or splitting their time equally between the two sectors. In our representative Indonesian sample working in both sectors in the same year is rare: only 9% of observations record income from a secondary job in a sector different from the main job. Furthermore, among those dual-sector workers the secondary sector accounts for only 13% of total income for a median worker (23% on average). Even restricting the sample to rural workers, only 13% report having jobs in both sectors. Thus the binary allocation to one of the two broad sectors is an accurate representation for the great majority of workers.

Another concern is that by focusing on income we are not taking into account home production which is not trivial in developing countries. If agricultural households do not include food produced and consumed in-house in their income then this could lead to an overstatement of the non-agriculture premium. The IFLS data allows us to assess how important this consideration is because it asks households to report the value of goods and services produced for own consumption. The average share of self-produced consumption is about 10%, but is predictably higher in rural areas (13%) than in urban areas (7%). As a robustness check we therefore scale up individual incomes (from both main and secondary job) by the inverse of the share of self-produced consumption in a household the individual belongs too. This effectively increases incomes of workers in rural and predominantly agricultural households. As columns 5 and 6 report, this has little effect on the estimated premia. Columns 7 and 8 consider an adjustment even more favorable for agriculture - scaling incomes by the inverse of the share of home-produced food in total food consumption. This again does not affect the estimated non-agriculture premium much, though the urban premium becomes insignificant.

A.5 Hours Worked

All the results so far show that workers in agriculture have lower annual income than workers in non-agriculture. One natural question is to what degree this income difference is driven by systematic differences in labor supply across sectors. To investigate this issue, Table A.7 adds hours worked per year to the set of individual controls. Controlling for hours worked reduces the non-agriculture premium by about a fifth. In particular, comparison of columns 2 and 4 shows that the premium identified from switchers falls from the baseline level of 33 lp to 27 lp. This reflects the fact that workers in non-agriculture work more hours, as illustrated in Table A.9. Column 2 of that table shows that the same workers supply on average 15% more hours when they switch to non-agriculture.

Whether one actually should condition on hours work in calculating the sectoral premia can be debated. The answer depends on the interpretation one wants to give to the premia and on the reason hours differ across sectors. In this paper, the non-agricultural premium is meant to capture an increase in the annual income that can be expected by a worker switching away from agriculture. To the extent that the switch is associated with higher labor supply, this increase in hours should be included as part of the benefit of switching. Our baseline measure therefore does not control for hours. In our view, thus calculated premium is a more interesting object than a premium netting out the effect of hours. The reason is that a sector of employment and supply of

[23] Using a continuous measure of the share of income a worker derives from non-agriculture instead of a dummy for the primary job leads to similar results, with a cross-sectional premium of 47 lp and 26 lp in the specification with worker fixed effects.

hours are best seen as a package. Our conjecture is that lower hours worked in agriculture observed for the same individuals are an indication that these individuals are frequently underutilized in agriculture, perhaps because of intrinsic seasonality of farm work.[24] If workers are forced to be idle for stretches of time in agriculture, then their low average utilization should be considered as a part of the productivity gap between agriculture and non-agriculture.

Another interesting feature seen in columns 3 and 4 in Table A.7 is that the elasticity of annual income with respect to annual hours worked is only about one half. This means that income per hour is declining in hours worked, consistent with diminishing returns to labor. Combining this observation with higher hours in non-agriculture explains why the non-agricultural premium in terms of income per hour (columns 5 and 6) is smaller than the premium controlling for hours (columns 3 and 4).[25] However, even when identified off switching workers income per hour is still significantly (19 lp) higher in non-agriculture. We report these numbers mainly because some of the literature interprets measures of income per hour as "wages" and uses them to calculate sectoral wage premia. In particular, in a concurrent paper also using the IFLS data Hicks et al. (2017) argue that non-agricultural premium in Indonesia largely disappears when they use their preferred regression of income per hour with worker fixed effects. There are two main reasons why our substantive findings are different. First, in our implementation we only rely on information on income and hours reported contemporaneously by the survey respondents. In contrast, Hicks et al. (2017) also rely on recall information for several years prior to the survey. As discussed in more detail in Appendix B, the recall information is likely subject to non-classical measurement error which can bias the estimated non-agricultural premium downwards. Second, even though our results are robust to controlling for hours and looking at hourly income, as argued earlier our conceptually preferred specification does not take hours into account. Comparing income per hour could indeed be preferable in a setting in which workers are offered constant hourly wages and freely choose the sector to which to allocate their marginal hour of work. But if hours are largely dictated by the nature of work in a sector then sector is the relevant "marginal" choice. Since we find the second case to be more plausible in the context of Indonesian labor markets we do not adjust our preferred non-agricultural premia for differences in hours.

A.6 Long-Run Income Growth

One of our most surprising findings is that workers who switch from non-agriculture to agriculture suffer an income loss of around 30%. To be more precise, an interpretation of coefficients in the first column of Table 4 is that a worker who switches away from non-agriculture between two survey waves has an income growth over that period 33 lp lower relative to what he would be expected to get if he remained in non-agriculture. Taking a large income cut could nevertheless be a rational decision for a worker maximizing his lifetime discounted income if he expects that the current loss of income will be compensated by higher future income growth in agriculture. This argument potentially has some merit because over our sample period average income growth was indeed higher in agriculture. To illustrate these differential trends, Table A.10 shows the evolution of the non-agricultural premium over time. While it is strong and statistically significant throughout, it does decline over our sample period, especially in the cross-section, consistent with agricultural incomes

[24]Table A.8 and column 3-4 of Table A.9 show that the results are robust to including the secondary job. This alleviates a concern that lower hours in the main job for agricultural workers are offset by having a second job.

[25]If we control for hours worked in the income per hour specifications in columns 5 and 6 then the premia would be identical to those in columns 3 and 4.

partially converging to those in non-agriculture.

However, if switching workers could accurately predict the future income path then we would expect that over a long period of time those who took a cut switching to agriculture are not worse off than workers who remained in non-agriculture. As a first test of this hypothesis we look at income growth over the entire 21-year period spanned by IFLS 1-5. Column 1 of Table A.11 shows that workers who started in non-agriculture in 1993 but switched to agriculture by 2014 had income growth over that period lower by 37 lp compared to those who began and finished in non-agriculture. This result suggests that switchers to agriculture do not make up their initial loss even after a prolonged time.

By using a single long time difference, the previous exercise identifies an average effect of switching among workers with diverse interim sectoral employment histories. Our second exercise exploits this interim information. For this purpose, we consider employment histories spanned by three observations at equal 7-year intervals (i.e. those individuals with data for 1993, 2000, 2007 or 2000, 2007, 2014). We are interested in comparing the change in income over the 14-year span for workers who made different sectoral decisions during that period. Figure A.1 shows the mean log wages for a few key histories. In particular, compare income of NAA-history workers (i.e. those who switched from non-agriculture to agriculture during the first 7-year period and stayed in agriculture during the second 7-year period) to income of NNN-history workers (who remained in non-agriculture throughout). Before the switch, NAA-workers had on average lower incomes, consistent with idea that those who switch are negatively selected from non-agricultural workers. More importantly, after the switch their incomes decline relative to NNN-workers. This is another reflection of the loss from switching emphasized in this paper. But the gap between NAA- and NNN-workers does not significantly narrow over the subsequent 7-year period. So crucially, over the entire 14-year period incomes of non-agricultural workers who permanently switched in the first half of the period fall back relative to those who stayed in non-agriculture.

Column 2 of Table A.11 casts this analysis into a regression framework with the usual controls. We find that workers who switched from non-agriculture to agriculture during the first 7-year period and were still in agriculture at the end of the second 7-year period had a cumulative growth over 14 years lower by 19 lp (significant at 0.05 level) than if they had remained in non-agriculture over this period. Similarly, workers who switched into non-agriculture in the first period and remained there had long-run income higher by 15 lp than if they had remained in agriculture, though that effect is less precisely estimated (significant at 0.10 level).[26] Overall we take these results as evidence that workers who chose agriculture have lower incomes even in the long run.

[26]In principle we could construct even longer histories which would allow us to control for pre- and post-trends of various groups. Unfortunately, between the number of possible histories increasing and the number of individuals with required data decreasing with history lengths, these longer histories would have limited statistical power.

Tables and Figures for Appendix A

Table A.1: Employment Shares by Job Type and Occupation

(A) Job Types

Job Type	Empl. share
Self-employed	0.471
Private worker	0.318
Government worker	0.068
Unpaid family worker	0.142

(B) Top Occupations

Top 10 Occupations	Empl. share
Agricultural and animal husbandry workers	0.352
Salesmen, shop assistants and related workers	0.136
Bricklayers, carpenters and other construction workers	0.038
Maids and related housekeeping service workers NEC	0.038
Working proprietors (catering and lodging services)	0.034
Transport equipment operators	0.032
Teachers	0.031
Food and beverage processors	0.027
Working proprietors (wholesale and retail trade)	0.026
Service workers NEC	0.025
Cumulative	0.739

Notes: Employment shares reported for IFLS 4 (2007).

Table A.2: Premia for Switchers and Stayers by Job Type

	(1) Self-employed	(2) Private Worker	(3) Government	(4) Unpaid Family
AN-AA	0.259***	0.245***	0.111	0.335
	18.31	11.98	0.43	1.21
NA-NN	-0.309***	-0.274***	-0.225	-0.871*
	33.61	17.89	1.02	3.79

Notes: Table presents tests based on results of a first-difference regression (2) (c.f. column 1 in Table 4) with direction of sectoral switch interacted with job type. Reported are the difference in coefficients of interest and the value of an *F(1,296)* test that the difference is zero. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.3: Wage Premia

	(1) Log Income	(2) Log Income	(3) Log Wage	(4) Log Wage
Non-Agriculture	0.574***	0.332***	0.490***	0.231***
	(0.036)	(0.033)	(0.051)	(0.050)
Urban	0.207***	0.084**	0.193***	0.119***
	(0.036)	(0.032)	(0.042)	(0.035)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes
Observations	44494	44497	23139	23140
R^2	0.503	0.518	0.556	0.601

Notes: Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.4: Consumption Premia

	(1) Log PCE	(2) Log PCE	(3) Log PCE	(4) Log PCI	(5) Log PCI	(6) Log PCI
NA sh. in HH income	0.305***			0.702***		
	(0.017)			(0.040)		
Non-Agr.		0.214***	0.075***		0.492***	0.197***
		(0.014)	(0.013)		(0.030)	(0.024)
Urban	0.315***	0.161***	0.095***	0.416***	0.225***	0.063*
	(0.029)	(0.024)	(0.026)	(0.043)	(0.034)	(0.037)
Non-Agr./$\overline{Y_{ih}/Y_h}$		0.382	0.134		0.884	0.352
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.		Yes	Yes		Yes	Yes
Individual FE			Yes			Yes
Observations	40168	53546	53550	38365	51690	51693
R^2	0.707	0.742	0.784	0.504	0.520	0.541

Notes: Specifications (1) and (4) estimated at a household level with observations weighted by longitudinal household survey weights. (1) also includes the number of household members (level and squared) as controls. *NA sh. in HH Income* is a continuous variable measuring the share of non-agriculture in household's income. Specifications (2)-(3) and (5)-(6) estimated at an individual level. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.5: Premia with Heterogeneity in Mincerian Returns

	(1)	(2)	(3)	(4)
	Log Income	Log Income	Log Income	Log Income
Non-Agriculture	0.574***	0.332***	0.625***	0.314***
	(0.036)	(0.033)	(0.039)	(0.034)
Urban	0.207***	0.084**	0.200***	0.074**
	(0.036)	(0.032)	(0.034)	(0.032)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Indiv. controls	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes
Het. in Mincer			Yes	Yes
Observations	44494	44497	44494	44497
R^2	0.503	0.518	0.506	0.520

Notes: Columns (3) and (4) allow for differences in Mincerian returns across sectors and locations. Average marginal effect for the population reported. Average effects for switchers are similar. Individual Mincerian controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.6: Premia with Additional Jobs and Home Production

	Base	Base	Add. Job	Add. Job	Add+HH TC	Add+HH TC	Add+HH FC	Add+HH FC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log Income	Log Income	Log Income	Log Income	Log Income	Log Income	Log Income	Log Income
Non-Agr.	0.574***	0.332***	0.501***	0.264***	0.462***	0.251***	0.447***	0.245***
	(0.036)	(0.033)	(0.034)	(0.032)	(0.033)	(0.032)	(0.032)	(0.032)
Urban	0.207***	0.084**	0.171***	0.063*	0.141***	0.057*	0.124***	0.051
	(0.036)	(0.032)	(0.034)	(0.034)	(0.033)	(0.034)	(0.033)	(0.034)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes		Yes		Yes
Observations	44494	44497	44489	44492	44489	44492	44489	44492
R^2	0.503	0.518	0.514	0.538	0.513	0.540	0.515	0.545

Notes: *Base* is the baseline specification involving primary job only. *Add. Job* also includes secondary job. *HH TC* scales income by the inverse of the share of self-produced consumption in household's overall consumption. *HH FC* scales income by the inverse of the share of self-produced food in household's food consumption. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.7: Premia with Hours Worked

	(1) Log Income	(2) Log Income	(3) Log Income	(4) Log Income	(5) Log Inc./Hour	(6) Log Inc./Hour
Non-Agriculture	0.574***	0.332***	0.441***	0.271***	0.297***	0.185***
	(0.036)	(0.033)	(0.034)	(0.032)	(0.036)	(0.036)
Urban	0.207***	0.084**	0.160***	0.084***	0.109***	0.076***
	(0.036)	(0.032)	(0.031)	(0.026)	(0.029)	(0.028)
Log Hours/Year			0.496***	0.432***		
			(0.011)	(0.011)		
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes		Yes
Observations	44494	44497	43841	43843	43841	43843
R^2	0.503	0.518	0.592	0.595	0.478	0.493

Notes: Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.8: Premia with Hours Worked: Additional Jobs

	(1) Log Income	(2) Log Income	(3) Log Income	(4) Log Income	(5) Log Inc./Hour	(6) Log Inc./Hour
Non-Agriculture	0.501***	0.264***	0.390***	0.216***	0.275***	0.150***
	(0.034)	(0.032)	(0.032)	(0.031)	(0.034)	(0.035)
Urban	0.171***	0.063*	0.143***	0.063**	0.112***	0.057**
	(0.034)	(0.034)	(0.030)	(0.026)	(0.028)	(0.028)
Log Hours/Year			0.509***	0.445***		
			(0.011)	(0.011)		
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes		Yes
Observations	44489	44492	43819	43821	43819	43821
R^2	0.514	0.538	0.603	0.615	0.495	0.514

Notes: Income and hours from both the main job and the secondary job. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.9: Hours Worked

	Base	Base	Add. Job	Add. Job
	(1)	(2)	(3)	(4)
	Log Hours	Log Hours	Log Hours	Log Hours
Non-Agriculture	0.286***	0.152***	0.234***	0.119***
	(0.024)	(0.029)	(0.022)	(0.027)
Urban	0.101***	0.014	0.062***	0.012
	(0.020)	(0.031)	(0.018)	(0.032)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes
Observations	43841	43843	43819	43821
R^2	0.053	0.023	0.052	0.026

Notes: *Base* is the baseline specification involving primary job only. *Add. Job* also includes secondary job. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.10: Premia over Time

(A) Cross-Sectional Premia

	Pooled	1993	1997	2000	2007	2014
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Income	Log Income	Log Income	Log Income	Log Income	Log Income
Non-Agriculture	0.574***	0.792***	0.721***	0.547***	0.461***	0.449***
	(0.036)	(0.070)	(0.052)	(0.051)	(0.048)	(0.058)
Urban	0.207***	0.388***	0.271***	0.227***	0.204***	0.097
	(0.036)	(0.057)	(0.051)	(0.051)	(0.049)	(0.062)
Year FE	Yes					
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE						
Observations	44494	5296	8548	10293	10619	9738
R^2	0.503	0.382	0.333	0.244	0.267	0.249

(B) Premia with Worker Fixed Effect

	Pooled	1993-97	1997-00	2000-07	2007-14
	(1)	(2)	(3)	(4)	(5)
	Log Income	Log Income	Log Income	Log Income	Log Income
Non-Agriculture	0.332***	0.339***	0.292***	0.303***	0.217***
	(0.033)	(0.071)	(0.052)	(0.056)	(0.059)
Urban	0.084**	0.210***	0.097	0.156***	0.144**
	(0.032)	(0.068)	(0.087)	(0.058)	(0.058)
Year FE	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes
Indiv. cont.	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Observations	44497	13844	18841	20912	20360
R^2	0.518	0.242	0.205	0.396	0.282

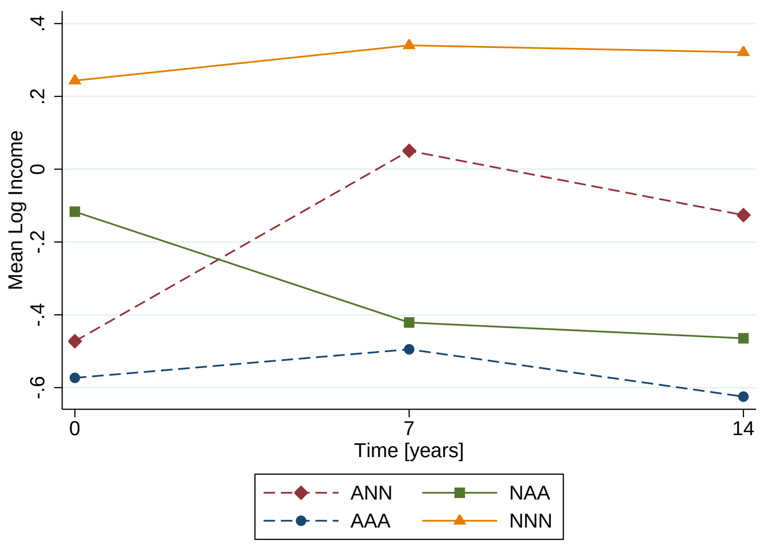
Notes: *Pooled* is the baseline sample with observations from IFLS 1-5. Panel A: cross-sectional regressions run separately for each survey wave. Panel B: panel regressions run separately for each two consecutive survey waves. Individual controls: education, experience, experience sq., and gender. Observations weighted by longitudinal survey weights. Standard errors clustered by enumeration areas (primary sampling units of the survey) in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.11: Long Run Premia

	1993-2014 (1) Δ Log Income	93-07/00-14 (2) Δ Log Income
AA-AN	0.172	
	1.38	
NA-NN	-0.369***	
	9.10	
ANN-AAA		0.147*
		2.79
NAA-NNN		-0.186**
		4.62
Observations	2567	7857
R^2	0.105	0.098

Notes: Column 1 presents tests based on results of a first-difference regression (2), where the difference is over the period 1993-2014. Reported are the difference in coefficients of interest and the value of an *F(1,288)* test that the difference is zero. Column 2 presents tests based on a first-difference specification over 14 years (1993-2007 or 2000-2014) controlling for direction of switch during the first and second 7-year period. Reported are the difference in coefficients of interest and the value of an *F(1,292)* test that the difference is zero. Other controls and weights are as in column 1 in Table 4. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Figure A.1: Mean Log Income by Employment History



Notes: Figure plots mean log income (after controlling for year and province fixed effects) by employment history spanned by three observations at 7-year intervals. XYZ indicates that worker was in sector X during the first observation (in 1993 or 2000), in sector Y during the second observation 7 years later (in 2000 or 2007), and in sector Z during the third observation 14 years later (in 2007 or 2014). A - Agriculture, N - Non-Agriculture. For clarity only histories of switchers who stick to their new sector and of always stayers are reported.

B Recall Bias

Each wave of the IFLS asks respondents about the income they earned over the past year. Throughout the paper we use this contemporaneously recorded income as our main dependent variable. In addition, the survey asks respondents to retrospectively recall employment information for several years prior to the survey. While this recall information can in principle be used to supplement the contemporaneous data and increase the sample size, retrospective survey data is known to raise serious quality concerns (cf. Bound, Brown and Mathiowetz, 2001). For this reason we do not use retrospective income information in our analysis. In this appendix we explain this choice in more detail and argue that it can largely explain why our results differ from those concurrently obtained by Hicks et al. (2017).

The first three columns of panel A of Table B.1 show the non-agricultural premia estimated on the contemporaneous data recorded by the IFLS. These numbers are similar to those reported in Table A.7 (columns 3, 4, and 6) in the main text, but not identical because the specifications and sample are modified to ease comparison with Hicks et al. (2017). In particular, we discard the information from the most recent wave of IFLS as it has not been incorporated by these authors. Columns 4-6 show the corresponding premia estimated on data from retrospective recall. Compared to the contemporaneous estimates, the cross-sectional premium (controlling for hours) drops from 71 lp to 53 lp, premium with worker fixed effects (controlling for hours) drops from 25 lp to 11 lp, and the 19 lp premium in terms of income per hour (with worker FE) disappears entirely.

These patterns are not surprising in light of research on biases arising in recall surveys. One such well documented bias is that past income reported by workers is biased towards their usual income.[27] For example, Gibson and Kim (2010) show the extent of this bias for US wage workers by comparing their self-reported retrospective earnings with administrative records. They also demonstrate that underreporting transitory income changes generates non-classical measurement error that biases the regression coefficients towards zero if the mismeasured variable is the dependent variable. This result is consistent with the reduced non-agricultural premia we find using recall data if workers cannot accurately recall how much higher their income was in years in which they worked in non-agriculture. Furthermore, the problem is likely to be exacerbated when the identifying variation comes from changes in income of individual workers over time. This would explain why the fall in the premium is proportionately much larger in the specification with worker fixed effects. Finally, the problems with measurement error are likely to be compounded when the dependent variable is constructed by dividing reported income by reported hours. That retrospectively recalled hours are unreliable is suggested by comparing coefficients on hours in columns 5 and 2. The elasticity of income with respect to hours implied by column 5 is less than 0.15, only 1/3 of the 0.44 elasticity implied by the corresponding column 2 for contemporaneous data. The implausibly low elasticity for recalled hours indicates that their relationship to income should be treated with great caution in recall data. In a rare validation study observing both hours worked and earnings, Duncan and Hill (1985) find that *"interview reports of average hourly earnings, obtained by dividing the interview reports of annual earnings by reports of annual work hours, appeared to be exceedingly unreliable"*

[27] Another bias with similar implications in this context is an anchoring bias, where respondents use an answer to a previously answered question as a mental anchor for subsequent answers. Godlonton, Hernandez and Murphy (2016) find strong evidence of this behavior in a survey of Central American farmers: retrospectively recalled income correlates more highly with current income (about which the respondents are asked first) than with income over the recall period that had been reported contemporaneously in the past. This type of cognitive bias is likely to be present in IFLS too, since IFLS also first asks about contemporaneous income and then asks respondents to retrospectively recall past income.

and caution against their use. The particularly low signal-to-noise ratio in income per hour derived from retrospective data can explain why the results become insignificant in column 6.

The take-away message from this discussion is that using data from retrospective recall in our application would introduces biases in our key results. These recall biases can be strong in IFLS since the respondents are asked retrospective questions about multiple years prior to the survey (up to a maximum of 10 years), and the quality of recall information deteriorates with time elapsed from the pertaining event (see, e.g. de Nicola and Gine (2014) in a developing country context). There are no obvious offsetting benefits to including the retrospective data. Statistical power, in particular, is not an issue, since the baseline sample of contemporaneous responses is large enough to allow us to estimate the key non-agricultural premium precisely.

We conclude this appendix by showing that the inclusion of retrospective data is likely the main reason why the substantive results on the strength of the non-agricultural premium reported by Hicks et al. (2017) are different than ours. In contrast to our results, they argue that the non-agricultural premium in Indonesia mostly disappears once individual fixed effects are allowed for. To aid comparison, columns 1-3 in panel B of Table B.1 repeat the same exercise as columns 1-3 and 4-6 in panel A, but now on a sample pooling the contemporaneous and retrospective responses. The estimates lie roughly half way between the two corresponding numbers reported in panel A. This means that the pooled-sample estimates are significantly attenuated relative to those based on better-measured contemporaneous data that we favor. For comparison, columns 4-6 copy the corresponding estimates from Hicks et al. (2017) (columns 2, 6, and 7 of their Table 5A), who use pooled contemporaneous and retrospective data. While we cannot replicate their results exactly without detailed knowledge of their data processing protocol, the estimates in columns 1-3 come close. Based on this exercise, we expect that their results would much have been much more in line with ours had they not used the retrospective data.[28]

[28] Furthermore, their headline result depends on using income per hour as their preferred measure. We do not use hours data in our preferred specifications, both because of measurement issues for hours described in this appendix and conceptual issues discussed in section A.5.

Table B.1: Retrospective Recall

(A) Contemporaneous vs. Recall Data

	Contemporaneous			Retrospective		
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Inc.	Log Inc.	Log Inc./Hr	Log Inc.	Log Inc.	Log Inc./Hr
Non-Agriculture	0.707***	0.245***	0.192***	0.525***	0.110***	-0.038
	(0.013)	(0.022)	(0.024)	(0.020)	(0.039)	(0.052)
Log Hours	0.604***	0.462***		0.140***	-0.012	
	(0.039)	(0.046)		(0.051)	(0.045)	
Log Hours Squared	0.000	-0.002		0.018***	0.016***	
	(0.005)	(0.005)		(0.006)	(0.005)	
Age squared		-0.000***	-0.000***		-0.001***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes	Yes		Yes	Yes
Observations	48626	48626	48626	63498	63498	63498
R-sq	0.423	0.540	0.433	0.161	0.192	0.158

(B) Pooled Data vs. Hicks et al. (2017)

	Pooled Data			Hicks et al. (2017)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Log Inc.	Log Inc.	Log Inc./Hr	Log Inc.	Log Inc.	Log Inc./Hr
Non-Agriculture	0.588***	0.173***	0.076***	0.514***	0.171***	0.047
	(0.015)	(0.019)	(0.021)	(0.016)	(0.025)	(0.031)
Log Hours	0.385***	0.206***		0.531**	0.323***	
	(0.040)	(0.037)		(0.025)	(0.034)	
Log Hours Squared	0.006	0.009**		-0.021***	-0.014**	
	(0.005)	(0.004)		(0.005)	(0.006)	
Age squared		-0.000***	-0.000***		-0.001***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes	Yes		Yes	Yes
Observations	107933	107933	107933	115897	115897	115897
R-sq	0.303	0.353	0.263			

Notes: *Contemporaneous* measures based on values reported for last year. *Retrospective* measures obtained from recall part of the survey. *Pooled Data* combines contemporaneous and retrospective observations. Sample restricted to IFLS 1-4. Sample includes all individuals with at least one observation of income and hours worked. *Income* is average monthly labor income from primary and secondary job. Contemporaneous income obtained by dividing annual income by 12. *Hours* are average monthly hours from primary and secondary job obtained as (weeks worked per year)*(normal hours per week)/12. Observations are not weighted. Standard errors clustered at the individual level in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Columns 4-7 in Panel B are columns 2, 6, 7, respectively, from Table 5A in Hicks et al. (2017).

C Estimation Procedure

In this Appendix we present some technical aspects of the estimation procedure. The vector of structural parameters in the frictionless economy, denoted by Θ, consists of the following set of 21 elements: $\{R_t^s, \beta, \sigma_{\theta^s}^2, \sigma_{\theta^{AN}}, \sigma_{\varepsilon^s}^2, \sigma_\nu^2\}$ for $t = 1, ..5$ and $s = A, N$. In this set, $\sigma_{\theta^s}^2$ and $\sigma_{\varepsilon^s}^2$ denote the variances in Σ_θ and Σ_ε, respectively, $\sigma_{\theta^{AN}}$ the covariance in Σ_θ, and β is comprised of the Mincerian returns to the five covariates, denoted β_{sex}, β_{loc}, β_{edu}, β_{exp}, and β_{exp2}. For the model with compensating differentials Θ is augmented by cd, and for the model with involuntary choices Θ is augmented by $\left\{p^T, p^S\right\}$. The indirect inference loss function, denoted by $Q(\Theta)$, is computed as:

$$Q(\Theta) = \left(\hat{\delta} - \hat{\delta}^s(\Theta)\right)' \Omega \left(\hat{\delta} - \hat{\delta}^s(\Theta)\right),$$

where $\hat{\delta}$ is the vector of coefficients of the auxiliary models estimated on the observed data, $\hat{\delta}^s(\Theta)$ is the corresponding vector of coefficients estimated on the data simulated in the structural model with parameters Θ, and Ω is a diagonal weighting matrix. For the weights, we use factors that represent the importance of the estimated coefficient in the identification of the structural parameters of the model. The values of those factors, displayed in column 1 of Table 7, were assigned after extensive experimentation with simulations of the model. We now comment on these weights, particularly those that differ from one (the reference weight).

As we argue in the main text, the within-individual variation is key for identification. It is thus important that the estimated model is able to match the transition probabilities and deliver the observed premia for switchers in the data. Matching the number of switchers and stayers in each year is especially crucial because it affects the precision of the obtained premia. For this reason, the coefficients of the linear probability models have the largest weights in the loss function, with a factor of 10. Moreover, we also assign larger weights (a factor of 5) to the two sets of switchers' sectoral premia. First, to the premia in the model iv, since this regression allows the structural model to deliver the gains for switchers to non-agriculture and the cuts in income for workers switching to agriculture of potentially different magnitudes, allowing the estimation procedure to identify any possible asymmetry in the barriers to mobility across sectors. Second, to the premia in model iii, which compares the average income of a switching worker to each sector with their peer group after the switch, providing information about the nature of sorting. In addition, the estimated coefficients of the interactions in model v help to identify the growth in relative human capital prices. Thus, the information about the full path of these prices can be recovered once the estimated growth rates are combined with levels in the first year. These levels are primarily affected by the constant and the interaction term of non-agriculture in the first year in model v, so those two coefficients get larger weights, with a factor of 5. Finally, given the importance of the residual variances in identifying the joint distribution of comparative advantage, they also have larger weights in the loss function, with a factor of 3.

We minimize $Q(\Theta)$ using, in each evaluation, H simulated samples, each with size equal to the number of observations in the balanced panel ($N \times T = 1752 \times 5 = 8760$)[29]. For observables, we take in each simulated sample the same values as observed in the data. To choose H, we numerically explore how $Q(\Theta)$ varies with the simulation sample size. We find that the range of variation of $Q(\Theta)$ starts to stabilize after we include 80000 individuals. Thus, we choose $H = 40 \approx 80000/1752$.

Optimizing $Q(\Theta)$ is challenging since the discrete choices in the selection model create a non-

[29] We compute the solution on the sample generated by $H \times N \times T$ observations, a method that is equivalent to but computationally more efficient than computing the average of H solutions on the sample of $N \times T$ observations.

smooth function that behaves as a step function in some regions of the parameter space. To deal with this problem, we use an algorithm with repeated iterations of an evolutionary method (particle-swarm optimization) to find the global solution.[30]. We start with 16 runs of the particle-swarm optimization over a wide range of the feasible parameter space. In each of these 16 runs we work with 96 particles that initially are randomly and uniformly distributed. In the second stage, we perform 8 additional runs of the particle-swarm optimization over a range of the parameter space bounded by the smallest and largest solution for each parameter in the first stage, plus a parameter-dependent margin. In this stage, we use the same distribution and twice the number of particles as in the first stage. Finally, we perform an additional optimization in which we initialize 8 of the 2*96 particles with the solutions found in the second stage. The estimate $\hat{\Theta}$ is the solution that minimizes $Q(\Theta)$ across all of the 25 described runs of the particle-swarm optimization. Using multiple numerical simulations, we find that our algorithm ensures a high degree of accuracy in finding the solution, outperforming common packaged optimization techniques.

D Identification

In this Appendix we demonstrate how the structural parameters in Θ are identified for the simplified version of our model with two periods and abstracting from the effect of observables on income. We first show how identification is achieved in the frictionless economy, and next we proceed to the model with barriers to sectoral mobility.

Denote $r_t^s = \ln R_t^s$, $r_t = r_r^N - r_t^A$, $u_{it}^s = \theta_i^s + \varepsilon_{it}^s$ for $s = A, N$ and $\tilde{\sigma}_{ks}^2 = \sigma_{ks}^2 - \sigma_{kAN}$ for $k = u, \theta$ and $s = A, N$. Notice that $\sigma_{us}^2 = \sigma_{\theta s}^2 + \sigma_{\varepsilon s}^2$ for $s = A, N, AN$. Further, denote the st. dev. of $u_{it} \equiv \left(u_{it}^A - u_{it}^N\right)$ as $\sigma_u^* = \sqrt{\tilde{\sigma}_{uA}^2 + \tilde{\sigma}_{uN}^2}$ and the st. dev. of $\theta_i \equiv \left(\theta_i^A - \theta_i^N\right)$ as $\sigma_\theta^* = \sqrt{\tilde{\sigma}_{\theta A}^2 + \tilde{\sigma}_{\theta N}^2}$. Let y_{it}^s denote the logarithm of income of individual i in sector s at time t.

Basic Frictionless Economy (Model from Section 4.1)

In the frictionless economy sectoral decisions do not depend on workers' histories, so the model in the cross-section in each period t looks much like the standard Roy model with comparative advantage determined by u_{it}^s. We can therefore identify the variance matrix Σ_u (up to the variance of the measurement error σ_ν^2) and the prices of human capital r_t^s from cross-sectional data. This is a consequence of the normality assumptions for the distribution of both $\left(\theta_i^A, \theta_i^N\right)$ and $\left(\varepsilon_{it}^A, \varepsilon_{it}^N\right)$, which imply that in each period $\left(u_{it}^A, u_{it}^N\right)$ is joint normally distributed with variance Σ_u. Hence, the standard arguments of Heckman and Honoré (1990) for the identification in the normal case can be applied. However, only with the panel data we can decompose Σ_u into Σ_θ and Σ_ε, the variances of the permanent and transitory components of productivity, and identify σ_ν^2, the variance of the measurement error, by using the information obtained from the switching workers. Specifically, we are able to find analytical expressions for the moments of the income distribution of the sector switchers by exploiting the property that the draws of u_{it}^s in the different periods are joint normally distributed, since each one is the sum of two normally distributed random variables. We can thus express the transition probabilities across periods and the observed moments of the switchers'

[30]Other possibilities recently developed in the literature include the use of a logistic-kernel of simulated latent utilities instead of endogenous variables (Bruins et al., 2018) or Monte Carlo importance sampling (Sauer and Taber, 2017). However, the applicability of each technique is model-dependent. Our algorithm is tailored to find an accurate solution to our model.

income growth using upper truncated multivariate normal distributions, where the prices of human capital in the two periods affect the truncation values.

Letting $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$, with ϕ and Φ denoting the PDF and CDF of the standard normal distribution, and using properties of normal random variables from Heckman and Honoré (1990), we can obtain the following expressions for the share of workers selecting non-agriculture and the first three moments of the income distribution conditional on sector in each period t:

$$P(t = N) = \Phi\left(\frac{r_t}{\sigma_u^*}\right) \tag{A.1}$$

$$E\left(y_{it}^N \mid t = N\right) = r_t^N + \frac{\tilde{\sigma}_{uN}^2}{\sigma_u^*}\lambda\left(\frac{r_t}{\sigma_u^*}\right) \tag{A.2}$$

$$E\left(y_{it}^A \mid t = A\right) = r_t^A + \frac{\tilde{\sigma}_{uA}^2}{\sigma_u^*}\lambda\left(\frac{-r_t}{\sigma_u^*}\right) \tag{A.3}$$

$$Var\left(y_{it}^N \mid t = N\right) = \sigma_{uN}^2 + \left(\frac{\tilde{\sigma}_{uN}^2}{\sigma_u^*}\right)^2\left[-\lambda\left(\frac{r_t}{\sigma_u^*}\right)\frac{r_t}{\sigma_u^*} - \lambda^2\left(\frac{r_t}{\sigma_u^*}\right)\right] + \sigma_\nu^2 \tag{A.4}$$

$$Var\left(y_{it}^A \mid t = A\right) = \sigma_{uA}^2 + \left(\frac{\tilde{\sigma}_{uA}^2}{\sigma_u^*}\right)^2\left[\lambda\left(\frac{-r_t}{\sigma_u^*}\right)\frac{r_t}{\sigma_u^*} - \lambda^2\left(\frac{-r_t}{\sigma_u^*}\right)\right] + \sigma_\nu^2 \tag{A.5}$$

$$E\left(\left[y_{it}^N - E\left(y_{it}^N \mid t = N\right)\right]^3 \mid t = N\right) = \left(\frac{\tilde{\sigma}_{uN}^2}{\sigma_u^*}\right)^3\lambda\left(\frac{r_t}{\sigma_u^*}\right)\left[2\lambda^2\left(\frac{r_t}{\sigma_u^*}\right) + 3\lambda\left(\frac{r_t}{\sigma_u^*}\right)\frac{r_t}{\sigma_u^*} + \left(\frac{r_t}{\sigma_u^*}\right)^2 - 1\right] \tag{A.6}$$

$$E\left(\left[y_{it}^A - E\left(y_{it}^A \mid t = A\right)\right]^3 \mid t = A\right) = \left(\frac{\tilde{\sigma}_{uA}^2}{\sigma_u^*}\right)^3\lambda\left(\frac{-r_t}{\sigma_u^*}\right)\left[2\lambda^2\left(\frac{-r_t}{\sigma_u^*}\right) - 3\lambda\left(\frac{-r_t}{\sigma_u^*}\right)\frac{r_t}{\sigma_u^*} + \left(\frac{r_t}{\sigma_u^*}\right)^2 - 1\right]. \tag{A.7}$$

With information for $T = 2$ repeated cross sections, this system of 14 equations allows us to identify r_t^N, r_t^A for $t = 1, 2$ and the elements of the matrix $\sigma_\nu^2\mathbb{I}_2 + \Sigma_u$ Let us now show that the additional information from the panel data allows us to further isolate Σ_θ, Σ_ε, and σ_ν^2. We exploit the property that $(u_{it}, u_{it'})$ for $t' \neq t$ and (u_{it}, u_{it}^s) for $s = A, N$ are joint normally distributed, since each element is the sum of two normally distributed random variables. Denoting the CDF of a bivariate normal distribution with mean $\mathbf{0'}$ and variance $\mathit{\Sigma}$ evaluated at the vector A as $\Phi(A, \Sigma)$, the probability of transition from N to N is given by:

$$\begin{aligned} P(2 = N, 1 = N) =& P\left(\left\{r_2^A + u_{i2}^A < r_2^N + u_{i2}^N\right\}, \left\{r_1^A + u_{i1}^A < r_1^N + u_{i1}^N\right\}\right) \\ =& P\left(\left\{u_{i2} < r_2\right\}, \left\{u_{i1} < r_1\right\}\right) \\ =& \Phi\left(\overrightarrow{r}_{NN}, \Sigma_T\right), \end{aligned} \tag{A.8}$$

where $\overrightarrow{r}_{NN} = [r_2, r_1]'$ and $\mathit{\Sigma}_T = \begin{bmatrix} \sigma_u^{*2} & \sigma_\theta^{*2} \\ \sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix}$. The probability of transition from N to A is given by:

$$\begin{aligned} P(2 = A, 1 = N) =& P\left(\left\{-u_{i2} < -r_2\right\}, \left\{u_{i1} < r_1\right\}\right) \\ =& \Phi\left(\overrightarrow{r}_{NA}, \Sigma_W\right), \end{aligned} \tag{A.9}$$

where $\overrightarrow{r}_{NA} = [-r_2, r_1]'$ and $\Sigma_W = \begin{bmatrix} \sigma_u^{*2} & -\sigma_\theta^{*2} \\ -\sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix}$. Similarly:

$$P(2 = N, 1 = A) = \Phi(\overrightarrow{r}_{AN}, \Sigma_W) \tag{A.10}$$
$$P(2 = N, 2 = A) = \Phi(\overrightarrow{r}_{AA}, \Sigma_T), \tag{A.11}$$

where $\overrightarrow{r}_{AN} = [r_2, -r_1]'$ and $\overrightarrow{r}_{AA} = [-r_2, -r_1]'$.

Now consider the values of the expected income in the second period for each transition group of workers. While we do not need these values directly in the case of the the frictionless economy, we illustrate here how to compute them to introduce some notation that we use later. The income of stayers in N in period 2 is given by:

$$\begin{aligned} E\left(y_{i2}^N | \, 2 = N, 1 = N\right) =& r_2^N + E\left(u_{i2}^N | \, 2 = N, 1 = N\right) \\ =& r_2^N + E\left(u_{i2}^N | \, \{u_{i2} < r_2\}, \{u_{i1} < r_1\}\right). \end{aligned}$$

Notice that the expected value in the second term of the RHS can be expressed as:

$$E\left(X_1^{k_1} X_2^{k_2} X_3^{k_3} | -\infty < X_i < b_i, i = 1, 2, 3\right), \tag{A.12}$$

where $X_1 = u_{i2}^N$, $X_2 = u_{i2}$, $X_3 = u_{i1}$, $k_1 = 1, k_2 = k_3 = 0$, $b_1 = \infty$, $b_2 = r_2$, and $b_3 = r_1$. This expected value is the moment of the upper truncated multivariate normal distribution with mean $\mathbf{0}$ and variance:

$$\Sigma_{NN} = \begin{bmatrix} \sigma_{uN}^2 & -\tilde{\sigma}_{uN}^2 & -\tilde{\sigma}_{\theta N}^2 \\ -\tilde{\sigma}_{uN}^2 & \sigma_u^{*2} & \sigma_\theta^{*2} \\ -\tilde{\sigma}_{\theta N}^2 & \sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix} = \begin{bmatrix} \sigma_{uN}^2 & \Lambda_{NN} \\ \Lambda_{NN}' & \Sigma_T \end{bmatrix},$$

where the vector Λ_{NN} is defined as $\Lambda_{NN} = \left[-\tilde{\sigma}_{uN}^2, -\tilde{\sigma}_{\theta N}^2\right]$. In general terms, we can denote the expected value in (A.12) for the particular case $k_2 = k_3 = 0$ and $b_1 = \infty$ as the function $\mathbf{M}_3(\cdot)$ of the variance matrix Σ, the elements b_2 and b_3 stacked up in a vector B and the coefficient $k = k_1$, that is:

$$\mathbf{M}_3(B, \Sigma, k) \equiv E\left(X_1^k | -\infty < X_1 < \infty, -\infty < X_2 < B_1, -\infty < X_3 < B_2\right),$$

where $\{X_1, X_2, X_3\} \sim \mathbf{N}(\mathbf{0}, \Sigma)$. Then we can rewrite:

$$E\left(y_{i2}^N \, | \, 2 = N, 1 = N\right) = r_2^N + \mathbf{M}_3(\overrightarrow{r}_{NN}, \Sigma_{NN}, 1).$$

To evaluate $\mathbf{M}_3(\cdot)$ we can use the recurrence relations developed by Kan and Robotti (2017) to compute numerically the moment generating function of the truncated multivariate normal distribution (first obtained by Tallis, 1961)[31]. Following similar arguments, we can show that the income

[31] We use the function *multivutmom* developed by Kan and Robotti (2017) in the Matlab package ftnorm.

of each remaining transition group in period 2 is given by:

$$\begin{aligned} E\left(y_{i2}^{A}|\, 2=N, 1=N\right) =& r_2^A + \mathbf{M}_3\left(\overrightarrow{r}_{NA}, \varSigma_{NA}, 1\right) \\ E\left(y_{i2}^{N}|\, 2=N, 1=A\right) =& r_2^N + \mathbf{M}_3\left(\overrightarrow{r}_{AN}, \varSigma_{AN}, 1\right) \\ E\left(y_{i2}^{A}|\, 2=A, 1=A\right) =& r_2^A + \mathbf{M}_3\left(\overrightarrow{r}_{AA}, \varSigma_{AA}, 1\right), \end{aligned}$$

where $\varSigma_{NA} = \begin{bmatrix} \sigma_{uA}^2 & \varLambda_{NA} \\ \varLambda_{NA}' & \varSigma_W \end{bmatrix}$, $\varSigma_{AN} = \begin{bmatrix} \sigma_{uN}^2 & \varLambda_{AN} \\ \varLambda_{AN}' & \varSigma_W \end{bmatrix}$, and $\varSigma_{AA} = \begin{bmatrix} \sigma_{uA}^2 & \varLambda_{AA} \\ \varLambda_{AA}' & \varSigma_T \end{bmatrix}$, and where $\varLambda_{NA} = \left[-\tilde{\sigma}_{uA}^2, \tilde{\sigma}_{\theta A}^2\right]$, $\varLambda_{AN} = \left[-\tilde{\sigma}_{uN}^2, \tilde{\sigma}_{\theta N}^2\right]$, and $\varLambda_{AA} = \left[-\tilde{\sigma}_{uA}^2, -\tilde{\sigma}_{\theta A}^2\right]$. Similar to the calculation of the first moments above, the second and third moments of income for each transition group can be computed as functions of $\mathbf{M}_3(\cdot,\cdot,2)$ and $\mathbf{M}_3(\cdot,\cdot,3)$, respectively.

Now let us compute the moments of the growth in income for switchers. For workers switching from A to N their expected growth in income is:

$$\begin{aligned} E\left(y_{i2}^{N} - y_{i1}^{A}|\, 2=N, 1=A\right) &= r_2^N - r_1^A + E\left(u_{i2}^{N} - u_{i1}^{A}|\, 2=N, 1=A\right) \\ &= r_2^N - r_1^A + E\left(u_{i2}^{N} - u_{i1}^{A}|\, \left\{u_{i2} < r_2\right\}, \left\{-u_{i1} < -r_1\right\}\right) \\ &= r_2^N - r_1^A + \mathbf{M}_3(\overrightarrow{r}_{AN}, \tilde{\varSigma}_{AN}, 1), \end{aligned} \tag{A.13}$$

where $\tilde{\varSigma}_{AN} = \begin{bmatrix} \sigma_{uA}^2 + \sigma_{uN}^2 - 2\sigma_{\theta AN} & \tilde{\varLambda}_{AN} \\ \tilde{\varLambda}_{AN}' & \varSigma_W \end{bmatrix}$ and $\tilde{\varLambda}_{AN} = \left[-\tilde{\sigma}_{uN}^2 - \tilde{\sigma}_{\theta A}^2, \tilde{\sigma}_{uA}^2 + \tilde{\sigma}_{\theta N}^2\right]$. Similarly, the expected value of the growth in income for switchers from N to A is:

$$E\left(y_{i2}^{A} - y_{i1}^{N}|\, 2=A, 1=N\right) = r_2^A - r_1^N + \mathbf{M}_3(\overrightarrow{r}_{NA}, \tilde{\varSigma}_{NA}, 1), \tag{A.14}$$

where $\tilde{\varSigma}_{NA} = \begin{bmatrix} \sigma_{uA}^2 + \sigma_{uN}^2 - 2\sigma_{\theta AN} & \tilde{\varLambda}_{NA} \\ \tilde{\varLambda}_{NA}' & \varSigma_W \end{bmatrix}$ and $\tilde{\varLambda}_{NA} = \left[-\tilde{\sigma}_{uA}^2 - \tilde{\sigma}_{\theta N}^2, \tilde{\sigma}_{uN}^2 + \tilde{\sigma}_{\theta A}^2\right]$. The variances of the growth in income for switchers are calculated as:

$$\begin{aligned} &Var\left(y_{i2}^{N} - y_{i1}^{A}|\, 2=N, 1=A\right) \\ &= E\left(\left(u_{i2}^{N} - u_{i1}^{A}\right)^2 |\, \left\{u_{i2} < r_2\right\}, \left\{-u_{i1} < -r_1\right\}\right) - E\left(u_{i2}^{N} - u_{i1}^{A}|\, 2=N, 1=A\right)^2 + 2\sigma_\nu^2 \\ &= \mathbf{M}_3(\overrightarrow{r}_{AN}, \tilde{\varSigma}_{AN}, 2) + \left(\mathbf{M}_3(\overrightarrow{r}_{AN}, \tilde{\varSigma}_{AN}, 1)\right)^2 + 2\sigma_\nu^2. \end{aligned} \tag{A.15}$$

Similarly:

$$Var\left(y_{i2}^{A} - y_{i1}^{N}|\, 2=A, 1=N\right) = \mathbf{M}_3(\overrightarrow{r}_{NA}, \tilde{\varSigma}_{NA}, 2) + \left(\mathbf{M}_3(\overrightarrow{r}_{NA}, \tilde{\varSigma}_{NA}, 1)\right)^2 + 2\sigma_\nu^2. \tag{A.16}$$

The system of 22 equations (A.1)-(A.7) (one set for each period), (A.8)-(A.11), and (A.13)-(A.16) has a unique solution for the 10 elements of Θ. We verified this after extensive experimentation using global solvers over a broad range of feasible values for Θ. This shows that the share of workers in each sector and the first three cross-sectional moments of income in each period, the transition probabilities across waves for each group of workers, and the first two moments of the income growth for switchers uniquely identify the full set of structural parameters.

Generalization I: Compensating Differential (Model from Section 4.2)

In the model with a compensating differential, we use exactly the same set of 22 moments computed above to identify the 11 elements of Θ (same parameters as in the basic frictionless economy plus the parameter cd). The only required change in the expressions for the moments is the substitution of $r_t = r_r^N - r_t^A - \ln cd$ for $r_t = r_r^N - r_t^A$. The compensating differential only affects the terms related to the selection bias. As in the frictionless case, we verified that the system of 22 moments has a unique solution for the 11 elements of Θ.

Generalization II: Barriers to Sectoral Mobility (Model from Section 4.3)

Now consider the model with involuntary switches. The objective is to find expressions for the same set of 22 moments as in the frictionless case (i.e., equivalents of equations (A.1)-(A.11), (A.13)-(A.16)) to identify the 12 elements of Θ (the same 10 parameters as in the frictionless economy plus the two probabilities of involuntary choices). The difficulty in obtaining expressions for those moments stems from the fact that sectoral decisions depend now on workers' histories, and hence all moments, including the cross-sectional ones, depend on the income distributions in the previous periods. We present here the general rules to derive expressions for those moments. Denote the CDF of a multivariate normal distribution with mean $\mathbf{0}'$ and variance Σ evaluated at the vector A as $\Phi(A, \Sigma)$. Moreover, define the probability of being forced to accept a job in a sector other than desired for workers who want to stay in the same sector as p^S, and for workers who want to switch sector as p^T. Denote $p_n^S = 1 - p^S$ and $p_n^T = 1 - p^T$. For the allocation of sectors in period 0 we assume that a fraction p^S of workers are forced to be out of their desired sector.

First, consider the cross-sectional moments in period 1. The probabilities of transition across sectors between period 0 and 1 can be computed as the sum of the products of the probabilities of occurrence of all four possible transition paths in the frictionless equilibrium (desired transitions) and the proportions of workers of each desired transition group that end up in each realized transition group. These proportions are displayed in Table D.1. Each cell can be read as the fraction of workers of the desired transition group (displayed in columns) that ends up in the realized transition group (displayed in rows).

Table D.1: Probabilities of ending up in a given transition group for period 1

		Desired transition*			
		NN	NA	AN	AA
Realized transition*	NN	$\left(p_n^S\right)^2$	$p_n^S p^T$	$p^S p_n^S$	$p^S p^T$
	NA	$p_n^S p^S$	$p_n^S p_n^T$	$\left(p^S\right)^2$	$p^S p_n^T$
	AN	$p^S p_n^T$	$\left(p^S\right)^2$	$p_n^S p_n^T$	$p_n^S p^S$
	AA	$p^S p^T$	$p^S p_n^S$	$p_n^S p^T$	$\left(p_n^S\right)^2$

*The first and second letters correspond to the sector in period 0 and in period 1, respectively.

Thus, denoting as P^d the desired transitions, the probability of observing a transition from non-agriculture in period 0 to non-agriculture (NN) in period 1 is:

$$\begin{aligned}P(0=N,1=N) &= \left(p_n^S\right)^2 P^d(0=N,1=N) + p_n^S p^T P^d(0=N,1=A)\\ &\quad + p^S p_n^T P^d(0=A,1=N) + p^S p^T P^d(0=A,1=A)\\ &= \left(p_n^S\right)^2 \Phi\left(\overrightarrow{r}_{NN},\Sigma_T\right) + p_n^S p^T \Phi\left(\overrightarrow{r}_{NA},\Sigma_W\right)\\ &\quad + p^S p_n^T \Phi\left(\overrightarrow{r}_{AN},\Sigma_W\right) + p^S p^T \Phi\left(\overrightarrow{r}_{AA},\Sigma_T\right),\end{aligned}$$

where $\overrightarrow{r}_{NN}$, $\overrightarrow{r}_{NA}$, $\overrightarrow{r}_{AN}$, $\overrightarrow{r}_{AA}$, Σ_W and Σ_T are as in the frictionless economy. Arrange the probabilities of all four possible transition paths in the frictionless economy in a column vector $\overrightarrow{P}_1^d$, i.e.:

$$\overrightarrow{P}_1^d = \left[\Phi\left(\overrightarrow{r}_{NN},\Sigma_T\right)\ \Phi\left(\overrightarrow{r}_{NA},\Sigma_W\right)\ \Phi\left(\overrightarrow{r}_{AN},\Sigma_W\right)\ \Phi\left(\overrightarrow{r}_{AA},\Sigma_T\right)\right]'.$$

The probability of observing a transition NN from period 0 to period 1 can then be written in a compact form as the cross product:

$$P(0=N,1=N) = \overrightarrow{F}_{NN} \times \overrightarrow{P}_1^d,$$

where $\overrightarrow{F}_{NN}$ is a (row) vector with the shares of workers corresponding to the realized transition group NN, displayed in the first row of Table D.1. Similarly,

$$P\left(0=s,1=s'\right) = \overrightarrow{F}_{ss'} \times \overrightarrow{P}_1^d$$

for $s,s'=A,N$, where $\overrightarrow{F}_{NA}$, $\overrightarrow{F}_{AN}$, $\overrightarrow{F}_{AA}$ are row vectors with the shares of workers displayed in the second, third and fourth rows of Table D.1, respectively. Hence:

$$P(1=s) = P(0=N,1=s) + P(0=A,1=s) = \left(\overrightarrow{F}_{Ns} + \overrightarrow{F}_{As}\right) \times \overrightarrow{P}_1^d$$

for $s=A,N$. This is the analogue of equation (A.1) for $t=1$ in the frictionless case.

Now consider the expressions for the values of the expected income in the first period. For sector N we have:

$$\begin{aligned}E\left(y_{i1}^N \middle| 1=N\right) &= r_1^N + E\left(u_{i1}^N \middle| 1=N\right)\\ &= r_1^N + \frac{E\left(u_{i1}^N \middle| 1=N,0=A\right)\overrightarrow{F}_{AN}\times\overrightarrow{P}_1^d + E\left(u_{i1}^N \middle| 1=N,0=N\right)\overrightarrow{F}_{NN}\times\overrightarrow{P}_1^d}{\left(\overrightarrow{F}_{NN}+\overrightarrow{F}_{AN}\right)\times\overrightarrow{P}_1^d}.\end{aligned} \tag{A.17}$$

To find expressions for $E\left(u_{i1}^N \middle| 1=N,0=A\right)$ and $E\left(u_{i1}^N \middle| 1=N,0=N\right)$ we need to compute the expected value of u_{i1}^N for the individuals in all possible desired transition groups. Such expected values are the elements of the column vector:

$$\overrightarrow{E}_{11}^N = \left[\mathbf{M_3}(\overrightarrow{r}_{NN},\mathit{\Sigma}_{NN},1)\ \mathbf{M_3}(\overrightarrow{r}_{NA},\Sigma_{NA|N},1)\ \mathbf{M_3}(\overrightarrow{r}_{AN},\mathit{\Sigma}_{AN},1)\ \mathbf{M_3}(\overrightarrow{r}_{AA},\Sigma_{AA|N},1)\right]'$$

with $\mathbf{M}_3(\cdot)$, Σ_{NN}, and Σ_{AN} as in the frictionless economy, and now $\Sigma_{NA|N} = \begin{bmatrix} \sigma^2_{uN} & -\Lambda_{AN} \\ -\Lambda'_{AN} & \Sigma_W \end{bmatrix}$ and $\Sigma_{AA|N} = \begin{bmatrix} \sigma^2_{uN} & -\Lambda_{NN} \\ -\Lambda'_{NN} & \Sigma_T \end{bmatrix}$. Thus, the conditional expected values of u^N_{i1} can be computed through the following weighted average:

$$E\left(u^N_{i1} | \, 1 = N, 0 = s\right) = \frac{\left[\overrightarrow{F}_{sN} \cdot \left(\overrightarrow{P}^d_1\right)'\right] \times \overrightarrow{E}^N_{11}}{\overrightarrow{F}_{sN} \times \overrightarrow{P}^d_1} \tag{A.18}$$

for $s = A, N$, where $\cdot$ denotes the dot product. Substituting (A.18) for $s = A, N$ in (A.17) we get the analogue of (A.2) for $t = 1$:

$$E\left(y^N_{i1} | \, 1 = N\right) = r^N_1 + \frac{\left[\left(\overrightarrow{F}_{NN} + \overrightarrow{F}_{AN}\right) \cdot \left(\overrightarrow{P}^d_1\right)'\right] \times \overrightarrow{E}^N_{11}}{\left(\overrightarrow{F}_{NN} + \overrightarrow{F}_{AN}\right) \times \overrightarrow{P}^d_1}.$$

Using a similar reasoning, we can obtain the analogue of (A.3) for $t = 1$:

$$E\left(y^A_{i1} | \, 1 = A\right) = r^A_1 + \frac{\left[\left(\overrightarrow{F}_{NA} + \overrightarrow{F}_{AA}\right) \cdot \left(\overrightarrow{P}^d_1\right)'\right] \times \overrightarrow{E}^A_{11}}{\left(\overrightarrow{F}_{NA} + \overrightarrow{F}_{AA}\right) \times \overrightarrow{P}^d_1},$$

where $\overrightarrow{E}^A_{11}$ is the column vector with elements:

$$\overrightarrow{E}^A_{11} = \left[\mathbf{M_3}(\overrightarrow{r}_{NN}, \Sigma_{NN|A}, 1) \; \mathbf{M_3}(\overrightarrow{r}_{NA}, \Sigma_{NA}, 1) \; \mathbf{M_3}(\overrightarrow{r}_{AN}, \Sigma_{AN|A}, 1) \; \mathbf{M_3}(\overrightarrow{r}_{AA}, \Sigma_{AA}, 1)\right]'$$

with Σ_{AA} and Σ_{NA} as in the frictionless economy, and now $\Sigma_{AN|A} = \begin{bmatrix} \sigma^2_{uA} & -\Lambda_{NA} \\ -\Lambda'_{NA} & \Sigma_W \end{bmatrix}$ and $\Sigma_{NN|A} = \begin{bmatrix} \sigma^2_{uA} & -\Lambda_{AA} \\ -\Lambda'_{AA} & \Sigma_T \end{bmatrix}$.

Following similar steps, the formulas for the second and third moments of the income distribution in period 1 can be obtained by substituting the expressions for the conditional expected values of u^{s2}_{i1} and u^{s3}_{i1} (for $s = A, N$), respectively, into the formulas for the corresponding moments derived in the frictionless case. We thus obtain the analogues of (A.4)-(A.7) for $t = 1$:

$$Var\left(y^s_{i1} | \, 1 = s\right) = \frac{\left[\left(\overrightarrow{F}_{Ns} + \overrightarrow{F}_{As}\right) \cdot \left(\overrightarrow{P}^d_1\right)'\right] \times \overrightarrow{E}^s_{12}}{\left(\overrightarrow{F}_{Ns} + \overrightarrow{F}_{As}\right) \times \overrightarrow{P}^d_1} - E\left(u^s_{i1} | \, 1 = s\right)^2 + \sigma^2_\mu$$

$$\begin{aligned} E\left(\left[y^s_{i1} - E\left(y^s_{i1} | \, 1 = s\right)\right]^3 | \, 1 = s\right) = & \frac{\left[\left(\overrightarrow{F}_{Ns} + \overrightarrow{F}_{As}\right) \cdot \left(\overrightarrow{P}^d_1\right)'\right] \times \overrightarrow{E}^s_{13}}{\left(\overrightarrow{F}_{Ns} + \overrightarrow{F}_{As}\right) \times \overrightarrow{P}^d_1} \\ & - 3E\left(u^s_{i1} | \, 1 = s\right)\left[Var\left(y^s_{i1} | \, 1 = s\right) - \sigma^2_\mu\right] - \left[E\left(u^s_{i1} | \, 1 = s\right)\right]^3 \end{aligned}$$

for $s = A, N$, where:

$$\overrightarrow{E}_{1j}^{A} = \left[\mathbf{M_3}(\overrightarrow{r}_{NN}, \Sigma_{NN|A}, j)\ \mathbf{M_3}(\overrightarrow{r}_{NA}, \Sigma_{NA}, j)\ \mathbf{M_3}(\overrightarrow{r}_{AN}, \Sigma_{AN|A}, j)\ \mathbf{M_3}(\overrightarrow{r}_{AA}, \Sigma_{AA}, j)\right]'$$
$$\overrightarrow{E}_{1j}^{N} = \left[\mathbf{M_3}(\overrightarrow{r}_{NN}, \Sigma_{NN}, j)\ \mathbf{M_3}(\overrightarrow{r}_{NA}, \Sigma_{NA|N}, j)\ \mathbf{M_3}(\overrightarrow{r}_{AN}, \Sigma_{AN}, j)\ \mathbf{M_3}(\overrightarrow{r}_{AA}, \Sigma_{AA|N}, j)\right]'$$

for $j = 2, 3$.

Now consider the transition probabilities between period 1 and 2 and the cross-sectional moments in period 2. Once again, we need the proportion of workers of each desired transition group that ends up in the corresponding realized transition groups. These shares are displayed in Table D.2, which can be read in the same way as Table D.1.

Table D.2: Probabilities of ending up in a given transition group for period 2

Realized transition* \ Desired transition*	NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA
NNN	$\left(p_n^S\right)^3$	$\left(p_n^S\right)^2 p^T$	$\left(p_n^S\right)^2 p^T$	$p_n^S \left(p^T\right)^2$	$p^S \left(p_n^S\right)^2$	$p^S p_n^S p^T$	$p^S p^T p_n^S$	$p^S \left(p^T\right)^2$
NNA	$\left(p_n^S\right)^2 p^S$	$\left(p_n^S\right)^2 p_n^T$	$p_n^S p^T p^S$	$p_n^S p^T p_n^T$	$\left(p^S\right)^2 p_n^S$	$p^S p_n^S p_n^T$	$\left(p^S\right)^2 p^T$	$p^S p^T p_n^T$
NAN	$p_n^S p^S p_n^T$	$p_n^S \left(p^S\right)^2$	$p_n^S \left(p_n^T\right)^2$	$p_n^S p_n^T p^S$	$\left(p^S\right)^2 p_n^T$	$\left(p^S\right)^3$	$p^S \left(p_n^T\right)^2$	$\left(p^S\right)^2 p_n^T$
NAA	$p_n^S p^S p^T$	$\left(p_n^S\right)^2 p^S$	$p_n^S p_n^T p^T$	$\left(p_n^S\right)^2 p_n^T$	$\left(p^S\right)^2 p^T$	$\left(p^S\right)^2 p_n^S$	$p^S p_n^T p^T$	$p^S p_n^T p_n^S$
ANN	$p^S p_n^T p_n^S$	$p^S p_n^T p^T$	$\left(p^S\right)^2 p_n^S$	$\left(p^S\right)^2 p^T$	$\left(p_n^S\right)^2 p_n^T$	$p_n^S p_n^T p^T$	$\left(p_n^S\right)^2 p^S$	$p_n^S p^S p^T$
ANA	$\left(p^S\right)^2 p_n^T$	$p^S \left(p_n^T\right)^2$	$\left(p^S\right)^3$	$\left(p^S\right)^2 p_n^T$	$p_n^S p_n^T p^S$	$p_n^S \left(p_n^T\right)^2$	$p_n^S \left(p^S\right)^2$	$p_n^S p^S p_n^T$
AAN	$p^S p^T p_n^T$	$\left(p^S\right)^2 p^T$	$p^S p_n^S p_n^T$	$\left(p^S\right)^2 p_n^S$	$p_n^S p^T p_n^T$	$p_n^S p^T p^S$	$\left(p_n^S\right)^2 p_n^T$	$\left(p_n^S\right)^2 p^S$
AAA	$p^S \left(p^T\right)^2$	$p^S p^T p_n^S$	$p^S p_n^S p^T$	$p^S \left(p_n^S\right)^2$	$p_n^S \left(p^T\right)^2$	$\left(p_n^S\right)^2 p^T$	$\left(p_n^S\right)^2 p^T$	$\left(p_n^S\right)^3$

*The first, second and third letters correspond to the sector in period 0, period 1 and period 2, respectively.

Arrange the probabilities of all eight possible transition paths in the frictionless economy in a column vector $\overrightarrow{P}_2^d$, i.e.:

$$\overrightarrow{P}_2^d = \left[\begin{array}{c} \Phi\left(\overrightarrow{r}_{NNN}, \Sigma_{TT}\right)\ \Phi\left(\overrightarrow{r}_{NNA}, \Sigma_{TW}\right)\ \Phi\left(\overrightarrow{r}_{NAN}, \Sigma_{WW}\right)\ \Phi\left(\overrightarrow{r}_{NAA}, \Sigma_{WT}\right)\ldots \\ \Phi\left(\overrightarrow{r}_{ANN}, \Sigma_{WT}\right)\ \Phi\left(\overrightarrow{r}_{ANA}, \Sigma_{WW}\right)\ \Phi\left(\overrightarrow{r}_{AAN}, \Sigma_{TW}\right)\ \Phi\left(\overrightarrow{r}_{AAA}, \Sigma_{TT}\right) \end{array}\right]',$$

where $\overrightarrow{r}_{ss's''}\ \forall\ s, s', s'' = N, A$ with:

$$\overrightarrow{r}_{NNN} = [r_2,\ \overrightarrow{r}'_{NN}]',\ \overrightarrow{r}_{NNA} = [-r_2,\ \overrightarrow{r}'_{NN}]',\ \overrightarrow{r}_{NAN} = [r_2,\ \overrightarrow{r}'_{AN}]'\ \overrightarrow{r}_{NAA} = [-r_2,\ \overrightarrow{r}'_{NA}]',$$
$$\overrightarrow{r}_{ANN} = [r_2,\ \overrightarrow{r}'_{AN}]',\ \overrightarrow{r}_{ANA} = [-r_2,\ \overrightarrow{r}'_{AN}]',\ \overrightarrow{r}_{AAN} = [r_2,\ \overrightarrow{r}'_{AA}]',\ \overrightarrow{r}_{AAA} = [-r_2,\ \overrightarrow{r}'_{AA}]',$$

and:

$$\Sigma_{TT} = \begin{bmatrix} \sigma_u^{*2} & \sigma_\theta^{*2} & \sigma_\theta^{*2} \\ \sigma_\theta^{*2} & \sigma_u^{*2} & \sigma_\theta^{*2} \\ \sigma_\theta^{*2} & \sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix},\ \Sigma_{TW} = \begin{bmatrix} \sigma_u^{*2} & -\sigma_\theta^{*2} & -\sigma_\theta^{*2} \\ -\sigma_\theta^{*2} & \sigma_u^{*2} & \sigma_\theta^{*2} \\ -\sigma_\theta^{*2} & \sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix},$$

$$\Sigma_{WW} = \begin{bmatrix} \sigma_u^{*2} & -\sigma_\theta^{*2} & \sigma_\theta^{*2} \\ -\sigma_\theta^{*2} & \sigma_u^{*2} & -\sigma_\theta^{*2} \\ \sigma_\theta^{*2} & -\sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix},\ \Sigma_{WT} = \begin{bmatrix} \sigma_u^{*2} & \sigma_\theta^{*2} & -\sigma_\theta^{*2} \\ \sigma_\theta^{*2} & \sigma_u^{*2} & -\sigma_\theta^{*2} \\ -\sigma_\theta^{*2} & -\sigma_\theta^{*2} & \sigma_u^{*2} \end{bmatrix}.$$

Thus, the probabilities of any of the eight possible transition paths can be written in a compact form as:

$$P\left(0=s,1=s',2=s''\right)=\overrightarrow{F}_{ss's''}\times\overrightarrow{P}_2^d\;\forall\;s,s',s''=N,A,$$

where $\overrightarrow{F}_{ss's''}$ is a row vector with the shares of workers displayed in the corresponding row of Table D.2 of the realized transition from sector s in period 0 to sector s' in period 1 and to sector s'' in period 2. Hence the probabilities for the observed transitions between period 1 and period 2 (analogues of (A.8)-(A.11)) are:

$$P\left(1=s',2=s''\right)=\left(\overrightarrow{F}_{As's''}+\overrightarrow{F}_{Ns's''}\right)\times\overrightarrow{P}_2^d$$

and the analogue of equation (A.1) for $t=2$ is:

$$P\left(2=s''\right)=\left(\overrightarrow{F}_{NNs''}+\overrightarrow{F}_{ANs''}+\overrightarrow{F}_{NAs''}+\overrightarrow{F}_{AAs''}\right)\times\overrightarrow{P}_2^d.$$

Similarly as for the cross-sectional moments in period 1, the expressions for all moments of the income distribution in period 2 can be obtained after substituting the probabilities of transition derived above and the expressions for the conditional expectations of u_{i2}^s, u_{i2}^{s2} and u_{i2}^{s3} into the corresponding formulas for the cross sectional moments in period 2. For this purpose, consider the moment $E\left(X_1^{k_1}X_2^{k_2}X_3^{k_3}X_4^{k_4}|-\infty<X_i<b_i,i=1,2,3,4\right)$ of the upper truncated multivariate normal distribution $\mathbf{N}\left(\mathbf{0},\Sigma\right)$ with $k_2=k_3=k_4=0$ and $b_1=\infty$ as a function $\mathbf{M}_4(B,\Sigma,k)$ of the variance Σ, the elements b_2, b_3 and b_4 stacked up in a vector B and the coefficient $k=k_1$, that is:

$$\mathbf{M_4}(B,\Sigma,k)\equiv E\left(X_1^k|-\infty<X_1<\infty,-\infty<X_2<B_1,-\infty<X_3<B_2,-\infty<X_4<B_3\right).$$

Define the column vectors:

$$\overrightarrow{E}_{2j}^A=\begin{bmatrix}\mathbf{M_4}(\overrightarrow{r}_{NNN},\Sigma_{NNN|A},j)\;\mathbf{M_4}(\overrightarrow{r}_{NNA},\Sigma_{NNA},j)\;\mathbf{M_4}(\overrightarrow{r}_{NAN},\Sigma_{NAN|A},j)\;\mathbf{M_4}(\overrightarrow{r}_{NAA},\Sigma_{NAA},j)...\\ \mathbf{M_4}(\overrightarrow{r}_{ANN},\Sigma_{ANN|A},j)\;\mathbf{M_4}(\overrightarrow{r}_{ANA},\Sigma_{ANA},j)\;\mathbf{M_4}(\overrightarrow{r}_{AAN},\Sigma_{AAN|A},j)\;\mathbf{M_4}(\overrightarrow{r}_{AAA},\Sigma_{AAA},j)\end{bmatrix}'$$

$$\overrightarrow{E}_{2j}^N=\begin{bmatrix}\mathbf{M_4}(\overrightarrow{r}_{NNN},\Sigma_{NNN},j)\;\mathbf{M_4}(\overrightarrow{r}_{NNA},\Sigma_{NNA|N},j)\;\mathbf{M_4}(\overrightarrow{r}_{NAN},\Sigma_{NAN},j)\;\mathbf{M_4}(\overrightarrow{r}_{NAA},\Sigma_{NAA|N},j)...\\ \mathbf{M_4}(\overrightarrow{r}_{ANN},\Sigma_{ANN},j)\;\mathbf{M_4}(\overrightarrow{r}_{ANA},\Sigma_{ANA|N},j)\;\mathbf{M_4}(\overrightarrow{r}_{AAN},\Sigma_{AAN},j)\;\mathbf{M_4}(\overrightarrow{r}_{AAA},\Sigma_{AAA|N},j)\end{bmatrix}'$$

for $j=1,2,3$, where:

$$\Sigma_{NNN}=\begin{bmatrix}\sigma_{uN}^2 & \Lambda_{NNN}\\ \Lambda_{NNN}' & \Sigma_{TT}\end{bmatrix},\;\Sigma_{NNA}=\begin{bmatrix}\sigma_{uA}^2 & \Lambda_{NNA}\\ \Lambda_{NNA}' & \Sigma_{TW}\end{bmatrix},\;\Sigma_{NAN}=\begin{bmatrix}\sigma_{uN}^2 & \Lambda_{NAN}\\ \Lambda_{NAN}' & \Sigma_{WW}\end{bmatrix},$$

$$\Sigma_{NAA}=\begin{bmatrix}\sigma_{uA}^2 & \Lambda_{NAA}\\ \Lambda_{NAA}' & \Sigma_{WT}\end{bmatrix},\;\Sigma_{ANN}=\begin{bmatrix}\sigma_{uN}^2 & \Lambda_{ANN}\\ \Lambda_{ANN}' & \Sigma_{WT}\end{bmatrix},\;\Sigma_{ANA}=\begin{bmatrix}\sigma_{uA}^2 & \Lambda_{ANA}\\ \Lambda_{ANA}' & \Sigma_{WW}\end{bmatrix},$$

$$\Sigma_{AAN}=\begin{bmatrix}\sigma_{uN}^2 & \Lambda_{AAN}\\ \Lambda_{AAN}' & \Sigma_{TW}\end{bmatrix}\;\Sigma_{AAA}=\begin{bmatrix}\sigma_{uA}^2 & \Lambda_{AAA}\\ \Lambda_{AAA}' & \Sigma_{TT}\end{bmatrix}$$

with:

$$\Lambda_{NNN}=\left[-\tilde{\sigma}_{uN}^2,-\tilde{\sigma}_{\theta N}^2,-\tilde{\sigma}_{\theta N}^2\right],\ \Lambda_{NNA}=\left[-\tilde{\sigma}_{uA}^2,\tilde{\sigma}_{\theta A}^2,\tilde{\sigma}_{\theta A}^2\right],\ \Lambda_{NAN}=\left[-\tilde{\sigma}_{uN}^2,\tilde{\sigma}_{\theta N}^2,-\tilde{\sigma}_{\theta N}^2\right],$$
$$\Lambda_{NAA}=\left[-\tilde{\sigma}_{uA}^2,-\tilde{\sigma}_{\theta A}^2,\tilde{\sigma}_{\theta A}^2\right],\ \Lambda_{ANN}=\left[-\tilde{\sigma}_{uN}^2,-\tilde{\sigma}_{\theta N}^2,\tilde{\sigma}_{\theta N}^2\right],\ \Lambda_{ANA}=\left[-\tilde{\sigma}_{uA}^2,\tilde{\sigma}_{\theta A}^2,-\tilde{\sigma}_{\theta A}^2\right],$$
$$\Lambda_{AAN}=\left[-\tilde{\sigma}_{uN}^2,\tilde{\sigma}_{\theta N}^2,\tilde{\sigma}_{\theta N}^2\right],\ \Lambda_{AAA}=\left[-\tilde{\sigma}_{uA}^2,-\tilde{\sigma}_{\theta A}^2,-\tilde{\sigma}_{\theta A}^2\right]$$

and:

$$\Sigma_{NNN|A}=\begin{bmatrix}\sigma_{uA}^2 & -\Lambda_{AAA}\\ -\Lambda_{AAA}' & \Sigma_{TT}\end{bmatrix},\ \Sigma_{NNA|N}=\begin{bmatrix}\sigma_{uN}^2 & -\Lambda_{AAN}\\ -\Lambda_{AAN}' & \Sigma_{TW}\end{bmatrix},\ \Sigma_{NAN|A}=\begin{bmatrix}\sigma_{uA}^2 & -\Lambda_{ANA}\\ -\Lambda_{ANA}' & \Sigma_{WW}\end{bmatrix},$$
$$\Sigma_{NAA|N}=\begin{bmatrix}\sigma_{uN}^2 & -\Lambda_{ANN}\\ -\Lambda_{ANN}' & \Sigma_{WT}\end{bmatrix},\ \Sigma_{ANN|A}=\begin{bmatrix}\sigma_{uA}^2 & -\Lambda_{NAA}\\ -\Lambda_{NAA}' & \Sigma_{WT}\end{bmatrix},\ \Sigma_{ANA|N}=\begin{bmatrix}\sigma_{uN}^2 & -\Lambda_{NAN}\\ -\Lambda_{NAN}' & \Sigma_{WW}\end{bmatrix},$$
$$\Sigma_{AAN|A}=\begin{bmatrix}\sigma_{uA}^2 & -\Lambda_{NNA}\\ -\Lambda_{NNA}' & \Sigma_{TW}\end{bmatrix},\ \Sigma_{AAA|N}=\begin{bmatrix}\sigma_{uN}^2 & -\Lambda_{NNN}\\ -\Lambda_{NNN}' & \Sigma_{TT}\end{bmatrix}.$$

In addition, define the (transition) event $\Omega^{ss'}\equiv\{2=s',1=s\}$ for any $s,s'=A,N$. Then, the formulas for the first, second, and third moments of income of individuals in each transition $\Omega^{ss'}$ are given by:

$$E\left(y_{i2}^{s'}|\,\Omega^{ss'}\right)=r_2^{s'}+E\left(u_{i2}^{s'}|\,\Omega^{ss'}\right)=r_2^{s'}+\frac{\left[\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\cdot\left(\overrightarrow{P}_2^d\right)'\right]\times\overrightarrow{E}_{21}^{s'}}{\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\times\overrightarrow{P}_2^d}$$

$$Var\left(y_{i2}^{s'}|\,\Omega^{ss'}\right)=\frac{\left[\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\cdot\left(\overrightarrow{P}_2^d\right)'\right]\times\overrightarrow{E}_{22}^{s'}}{\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\times\overrightarrow{P}_2^d}-E\left(u_{i2}^{s'}|\,\Omega^{ss'}\right)^2+\sigma_\mu^2$$
$$E\left(\left[y_{i2}^{s'}-E\left(y_{i2}^{s'}|\,\Omega^{ss'}\right)\right]^3|\,\Omega^{ss'}\right)=\frac{\left[\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\cdot\left(\overrightarrow{P}_2^d\right)'\right]\times\overrightarrow{E}_{23}^{s'}}{\left(\overrightarrow{F}_{Nss'}+\overrightarrow{F}_{Ass'}\right)\times\overrightarrow{P}_2^d}$$
$$-3E\left(u_{i2}^{s'}|\,\Omega^{ss'}\right)\left[Var\left(y_{i2}^{s'}|\,\Omega^{ss'}\right)-\sigma_\mu^2\right]-\left[E\left(u_{i2}^{s}|\,\Omega^{ss'}\right)\right]^3.$$

Combining those formulas with the probabilities of each transition group, it is straightforward to derive the expressions for the first, second, and third moments of income for the pool of workers in each sector in period 2 (i.e., the analogues of (A.2)-(A.7) for $t=2$).

Finally, we need to derive the expressions for the first and second moments of the growth of income for switchers. Since we already have the mapping of probabilities between the actual and the desired transitions, we only need to find the expected values of the differences $u_{i2}^N-u_{i1}^A$ and $u_{i2}^A-u_{i1}^N$ for individuals in all desired transitions groups. For this purpose, similarly as above define the (column) vectors:

$$\overrightarrow{E}_{2j}^{AN}=\begin{bmatrix}\mathbf{M_4}(\overrightarrow{r}_{NNN},\tilde{\Sigma}_{NNN|AN},j)\ \mathbf{M_4}(\overrightarrow{r}_{NNA},\tilde{\Sigma}_{NNA|AN},j)\ \mathbf{M_4}(\overrightarrow{r}_{NAN},\tilde{\Sigma}_{NAN},j)\ \mathbf{M_4}(\overrightarrow{r}_{NAA},\tilde{\Sigma}_{NAA|AN},j)...\\ \mathbf{M_4}(\overrightarrow{r}_{ANN},\tilde{\Sigma}_{ANN|AN},j)\ \mathbf{M_4}(\overrightarrow{r}_{ANA},\tilde{\Sigma}_{ANA|AN},j)\ \mathbf{M_4}(\overrightarrow{r}_{AAN},\tilde{\Sigma}_{AAN},j)\ \mathbf{M_4}(\overrightarrow{r}_{AAA},\tilde{\Sigma}_{AAA|AN},j)\end{bmatrix}'$$
$$\overrightarrow{E}_{2j}^{NA}=\begin{bmatrix}\mathbf{M_4}(\overrightarrow{r}_{NNN},\tilde{\Sigma}_{NNN|NA},j)\ \mathbf{M_4}(\overrightarrow{r}_{NNA},\tilde{\Sigma}_{NNA},j)\ \mathbf{M_4}(\overrightarrow{r}_{NAN},\tilde{\Sigma}_{NAN|NA},j)\ \mathbf{M_4}(\overrightarrow{r}_{NAA},\tilde{\Sigma}_{NAA|NA},j)...\\ \mathbf{M_4}(\overrightarrow{r}_{ANN},\tilde{\Sigma}_{ANN|NA},j)\ \mathbf{M_4}(\overrightarrow{r}_{ANA},\tilde{\Sigma}_{ANA},j)\ \mathbf{M_4}(\overrightarrow{r}_{AAN},\tilde{\Sigma}_{AAN|NA},j)\ \mathbf{M_4}(\overrightarrow{r}_{AAA},\tilde{\Sigma}_{AAA|NA},j)\end{bmatrix}'$$

for $j = 1, 2$, where:

$$\tilde{\Sigma}_{AAN} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{AAN} \\ \tilde{\Lambda}'_{AAN} & \Sigma_{TW} \end{bmatrix}, \ \tilde{\Sigma}_{NAN} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{NAN} \\ \tilde{\Lambda}'_{NAN} & \Sigma_{WW} \end{bmatrix}$$

$$\tilde{\Sigma}_{ANA} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{ANA} \\ \tilde{\Lambda}'_{ANA} & \Sigma_{WW} \end{bmatrix}, \ \tilde{\Sigma}_{NNA} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{NNA} \\ \tilde{\Lambda}'_{NNA} & \Sigma_{TW} \end{bmatrix}$$

with:

$$\tilde{\Lambda}_{AAN} = \left[-\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, \tilde{\sigma}^2_{\theta A} + \tilde{\sigma}^2_{\theta N}\right], \ \tilde{\Lambda}_{NAN} = \left[-\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{\theta A} - \tilde{\sigma}^2_{\theta N}\right]$$

$$\tilde{\Lambda}_{ANA} = \left[-\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}\right]; \ \tilde{\Lambda}_{NNA} = \left[-\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}\right]$$

and:

$$\tilde{\Sigma}_{ss's''|AN} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{ss's''|AN} \\ \tilde{\Lambda}'_{ss's''|AN} & \tilde{\Sigma}_{ss's''}(2,2) \end{bmatrix}, \ \tilde{\Sigma}_{ss's''|NA} = \begin{bmatrix} \sigma^2_{uA} + \sigma^2_{uN} - 2\sigma_{\theta AN} & \tilde{\Lambda}_{ss's''|NA} \\ \tilde{\Lambda}_{ss's''|NA} & \tilde{\Sigma}_{ss's''}(2,2) \end{bmatrix}$$

$\forall \ s, s', s'' = N, A$ (that is, the matrix in the position (2,2) of $\tilde{\Sigma}_{ss's''|AN}$ and $\tilde{\Sigma}_{ss's''|NA}$ is the same matrix of the position (2,2) of $\tilde{\Sigma}_{ss's''}$ in the expressions above, where:

$$\begin{aligned}
\tilde{\Lambda}_{AAA|AN} &= [\ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, \ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{AAA|NA} &= [-\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] \\
\tilde{\Lambda}_{ANA|AN} &= [\ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{AAN|NA} &= [\ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] \\
\tilde{\Lambda}_{ANN|AN} &= [-\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{ANN|NA} &= [\ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, \ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] \\
\tilde{\Lambda}_{NAA|AN} &= [\ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, \ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{NAA|NA} &= [-\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}] \\
\tilde{\Lambda}_{NNA|AN} &= [\ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{NAN|NA} &= [\ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}] \\
\tilde{\Lambda}_{NNN|AN} &= [-\tilde{\sigma}^2_{uN} - \tilde{\sigma}^2_{\theta A}, -\tilde{\sigma}^2_{uA} - \tilde{\sigma}^2_{\theta N}, -\tilde{\sigma}^2_{\theta N} - \tilde{\sigma}^2_{\theta A}] & \tilde{\Lambda}_{NNN|NA} &= [\ \tilde{\sigma}^2_{uA} + \tilde{\sigma}^2_{\theta N}, \ \tilde{\sigma}^2_{uN} + \tilde{\sigma}^2_{\theta A}, \ \tilde{\sigma}^2_{\theta N} + \tilde{\sigma}^2_{\theta A}].
\end{aligned}$$

Using the definitions of $\overrightarrow{E}^{AN}_{2j}$ and $\overrightarrow{E}^{NA}_{2j}$ for $j = 1, 2$, it is straightforward to derive the expressions for the first and second moments of the growth in income for switchers. For example, the corresponding expressions for switchers from N to A (analogues of (A.13) and (A.15)) are:

$$E\left(y^A_{i2} - y^N_{i1} |\, \Omega^{NA}\right) = r^A_2 - r^N_1 + E\left(u^A_{i2} - u^N_{i1} |\, \Omega^{NA}\right) = r^A_2 - r^N_1 + \frac{\left[\left(\overrightarrow{F}_{NNA} + \overrightarrow{F}_{ANA}\right) \cdot \left(\overrightarrow{P}^d_2\right)'\right] \times \overrightarrow{E}^{NA}_{21}}{\left(\overrightarrow{F}_{NNA} + \overrightarrow{F}_{ANA}\right) \times \overrightarrow{P}^d_2}$$

$$Var\left(y^A_{i2} - y^N_{i1} |\, \Omega^{NA}\right) = \frac{\left[\left(\overrightarrow{F}_{NNA} + \overrightarrow{F}_{ANA}\right) \cdot \left(\overrightarrow{P}^d_2\right)'\right] \times \overrightarrow{E}^{NA}_{22}}{\left(\overrightarrow{F}_{NNA} + \overrightarrow{F}_{ANA}\right) \times \overrightarrow{P}^d_2} - E\left(u^A_{i2} - u^N_{i1} |\, \Omega^{NA}\right)^2 + 2\sigma^2_{\mu}.$$

Similar expressions can be derived for switchers from N to A (analogues of (A.14) and (A.16)).

As in the cases of the frictionless economy and the model with a compensating differential, with the explicit formulas for the set of 22 moments of interest we numerically verified that the system has an unique solution for the 12 elements of Θ.

Finally, in the most general model with both barriers to sectoral mobility and a compensating differential, the same formulas for the set of 22 moments mentioned in this section apply when substituting $r_t = r_r^N - r_t^A - \ln cd$ for $r_t = r_r^N - r_t^A$. As in previous cases, we verified the system of 22 moments has a unique solution for the 13 elements of Θ (the same 10 parameters as in the frictionless economy, the two probabilities of involuntary choices, and the compensating differential).

E Proofs

Proof of Proposition 1

Under the assumptions of Proposition 1, the expression for expected log income growth of switchers from agriculture to non-agriculture given in (A.13) simplifies to:

$$E\left(y_{i2}^N - y_{i1}^A | \, 2 = N, 1 = A\right) = \mathbf{M}_3(\overrightarrow{0}, \tilde{\Sigma}_{AN}, 1),$$

and where $\tilde{\Sigma}_{AN}$ can be simplified to

$$\tilde{\Sigma}_{AN} = \begin{bmatrix} \sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2 & -\left(\sigma_\theta^{*2} + \sigma_{\varepsilon N}^2\right) & \sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 \\ -\left(\sigma_\theta^{*2} + \sigma_{\varepsilon N}^2\right) & \sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2 & -\sigma_\theta^{*2} \\ \sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 & -\sigma_\theta^{*2} & \sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2 \end{bmatrix}.$$

Re-write $\tilde{\Sigma}_{AN}$ in terms of the correlation matrix C_{AN} :

$$C_{AN} = \begin{bmatrix} 1 & \frac{-\left(\sigma_\theta^{*2}+\sigma_{\varepsilon N}^2\right)}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} & \frac{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} \\ \frac{-\left(\sigma_\theta^{*2}+\sigma_{\varepsilon N}^2\right)}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} & 1 & \frac{-\sigma_\theta^{*2}}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} \\ \frac{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} & \frac{-\sigma_\theta^{*2}}{\sigma_\theta^{*2}+\sigma_{\varepsilon A}^2+\sigma_{\varepsilon N}^2} & 1 \end{bmatrix}$$

and denote ρ_{ij}^{AN} the (i, j) element of C_{AN}. Using our definition of $\mathbf{M}_3(\cdot)$ and explicit formulas for the moments of the upper-truncated multivariate normal distribution in the trivariate case (derived from recurrence relations) from Kan and Robotti (2017), $\mathbf{M}_3(\overrightarrow{0}, \tilde{\Sigma}_{AN}, 1)$ can be re-written as:

$$\mathbf{M}_3(\overrightarrow{0}, \tilde{\Sigma}_{AN}, 1) = -\sqrt{\sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2}\phi(0) \left[\frac{\rho_{12}^{AN}\Phi_2\left([\infty, 0]\,; \rho_{13\cdot2}^{AN}\right)}{\Phi_3\left([\infty, 0, 0]\,; C_{AN}\right)} + \frac{\rho_{13}^{AN}\Phi_2\left([\infty, 0]\,; \rho_{12\cdot3}^{AN}\right)}{\Phi_3\left([\infty, 0, 0]\,; C_{AN}\right)}\right]$$

with $\rho_{ij\cdot k}^{AN} = \frac{\rho_{ij}^{AN} - \rho_{ik}^{AN}\rho_{jk}^{AN}}{\sqrt{\left(1-\left(\rho_{ik}^{AN}\right)^2\right)\left(1-\left(\rho_{jk}^{AN}\right)^2\right)}}$. Noticing that in our case $\Phi_2\left([\infty, 0]\,; \rho_{ij\cdot k}^{AN}\right) = \frac{1}{2}\ \forall\ i, j, k$, we have

$$\mathbf{M}_3(\overrightarrow{0}, \tilde{\Sigma}_{AN}, 1) = \frac{\phi(0)\left[\sigma_{\varepsilon N}^2 - \sigma_{\varepsilon A}^2\right]}{2\sqrt{\sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2}\Phi_3\left([\infty, 0, 0]\,; C_{AN}\right)},$$

which is positive if and only if $\sigma_{\varepsilon N}^2 > \sigma_{\varepsilon A}^2$. Following the same steps we find the expected log income growth of switchers to from non–agriculture to agriculture as:

$$\mathbf{M}_3(\overrightarrow{0}, \tilde{\Sigma}_{NA}, 1) = -\frac{\phi(0)\left[\sigma_{\varepsilon N}^2 - \sigma_{\varepsilon A}^2\right]}{2\sqrt{\sigma_\theta^{*2} + \sigma_{\varepsilon A}^2 + \sigma_{\varepsilon N}^2}\Phi_3\left([\infty, 0, 0]\,; C_{NA}\right)}.$$

Furthermore, it can be verified that $\Phi_3([\infty,0,0];C_{NA}) = \Phi_3([\infty,0,0];C_{AN})$, which implies that $\mathbf{M}_3(\overrightarrow{0},\tilde{\Sigma}_{AN},1)$ and $\mathbf{M}_3(\overrightarrow{0},\tilde{\Sigma}_{NA},1)$ have the opposite sign but the same magnitude. QED.

F Alternative Specification of Mobility Barriers

F.1 Model with Switching Costs

In our baseline model barriers to mobility across sectors are captured by a probability of being allocated to a different sector than the worker would prefer. In this appendix we consider an alternative type of barrier to sectoral mobility: a utility cost of switching across sectors. This cost could reflect tangible expenditures such as training costs, or intangibles such as social adjustment costs. There is a large literature estimating switching costs in other settings (Dixit and Rob, 1994; Artuc, Chaudhuri and McLaren, 2010; Dix-Carneiro, 2014).

Denote by $\phi^{s's}$ the utility cost for switching from sector s' (which was chosen in $t-1$) to sector s, common to all individuals. The value of working in sector s is then:

$$V^s_{sc}(\Omega_{it}) = \ln y^s_t(\Omega_{it}) - \ln C^{s_{t-1}s_t}(\Omega_{it}),$$

where

$$C^{s_{t-1}s_t}(\Omega_{it}) = C^{s's} = \begin{cases} \phi^{s's} & \text{if } s \neq s' \\ 1 & \text{if } s = s' \end{cases}.$$

The problem of the worker is the same as in (3) but with $V^s(\Omega_{it}) = V^s_{sc}(\Omega_{it})$. The switching costs act as if proportionally scaling down or up income, so they can be interpreted in terms of annual earnings. Switching costs adds two parameters (ϕ^{AN},ϕ^{NA}) to Θ.[32]

Perhaps a natural case is to think about switching costs as indeed costs, i.e. imposing $\ln\phi^{s's}$ to be non-negative. However, we also consider a more general case in which switching "costs" could measure a utility compensation (for $\ln\phi^{s's} < 0$). There is a link between this more general case and a specification with compensating differential, in that the model with a compensating differential cd is observationally equivalent to a model with switching costs $\phi^{AN} = cd$, $\phi^{NA} = 1/cd$.

F.2 Results

When we restrict the switching costs to be non-negative we find that they have little impact on the estimates relative to the basic frictionless case (cf. column 1 in panel A of Table F.1 with column 2 in Table 5 for structural parameters and column 3 in panel B of Table F.1 with column 4 in Table 7 for auxiliary model coefficients and fit). The reason is that the estimated costs are small in magnitude, and in particular, the zero bound for the cost of switching from non-agriculture to agriculture is binding. This result might seem surprising, given that the literature estimating utility costs of switching sectors typically finds them to be large, often equivalent to multiples of a worker's annual income (e.g. Artuc, Lederman and Porto, 2015). But the magnitudes might not be easily comparable across studies, as they depend on what other mechanisms of sector determination are built into the respective models. In our case, when we allow for self-selection according to comparative advantage then positive switching costs do not have much additional explanatory

[32]Following similar steps as in Appendix D we can demonstrate how the structural parameters are identified. These derivations are available upon request.

power. In particular, if switching to agriculture was costly then it would be even more puzzling why so many workers make the move.

The situation is different if we remove the restriction on the sign of the switching costs. Column 3 in panel A of Table F.1 shows the estimates for the model with unrestricted switching costs, and column 4 in panel B of Table F.1 the corresponding coefficients of the auxiliary models. The switching costs are of opposite signs, approximately symmetric in magnitude, and of a large magnitude. A worker switching from agriculture to non-agriculture faces a cost of 64 lp of annual income equivalent (i.e. roughly equivalent to her annual income). That is, a worker who actually moves from agriculture to non-agriculture, must have a value of her human capital in non-agriculture at least 90% largeer than in agriculture. For smaller differences, the worker remains in agriculture. A worker switching towards agriculture receives a utility *compensation* equivalent to almost doubling her new agricultural income. That is, a worker who actually switches from non-agriculture to agriculture, could have a value of her human capital in agriculture as much as 47% smaller than in non-agriculture. Because of this implied compensation, the model now has an easier time justifying why workers switch to agriculture. The overall fit of the model with unrestricted switching costs is thus substantially better than for the basic frictionless model.

However, since the estimated switching costs are nearly symmetric (i.e. $\phi^{AN}\phi^{NA}$ is close to 1), the model with unrestricted switching costs is nearly identical to the compensating differential model from section 4.2. This can by seen by comparing results in Table F.1 with those in columns 3 in Table 5 and 5 in Table 7. Despite having one more free parameter, the model with unrestricted switching costs offers only a negligible improvement over the compensating differential model in terms of model fit and it offers basically the same interpretation of the data - workers need to be compensated in utility terms to switch to agriculture in the observed numbers.

Finally, note that the model with utility switching costs fits the data substantially worse than our baseline model with forced sectoral choices. This is a fair comparison as both specifications have the same number of free parameters.

Table F.1: Results for Model with Switching Costs

(A) Parameter Estimates

(1) Parameter	(2) Non-negative switching costs	(3) Unrestricted switching costs
$\sigma^2_{\theta^A}$	0.27	0.50
	(0.02)	(0.05)
$\sigma^2_{\theta^N}$	0.50	0.45
	(0.04)	(0.04)
$\sigma_{\theta^{AN}}$	0.30	0.16
	(0.04)	(0.04)
$\sigma^2_{\varepsilon^A}$	0.00	0.12
	(0.01)	(0.03)
$\sigma^2_{\varepsilon^N}$	0.04	0.00
	(0.01)	(0.01)
σ^2_{ν}	0.74	0.72
	(0.01)	(0.01)
R^A_1	0.92	0.45
R^A_2	1.37	0.73
R^A_3	1.27	0.57
R^A_4	1.63	0.81
R^A_5	1.90	1.07
R^N_1	1.27	1.33
R^N_2	2.05	1.95
R^N_3	1.70	1.69
R^N_4	2.08	2.21
R^N_5	2.64	2.42
Switching cost of moving from sector s to sector s' ($\phi^{ss'}$)		
$\ln \phi^{AN}$	0.12	0.64
	(0.03)	(0.04)
$\ln \phi^{NA}$	0.00	-0.63
	(0.00)	(0.03)

(B) Coefficients of Auxiliary Models

(1) Coefficient δ_i (weight Ω_i)	(2) Data ($\hat{\delta}_i$)	(3) Non-negative switching costs	(4) Unrestricted switching costs
δ_1 (1)	0.57	0.62	0.60
δ_2 (1)	0.40	0.23	0.35
δ_3 (5)	-0.05	0.05	-0.10
δ_4 (5)	-0.31	-0.34	-0.36
δ_5 (5)	0.15	0.26	0.29
δ_6 (5)	-0.42	-0.22	-0.34
δ_7 (5)	-0.17	-0.16	-0.20
δ_8 (1)	0.38	0.41	0.45
δ_9 (1)	0.34	0.34	0.27
δ_{10} (1)	0.63	0.60	0.55
δ_{11} (1)	0.85	0.74	0.77
δ_{12} (5)	0.76	0.65	0.66
δ_{13} (1)	1.10	1.07	1.04
δ_{14} (1)	0.89	0.95	0.89
δ_{15} (1)	1.05	1.16	1.17
δ_{16} (1)	1.27	1.38	1.30
δ_{17} (10)	0.70	0.73	0.70
δ_{18} (10)	0.01	0.04	-0.04
δ_{19} (10)	-0.02	-0.02	0.00
δ_{20} (10)	-0.03	-0.06	-0.03
δ_{21} (10)	-0.04	-0.01	-0.12
δ_{22} (10)	0.21	0.22	0.22
δ_{23} (10)	0.68	0.69	0.63
δ_{24} (3)	1.24	1.00	1.13
δ_{25} (3)	0.95	1.14	1.10
δ_{26} (3)	1.43	1.46	1.59
δ_{27} (3)	1.08	1.56	1.45
δ_{28} (3)	1.73	1.58	1.52
δ_{29} (3)	1.86	1.53	1.52
Overall fit (loss function)		1.866	1.439

Notes: Panel A: A description of the structural parameters in available in Table 5. Standard errors from 100 bootstraps in parentheses. All standard errors for R^s_t are smaller than 0.1 times the point estimate value. Panel B: A description of the auxiliary model coefficients is available in Table 7. Ω_i refers to the $i-$th element of the diagonal of the matrix Ω. Overall fit is the value of the loss function being minimized by the estimation procedure.