

Risk and Rural-Urban Structural Change^{*}

Florian Trouvain
University of Oxford

Current Version: September 8, 2026

First Version: September 1, 2025

Preliminary – Comments Welcome!

[\[Please click here for most recent version\]](#)

Abstract

Why do so few rural households in developing economies migrate to high-wage urban areas? I build a tractable dynamic general equilibrium model with endogenous migration to quantify the role of risk as a barrier to rural-urban migration. Migration is risky in two ways. First, households may not know their initial urban income before they move. Second, households in the urban sector are more exposed to persistent income shocks. I discipline the model using South African panel data on income and consumption. Uncertainty about initial urban income disproportionately deters high-income rural households and generates negative selection into migration, while persistent income shocks instead deter rich and poor alike. Together, these risks hold the urban population share 16 percentage points below where it would otherwise be. Wealth transfers do not raise migration because savings provide little protection against persistent shocks. High levels of wealth even discourage migration. A model that ignores initial uncertainty and persistent shocks overstates the residual migration wedge by 80 percent.

^{*}E-mail: florian.trouvain@economics.ox.ac.uk. I thank the STEG research initiative for funding under grant 1030. For their insights I thank Wyatt Brooks, Doug Gollin, Joe Kaboski, Fernando Leibovici, Mushfiq Mobarak, B. Ravikumar, Brit Sharoni, and Gianluca Violante. I am indebted to Qingyu Chen and Pedro Almas for outstanding research assistance.

1 Introduction

A striking regularity across developing economies is the large income gap between rural and urban households. These gaps raise a natural question: would rural households gain from moving to the urban sector? A common view is that the gains are limited because observed wage gaps largely reflect efficient sorting across sectors. This view is usually based on settings where households have perfect information about their long-run urban productivity. In this paper, I study permanent rural-urban migration in a dynamic framework with incomplete markets, where long-run urban productivity is potentially unknown, and subject to persistent shocks. Initial uncertainty and persistent income risk turn out to be first-order barriers to migration.

I emphasize two distinct sources of risk. First, households may face uncertainty regarding their initial urban earnings if they migrate. This is especially relevant for the case of permanent rural-urban migration usually involving a move from agricultural to non-agricultural work. Second, permanent migration remains risky even when households know their initial urban income because persistent shocks after migration affect long-run income trajectories. Both types of risk are needed to match the South African panel data, which form the main application.

Removing initial uncertainty and persistent shocks raises the urban population share from 65 to 81 percent. An immediate implication is that models that ignore initial uncertainty and persistent income shocks substantially overstate the residual migration wedge. My framework suggests that this commonly inferred wedge is inflated by as much as 80 percent.

Both risks retard rural-urban structural change substantially, but they affect high- and low-income rural households differentially. Uncertainty about the initial urban income draw holds back high-income rural households, who have the most to lose from a low draw. This feature generates negative selection into migration. Persistent income shocks after migration, by contrast, deter high- and low-income households alike.

So why do rural households not simply accumulate wealth in the form of buffer-stock savings to self-insure against these risks? The problem is that savings provide little protection against persistent shocks, which shift the entire path of future labor income. Moreover, greater wealth can even reduce the incentive to migrate. Migration raises labor income, not asset income. As wealth increases, labor income becomes a smaller share of total household income, chipping away at the gains from migration.

My analysis proceeds in three steps. I start out by building a theoretical framework that embeds an endogenous migration choice into an incomplete market model where households have access to a risk-free bond, and make forward-looking consumption-saving choices in the presence of income risk. I build on previous work, in particular Lagakos, Mobarak, and Waugh (2023) (henceforth LMW), but depart from it in three crucial ways that are central for studying permanent rural-urban migration.

First, I develop a model with birth and death dynamics. This allows for fully persistent income shocks, which are ruled out in a stationary infinite-horizon environment. Moreover, newborn households need not be born in the sector where their productivity is highest. Contrast this with standard infinite-horizon models with fixed long-run sectoral productivity. In such frameworks, households eventually sort into their preferred sector, and observed migration is concentrated among households that are largely indifferent between sectors. These features change the role of risk. Without persistent shocks, migration is primarily a response to adverse transitory rural shocks and can itself serve as a form of insurance, as in LMW. With persistent urban risk, migration instead exposes households to uncertainty about their long-run income.

Second, I allow migrants to have imperfect information about initial urban labor market outcomes. I incorporate an information friction that nests two limiting cases: full and partial information. In the full-information case, households observe their urban outcome before migrating. In the partial-information case, they do not observe their urban outcome ex ante and instead form expectations based on the statistical relationship between rural and urban income. This departs from standard Roy models with perfect information and is especially relevant for permanent rural-urban migration, where households enter a new economic environment for the first time.

Third, I introduce a tractable way to model selection into migration in dynamic Roy models that sidesteps the curse of dimensionality. When a migration opportunity arrives, households receive an urban productivity draw that can be arbitrarily correlated with their rural type. This allows for rich patterns of comparative advantage and selection into migration, which I discipline directly using the data. At the same time, households need not carry urban productivity as an additional state variable while in the rural sector. The dimension of the household state space therefore does not grow with the number of potential sectors. This tractability allows me to embed the migration problem in general equilibrium with an endogenous rural-urban wage gap and fixed migration costs.¹

In the second step, I quantify the model using household-level panel data on income and consumption from South Africa. I use the model to motivate a novel set of easily computable empirical moments that are informative about the riskiness of migration.

A first set of moments concerns the nature of income risk within each sector, which requires estimating sector-specific income processes. I follow Crawley, Holm, and Tretvoll (2026), which proposes a tractable income process that generalizes the canonical persistent-transitory model to include imperfectly transitory “passing shocks”. These are lumpy shocks that eventually mean-revert. While this generalization matters little for urban households, it is important for rural households where imperfectly transitory shocks are the dominant source of income volatility. In contrast, fully persistent shocks are more important in the urban sector, which makes long-term urban earnings uncertain, even in the case where households have full information about their initial urban productivity draw. I confirm that these sectoral differences in income risk hold more broadly across developing countries, which represents a standalone empirical contribution that should be of interest to researchers using heterogeneous agent models to study developing economies.

A second set of moments concerns the uncertainty associated with rural-urban migration around the time of migration. Intuitively, migrants may not know whether they land a high or low paying first job in the urban sector, reminiscent of the role of urban unemployment in Harris and Todaro (1970). To discipline the information structure that governs what fraction of migrants have advance information, I measure how frequently rural-urban migrants experience income losses after moving. Frequent post-migration income losses indicate limited ex-ante information.

Using these empirical moments, I estimate the structural parameters of the model. Only about one in five potential migrants have full information about their initial persistent urban income type, a key determinant of success in the urban sector. This uncertainty generates negative selection on rural income, consistent with the South African data.

Negative selection has sometimes been interpreted as evidence against risk as a barrier to migration. The argument is straightforward. If risk deters migration, it should primarily hold back low-income households. High-income households can accumulate savings to self-insure against urban risk and should there-

¹I also develop a new technique to solve continuous-time incomplete-market models on a non-linear grid that is particularly useful for models with non-convexities, such as fixed migration costs. See appendix B.

fore be more likely to migrate. Risk should thus generate positive selection on income and wealth. This logic breaks down with persistent shocks. Buffer-stock savings provide little protection against a persistent fall in income. Moreover, wealth can even reduce the gains from migration. Migration raises labor income, not asset income. As wealth increases, labor income becomes a smaller share of total household income, chipping away at the gains from migration. The model can therefore generate negative selection on both income and wealth even when risk is a major barrier to migration.

I also explore the implications of the model for consumption risk, since income risk only deters migration when it translates into consumption risk. Rural households appear well insured: the passthrough from income shocks to consumption is low, consistent with a large literature following Townsend (1994). Surprisingly, the model reproduces these low passthrough estimates despite containing no insurance beyond buffer-stock savings. The reason is that rural income shocks are largely mean-reverting, and buffer-stock savings insure against such shocks effectively.

The same mechanism explains why consumption insurance appears to worsen after rural-urban migration, an influential theme in the literature (Munshi and Rosenzweig, 2016). In my framework, households need not lose access to informal insurance when they move. They instead become exposed to more persistent income shocks, which are harder to self-insure. What appears to be a loss of insurance therefore reflects a change in the nature of income risk.

In the final step, I quantify the aggregate impact of risk on rural-urban structural change. Risk requires heterogeneous outcomes, uncertainty about those outcomes, and incomplete markets. I remove these ingredients in turn to study their implications for aggregate migration, selection, and welfare.

Providing all potential migrants with information about their initial urban persistent type raises the urban population share from 0.65 to 0.76 in general equilibrium. Information makes selection into migration more positive: households further up the rural income distribution migrate when they learn that their urban draw is high.

Insurance raises migration by a similar amount but operates differently. Insuring income shocks after migration raises the urban share to 0.73; insuring the initial draw as well raises it to 0.79. Unlike information, insurance protects rich and poor households alike against urban income risk. The two can therefore produce similar aggregate migration responses but very different selection and welfare effects.

General equilibrium dampens these migration responses by about one-third. As migration rises, land becomes more abundant and rural wages increase, reducing the incentive for further migration. The flipside is that higher rural wages deliver welfare gains even to households that never migrate.

Next, I contrast risk with liquidity constraints, another prominent explanation for low rural-urban migration in the presence of fixed moving costs. Relaxing the borrowing constraint raises the urban population share by only one percentage point, despite the model featuring a poverty trap consistent with Banerjee, Duflo, and Sharma (2021). Two features of the income process explain the small effect. Urban shocks are persistent, so borrowing provides little protection against them. Rural shocks, by contrast, are volatile but mean-reverting. Rural households therefore do not remain ultra-poor for long and escape the poverty trap quickly.² Risk is therefore a much larger barrier to rural-urban migration than liquidity.

Lastly, I ask by how much residual migration wedges are inflated when initial uncertainty and persistent income shocks are omitted from the analysis. A researcher using my framework but assuming full information and complete insurance in the urban sector must raise the migration wedge by 78.66% to repro-

²I also consider a more persistent rural income process, which substantially increases the welfare cost of poverty traps. This suggests that the importance of poverty traps depends critically on the persistence of low-income states, rather than on the presence of non-convex costs.

duce observed migration. In a complementary exercise, I instead remove initial uncertainty and persistent shocks and reestimate the model, bringing the framework closer to common practice. Both approaches reach the same conclusion: a large share of the cost of moving is the risk of moving.

Contribution to the Literature. The paper contributes to a large literature on rural-urban migration and structural change. The classic references are Lewis (1954)’s theory of rural surplus labor, and Harris and Todaro (1970)’s two-sector model of urban unemployment.

The paper is most closely related to the influential work of LMW, which embeds endogenous migration in a heterogeneous-agent model to study the migration experiment of Bryan, Chowdhury, and Mobarak (2014), and shares important features with Kleemans (2015). In LMW, migration is largely temporary and serves as a coping strategy after negative rural income shocks. I instead study permanent migration, where uncertainty about long-run urban outcomes makes migration itself risky.³

A related literature studies whether sorting across sectors explains the agricultural productivity gap (Lagakos and Waugh, 2013; Alvarez-Cuadrado, Amodio, and Poschke, 2026). These models typically assume that households know their productivity in each sector when choosing where to work. I show that uncertainty about urban productivity fundamentally changes selection.⁴ In the model and the South African data, rural-urban migrants are negatively selected on income, in contrast to Young (2013)’s influential interpretation of rural-urban gaps through positive sorting on unobserved skill. More generally, the framework quantifies how risk shapes the allocation of unobserved human capital across sectors.⁵

The paper also relates to work emphasizing informal insurance as a determinant of migration, particularly Munshi and Rosenzweig (2016) and Morten (2019), building on Ligon, Thomas, and Worrall (2002). I offer a different interpretation of why consumption insurance appears to worsen after rural-urban migration. Households need not lose access to informal insurance when they move; they simply become exposed to more persistent income risk. A standard buffer-stock model reproduces the low consumption passthrough observed for rural households once the income process allows for mean-reverting shocks. This connects the paper to work estimating persistent and transitory income risk, especially Blundell, Pistaferri, and Preston (2008) (henceforth BPP). Income dynamics in developing economies remain relatively unexplored, with notable exceptions in De Magalhaes and Santaaulalia-Llopis (2018), De Magalhaes, Martorell, and Santaaulalia-Llopis (2024), and Carli, Rodriguez-Sala, and Santaaulalia-Llopis (2025). I contribute to this literature by showing that the nature of income risk differs systematically between rural and urban sectors in developing economies.⁶

Recent experimental work studies information and insurance as barriers to rural-urban migration (Baseler,

³Treating permanent and temporary migration separately is justified by previous work: the RCT of Bryan, Chowdhury, and Mobarak (2014) raised temporary rural-urban migration in Bangladesh but had no impact on permanent migration; see Mobarak and Reimão (2020).

⁴Imperfect information as a barrier to migration is also studied in Porcher, Morales, and Fujiwara (2024) and Porcher (2025), who recover information frictions from revealed preference in a quantitative spatial framework. Other relevant static Roy models that explicitly deal with selection into migration are Borjas (1987) and Borjas, Kauppinen, and Poutvaara (2019). Bertoli (2010) argues that imperfect information can lead to more negative selection in the context of international migration. Bryan and Morten (2019) quantify how migration costs suppress aggregate productivity.

⁵See also the large empirical literature on rural-urban wage gaps. Beegle, De Weerdt, and Dercon (2011), Hamory et al. (2020), Alvarez (2020), Pulido and Swiecki (2018), and Lagakos et al. (2020) for micro-economic estimates of rural-urban wage gaps. See Caselli (2005), Restuccia, Yang, and Zhu (2008), and Gollin, Lagakos, and Waugh (2014) for the role of rural-urban wage gaps and agricultural vs. non-agricultural productivity gaps in accounting for cross-country income differences.

⁶A large empirical literature has studied the degree of informal insurance in developing economies. See Alderman and Paxson (1994), Udry (1994), Morduch (1995), Ligon (1998), Angelucci and De Giorgi (2009), Mazzocco and Saini (2012), Kinnan and Townsend (2012), Chiappori et al. (2014), Jack and Suri (2014), Samphantharak and Townsend (2018), Santaaulalia-Llopis and Zheng (2018), Attanasio et al. (2025), Meghir et al. (2022), and Ligon (2025). An even larger literature has estimated persistent and transitory income shocks, see MaCurdy (1982), Abowd and Card (1989), Gottschalk et al. (1994), Deaton and Paxson (1994), Storesletten, Telmer, and Yaron (2004), Kaplan and Violante (2010), Guvenen (2009), Low, Meghir, and Pistaferri (2010), Huggett, Ventura, and Yaron (2011), Guvenen and Smith (2014), Guvenen et al. (2021) and Braxton et al. (2025).

2023; Barnett-Howell et al., 2025; Miner, 2025). Most closely related is contemporaneous work by Baseler and Gordeev (2026), who build a structural model of Kenyan migration to jointly interpret these experiments. They find that urban unemployment and biased expectations explain relatively little migration, leaving a large residual migration wedge. My paper unpacks this residual. Once migration involves initial uncertainty and persistent shocks, much of the residual wedge is accounted for by risk.⁷

Lastly, I complement the experimental literature by incorporating general-equilibrium feedback effects from migration, which are quantitatively important and consistent with recent evidence in Madhok et al. (2025).

The rest of the paper proceeds as follows. Section 2 develops the model. Section 3 quantifies the model using South African panel data. Section 4 studies the role of risk as a barrier to structural change and explores additional counterfactuals. A final section concludes.

2 Theory

I next develop a model of risky permanent rural-urban migration. Additional details, derivations, and proofs are deferred to appendix C.

2.1 Environment

Time is continuous and markets are competitive. I consider a small open economy setup where the only tradable asset is a risk-free bond with exogenous interest rate r . A unit-mass of households indexed by $i \in [0, 1]$ make consumption-saving choices, and supply their labor inelastically. Households are either rural (R) or urban (U), and rural households additionally face a choice to migrate to the urban sector.

I consider stationary equilibria, which require reverse urban-rural flows. I detail how to embed reverse migration flows tractably into my framework at the end of this section. One unit of time equals one year. I omit time subscripts, and use $\dot{x}$ as shorthand for $\frac{dx}{dt}$ and $\partial_b v$ as shorthand for $\frac{\partial v}{\partial b}$.

Income Risk. Household labor income within each sector $j \in \{R, U\}$ equals

$$W_i(p, \epsilon) = w_{j(i)} e^{p_i + \epsilon_i}, \quad (1)$$

where w_j is the sectoral efficiency wage, and the random variables p_i and ϵ_i are idiosyncratic state variables. These state variables follow persistent and transitory stochastic processes reflecting sector-specific income risk.

The persistent and transitory processes are as follows:

$$dp = \sigma_{p,j} dZ, \quad (2)$$

$$d\epsilon = (\epsilon' - \epsilon) dJ_j, \quad \epsilon' \sim N(0, \sigma_{\epsilon,j}), \quad (3)$$

where dZ is a Wiener process,⁸ and dJ_j is a Poisson counter with arrival rate $\lambda_{\epsilon,j}$. In words, the transitory process consists of i.i.d. draws from a mean zero normal distribution that arrive at rate $\lambda_{\epsilon,j}$. These shocks

⁷Note that information frictions differ starkly: Baseler and Gordeev (2026) assume a common bias in the level of expected urban income, whereas I introduce idiosyncratic uncertainty about a household's own urban type affecting only a subset of households. See also Porcher (2025) for recent work endogenizing the information set of migrants using a model of rational inattention.

⁸The absence of a drift term is consistent with the empirical application later on where drift terms are partialled out in a first-stage regression procedure.

are sometimes described as “passing shocks” as outcomes eventually revert back. The setup builds directly on Crawley, Holm, and Tretvoll (2026), and nests the canonical persistent-transitory model of MaCurdy (1982) when $\lambda_{\epsilon,j}$ becomes large.⁹

Households die at Poisson rate ϕ and are replaced by a newborn that inherits their assets. Newborns draw their persistent type from a normal distribution $p \sim N(c_{0,j}, \sigma_{0,j})$, and their transitory type from a mean zero normal distribution $\epsilon' \sim N(0, \sigma_{\epsilon,j})$. This perpetual youth structure (Yaari, 1965; Blanchard, 1985) offers two benefits relative to infinite-horizon models. First, it allows for a random walk in income while preserving the stationarity of the income distribution. Second, newborns need not start in the sector where they are most productive so there is potential for observed migration based on comparative advantage, whereas infinitely lived households have already sorted in steady state and so never actually migrate.¹⁰

Urban Household Problem. Urban households discount the future at rate ρ , and have CRRA preferences with coefficient of risk aversion γ . Urban households solve the following dynamic consumption-saving problem

$$\begin{aligned} V_U(b, p, \epsilon) = \max_{\{c_s\}} \mathbb{E}_t \left[\int_t^\infty e^{-(\rho+\phi)(s-t)} \frac{c_s^{1-\gamma}}{1-\gamma} ds \right] \\ \text{s.t.} \quad \dot{b} = rb + W_U(p, \epsilon) - c, \quad b \geq 0, \end{aligned} \quad (4)$$

subject to a standard intertemporal budget constraint where c is household consumption and b is a risk-free bond paying return r . Moreover, households face a zero-borrowing constraint.

Rural Household Problem. The rural household problem is identical to the urban household problem except for an additional migration choice.

Migration. Rural-Urban migration works as follows.

- Rural households receive migration opportunities at Poisson rate λ_m , and move when the expected value of moving to the urban sector exceeds staying put.
- Migration entails a one-time monetary cost $f_{RU} \geq 0$ paid in rural labor such that urban assets after migration equal $b_U = b_R - f_{RU}w_R$.
- Migration includes an idiosyncratic taste shock φ , and a disutility wedge $\tau_{RU} > 0$, following a growing literature in dynamic spatial economics (Kennan and Walker, 2011; Caliendo, Dvorkin, and Parro, 2019). There is no disutility of staying in rural, $\tau_{RR} = 1$.
- Taste shocks φ_j are i.i.d., and drawn from a Frechet distribution with scale parameter μ and shape parameter s .

Both the taste shocks and the location-specific disutility are raised to the power $1-\gamma$, which provides the appropriate rescaling under household risk aversion so that the resulting utility wedge admits a consumption-equivalent interpretation.¹¹ Because the wedge admits this consumption-equivalent interpretation, the

⁹The implied autocovariance structure is identical to that of an Ornstein-Uhlenbeck (continuous-time AR(1)) process; however, the passing shocks better reflect the jump-like nature of observed income dynamics.

¹⁰The structure also reduces the computational burden, and provides a more appropriate benchmark in rural settings where an explicit life-cycle framework is less compelling due to widespread intergenerational co-residence.

¹¹The issue here is that when $\gamma > 1$, the value function is negative. Since the shocks are multiplicative, care must be taken in how to normalize the raw shocks. The normalization I have chosen ensures that the Frechet shape parameter preserves its usual meaning. It also makes clear that risk aversion and preference heterogeneity are not separately identified from migration flows.

analysis is robust to mis-measured price gaps across sectors.¹²

A rural household thus solves the following discrete choice problem upon receiving a migration opportunity

$$V_m = \max_{R,U} \left\{ \mathbb{E} [V_U(b_U, p_U, \epsilon_U) | p_R, \epsilon_R] \cdot \left(\frac{\varphi_U}{\tau_{RU}} \right)^{1-\gamma}, V_R(b_R, p_R, \epsilon_R) \cdot \varphi_R^{1-\gamma} \right\}, \quad (5)$$

where V_m denotes the value of the migration opportunity, and the expectation in (5) is taken with respect to persistent and transitory urban type (p_U, ϵ_U) . The conditional expectation in (5) captures the first of the two sources of risk, uncertainty about the initial urban draw. How much risk it carries depends crucially on (i) the stochastic relationship between rural and urban income, and (ii) the information structure governing which variables are observed at the time of migration. These elements are introduced through the following two modeling choices.

First, the relationship between rural and urban persistent type follows a draw from a normal distribution

$$p_U | p_R \sim N(\rho_{RU} \cdot p_R + c_{RU}, \sigma_{RU}), \quad (6)$$

where the parameter $\rho_{RU} > 0$ induces a positive correlation between rural and urban type, σ_{RU} is the standard deviation around the conditional mean, and $c_{RU} \leq 0$ captures skill downgrading,¹³ where $F(\cdot | p_R)$ denotes the CDF associated with (6).

In contrast, I assume that transitory types are independent across sectors, $\epsilon_U \perp \epsilon_R$, so ϵ_R disappears from the conditional expectation in (5). This is a relatively unimportant assumption relaxed at a later point.

The second, and crucial, modeling choice is whether the household learns their persistent urban type before deciding whether to migrate. I consider two limiting cases, one of partial and one of full information regarding p_U .

With probability ζ , households have partial information (pi) and use (6) to compute the expected value of migration

$$\mathbb{E} [V_U(b_U, p_U, \epsilon_U) | p_R] = \int \int_{\epsilon' p_U} V_U(b_U, p_U, \epsilon') dF(p_U | p_R) dG_U(\epsilon') := \bar{V}_U^{pi}(b_R, p_R). \quad (\text{partial information})$$

This partial information case serves as a natural lower bound, and an econometrician with panel data can, in principle, recover the information necessary to compute this expectation.

At the same time, households may have superior information sets relative to an econometrician with access to panel data. I assume that with probability $1 - \zeta$, potential migrants have full information (fi) about their urban persistent type around the time of migration so the only unknown around the time of migration is the urban transitory component,

$$\mathbb{E} [V_U(b_U, p_U, \epsilon_U) | p_R] = \int_{\epsilon'} V_U(b_U, p_U, \epsilon') dG_U(\epsilon') := \bar{V}_U^{fi}(b_R, p_U). \quad (\text{full information})$$

Even though the urban persistent type always evolves stochastically, its realization is revealed right

¹²That is to say, suppose prices are higher in the urban sector, then the wedge τ_{RU} would partly reflect a price wedge, but the analysis would remain otherwise unchanged, and risk plays the same role as before. See Gollin, Kirchberger, and Lagakos (2021), which document surprisingly small price gaps across sectors.

¹³This reflects an imperfect overlap between skills rewarded in rural vs. urban production. While I don't model occupations explicitly, this assumption is consistent with the large literature of occupational downgrading of migrants from poorer to richer locations, see the discussion in Hendricks and Schoellman (2018).

around the time of migration in the full information case, while it is uncertain in the partial information case, and hence a source of risk. Consequently, the expected value in the full information case, $\bar{V}_U^{fi}$, depends on urban persistent type and assets, while the partial information case depends on rural persistent type and assets.

To be clear, even in the full information case, migration may be risky due to shocks that continue to arrive after the household has moved. The setup thus cleanly separates two distinct margins related to initial uncertainty around the time of migration, and continued risk exposure thereafter.

Before deriving the usual discrete-choice migration probability formulas, I make one adjustment to skill-downgrading that allows me to match the overlap of rural and urban earnings of the least well-off households. I assume that migrants' urban earnings are scaled by a factor $\Lambda \leq 1$, so a rural-urban migrant with urban type p_U earns $w_U \Lambda e^{p_U + \epsilon}$, and total downgrading in log efficiency units is $c_{RU}^+ = c_{RU} + \log \Lambda$. The difference is that c_{RU} only operates on the p -space and thus represents skill-downgrading, which doesn't impact the least-skilled households sitting on the lowest possible realization denoted by $\underline{p}$. In contrast, the factor Λ shifts expected earnings of high and low-skilled households by the same amount.¹⁴

Migration Choice Probabilities. The probability of migrating for the imperfect and full information case takes the convenient logit form (McFadden, 1974)

$$\pi_{RU}^{\mathcal{I}}(b_R, p_U, p_R, \epsilon_R) = \frac{\tau_{RU}^{-s} \left(-\bar{V}_U^{\mathcal{I}}(b_R, p_{j(\mathcal{I})}) \right)^{-\frac{s}{\gamma-1}}}{\left(-V_R(b_R, p_R, \epsilon_R) \right)^{-\frac{s}{\gamma-1}} + \tau_{RU}^{-s} \left(-\bar{V}_U^{\mathcal{I}}(b_R, p_{j(\mathcal{I})}) \right)^{-\frac{s}{\gamma-1}}} \text{ for } \mathcal{I} \in \{pi, fi\}, \quad (7)$$

where the minus term accounts for the fact that the value function is negative for $\gamma > 1$, and $p_{j(\mathcal{I})} = p_U$ in the full information case and p_R otherwise.¹⁵

A useful property of the Frechet distribution is that the expected value of the migration opportunity admits a closed-form expression

$$\mathbb{E} [V_m^{pi} | b_R, p_R, \epsilon_R] = - \left[\tau_{RU}^{-s} \left(-\bar{V}_U^{pi}(b_R, p_R) \right)^{-\frac{s}{\gamma-1}} + \left(-V_R(b_R, p_R, \epsilon_R) \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} := \tilde{V}_m^{pi}(b_R, p_R, \epsilon_R), \quad (8)$$

where the tilde notation $\tilde{V}_m^{pi}$ makes clear that this is an unconditional expectation for the partial information case. In the full information case, I have

$$\mathbb{E} [V_m^{fi} | p_U, b_R, p_R, \epsilon_R] = - \left[\tau_{RU}^{-s} \left(-\bar{V}_U^{fi}(b_R, p_U) \right)^{-\frac{s}{\gamma-1}} + \left(-V_R(b_R, p_R, \epsilon_R) \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} := \bar{V}_m^{fi}(p_U, b_R, p_R, \epsilon_R), \quad (9)$$

where a meaningful distinction can be drawn between conditional, $\bar{V}_m^{fi}$, and unconditional expected value given by

$$\tilde{V}_m^{fi}(b_R, p_R, \epsilon_R) = \int_{p_U} \bar{V}_m^{fi}(p_U, b_R, p_R, \epsilon_R) dF(p_U | p_R). \quad (10)$$

¹⁴Section 3 disciplines Λ to match the worst urban outcome for households at the bottom of the rural income distribution. Appendix B offers additional details on this issue and how it informs the grid-space in the quantitative model.

¹⁵I also impose the normalization on the Frechet scale parameter $\mu = \Gamma \left(1 + \frac{\gamma-1}{s} \right)^{\frac{1}{\gamma-1}}$ to purge the continuation value from the mean effect induced by the randomness of the taste shocks. This normalization ensures that the value function in a model with infinite migration costs coincides with the value function of the same model but without migration ($\lambda_m = 0$). In general, this would not be the case because the taste shock would shift the continuation value of staying in the rural sector relative to a model without migration.

Since the expected value of migration in the full information case represents a convex combination of random outcomes in p_U , one can use Jensen's inequality to show that the full information case dominates the imperfect information case, $\tilde{V}_m^{fi} > \tilde{V}_m^{pi}$. This is intuitive: households get to decide whether they want to migrate or not after they see their urban persistent type, which makes them better off compared to households that have to commit to migration before they know their urban persistent type with certainty.

Closing the Migration Problem. Any stationary model with rural-urban migration requires reverse flows back into the rural sector. I introduce this flow in a way that makes the model particularly tractable. With probability δ_{UR} , newborn urban households are reassigned to the rural sector, where they enter with zero assets and draw their persistent and transitory productivity according to,

$$p \sim N(c_{0,R}, \sigma_{0,R}), \quad \epsilon \sim N(0, \sigma_{\epsilon,R}).$$

Assigning the reverse flow at birth generates a block-recursive structure: the urban household problem can be solved independently of the rural household problem and of equilibrium objects such as the rural-urban wage gap.¹⁶

A second notable feature makes the rural household problem particularly tractable. Standard dynamic Roy models require households to track their evolving productivity in every potential sector. Here, households track productivity only in their current sector, sidestepping the curse of dimensionality. Urban productivity is drawn only when a migration opportunity arrives, with its distribution conditional on current rural productivity. This retains rich selection into migration without adding urban productivity to the rural state space.

To see this, consider the recursive formulation of the rural household problem,

$$(\rho + \phi) V_R(b_R, p_R, \epsilon_R) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_R) \dot{b} + \mathcal{A}V_R + \lambda_m \left(\zeta \tilde{V}_m^{pi} + (1 - \zeta) \tilde{V}_m^{fi} - V_R \right), \quad (11)$$

where $\mathcal{A}$ is the infinitesimal generator of the income process, and $\zeta \tilde{V}_m^{pi} + (1 - \zeta) \tilde{V}_m^{fi}$ is the expected value associated with a migration opportunity.

Crucially, persistent urban productivity does not enter the rural state space. Urban outcomes enter only through the migration values in (8) and (10), which reduce to functions of current rural productivity through the statistical link in (6). The state space therefore does not grow with the number of potential sectors: households track only the permanent and transitory productivity components of their current sector. Destination draws can nevertheless be arbitrarily correlated with current productivity, allowing for rich selection into migration and sectoral comparative advantage.¹⁷

The price of this tractability is that destination draws cannot depend on previous destination draws, since these are not carried as state variables. I view this as a reasonable price to pay when sectoral switches are rare and destination productivity is uncertain, as in the case of permanent rural-urban migration studied here.

Aggregates and Production. I define the share of urban households by M , and collapse the state variables

¹⁶One could instead model endogenous urban-rural migration using the same machinery. I do not pursue this extension because urban-rural migration appears to reflect different forces, such as retirement and family considerations. Appendix A documents systematic differences between urban-rural and rural-urban migrants in income, assets, age, and education.

¹⁷The full-information case can also be interpreted as allowing costless return to the rural sector immediately after the urban draw is observed. A common concern with models of urban income risk is that unsuccessful migrants could simply return, thereby limiting the importance of such risk. If this were the relevant environment, the structural model would infer that the full-information case provides the appropriate description.

into the triplet $x = (b, p, \epsilon)$. After adding up migration probabilities for partial and full information case, the aggregate rural-urban migration rate Π_{RU} follows

$$\Pi_{RU} = \lambda_m \int \pi_{RU}(x) dG_R(x), \quad \pi_{RU}(x) := \zeta \tilde{\pi}_{RU}^{pi}(x) + (1 - \zeta) \tilde{\pi}_{RU}^{fi}(x),$$

where $G_R(x)$ represents the stationary rural distribution. In the full information case, there is a meaningful distinction between the probability of migrating conditional on the urban type draw (which is known), and the unconditional probability, given by $\tilde{\pi}_{RU}^{fi}(b_R, p_R, \epsilon_R) = \int \pi_{RU}^{fi}(b_R, p_U, p_R, \epsilon_R) dF(p_U|p_R)$. In the partial information case, there are only unconditional probabilities as realized urban productivity p_U is unknown before migration so $\tilde{\pi}_{RU}^{pi}(b_R, p_R, \epsilon_R) = \pi_{RU}^{pi}(b_R, p_R, \epsilon_R)$.

The total number of rural-urban migrants thus equals $\Pi_{RU}(1 - M)$. The law of motion for the mass of households in the urban sector reads

$$\dot{M} = \Pi_{RU}(1 - M) - \phi \delta_{UR} M, \quad (12)$$

which depends on endogenous rural-urban migration, as well as the reverse flow $\phi \delta_{UR} M$. These reverse flows combine the death shock ϕ with the exogenous probability of migrating δ_{UR} to the rural sector at birth.

Rural production combines land T , which is in fixed supply and normalized to unity, with total labor $H_R = \int_{i \in R} e^{p_i + \epsilon_i} di$ and reproducible capital K_R using a constant-returns-to-scale Cobb-Douglas production function

$$Y_R = (T)^\chi (K_R)^\nu (H_R)^{1-\chi-\nu}, \quad (13)$$

where χ and ν represent the land and physical capital share in production. Physical capital depreciates at rate δ_k in both sectors.

Urban production also follows a constant-returns-to-scale Cobb-Douglas production function but with zero land ($\chi = 0$) and an additional labor-augmenting productivity term A_U ,

$$Y_U = (K_U)^\theta (A_U H_U)^{1-\theta}. \quad (14)$$

2.2 Equilibrium

I focus on a stationary equilibrium with constant rural-urban share.

Definition 1. A stationary equilibrium consists of aggregate output, capital, factor prices, consumption and migration policy functions within each sector, and an exogenous interest rate r such that

- households maximize utility subject to constraints and transversality condition;
- final goods firms maximize profits, taking factor prices as given;
- labor markets clear within each sector;
- a stationary distribution g over household income and wealth obtains respecting aggregation constraints;
- with constant population shares in each sector.

I next explore some equilibrium properties of the model, which highlight the dependence of rural-urban migration on the rural-urban wage gap, and lead to a convenient solution algorithm detailed in appendix B.

Rural-Urban Wage Gap. The urban-rural wage gap defined as $\omega := \frac{w_U}{w_R}$ is the key equilibrium object that governs the flow of migrants, which becomes apparent after studying the migration problem in normalized form using value functions that are re-scaled by sectoral wages.

Lemma 1. *Conditional on a fixed rural-urban wage gap ω , the value function in both sectors admits a normalized representation $v_j \left(\frac{b}{w_j}, p, \epsilon \right) = \frac{V_j(b, p, \epsilon)}{(w_j)^{1-\gamma}}$.*

The lemma applies to any household whose labor income is proportional to a constant scale factor. A rural-urban migrant earns $w_U \Lambda e^{p+\epsilon}$, so the relevant scale factor is the effective urban wage $w_U \Lambda$ rather than w_U . Using the normalized value function in the migration probability is helpful as it shows that the migration probability is homogeneous of degree zero in wages and assets, which leads to a convenient computational algorithm for the migration block that depends only on the relative wage.¹⁸ Unsurprisingly, rural-urban migration is increasing in the rural-urban wage gap summarized in the following proposition.

Proposition 1. *Using Lemma 1 together with the conditional migration probabilities in (7) delivers*

$$\pi_{RU}^{\mathcal{I}} = \frac{(\omega \Lambda)^s \cdot \tau_{RU}^{-s} (-\bar{v}_U^{\mathcal{I}})^{-\frac{s}{\gamma-1}}}{(-v_R)^{-\frac{s}{\gamma-1}} + (\omega \Lambda)^s \cdot \tau_{RU}^{-s} (-\bar{v}_U^{\mathcal{I}})^{-\frac{s}{\gamma-1}}} \text{ for } \mathcal{I} \in \{pi, fi\},$$

which is strictly increasing in the rural-urban wage gap ω for any $\pi_{RU}^{\mathcal{I}}$. This relationship carries over to the unconditional probabilities $\tilde{\pi}^{\mathcal{I}}$.¹⁹

The steady-state urban population share follows from setting $\dot{M}$ in (12) to zero,

$$M = \frac{\Pi_{RU}}{\Pi_{RU} + \phi \delta_{UR}}. \quad (15)$$

Equation (15) establishes a positive link between the rural-urban wage gap and the share of the urban population as a larger wage gap induces more migration.

In contrast, a negative link between the rural population share and the rural-urban wage gap can be derived from labor demand in each sector

$$w_R = (1 - \chi - \nu) \left(\frac{\nu}{r + \delta_k} \right)^{\frac{\nu}{1-\nu}} \left((1 - M) \tilde{h}_R \right)^{-\frac{\chi}{1-\nu}} \quad (16)$$

$$w_U = (1 - \theta) \left(\frac{\theta}{r + \delta_k} \right)^{\frac{\theta}{1-\theta}} A_U, \quad (17)$$

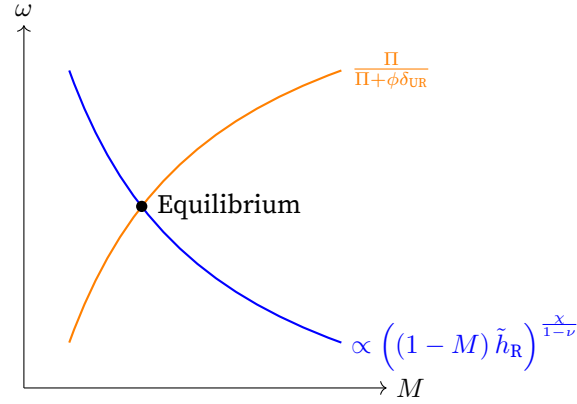
where $\tilde{h}_R = \int h_R(x) dG_R(x)$ is the average human capital level in rural.

Combining (16) and (17), it becomes clear that the rural-urban wage gap implied by labor demand from the production side is falling in the urban population. The reason is that land becomes relatively more abundant when there is more rural-urban migration, driving up rural wages. This gives rise to a unique equilibrium depicted in 1.

¹⁸This intuitive result, that renders the computational algorithm more tractable, would not be true in the case of two-way migration.

¹⁹In a slight abuse of notation, let $\bar{v}_U^{pi} = \bar{v}_U^{fi}$, since there are only unconditional expectations in the partial information case.

Figure 1. Equilibrium Rural-Urban Wage Gap



I defer the computational algorithm to solve the model to appendix B, and turn to the household-level panel data to discipline the model next.

3 Quantification

The quantitative section is split into three subsections. The first subsection discusses the raw data and measurement context, estimates income processes, and computes additional empirical moments relevant for model estimation. The second subsection discusses how the model is identified and uses a minimum distance routine to find the set of structural parameters that best fit the data. The final subsection explores the model fit with an eye toward selection into migration and consumption volatility.

3.1 Empirics

3.1.1 Data Sources and Measurement

South African Panel Data. I use data from the National Income Dynamics Study (NIDS), a household-level panel dataset in South Africa, run by the Southern Africa Labour and Development Research Unit based at the University of Cape Town’s School of Economics. The data appear in biennial fashion, and comprise information on income, consumption expenditure, and demographics on a monthly level from 2008 to 2018. Income and consumption are deflated using the CPI from the South African Statistical Office.

Household Income & Household Consumption. I measure income and consumption on the household level. Focusing on the household as organizing unit is sensible since individual labor income is often not well defined in rural areas, especially in agriculture where the entire family shares the burden of and income from farm work. Consumption expenditure includes food and non-food items but excludes large durable purchases like cars.²⁰

²⁰I drop several additional expenditure classes such as spending on jewelry or wedding and funeral related expenses from the baseline consumption measure, which would lead to spikes in household expenditure beyond the scope of the model. My preferred measure is relatively generous towards semi-durable consumption goods in the sense that educational spending, healthcare spending, clothing, and smaller kitchen items are included. Restricting consumption expenditure to only feature narrowly defined non-durables such as food leads to artificially low consumption inequality. This is unsurprising since food is a necessary good with income elasticity below unity. See Ligon (2025) for recent work on how non-homothetic preferences impact the measurement of risk-sharing arrangements.

I define labor income as net human capital income broadly construed, including business profits, own consumption of agricultural output evaluated at market prices, and government transfers. I exclude asset income and remittances so that measured fluctuations primarily reflect household-specific income risk.²¹

Migration. Table 1 reports rural-urban migration rates where households fall into one of five mutually exclusive categories: rural stayers, urban stayers, rural-urban movers, urban-rural movers, and households that move multiple times, which I call two-way movers. These categories are fixed over time, i.e., a household is always a rural-urban mover if it moves at some point in the future. The table shows that rural-urban migration is much larger than urban-rural one, and two-way migration of household heads is rare, justifying the focus of the paper on permanent migration.²²

Table 1. Rural-Urban migration flows in South Africa

	(1) Share
Rural stayer	0.238
Urban stayer	0.611
Rural-Urban mover	0.086
Urban-Rural mover	0.037
Two-way mover	0.027
Observations	26245

The other key takeaway is that annualized migration rates are small, about 2%, which follow approximately from dividing the share of rural-urban migrants by the length of the time interval. An implication is that the migrant sample is too small for inference on income dynamics, so I focus on stayers within each sector to estimate income processes.

3.1.2 Income Risk

To estimate income processes across sectors, I follow the approach in BPP and study residualized income and consumption dynamics with a focus on stable households attached to the labor market. I restrict the sample to households with heads between the age of 27 and 62 and strong attachment to the labor market. Table 20 in the appendix contains a detailed breakup of sample selection by sector, and table 21 reports summary statistics for both the raw sample and the selected sample.²³ I residualize the data to take out predictable life-cycle trends and other features absent from my model (sex, age, race, marriage, household size, time effects, and years of schooling) estimated for each sector separately. Details are in appendix A.3.

²¹Table 37 in the appendix summarizes the mean share of remittances and government income in rural (3.7% and 26.4%) and urban (2.4% and 10.3%), highlighting that remittances constitute a negligible share of income, while income provided by the government plays a more sizable role in rural South Africa. This latter point has been emphasized as a major achievement of successive post-apartheid governments in South Africa (see World Bank (2021)).

²²I have explored whether permanent rural-urban migrants may be mis-classified in the sense that these migrants are seasonal migrants found repeatedly in the urban sector. This is unlikely to be the case for two reasons. First, the overall migration rates line up nicely with the decline of the rural share, which drops by about 6% in the aggregate data, see figure 13 in the appendix. Stable seasonal migration would not deliver this decline. Furthermore, it is not the case that households classified as permanent rural-urban migrants are more likely to be observed in the urban sector during months associated with the lean season.

²³I drop economically inactive household heads, and households with ultra-low monthly income below 30 US Dollars. I drop households experiencing splits, which could happen for example when a child marries and starts their own family inducing income and consumption dynamics beyond the scope of the model, and households with large swings in the number of adults. Following Daly, Hryshko, and Manovskii (2022), I also drop households that appear irregularly in the sample. Lastly, I drop the lowest 5% of income observations for each sector-wave, which turns out to be only a few observations once I condition on the previous sample selection criteria.

Adding classical measurement error $x_{i,t}$ gives observed residual log labor income

$$y_{i,t}^* = p_{i,t} + \epsilon_{i,t} + x_{i,t}. \quad (18)$$

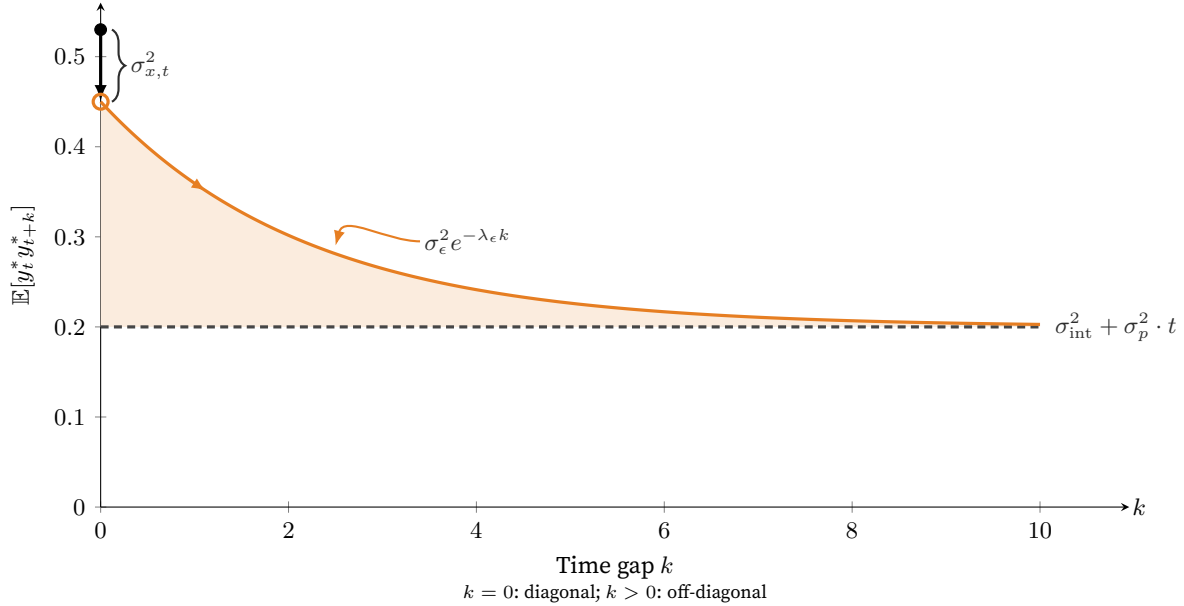
Conditional on household survival, the implied autocovariance is

$$\mathbb{E}[y_{t_1}^* y_{t_2}^*] = \sigma_{\text{int}}^2 + \sigma_p^2 t_1 + e^{-(t_2 - t_1)\lambda_\epsilon} \sigma_\epsilon^2 + \mathbb{1}_{\{t_1 = t_2\}} \sigma_{x,t_1}^2, \quad (19)$$

where $t_1 \leq t_2$ index time relative to an initial period normalized to zero. Classical measurement error enters only on the diagonal of the covariance matrix. The autocovariances are therefore governed by $\theta = \{\sigma_{\text{int}}^2, \sigma_p^2, \lambda_\epsilon, \sigma_\epsilon^2, \{\sigma_{x,t}^2\}\}$; see [Crawley, Holm, and Tretvoll \(2026\)](#).

Figure 2 illustrates identification by plotting the autocovariance against the time gap $k = t_2 - t_1$. Persistent shocks pin down the covariance that remains at long horizons. Passing shocks generate the decline toward this level: σ_ϵ^2 determines how far above it the covariance starts, while λ_ϵ determines how quickly it falls. Finally, classical measurement error affects only $k = 0$, generating a discrete drop between the diagonal and the first off-diagonal covariance.

Figure 2. Identification of Income Process



Dynamic autocovariance, k periods ahead. The dashed line is inequality due to persistent shocks, $\sigma_{\text{int}}^2 + \sigma_p^2 \cdot t$. The shaded area is the passing-shock component, $\sigma_\epsilon^2 e^{-\lambda_\epsilon k}$, which decays with the horizon k . Measurement error, $\sigma_{x,t}^2$, appears only at $k = 0$ and drops out immediately for off-diagonal covariances.

When estimating income risk, I focus on households that do not move across sectors, based on the idea that different economic environments entail different risks. This raises a potential selection concern. In practice, migration is sufficiently rare that selection has a negligible effect on the estimated income process: model simulations with endogenous migration generate autocovariance moments within 1% of those from the GMM procedure that ignores selection. Selection does matter, however, for decomposing

initial inequality, σ_{int}^2 , into inequality at birth and persistent shocks accumulated with age.²⁴ I therefore infer inequality at birth, $\{\sigma_{0,R}, \sigma_{0,U}\}$, using the full equilibrium model in the next section.

Using household level panel data on income, I compute the data counterpart

$$\hat{V} = \hat{\mathbb{E}}_i [\mathbf{y}_i^* \mathbf{y}_i^{*'}]$$

where $\mathbf{y}_i^* = [y_{i,0}^* y_{i,2}^* \dots y_{i,T}^*]'$ is a column vector representing log household income in different periods, and the mean term is omitted since all means are zero after controlling for time-sector fixed effects.

Empirical Autocovariances and GMM. Table 2 reports the empirical autocovariance matrices. Reading down a column increases the time gap between observations, while moving down a sub-diagonal holds the gap fixed and shifts both observations forward in time.

First, rural autocovariances decline sharply with the time gap, while urban autocovariances decline much less. This decay is the signature of passing shocks: its magnitude identifies σ_ϵ^2 and its speed λ_ϵ . Passing shocks are therefore more important in the rural sector.

Second, urban autocovariances rise as a given pair of observations is shifted forward in time, while rural autocovariances remain roughly flat. This fanning out of inequality identifies σ_p^2 and points to a larger role for persistent shocks in the urban sector.

Table 2. Income Covariance Matrix – South Africa

(a) Rural						(b) Urban					
	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}		y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.52					y_t	0.52				
y_{t+2}	0.27	0.55				y_{t+2}	0.27	0.63			
y_{t+4}	0.23	0.27	0.49			y_{t+4}	0.24	0.26	0.58		
y_{t+6}	0.22	0.18	0.25	0.48		y_{t+6}	0.26	0.25	0.32	0.53	
y_{t+8}	0.18	0.16	0.23	0.34	0.57	y_{t+8}	0.23	0.26	0.32	0.34	0.56

Log household income is total earned labor income broadly defined plus net government transfers, i.e., welfare payments minus taxes. This measure includes income from own consumption in subsistence agriculture as well as profits but excludes investment, rental income, and remittances. Income is residualized using a first-stage regression controlling for age, education, sex, race, and household size as well as time effects, separately for urban and rural households. Panel weights are applied.

I estimate the income process by GMM using the autocovariance moments in (19). I use an optimal weighting matrix and exclude diagonal moments, which are contaminated by time-varying measurement error.²⁵

The estimates confirm the patterns visible in the autocovariances. The variance of persistent shocks is only 0.003 in the rural sector, compared with 0.0117 in the urban sector. The latter is roughly four times larger and comparable to estimates for the United States (Crawley, Holm, and Tretvoll, 2026).

Passing shocks show the opposite pattern. In the rural sector, $\sigma_\epsilon^2 = .26$ and $\lambda_\epsilon = .5$, implying large and persistent, but ultimately mean-reverting, income shocks.²⁶ Passing shocks play a much smaller role in the

²⁴Absent migration, initial inequality would satisfy

$$\sigma_{\text{int}}^2 = \sigma_0^2 + \sigma_p^2 \widehat{age}.$$

With migration, this decomposition no longer holds because urban households include migrants who entered the sector at different ages and drew their urban type from a distribution different from that of urban newborns. Restricting the sample to stayers does not eliminate this problem, since households observed as urban stayers may have migrated before the sample begins.

²⁵Income is measured monthly in the South African sample, so time-aggregation error is negligible. The code also accommodates time-aggregated data when income or consumption is measured over longer intervals.

²⁶The implied one-year autocorrelation is $e^{-.5} \approx .6$.

urban sector. Accordingly, the urban data have little power to distinguish a high arrival rate λ_ϵ from a low shock variance σ_ϵ^2 , since either implies little contribution from passing shocks.

Finally, I recover measurement error from the diagonal of the covariance matrix after estimating the income process. Measurement error is sizable, although it cannot be distinguished from fully transitory income shocks (Friedman, 1957). This distinction will matter when I study the passthrough of income shocks to consumption.

Table 3. Income Risk in South Africa

	Rural		Urban	
	Estimate	SE	Estimate	SE
σ_p^2	0.003	0.0065	0.0117	0.0045
λ_ϵ	0.4987	0.2611	0.3333	0.7473
σ_ϵ^2	0.2602	0.0987	0.05	0.0415
σ_{int}^2	0.1795	0.0306	0.18	0.0365

Results are based on a gmm routine where the simulations account for time aggregation in income. Asymptotic standard errors are clustered on the household level. After estimating the key parameters, I infer the measurement error on the diagonal as a residual with $\{\sigma_{x,0}^2, \sigma_{x,2}^2, \sigma_{x,4}^2, \sigma_{x,6}^2, \sigma_{x,8}^2\}^R = \{0.1088, 0.1213, 0.0515, 0.0482, 0.1161\}$, and $\{\sigma_{x,0}^2, \sigma_{x,2}^2, \sigma_{x,4}^2, \sigma_{x,6}^2, \sigma_{x,8}^2\}^U = \{0.1678, 0.2583, 0.17, 0.0902, 0.0962\}$.

The difference in persistent income risk across sectors is an important input into the quantification, so I do not want it to rest on noisy panel estimates alone. I provide two pieces of additional evidence. First, following Deaton and Paxson (1994), I exploit the fact that persistent shocks generate a systematic fanning out of within-cohort inequality over the life cycle. This approach requires only cross-sectional data. Consistent with the South African panel evidence, within-cohort inequality fans out faster in urban areas across several other developing economies, including Mali and Côte d'Ivoire; see appendix A.4. Second, panel data from China yield very similar rural-urban differences.²⁷

An implicit assumption underlying the previous exercises is that income risk is a property of the sector rather than the household. A rural household that migrates therefore takes on the urban income process. The migrant sample is too small to test this assumption directly, but two pieces of indirect evidence support it. First, the urban income process in South Africa closely resembles that of the United States, despite the much larger share of recent rural-urban migrants in South African cities. If migrants carried their rural income process with them, this similarity would be surprising. Second, recent migrants in South Africa do not experience faster income growth than long-term urban residents, suggesting that the two groups face similar mean income trajectories.

Summary. Rural income shocks are large but mean-reverting, while urban shocks are smaller but more persistent. A natural interpretation is that urban earnings depend more heavily on differentiated skills, whose accumulation and depreciation generate persistence over time, while rural incomes depend more on agriculture and are exposed to idiosyncratic shocks that eventually pass.

The analysis so far has focused on income dynamics within sectors. I next turn to rural-urban migration and income gaps, which provide additional moments for the quantification.

²⁷Results available upon request.

3.1.3 Rural-Urban Wage Gaps and Selection into Migration

I now use less strict sample restrictions to keep the sample size large given relatively few migration events.²⁸

Rural-Urban Wage Gaps. The raw rural-urban wage gap is about 70 log points. Controlling for observables reduces it to about 45 log points, the relevant target because the model abstracts from endogenous schooling choices.²⁹ Household fixed effects reduce the gap further to about 30 log points, in line with previous estimates, see Lagakos et al. (2020).³⁰

Table 4. Rural-Urban Wage Gaps

Estimate	(SE)	Description
0.45	(0.04)	Log wage gap (cross-section)
0.34	(0.09)	Log wage gap (fixed effects)

Average wage gaps mask substantial overlap between the two income distributions. Figure 3 plots residual income by sector and shows that many rural households earn more than observationally similar urban households. A positive average urban premium therefore does not imply that migration is a guaranteed income gain.

Figure 3. Income distribution rural vs urban



This figure plots the distribution of residual log income in rural and urban areas. The sample is restricted to households with income within the 3rd–99th percentile of the income distribution and to individuals of working age (23–64). Income is residualized on age and age squared, education, household size, gender, and year fixed effects; Both the rural and urban distributions are evaluated against the same set of coefficients and only the location-specific level is left to differ. Cross-sectional survey weights are applied.

Selection into Migration. I next explore how rural-urban migrants are selected relative to rural stayers. To this end, I run the following regression model amongst rural households

$$y_{i,t} = \alpha + \beta D(i \in RU) + \gamma' \mathbf{x}_{i,t} + u_{i,t}$$

where $D(i \in RU)$ is a dummy variable that takes a value of one if the household migrates permanently to the

²⁸I focus on households aged 23–64 with attachment to the labor market. Moreover, I trim the sample dropping the lowest 3% and highest 1% of log income and log consumption for each sector. Overall, cross-sectional moments are not sensitive to which sample restrictions are applied.

²⁹A natural question is to what extent nominal wage differences reflect real wage differences, i.e., how sensitive are the results if I deflate wages using spatial price indices. Young (2013) sidesteps this issue by using direct measures of consumption, and finds gaps that, if anything, are larger than the income gaps reported here. As discussed in Section 2, any remaining price wedge is absorbed into τ_{RU} and leaves the role of risk unchanged.

³⁰See also Hamory et al. (2020), Alvarez (2020), and Pulido and Swiecki (2018), which find even larger differences between the cross-sectional and the fixed effects estimates. Note that my fixed effects estimate is relatively large because I control for household size and focus only on rural-urban migration. Both are purposeful choices. Household size often falls around migration as some household members are left behind for some time, understating the gains when not controlling for size. In addition, the gains for rural-urban movers tend to be larger than the losses for urban-rural movers, so pooling the two directions attenuates the estimate; one more reason to treat the two kinds of migrants separately.

urban sector at some point in the future and zero otherwise, and $\mathbf{x}_{i,t}$ is a vector of controls. The regression model thus compares stayers with future migrants when both groups are still in the rural sector.

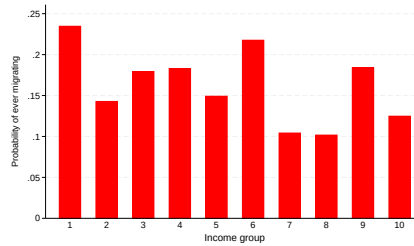
Table 5 shows strong negative selection. Future migrants earn about 30% less than stayers before moving; controlling for observables reduces the gap to about 15%. Consumption yields a similar pattern, suggesting that selection reflects persistent differences in household income. Figure 4 complements the regression results and computes the likelihood of migration along the household income distribution.

Table 5. Selection into Migration

Estimate	(SE)	Description
-0.28	(0.08)	log income gap (raw)
-0.14	(0.06)	log income gap (controls)

Control variables include a second order polynomial in age, years of schooling, household size, and fixed effects for race, gender, year, and marital status. Standard errors are clustered on the household level.

Figure 4. Migration Probability amongst Rural Households



This figure plots the share of rural households that eventually move to the urban sector, by decile of their rural income. The sample is restricted to heads of working age (23–64) with income within the 3rd–99th percentile of the income distribution. Income is residualized on household size, education, age and age squared, gender, and year fixed effects. Cross-sectional weights are applied. The slope of a bivariate regression of migration probability on decile equals -0.0075.

This negative selection contrasts with Young (2013), who emphasize positive sorting as an explanation for rural-urban gaps. A key piece of evidence is that migrants are more educated than rural stayers. I find the same raw schooling gap in South Africa: migrants have roughly two additional years of education. The difference, however, is fully accounted for by a cohort effect. Migrants are younger, and younger cohorts are more educated. Once age is controlled for, the schooling gap disappears, while migrants remain negatively selected on income and consumption; see table 21b.

Rural-Urban Income Correlation.

If households do not know their urban income before moving, the statistical link between rural and urban income becomes an important input to predict urban outcomes. I estimate this link by regressing urban income on lagged rural income,

$$y_{i,U,t} = \alpha + \beta y_{i,R,t-2} + u_{i,t}. \quad (20)$$

Measurement error and transitory shocks attenuate the OLS coefficient, so I instrument $y_{i,t-2}$ with $y_{i,t-4}$. Table 6 reports an IV estimate of about .6, roughly twice the OLS estimate. The structural model does not interpret this coefficient directly as ρ_{RU} because it is still somewhat confounded by selection into migration of informed households, but uses it as an informative moment in the structural estimation.

Income Gains, Losses, and Volatility. I next document the probability that migrants gain or lose income

Table 6. Rural-Urban Wage Gaps

Estimate	(SE)	Description
0.28	(0.052)	OLS
0.58	(0.186)	IV

Table 27 in the appendix reports additional details, and implements a placebo test for stayers. The IV estimate for stayers is substantially higher, at around .8.

and the volatility of income around migration. These moments are rarely reported but, interpreted through the structural model, are informative about migration risk.

Figure 5 compares income and consumption before and after migration. Two facts stand out. First, migration is not a guaranteed gain: a substantial share of migrants experience income and consumption losses. Second, urban outcomes vary widely even among households with similar pre-migration income or consumption.

Figure 5. Migration and Income/Consumption Gain



This figure plots pre- versus post-migration log income for rural-to-urban migrants and rural stayers, restricted to households with income within the 3rd–99th percentile of the income distribution and to heads of working age (23–64). For each household I compute the average log income while in rural and urban separately; households are included only if both a rural- and an urban-period average is observed. Cross-sectional panel weights are applied. Same restrictions apply to the consumption plot with trimming based on consumption.

To explore the probability of income gains and losses more systematically that controls for time effects and family composition, I estimate a linear probability model for the likelihood that household income per adult exceeds its previous level,

$$D_{i,t}(y_{i,t} > y_{i,t-2}) = \alpha + \beta D_{i,t}(U) + \gamma' \mathbf{x}_{i,t} + u_{i,t},$$

where $D_{i,t}(U)$ indicates the first period in which the household is observed in the urban sector, and $\mathbf{x}_{i,t}$ is a set of control variables. The first row of table 7 shows that migration raises the probability of an income gain by about 20 percentage points. Since 55% of rural stayers experience income growth over the same horizon, roughly one-quarter of migrants still experience an income loss.

I also estimate the effect of migration on mean income growth and its residual variance using

$$\begin{aligned} \Delta y_{i,t} &= \alpha + \beta D_{i,t}(U) + \gamma' \mathbf{x}_{i,t} + u_{i,t}, \\ \left(\widetilde{\Delta y_{i,t}} \right)^2 &= \alpha + \beta D_{i,t}(U) + \gamma' \mathbf{x}_{i,t} + u_{i,t}. \end{aligned}$$

The first regression measures annualized income growth around migration; the second measures the corresponding increase in dispersion after removing group-specific mean growth.

Table 7. Income Gains and Income Volatility

Baseline	(SE)	Migration Dummy	(SE)	Description
0.56	(0.017)	0.21	(0.042)	Probability of income improvement
0.03	(0.009)	0.32	(0.044)	Income growth
0.12	(0.007)	0.048	(0.025)	Income volatility

Table 28 and 30 in the appendix reports additional results and details, including consumption and specifications without controls. In terms of growth and volatility, consumption the migration dummy is slightly lower. I will return to the consumption moments later on, and use them to see how well the model performs for non-targeted moments.

Income volatility rises around migration. Yet volatility is already high among rural stayers ($SD = \sqrt{0.12} \approx .35$), where it is driven largely by passing shocks that can be smoothed through buffer-stock savings.³¹ Volatility alone is therefore not a measure of risk exposure; what is key is the persistence of the underlying shocks. Consumption provides some evidence on this margin. Consumption volatility rises by a similar amount around migration, as one would expect if the additional income shocks are persistent. I return to these consumption moments in section 3.3, where the structural model decomposes the increase in volatility into transitory and persistent components.

3.2 Model Calibration and Estimation

I discipline the model using a combination of calibration and indirect inference. A small set of parameters is fixed using external evidence or accounting normalizations; the remaining parameters are jointly estimated by simulated method of moments (SMM). I describe each in turn and highlight the moments most informative about the internally estimated parameters.

3.2.1 Calibrated and Externally Estimated Parameters

Preferences and Production. The interest rate of the risk-free bond is set to $r = 2.7\%$, the discount factor equals $\rho = .02$, and the coefficient of relative risk aversion is set to $\gamma = 2.0$, consistent with Choukhmane and Silva (2026).

Stationarity requires rural outflows to equal inflows. Given ϕ , this pins down $\delta_{UR} = \frac{\Pi_{RU}}{\phi} \left(\frac{1-M}{M} \right)$. An annual migration rate of roughly 2% and an urban share of .65 implies $\delta_{UR} = .35$.

Urban production has capital and labor shares of 40 and 60 percent and capital depreciation of 5%. Rural production has land, capital, and labor shares of 30, 10, and 60 percent. The land share χ is particularly important as it governs the rural wage response to migration. Holding the rural-urban wage gap ω and average rural skill $\tilde{h}_R$ fixed, a link between urban productivity growth g_{Au} and rural-urban structural change emerges,

$$\frac{d \log(1 - M)}{dt} = -\frac{1 - \nu}{\chi} g_{Au}, \quad (21)$$

which is reminiscent of Lucas (2004). Equation (21) shows that a smaller land share requires faster rural out-migration for rural wages to keep pace with urban productivity growth. I set χ to approximately one-

³¹The volatility of stayers closely matches the volatility implied by the stochastic process estimated earlier. To see this, use the stochastic process to derive volatility of the growth rate, and plug in numbers close to the earlier estimates: $\mathbb{V}(\Delta y/2) = \sigma_p^2/2 + (1 - e^{-2\lambda\epsilon}) \sigma_\epsilon^2/2 + \sigma_x^2/2 \approx 0 + (1 - e^{-1}) * 0.26/2 + .1/2 \approx .13$.

third following Hansen and Prescott (2002).³² The resulting pace of structural change is consistent with the historical experience of advanced economies and with the cross-country relationship between productivity growth and rural population decline (Trouvain, 2025).

Income Process. I set the income-process parameters to the estimates from the previous section. The death rate is $\phi = .03$, corresponding to an average working life of 33 years (Gabaix et al., 2016).³³ I normalize average human capital of urban newborns to $c_{0,U} = -.4$. This level has little bearing on the results because the relevant object is the rural-urban gap in initial human capital, $c_{0,U} - c_{0,R}$, with $c_{0,R}$ estimated internally.

Migration Arrival Rate and Fixed Cost of Migration. In general, the migration wedge τ_{RU} and the arrival rate λ_m are not separately identified as a higher disamenity or fewer migration opportunities can rationalize the same steady-state migration flow.³⁴ I calibrate these two parameters as follows. First, I choose an arrival rate of migration of $\lambda_m = .2$. This means that households obtain opportunities to permanently migrate every five years, on average. Given λ_m , I can then infer τ_{RU} to match moments in the data.

Regarding the fixed cost of migration, I choose a value consistent with experimental evidence on poverty traps in Banerjee, Duflo, and Sharma (2021).³⁵ The authors show that a large one-time transfer to ultra-poor households in rural India, amounting to roughly 80% of annual household income, generates substantial and persistent labor income gains driven by rural-urban migration to higher-wage destinations. A fixed cost consistent with such poverty traps must be somewhat close to the size of the transfer, which leads to a value of $f_{RU} = .046$ in my quantification. Later on, I will use the model to study how important micro-level poverty traps are for aggregate rural-urban structural change.

Table 8 summarizes the calibrated parameters.

3.2.2 Internally Estimated Parameters

I estimate the remaining parameters jointly by SMM,

$$\Theta = \{c_{0,R}, c_{RU}, \Lambda, \rho_{RU}, \sigma_{RU}, \zeta, s, \tau_{RU}, \sigma_{R,0}, \sigma_{U,0}\}.$$

Although all parameters are jointly disciplined by the full set of moments, particular moments are especially informative about individual parameters. I highlight these mappings below.

The cross-sectional rural-urban wage gap is informative about average human capital among rural newborns, $c_{0,R}$. The fixed-effects wage gap instead helps pin down the migration wedge τ_{RU} : a larger wedge requires larger realized wage gains to induce migration. Negative selection into migration is particularly informative about skill downgrading, c_{RU} . Because the urban productivity draw is bounded below, skill downgrading primarily affects higher-productivity rural households and therefore shapes selection into migration.

In contrast, the parameter Λ governs downside risk at the bottom of the rural income distribution. Without this adjustment, the rural-urban wage gap $\omega > 1$ implies that even the worst urban draw leaves

³²This is larger than the land share of approximately ten percent implied by the village-level estimates of Akram, Chowdhury, and Mobarak (2017). These estimates exploit village-level variation and need not identify the aggregate general-equilibrium elasticity relevant for a two-sector model. Extrapolating this estimate to the aggregate economy would imply a much faster decline in the rural population than is observed over the long run.

³³This stabilizes the income distribution and implies a tail-coefficient of log permanent income of $\sqrt{2\phi/\sigma_p^2}$. While not targeted, the model-based tail coefficient is consistent with its empirical counterpart, in particular a thicker income tail in the urban sector, see table 36 in the appendix.

³⁴The fact that the arrival rate of migration opportunities and migration disutility wedges are not separately identified is more generally true in discrete choice models of migration, see the appendix of Caliendo, Dvorkin, and Parro (2019).

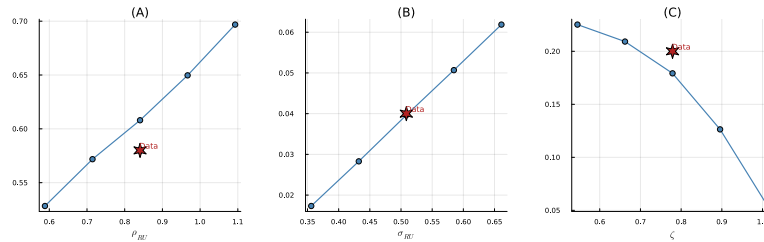
³⁵See also Bryan, Chowdhury, and Mobarak (2014) and Balboni et al. (2022).

Table 8. Calibrated or Externally Estimated Parameters

Parameters	Value	Reference
Preferences		
ρ	0.0200	standard value
ϕ	0.0300	avg. length of work life
γ	2.0000	Choukhmane and Silva (2026)
Technology		
r	2.7%	see text
δ_{UR}	0.3500	see text
χ	0.3000	Hansen and Prescott (2002)
ν	0.1000	Hansen and Prescott (2002)
θ	0.4000	Gollin (2002)
λ_m	0.2000	see text
f_{RU}	0.0460	see text
Income Process		
$\sigma_{p,R}^2$	0.0030	own estimation
$\lambda_{\epsilon,R}$	0.4987	own estimation
$\sigma_{\epsilon,R}^2$	0.2602	own estimation
$\sigma_{p,U}^2$	0.0100	own estimation
$\lambda_{\epsilon,U}$	0.3333	own estimation
$\sigma_{\epsilon,U}^2$	0.0500	own estimation
$c_{0,U}$	-0.40	urban income distribution

the poorest rural household better off, i.e., migration would be a guaranteed gain. I choose Λ such that the lowest-income rural household loses 5% of income under the worst urban outcome.

Relative to the previous literature, there are three novel parameters: the average link between rural and urban productivity, ρ_{RU} ; the dispersion of urban outcomes around this conditional mean, σ_{RU} ; and the share of households that do not know their urban persistent draw ex ante, ζ . All parameters are jointly estimated by SMM, but three moments are especially informative about these parameters. First, the reduced-form link between rural and urban income, estimated by IV at about .6, helps pin down ρ_{RU} . Second, the excess income volatility of movers relative to stayers, about .04, is particularly informative about σ_{RU} . Finally, the share of migrants whose income improves after moving helps identify ζ . Intuitively, if more households knew their urban type before moving, fewer would choose to migrate only to experience an income loss afterward. Figure 6 illustrates these mappings by varying each parameter while holding all others at their estimated values.

Figure 6. Identification of ρ_{RU} , σ_{RU} and ζ 

The plots hold all other structural parameters used to solve the model fixed, and vary the parameter in question by up to 20%.

The remaining parameters govern preference heterogeneity s and initial inequality $\{\sigma_{0,R}, \sigma_{0,U}\}$. Pref-

erence heterogeneity is disciplined by the slope of migration rates across rural income deciles. As taste shocks become more important, migration becomes less responsive to income and this slope approaches zero. Initial inequality is chosen to reproduce observed inequality in each sector. Because these moments are affected by endogenous migration, they are estimated jointly within the full equilibrium model.

Finally, I minimize the loss function in terms of absolute percent deviations

$$\hat{\theta} = \min_{\theta \in \Theta} \sum_k \frac{|\text{model}_k - \text{data}_k|}{\frac{1}{2}(|\text{model}_k| + |\text{data}_k|)} \quad (22)$$

across all $k \in K$ moments. Appendix B.4 contains details on the exact simulation procedure used to compute simulated moments and minimize the loss function.

3.3 Model Fit

Table 9 compares the targeted empirical moments with their simulated counterparts. The model fits the data closely, including the moments most informative about migration risk: the rural-urban income correlation, excess volatility around migration, and the probability of an income gain after moving. Table 10 shows the estimated parameter values that minimize the loss function.

Table 9. Moments Targeted in the Estimation

Moments	data SA	model SA
Log rural-urban wage gap (cross-section)	0.45	0.45
Log rural-urban wage gap (fixed effects)	0.34	0.35
Autocorrelation urban-rural income (IV)	0.58	0.61
Volatility of log income around migration rel. to no migration	0.04	0.04
Log(migrant income in rural/stayer income in rural)	-0.14	-0.16
Slope migration rate on income decile	-0.0075	-0.0075
Probability of income around migration rel. to no migration	0.21	0.18
Initial inequality rural	0.18	0.17
Initial inequality urban	0.18	0.21

The routine is solved such that migration rates exactly agree with the urban population share of 0.65. Initial inequality is the degree of inequality in the simulated data in the first wave for each sector net of measurement error and transitory shock. I further enforce that the worst outcome in urban leaves the lowest income rural-urban migrant 5% worse off, which pins down Λ .

Table 10. Internally Estimated Parameters

Parameter	$c_{0,R}$	c_{RU}	Λ	ρ_{RU}	σ_{RU}	ζ	s	τ_{RU}	$\sigma_{R,0}$	$\sigma_{U,0}$
Estimate	-0.92	-0.14	0.90	0.84	0.51	0.78	8.62	1.08	0.14	0.0100

The urban TFP shifter $A_U = 0.64$ is uniquely pinned down as a residual to ensure urban labor demand is consistent with urban labor supply. I thus don't have to estimate this parameter directly and simply infer it by combining the rural wage and the endogenous rural-urban wage gap. This convenient simplification works because both rural and urban value function are homogeneous of degree $1 - \gamma$ in the sectoral wage.

Three estimates are central to the riskiness of migration: ρ_{RU} , σ_{RU} , and ζ . The estimated ρ_{RU} is substantially larger than the reduced-form IV coefficient. The difference reflects three forces that prevent the IV coefficient from mapping directly into the structural parameter. Passing shocks attenuate the empirical relationship between rural and urban income. Skill downgrading affects higher-income migrants more strongly because urban productivity is bounded below. Finally, informed households select into migration based on their urban draw, further changing the observed rural-urban income relationship

The estimated dispersion of urban persistent types is large, at $\sigma_{RU} \approx .5$. This may appear difficult to reconcile with the modest increase in observed income volatility around migration, only about .04. The distinction between persistent and passing shocks resolves the apparent tension. Rural income is already highly volatile, but much of this volatility is mean-reverting. Migration adds relatively little total volatility, but substantially more persistent volatility. Indeed, when I apply the same volatility calculation to persistent income alone, the model implies an increase of 0.06. This magnitude is consistent with $\sigma_{RU} \approx .5$, since $\mathbb{V}(\Delta p/2) = .5^2/4 = 0.0625$.³⁶

Regarding information about initial urban persistent type, I estimate $\zeta = 0.78$, so roughly four in five potential migrants move without knowing their long-run urban productivity. This is a large degree of uncertainty but consistent with the large number of households that end up not improving their income.

The remaining estimates are also consistent with the targeted moments. The model requires relatively little taste heterogeneity to reproduce the strong negative selection into migration, with $s \approx 8.6$. Initial rural inequality is 0.14, close to its empirical counterpart. Initial inequality among urban newborns is much lower, at 0.01. This reflects endogenous migration: risky rural-urban migration itself contributes substantially to observed urban inequality, leaving less inequality to be accounted for at birth.³⁷

3.4 Selection into Migration

Income. While the model matches negative selection on income by design, it is instructive to decompose this selection into persistent and transitory type. Figure 7 plots the migration choice probabilities π_{RU} as a function of these types, i.e.,

$$\pi_{RU}(p) := \frac{\int_{\epsilon} \int_b \pi_{RU}(b, p, \epsilon) dG_R(b, p, \epsilon)}{\int_{\epsilon} \int_b dG_R(b, p, \epsilon)},$$

which is the average migration probability among rural households of persistent type p , and analogously for ϵ .

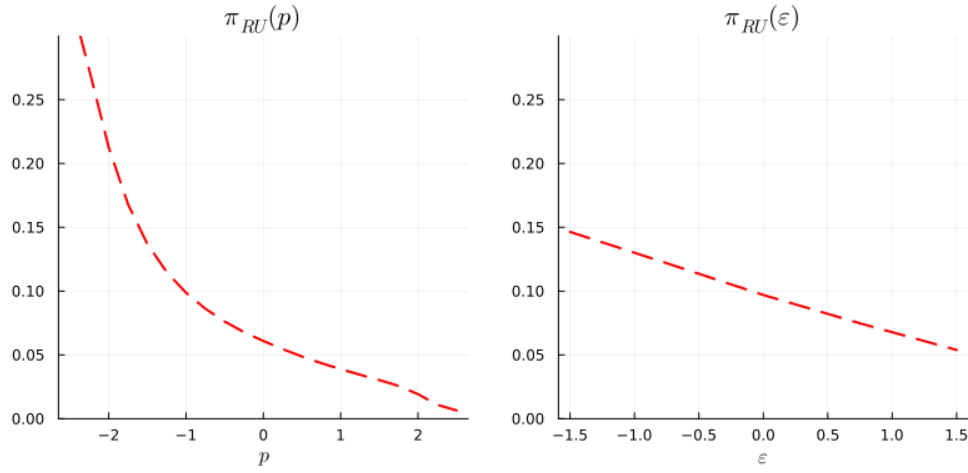
The left panel illustrates a steep negative migration gradient in the persistent component p . This negative selection in p follows from both the mean reversion in the urban draw, and risk in the form of imperfect information. A low-income rural household expects to draw a higher urban type than it holds today. The opposite is true for high-income rural households, which generates the negative selection into migration in terms of income.

The right panel shows that migration also falls with transitory income, but much less steeply. Since transitory shocks are uncorrelated across sectors, a low rural draw raises the incentive to move. But these shocks mean-revert quickly, so their effect on long-run consumption is small. Persistent income therefore plays a much larger role in migration decisions.

³⁶Allowing rural and urban transitory types to be correlated only strengthens this result. In the limiting case of perfect correlation, transitory types contribute nothing to income growth around migration, so an even larger share of observed volatility must come from persistent income differences.

³⁷Appendix B.3 offers more detail on this point. There I plot the simulated autocovariance matrices, which line up nicely with the empirical ones, and illustrate the confounding impact of migration on initial inequality across sectors.

Figure 7. Selection into Migration – Income



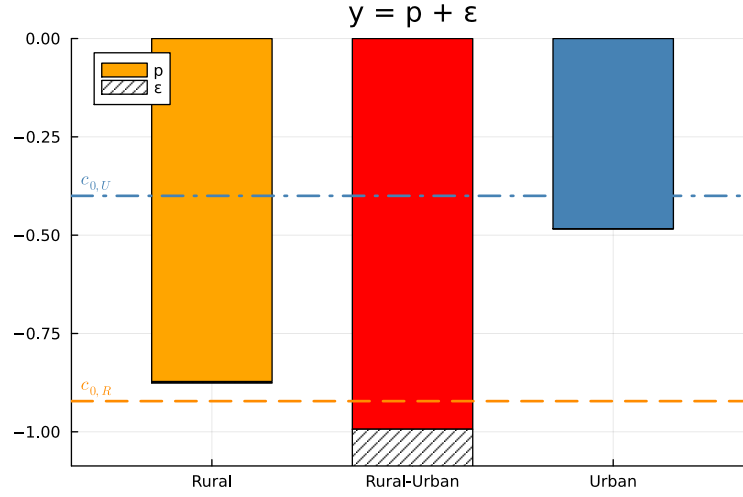
Average migration probability by persistent type p (left) and transitory type ε (right), integrating over the remaining rural state variables as defined in the text.

Accounting for negative selection into migration is important for measuring human-capital differences across sectors. Figure 8 plots average persistent and transitory types for rural households, rural-urban migrants while still rural, and urban households. The observed rural-urban human-capital gap understates the gap between newborn rural and urban households, $c_{0,U} - c_{0,R}$. Negative selection compresses the observed gap from both sides: low-income households leave the rural sector, raising average rural human capital, and enter the urban sector, lowering average urban human capital. Adding the sectoral wage gap to the human-capital gap at birth gives the expected income difference between a newborn urban and rural household,

$$c_{0,U} - c_{0,R} + \log \omega = 0.58.$$

This gap is well above both the measured cross-sectional wage gap of 0.45 and the fixed-effects gap of 0.35. Thus, endogenous migration masks part of the disadvantage of being born in the rural sector.

Figure 8. Unobserved log human capital

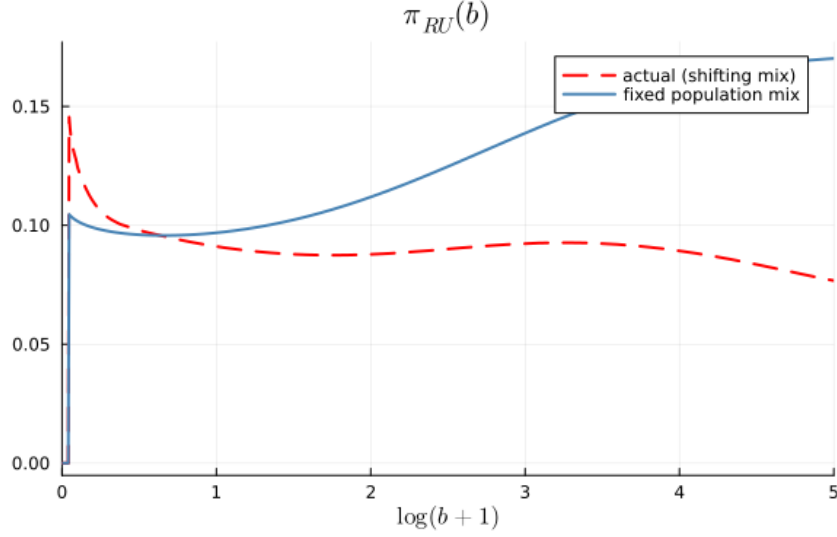


To compute the average amongst migrants, note that the distribution of rural-urban migrants follows from Bayes' law $dG_{RU}(x) = dG_R(x) \pi_{RU}(x) / \int \pi_{RU}(x) dG_R(x)$. Averages of rural and urban are computed against the stationary rural and urban distribution. The gap in human capital is smaller than the income gap due to a positive wage premium, $\log(\omega) > 0$.

Wealth. Turning to the link between wealth and migration, the red dashed line in the left panel of figure 9 plots the migration probability by household assets. Migration jumps once households can cover the fixed cost, but is non-monotone in wealth thereafter. This pattern could reflect changes in the composition of income types across the wealth distribution. The solid blue line therefore holds the distribution of income types fixed while varying assets, yet the same non-monotone patterns remain.

The non-monotonic link between wealth and migration reflects three countervailing channels. First, wealth allows households to pay the fixed cost of migration. This generates the jump in migration once household assets are sufficient to cover the fixed cost. See the left panel in figure 10 that plots the migration rate of the lowest income rural household as a function of wealth.

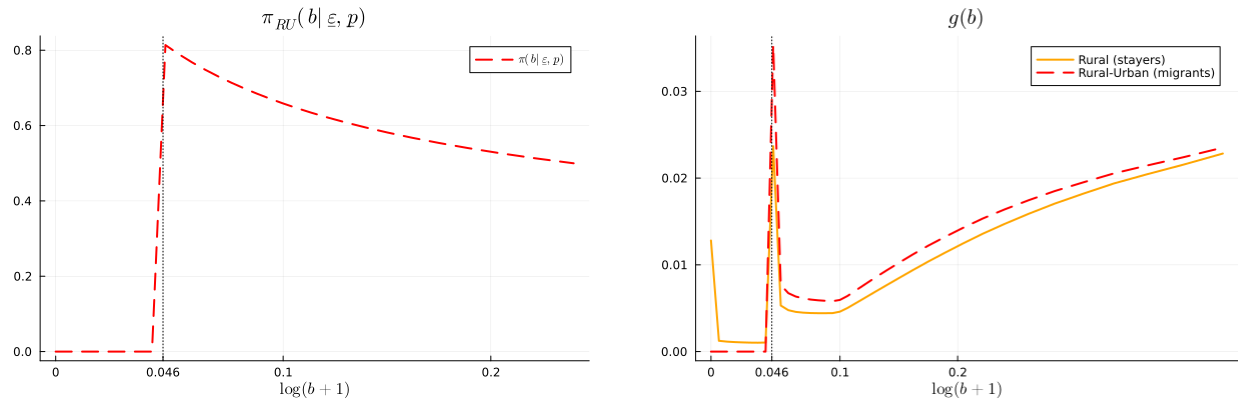
Figure 9. Selection into Migration – Wealth



I compute unconditional migration probabilities using $\pi_{RU}(b) = \int_p \int_\epsilon \pi_{RU}(x) dG_R(p, \epsilon)$, which gives the unconditional probability that a household with assets b seizes the migration opportunity. One issue that makes interpretation hard is that the share of high and low income types shifts with assets. The blue line deals with this issue by holding the population type fixed using the marginal distribution over p and ϵ , and then studies how migration probabilities change as assets increase. The right panel computes the distribution of rural-urban migrants using Bayes' rule, i.e., $g_{RU}(x) = \pi_{RU}(x) g_R(x) / \int \pi_{RU}(x) dG_R(x)$.

Second, migration is risky, and wealth reduces the impact of income risk on consumption. This is not the standard buffer-stock channel. Wealth cannot smooth away a persistent shock, but asset income is unaffected by labor income shocks. Persistent labor-income risk therefore matters less for consumption as the labor income share, $\theta_L = wh/(wh + \tau b)$, falls.

Figure 10. Migration and Wealth



Third, the same force works against migration. Migration raises labor income, not asset income. As wealth increases and the labor income share falls, persistent labor-income risk matters less for consumption, but so do the labor-income gains from migration. Greater wealth therefore makes migration safer while simultaneously making it less valuable. This is what drives the decline in migration on left panel of

figure 10.³⁸

A crucial point follows. The model replicates negative selection on wealth, driven by low-income households for whom labor-income gains from migration are particularly important. For these households, greater wealth reduces the labor income share and, with it, the gains from migration. Thus, migrants can be negatively selected on wealth even when risk is a major barrier to migration.³⁹

3.5 Consumption Insurance

I next analyze consumption risk. Income risk matters for migration only insofar as it translates into consumption risk. I therefore measure the passthrough of income shocks to consumption, a common measure of consumption insurance, using two estimators. The first, following Townsend (1994), measures the response of consumption to idiosyncratic income variation. The second, following BPP, isolates the response to persistent income shocks:

$$\beta_{c-y}^T = \frac{Cov(c_{i,t}, y_{i,t} | \alpha_i, \gamma_t, x_{it})}{\mathbb{V}(y_{i,t} | \alpha_i, \gamma_t, x_{it})}, \quad (\text{Townsend})$$

$$\beta_{c-p}^{BPP} = \frac{Cov(\Delta c_{i,t}, \Delta y_{i,t-2} + \Delta y_{i,t} + \Delta y_{i,t+2})}{Cov(\Delta y_{i,t}, \Delta y_{i,t-2} + \Delta y_{i,t} + \Delta y_{i,t+2})}, \quad (\text{BPP})$$

where c and y are log consumption and log income. The Townsend approach uses a simple linear regression, controlling for individual and time effects and other time-varying observables $(\alpha_i, \gamma_t, x_{it})$. The BPP approach instead derives a moment restriction from a forward-looking consumption-saving problem and a standard persistent-transitory income process, adjusted to biennial data.

Table 11 reports the empirical passthrough estimates. The Townsend estimator implies low consumption passthrough in both sectors, at 0.15 in rural areas and 0.20 in urban areas. The BPP estimator reveals a sharper sectoral difference: passthrough of persistent shocks is only 0.22 in the rural sector, compared with 0.62 in the urban sector. Rural households therefore appear remarkably well insured against income risk. This finding is consistent with a large literature on developing economies,⁴⁰ while the urban BPP estimate is close to that found for U.S. households by BPP.

I next apply the same estimators to model-simulated data, reported in table 12. Despite containing no informal insurance, the model reproduces the low rural passthrough and the large rural-urban gap in the BPP coefficient. It also matches the urban BPP estimate almost exactly. This raises a natural question: why

³⁸This labor-income-share channel depends on two features: preferences are CRRA, and the migration wedge takes the form of urban consumption disutility. Under CRRA, migration decisions depend on proportional changes in lifetime consumption. As wealth increases, a given wage gain represents a progressively smaller proportional improvement in consumption, reducing the incentive to migrate. Contrast this with constant absolute risk aversion (CARA) preferences, as in Dvorkin and Greaney (2025). In that case, migration decisions depend on the level of the consumption gain, and are thus largely invariant to a household's wealth. The migration wedge discounts consumption irrespective of whether it is financed by labor or asset income, yet the gains from migration for high wealth households are much smaller, so wealthy households with low labor income are better off staying rural. To see the countervailing effects of wealth on migration, I derive analytical expressions in appendix C under the simplifying assumption that households consume all labor and asset income, $c = wh + rb$, which make the ambiguity explicit.

³⁹The data suggest a negative link as well, the empirical density of wealth across rural-urban migrants and rural-urban stayers in appendix A exhibits negative selection on wealth. The wealth module of the South African data is less detailed, which is why I don't lean heavily on it. More generally, the model predicts that the effect of wealth for low-income households is non-monotone: positive for households just below the fixed cost, negative for those above it. Estimates that average across the wealth distribution will therefore vary with the sample. See Bazzi (2017) or Gazeaud, Mvukiyehe, and Sterck (2023), for examples of a large literature documenting disparate effects of wealth on international migration, consistent with the many different channels through which wealth impacts migration: covering fixed costs, reducing risk, and reducing the relative expected gains from migration.

⁴⁰Both Townsend-style and BPP regressions consistently produce very low passthrough coefficients in rural environments in developing economies. See De Magalhaes and Santaaulalia-Llopis (2018), Santaaulalia-Llopis and Zheng (2018), Morten (2019), Attanasio et al. (2025), Meghir et al. (2022), and Ligon (2025) for recent work.

Table 11. Empirical passthrough estimates

	South Africa	
	Rural	Urban
$\hat{\beta}_{c-y}^T$	0.15	0.20
$\hat{\beta}_{c-p}^{BPP}$	0.22	0.62

The Townsend-passthrough regression is run on the entire sample, and includes standard controls such as age, education, household size, and individual as well as sector-year-month fixed effects. The BPP passthrough is based on the selected sample and same residualization procedure as the income process. The standard errors of the rural and urban estimates are 0.26 and 0.32 respectively.

does a model with only self-insurance make rural households appear so well insured?

Table 12. Simulations

	South Africa	
	Rural	Urban
β_{c-y}^T	0.19	0.13
β_{c-p}^{BPP}	0.31	0.61
β_{c-p}	0.80	0.89

The first two rows are the simulated counterparts of the estimators reported in table 11. The third row, β_{c-p} , is the true passthrough of persistent shocks to consumption, which is only observable in the model.

The answer lies in the nature of rural income risk. First, measurement error and fully transitory shocks lower the Townsend coefficient because they generate income variation with little corresponding movement in consumption. This familiar force is quantitatively important because measurement error is large in household surveys, particularly in developing economies.

Second, imperfectly transitory shocks bias the BPP coefficient downward. BPP is based on a persistent-transitory income process and therefore attributes some imperfectly transitory variation to persistent shocks. This matters especially in the rural sector, where passing shocks account for much of income volatility. The estimator consequently overstates the variance of persistent income risk and understates its passthrough to consumption.

This sectoral difference matters for how we interpret the apparent deterioration in consumption insurance after rural-urban migration, emphasized by Munshi and Rosenzweig (2016). In my framework, measured insurance deteriorates even though households lose no access to informal insurance.

To see how imperfectly transitory shocks affect the BPP estimator, consider the following approximation of log consumption:

$$dc = \beta_p dp + \beta_\epsilon d\epsilon, \quad (23)$$

with $\beta_p > \beta_\epsilon$ as the passthrough from permanent shocks is larger than from transitory shocks. In this case, the BPP estimator using (23) yields an estimate between β_p and β_ϵ given by

$$\beta_{c-p}^{BPP} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})} \beta_p + \frac{\sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})} \beta_\epsilon. \quad (24)$$

Since $\beta_\epsilon < \beta_p$, the BPP coefficient lies below the true passthrough of persistent shocks, β_p . The bias grows with the importance of imperfectly transitory shocks.⁴¹ Equivalently, the BPP estimator attributes

⁴¹The bias here is distinct from the ones explored in Kaplan and Violante (2010), and in particular from replacing the unit root in the income process with a highly persistent AR(1) process. Here, both a permanent and an imperfectly transitory component are present at the same time.

$\sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})$ to persistent income risk. This bias is quantitatively important in the rural sector, where σ_ϵ^2 is large and λ_ϵ is small.

The key point is that this misspecification is much more severe in the rural sector. The true passthrough of persistent shocks is high in both sectors, at 0.80 in rural areas and 0.89 in urban areas, despite the large gap in measured BPP coefficients. Given the amount of wealth households choose to hold, buffer-stock savings provide little protection against persistent shocks.⁴² The rural-urban gap in measured insurance therefore reflects differences in the nature of income risk rather than differences in households' ability to insure.

4 Risk and Rural-Urban Structural Change

This section quantifies by how much risk retards rural-urban structural change. Section 4.1 studies the impact of risk on aggregate rural-urban structural change and welfare. Section 4.2 investigates the implication for selection into migration. Section 4.3 benchmarks risk against liquidity. Finally, section 4.4 shows what a model without risk must wrongly attribute to the migration disamenity τ_{RU} .

4.1 Aggregate Impact of Risk

Welfare and Insurance Definition. Before studying the implications of risk for rural-urban structural change and welfare, I first need to define my notion of welfare and insurance.

To measure welfare, I follow the consumption-equivalent variation of Lucas (1987): the uniform proportional change in baseline consumption that makes a household indifferent between the baseline (bl) and counterfactual (cf) economies. Welfare effects differ across households, so I have to take a stand on how to aggregate them. This requires computing rural welfare as the expectation of the value function against some distribution, $\tilde{V}^n := \int V(x) dG^n(x)$. The consumption-equivalent welfare change is then

$$\lambda_W^n := \left(\frac{\tilde{V}_{cf}^n}{\tilde{V}_{bl}^n} \right)^{\frac{1}{1-\gamma}} - 1.$$

The choice of $G^n(x)$ is key. In the main text, I lead with the simplest one: the distribution of newborn rural types, evaluated at zero assets. Because this distribution is unaffected by any counterfactual, the statistic is free of confounding compositional effects and answers a simple question: how much better off a wealth-poor rural household would be if it were born into the counterfactual economy.

I define insurance as follows. Households arriving in the urban sector with the same types (p, ϵ) form a cohort and insure each other against income shocks by pooling their income. At every date, each household within the cohort receives the average income,

$$\bar{y}_t(p, \epsilon) = \mathbb{E}[(w_U \Lambda) e^{p_t + \epsilon_t} | p, \epsilon], \quad (25)$$

which remains a function of initial types since income processes are persistent. This setup isolates the role of urban income risk, while leaving the borrowing constraint and mortality risk in place.⁴³

⁴²These estimates are in line with standard buffer-stock models; see Blundell, Low, and Preston (2013) and Kaplan and Violante (2010). In the urban sector, the coefficient is additionally pushed up by migration, as many migrants arrive with little wealth.

⁴³In the appendix, I show how to generalize the setup with constructions that insure mortality risk or extend the borrowing constraint. The reason this is not a suitable choice of the main counterfactual is that it introduces asymmetries across sectors beyond the

I distinguish between insurance provided before and after the initial urban draw is realized. Insurance provided after migration pools subsequent income shocks but leaves the initial urban draw uninsured. I label this economy *ni*, for no income risk.

Insurance provided before migration also pools uncertainty about the initial urban draw. Households with the same rural type p_R contract before their urban types are revealed and receive the deterministic income path

$$\bar{y}_t^m(p_R) = \mathbb{E}[\bar{y}_t \mid p_R],$$

regardless of the urban type they subsequently draw. This removes both initial migration risk and subsequent income risk. I label this economy *nm-ni*, for no migration risk and no income risk.⁴⁴

Two features of this construction are deliberate. First, insurance is asymmetric: it covers urban income while leaving rural income risky. This isolates the object of interest—the risk associated with migration. Second, the exercise is an accounting device, not a policy proposal. It asks how much risk deters migration and how much of that risk is absorbed by the residual migration wedge when risk is omitted from the model. It provides a building block for a richer optimal policy exercise.⁴⁵

Results. I consider four counterfactuals. The first removes uncertainty about the initial urban draw by setting the share of imperfectly informed households to zero, $\zeta = 0$. The second insures income shocks after arrival, corresponding to the *ni* economy defined above. The third combines this insurance with full information about the initial draw, $\zeta = 0$. The fourth, the *nm-ni* economy, insures the initial urban draw on top of subsequent income shocks for the uninformed share $\zeta = 0.78$ of households.

Table 13 reports the impact of each scenario on the aggregate urban population share, distinguishing partial from general equilibrium results, the latter letting the rural wage adjust to its market-clearing level.

Table 13. Risk and the Urban Population Share

	Urban population share		$\Delta \log w_R$
	GE	PE	
Baseline	0.65	0.65	–
Information ($\zeta = 0$)	0.76	0.80	0.1373
Insurance, income risk (<i>ni</i>)	0.73	0.77	0.0845
Information and insurance ($\zeta = 0$, <i>ni</i>)	0.81	0.85	0.2088
Insurance, both (<i>nm-ni</i>)	0.79	0.86	0.1778
Liquidity (section 4.3)	0.66	0.66	0.0068

GE lets the rural-urban wage gap adjust, with urban productivity pinned at its baseline value throughout; PE holds the wage gap at its baseline value. The last column reports the rise in the log rural wage relative to baseline. Because urban wages are fixed, the increase in the rural wage coincides exactly with the fall in the rural-urban wage gap ω . The last two rows remove both margins of migration risk, but not in the same way. In ($\zeta = 0$, *ni*) the persistent part of the draw is revealed before the decision while the transitory part remains unknown. In *nm-ni* the whole draw is insured. The *nm-ni* row reports the scenario in which insurance against the initial draw is offered only to households without ex-ante information; providing it to all households changes little (appendix B.5).

First, both initial uncertainty and subsequent income risk are quantitatively important. Removing uncertainty about the initial persistent draw raises the urban share by eleven percentage points. Insuring income risk after migration raises it by eight points; insuring the initial draw as well raises it by 14. Combining

role of income risk alone, see the appendix for more details.

⁴⁴Insurance against the initial urban draw must be contracted before information about p_U is revealed. Information and insurance are therefore substitutes on this margin. If p_U is publicly revealed before contracting, the value of ex-ante insurance disappears (Hirshleifer, 1971); if it is learned privately, insurance is subject to adverse selection and may unravel (Hendren, 2013). The insurance contracts leave idiosyncratic migration taste shocks unchanged, so households continue to migrate partly for idiosyncratic reasons. See Donald, Fukui, and Miyauchi (2026) for recent work on insurance against such taste risk and appendix B.5 for additional details.

⁴⁵Solving the planner problem is a different exercise that requires taking a stand on several additional forces. A planner would internalize the effect of migration choices on future generations, specify how welfare is aggregated across cohorts, and internalize pecuniary externalities because rural wages respond to aggregate migration. Extending insurance to rural income raises a separate issue. Because migration is endogenous, pooling rural income changes the incentive to migrate and hence the composition of the insurance pool itself.

full information with insurance against subsequent income shocks raises the urban share by 16 percentage points, from 65 to 81 percent.

Second, general-equilibrium feedback substantially dampens the migration response. Partial-equilibrium effects are as much as 50 percent larger, as can be seen by comparing the first two columns. As migration rises, land per rural worker increases and pushes up rural wages, reducing the incentive for further migration. The strength of this feedback depends on the importance of land in rural production, and increases with the land share χ .

Table 14 reports consumption-equivalent welfare gains for rural households. I focus first on newborn households, whose welfare comparisons are free of composition effects. Once the rural wage response is taken into account, the welfare gains for newborn rural households range from 11.38 to 31.16 percent of consumption across counterfactuals. Welfare gains are similarly large for the average and poorest rural households. Their relative gains differ across counterfactuals, however: information benefits the average household more than the poorest, while the gains from insurance are more evenly distributed. Section 4.2 explains this incidence.

Table 14. Welfare Effects, Rural Households

	Newborn		Average	Poorest
	GE	PE	GE	GE
Information ($\zeta = 0$)	17.80	6.85	16.91	15.41
Insurance, income risk (ni)	11.38	5.28	10.79	10.42
Information and insurance ($\zeta = 0, ni$)	31.16	13.53	30.15	26.68
Insurance, both ($nm-ni$)	23.22	10.49	22.58	21.33

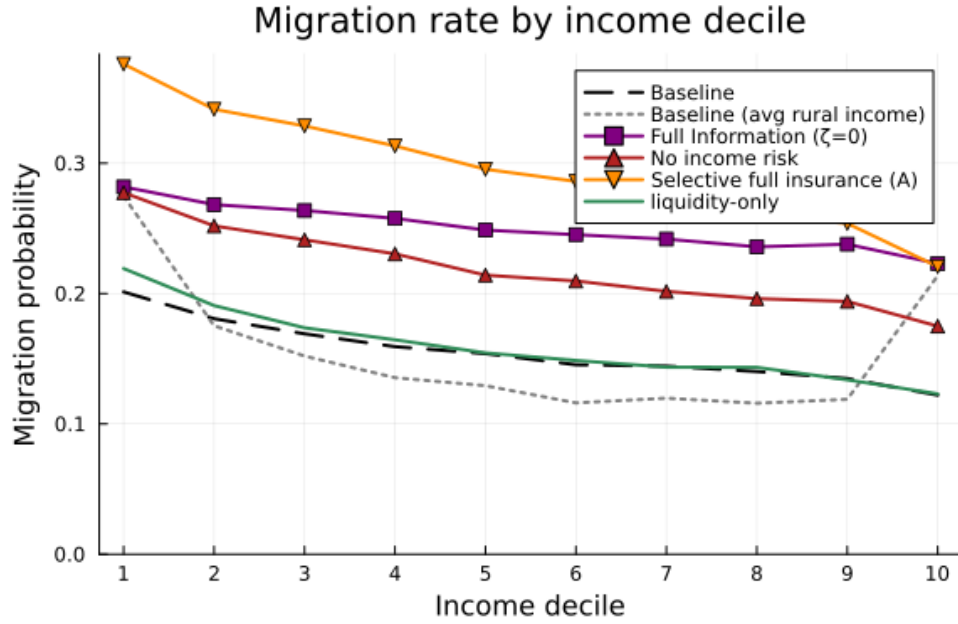
Consumption-equivalent welfare changes for rural households, in percent. GE lets the rural-urban wage gap adjust; PE holds it at its baseline value. Rows correspond to the economies of table 13. The newborn columns evaluate the counterfactual value function at zero assets against the fixed distribution of newborn rural types and therefore contain no composition effect. The average and poorest rural household are not composition-free, since who remains rural differs across economies; they are reported for what their ordering reveals about incidence rather than for their levels. I report welfare for rural households throughout, since they are the population whose migration decisions drive structural transformation.

The general equilibrium welfare effects are roughly twice as large as the partial equilibrium effects. The reason is that while most households do not migrate, they share in the welfare gains through rural wage increases. These wage increases originate from increasing land abundance when out-migration picks up, consistent with empirical evidence in Madhok et al. (2025).

4.2 Selection into Migration: Information vs. Insurance

Selection into Migration. I next analyze the microeconomic implications of each counterfactual, stressing that information provision draws in different migrants than insurance provision.

Figure 11. Selection into Migration – Counterfactuals



Migration probability, conditional on having a migration opportunity and by rural income decile. Income deciles are based on simulated income in the first wave inclusive of measurement error. The dashed line with diamond-shaped markers uses average rural income across all waves before migration instead, which produces the U-shape discussed in the text.

Figure 11 plots migration rates by income deciles, where income is measured in the first wave of the simulated sample.⁴⁶ The baseline selection pattern is given by the black dashed line. Providing information, the purple line, raises migration rates substantially. Moreover, it rotates the income-migration profile counter-clockwise – migration rises most at the top. The intuition is that a household with a high rural persistent type has much to lose from a low urban draw. For the lowest-income rural households the draw is mostly upside, since mean reversion pulls it up, so they move at high rates with or without information.

Insurance, the red line, also raises migration, but shifts migration rates up more evenly. Both high- and low-income households are similarly deterred by persistent income shocks, and I show in appendix B.6 how positive welfare effects are evenly distributed and hence flat along the income distribution in this counterfactual.

Rural-Urban Wage Gaps. The first result in table 15 is that information does not close measured rural-urban wage gaps. It widens them, even as the underlying efficiency wage gap falls substantially; see row two of table 13. The cross-sectional wage gap increases, while the fixed-effects gap rises by about 20 log points.

Both results reflect the more positive selection into migration shown in figure 11. With information, households migrate when their urban draw is high and stay rural when it is low. This raises the average urban type among migrants. Because rural and urban types are positively correlated, it also lowers the average type remaining in the rural sector, widening the cross-sectional wage gap. The fixed-effects gap

⁴⁶It turns out that mean-reversion induced by imperfectly-transitory shocks leads to non-monotone migration patterns when income deciles are based on average rural income before migration. While migration rates are always monotonically falling in p and ϵ , see figure 7, averaging across different periods can confound that monotone relationship. The dashed line with diamond-shaped markers illustrates this U-shape. Using first-wave income instead leads to well-behaved monotonically declining migration rates along the income distribution, consistent with the underlying migration motive.

risers even more because information selects precisely those households with large realized gains from migration.

Table 15. Risk and Selection into Migration

	log rural-urban wage gap		log(migrant/stayer income)
	cross-section	fixed effects	
Baseline	0.45	0.35	-0.16
Information ($\zeta = 0$)	0.46	0.55	-0.11
Insurance, income risk (ni)	0.34	0.19	-0.15
Information and insurance ($\zeta = 0, ni$)	0.37	0.41	-0.09
Insurance, both ($nm-ni$)	0.21	0.02	-0.17

All moments are simulated counterparts of the empirical moments reported in section 3, computed in general equilibrium.

The opposite occurs with insurance; see rows three and five. Insuring post-migration income shocks (ni) reduces the cross-sectional wage gap by about a quarter and roughly halves the fixed-effects gap, from 35 to 19 log points. Insurance raises migration among both low and high types without selecting households on their realized urban draw. Insuring the initial draw as well eliminates the fixed-effects wage gap entirely.

4.3 Liquidity and Poverty Traps

It is useful to benchmark risk against liquidity constraints, a traditional focus of development policy. The comparison is natural because fixed migration costs generate a poverty trap in the model: the poorest rural households cannot afford to move. I therefore relax the urban borrowing constraint to the natural borrowing limit,

$$b_U \geq -\frac{w_U e^{p+\epsilon}}{r},$$

and allow rural households to borrow against this limit to finance migration. Income risk remains in place, but the counterfactual removes the liquidity-driven poverty trap: even the poorest household can now finance the fixed cost.

The aggregate effect of liquidity is surprisingly small. Relaxing the borrowing constraint raises the urban population share by only one percentage point, compared with eight points from insuring post-migration income risk. Liquidity removes the fixed-cost barrier at the very bottom of the distribution, but does nothing about the persistent urban shocks that deter households throughout the income distribution.

Liquidity nevertheless matters for households currently at the bottom: relaxing the borrowing constraint raises welfare for the poorest rural household by 16.60%, more than insurance is worth to the same household. But few households remain at the bottom for long. Rural income is dominated by passing shocks arriving at rate $\lambda_{\epsilon,R} = 0.4987$, so low income states last only about two years. The gains are correspondingly smaller for newborn households, at 7.73%, and for the average rural household, at 1.07%.

To show that this result is driven by the income process rather than by an inability of the model to generate poverty traps, I replicate the experiment of Banerjee, Duflo, and Sharma (2021). The fixed cost f_{RU} is calibrated to be consistent with their setting, where a one-time transfer of roughly this size generated large and persistent income gains through rural-urban migration. Giving the same transfer to the poorest rural households in the baseline economy raises labor income by only 1.17, 2.19, and 1.44 log points after two, six, and ten years. After ten years, the share that has migrated rises by only 0.03, from a baseline of 0.15. The treatment effect is small because untreated households eventually receive better rural income draws and finance migration on their own.

In contrast, when rural low-income states are more persistent, the model replicates the empirical findings extremely well. I halve the arrival rate of passing shocks while holding rural income variance fixed, increasing the expected duration of a low-income state from roughly two to four years. Labor income now rises by 24.65 log points after six years, and the share who have moved after ten years rises by 0.31, from a base of 0.13 among the untreated. The resulting dynamics are strikingly close to Banerjee, Duflo, and Sharma (2021), who estimate labor-income gains of 27% after seven years.⁴⁷ The quantitative importance of poverty traps therefore depends critically on the persistence of low-income states, and not so much on the presence of non-convex costs.

In sum, risk is a much larger barrier to rural-urban migration than liquidity, echoing Karlan et al. (2014), who reach a similar conclusion for agricultural investment.

4.4 Migration Wedges and Risk

In a final exercise, I ask how much of the large residual migration wedge commonly inferred in models of rural-urban migration is accounted for by risk.⁴⁸ Models that abstract from initial uncertainty and persistent income shocks can still match observed migration by adjusting the residual migration wedge. I ask how much of this wedge is absorbing omitted risk.

I answer this question in two complementary ways. First, I treat my estimated model as the ground truth and ask how much the migration wedge must rise when different sources of risk are removed. Second, I reestimate a restricted model that eliminates initial uncertainty and persistent shocks, allowing the remaining structural parameters to adjust. Both exercises suggest that much of the residual cost of moving is the risk of moving.

The Migration Wedge: Model as Ground Truth. I first treat the estimated model as the ground truth and ask how much the migration wedge must rise to keep the urban population share at 65% as risk is removed. Table 16 reports the results. The required wedge rises by as much as 80 percent.

A researcher who assumes households know their initial urban productivity before moving must raise the wedge by 43.78%. Assuming instead that urban income is fully insured raises it by 20.92%. Imposing both full information and safe income raises the wedge by 78.66%. The required wedge is somewhat smaller when the initial draw is insured rather than revealed. The reason is selection: once households know they have received a high urban draw, a particularly large wedge is needed to deter them from moving. This result is consistent with Porcher, Morales, and Fujiwara (2024), who also find that ignoring information frictions inflates inferred migration wedges in a different setting.

Table 16. Risk and the Inferred Migration Wedge

	baseline	$\zeta = 0$	ni	$\zeta = 0, ni$	$nm-ni$
$\Delta\tau_{RU}$ (%)	0.0	43.78	20.92	78.66	43.37

The table reports the percentage increase in the migration wedge τ_{RU} required to reproduce the baseline urban share of 65% in each counterfactual economy. Each column corresponds to a row of table 13: the $\zeta = 0$ column is the full-information economy, the ni column the economy with cohort pooling of urban income, and the $(\zeta = 0, ni)$ column the economy that imposes both. The $nm-ni$ column also removes both margins, but it insures the initial draw for uninformed households instead of revealing it.

The Migration Wedge: Reestimating the Model. I next drop the assumption that my baseline model is the ground truth. Instead, I reestimate a restricted model that removes initial uncertainty and persistent income shocks, bringing the framework closer to standard models such as LMW and Kleemans (2015).

⁴⁷The welfare gain for the poorest rural household rises accordingly, from 16.60% to 28.69%.

⁴⁸See, for instance, Bryan and Morten (2019), Lagakos et al. (2020), LMW, and Baseler and Gordeev (2026).

I set persistent income shocks to zero in both sectors and recalibrate the remaining mean-reverting process to match income growth volatility and autocorrelation. I also eliminate uncertainty about urban persistent productivity by imposing $\rho_{RU} = 1$, $\sigma_{RU} = 0$, and $c_{RU} = \log \Lambda = 0$. The remaining parameters are reestimated to match the cross-sectional and fixed-effects rural-urban wage gaps, selection into migration, and urban inequality.⁴⁹ I hold the fixed migration cost f_{RU} and opportunity rate λ_m at their baseline values so that the implied migration wedge remains comparable across models.

Table 17 shows how the income process adjusts under these restrictions. The rural process changes little because persistent shocks were already unimportant. The urban process changes substantially: without permanent shocks, the mean-reverting component must become more persistent and more volatile to reproduce the observed income dynamics.

Table 17. The Income Process Under the Restriction

	Rural		Urban	
	Baseline	Restricted	Baseline	Restricted
σ_p^2	0.003	0	0.010	0
λ_ϵ	0.4987	0.453	0.3333	0.216
σ_ϵ^2	0.2602	0.320	0.0500	0.100

Permanent shocks are switched off in both sectors by construction, so the mean-reverting shock is the only source of income risk that survives. It is recalibrated on the variance of the change in log income and its first-order autocorrelation. The rural process barely moves, since rural risk was mean-reverting to begin with. The urban process changes sharply: the mean-reverting shock has to become both more persistent and considerably more volatile to stand in for the permanent shocks that have been switched off.

Tables 18 and 19 report the model fit and estimated parameters. The restricted model matches the targeted rural-urban wage gaps and selection into migration.

Table 18. Moments Targeted, Restricted Model

Moments	data SA	model SA
Log rural-urban wage gap (cross-section)	0.45	0.46
Log rural-urban wage gap (fixed effects)	0.34	0.36
Log(migrant income in rural/stayer income in rural)	-0.14	-0.14
Initial inequality urban	0.18	0.180

The four moments targeted in the restricted model, against the same data counterparts used in table 9. The system is exactly identified, so these four are the entire target vector. As in the baseline, the urban population share is matched by construction rather than targeted, since A_U adjusts to clear the urban labor market. The remaining moments of table 9 are not targeted here: the arrival rate and variance of the mean-reverting shock are calibrated outside the estimation, and the moments that discipline ρ_{RU} , σ_{RU} and ζ have no counterpart once those parameters are imposed.

Table 19. Estimated Parameters, Restricted Model

Parameter	$c_{0,R}$	$\sigma_{U,0}$	s	τ_{RU}
Estimate	-0.76	0.20	7.60	1.51

The remaining parameters of the full model are imposed by the restriction rather than estimated, as described above. Urban productivity adjusts to $A_U = 0.74$ so that the urban population share matches its baseline value, and the implied rural-urban wage ratio is $\omega = 1.26$.

The same moments are now generated by a very different migration strategy. Future migrants earn -0.14 log points less than stayers before moving because they have received adverse transitory rural shocks. Migration allows them to draw a new urban shock, so measured wage gains combine the equilibrium wage

⁴⁹I restrict heterogeneity in persistent type amongst households born in the rural sector to zero, i.e., $\sigma_{0,R}^2 = 0$, while still allowing for differences in persistent urban productivity amongst urban natives, i.e., $\sigma_{0,U}^2 > 0$, just as in LMW.

gap, $\omega = 1.26$, with mean reversion. Migration has become a coping strategy for temporary rural income shocks rather than a response to differences in expected lifetime income.⁵⁰⁵¹

The restricted model requires a substantially larger migration wedge. It rises by 40.07% relative to baseline, from 1.08 to 1.51. The consumption-equivalent interpretation puts this number into perspective: in the baseline model, absent preference shocks, a migrant requires only eight percent more urban consumption to be indifferent between moving and staying. Once initial uncertainty and persistent shocks are omitted, this rises to 53 percent. The model compensates for missing risk by assigning households a much larger residual cost of moving.

The two exercises therefore reach the same conclusion: much of what appears to be the cost of moving is the risk of moving.

5 Conclusion

Rural-urban income gaps are large, and a common interpretation is that they reflect efficient sorting. This paper takes a different view. Permanent migration is a dynamic and risky decision: most households do not know where they will land on the urban income distribution and are exposed to persistent income shocks after they move.

This risk represents a quantitatively important barrier to rural-urban structural change. Removing risk raises the urban population share by about 16 percentage points in the South African context. A researcher who ignores risk overstates the residual migration wedge by as much as 80 percent.

The two risks operate differently across households. Uncertainty about initial urban income holds back high-income rural households and generates negative selection into migration, while persistent income shocks deter rich and poor alike. Wealth is not a substitute for insurance. Buffer-stock savings smooth transitory shocks but provide little protection against persistent ones. Moreover, the gains from migration are tied to labor income. As households become wealthier, labor income becomes a smaller share of total income, chipping away at the gains from migration. Wealth can therefore even reduce incentives to migrate.

Several open questions remain. How could insurance be provided in the presence of moral hazard, which I have abstracted from throughout? Can better information, perhaps through improved search and matching, reduce uncertainty about urban outcomes? And how do these forces shape structural change along the transition path, where rural-urban migration necessarily generates unbalanced growth?

References

- Abowd, John M and David Card (1989). "On the Covariance Structure of Earnings and Hours Changes". In: *Econometrica* 57.2, pp. 411–445.
- Achdou, Yves et al. (2022). "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach". In: *Review of Economic Studies* 89.1, pp. 45–86.

⁵⁰This mechanism is consistent with the large IV estimates in Bryan, Chowdhury, and Mobarak (2014) and is developed in detail by LMW.

⁵¹In this restricted economy, I can also ask: how much more migration do we get if we insure against initial uncertainty and subsequent income shocks in the urban sector. This raises the urban share from 0.65 to 0.69, which is substantially lower than in the baseline economy. This highlights that the strong response in the baseline economy isn't trivially implied, but depends on the large risk due to persistent shocks.

- Aguiar, Mark, Mark Bilal, and Corina Boar (2025). “Who Are the Hand-to-Mouth?” In: *Review of Economic Studies* 92.3, pp. 1293–1340. DOI: [10.1093/restud/rdae056](https://doi.org/10.1093/restud/rdae056).
- Akram, Agha Ali, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2017). *Effects of Emigration on Rural Labor Markets*. Working Paper 23929. National Bureau of Economic Research.
- Alderman, Harold and Christina H. Paxson (1994). “Do the Poor Insure? A Synthesis of the Literature on Risk and Consumption in Developing Countries”. In: *Economics in a Changing World*. Macmillan, pp. 48–78.
- Alvarez, Jorge A (2020). “The Agricultural Wage Gap: Evidence from Brazilian Micro-Data”. In: *American Economic Journal: Macroeconomics* 12.1, pp. 153–173.
- Alvarez-Cuadrado, Francisco, Francesco Amodio, and Markus Poschke (2026). “Selection, Absolute Advantage, and the Agricultural Productivity Gap”. In: *Journal of Political Economy Macroeconomics* 4.3, pp. 513–562.
- Angelucci, Manuela and Giacomo De Giorgi (2009). “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption?” In: *American Economic Review* 99.1, pp. 486–508.
- Attanasio, Orazio et al. (2025). “Growing Apart: Declining Within- and Across-Localities Insurance in Rural China”. In: *Journal of Development Economics* 172, p. 103366. DOI: [10.1016/j.jdeveco.2024.103366](https://doi.org/10.1016/j.jdeveco.2024.103366).
- Balboni, Clare et al. (2022). “Why Do People Stay Poor?” In: *Quarterly Journal of Economics* 137.2, pp. 785–844.
- Banerjee, Abhijit, Esther Duflo, and Garima Sharma (2021). “Long-Term Effects of the Targeting the Ultra Poor Program”. In: *American Economic Review: Insights* 3.4, pp. 471–486.
- Barnett-Howell, Zachary et al. (2025). “Reaching the Novice or Nudging the Expert? Networks, Information, and the Experimental Returns to Migration”. Working paper.
- Baseler, Travis (2023). “Hidden Income and the Perceived Returns to Migration”. In: *American Economic Journal: Applied Economics* 15.4, pp. 321–352.
- Baseler, Travis and Stepan Gordeev (2026). “The Macroeconomic Effects of Lifting Barriers to Migration”. Mimeo.
- Bazzi, Samuel (2017). “Wealth Heterogeneity and the Income Elasticity of Migration”. In: *American Economic Journal: Applied Economics* 9.2, pp. 219–255.
- Beegle, Kathleen, Joachim De Weertdt, and Stefan Dercon (2011). “Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey”. In: *Review of Economics and Statistics* 93.3, pp. 1010–1033.
- Bertoli, Simone (2010). “The Informational Structure of Migration Decision and Migrants Self-Selection”. In: *Economics Letters* 108.1, pp. 89–92.
- Blanchard, Olivier J (1985). “Debt, Deficits, and Finite Horizons”. In: *Journal of Political Economy* 93.2, pp. 223–247.
- Blundell, Richard, Hamish Low, and Ian Preston (2013). “Decomposing Changes in Income Risk Using Consumption Data”. In: *Quantitative Economics* 4.1, pp. 1–37.
- Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). “Consumption Inequality and Partial Insurance”. In: *American Economic Review* 98.5, pp. 1887–1921.
- Borjas, George J (1987). “Self-Selection and the Earnings of Immigrants”. In: *American Economic Review* 77.4, pp. 531–553.
- Borjas, George J, Ilpo Kauppinen, and Panu Poutvaara (2019). “Self-Selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics”. In: *Economic Journal* 129.617, pp. 143–171.

- Braxton, J. Carter et al. (2025). “Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter”. In: *American Economic Review* 115.12, pp. 4438–4475. DOI: [10.1257/aer.20231656](#).
- Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2014). “Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh”. In: *Econometrica* 82.5, pp. 1671–1748.
- Bryan, Gharad and Melanie Morten (2019). “The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia”. In: *Journal of Political Economy* 127.5, pp. 2229–2268.
- Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019). “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock”. In: *Econometrica* 87.3, pp. 741–835.
- Carli, Francesco, Albert Rodriguez-Sala, and Raül Santaaulalia-Llopis (2025). “Social Norms and Excess of Transfer Progressivity in the Village”. Work in progress.
- Caselli, Francesco (2005). “Accounting for Cross-Country Income Differences”. In: *Handbook of Economic Growth*. Ed. by Philippe Aghion and Steven N. Durlauf. Vol. 1A. Elsevier North-Holland. Chap. 9, pp. 679–741.
- Chiappori, Pierre-André et al. (2014). “Heterogeneity and Risk Sharing in Village Economies”. In: *Quantitative Economics* 5.1, pp. 1–27.
- Choukhmane, Taha and Tim de Silva (2026). “What Drives Investors’ Portfolio Choices? Separating Risk Preferences from Frictions”. In: *Journal of Finance* 81.1, pp. 5–48. DOI: [10.1111/jofi.70013](#).
- Collins, Daryl et al. (2009). *Portfolios of the Poor: How the World’s Poor Live on \$2 a Day*. Princeton University Press.
- Crawley, Edmund (2020). “In Search of Lost Time Aggregation”. In: *Economics Letters* 189, p. 108998.
- Crawley, Edmund, Martin B. Holm, and Håkon Tretvoll (2026). “A Parsimonious Model of Idiosyncratic Income”. In: *International Economic Review* 67.2, pp. 533–548. DOI: [10.1111/iere.70041](#).
- Daly, Moira, Dmytro Hryshko, and Iouri Manovskii (2022). “Improving the Measurement of Earnings Dynamics”. In: *International Economic Review* 63.1, pp. 95–124.
- De Magalhaes, Leandro, Enric Martorell, and Raül Santaaulalia-Llopis (2024). “Transfer Progressivity and Development”. Unpublished manuscript.
- De Magalhaes, Leandro and Raül Santaaulalia-Llopis (2018). “The Consumption, Income, and Wealth of the Poorest: An Empirical Analysis of Economic Inequality in Rural and Urban Sub-Saharan Africa for Macroeconomists”. In: *Journal of Development Economics* 134, pp. 350–371.
- Deaton, Angus and Christina Paxson (1994). “Intertemporal Choice and Inequality”. In: *Journal of Political Economy* 102.3, pp. 437–467.
- Donald, Eric, Masao Fukui, and Yuhei Miyauchi (2026). “Unpacking Aggregate Welfare in a Spatial Economy”. In: *Review of Economic Studies*. Forthcoming. DOI: [10.1093/restud/rdag021](#).
- Dvorkin, Maximiliano and Brian Greaney (2025). *The Geography of Wealth: Shocks, Mobility, and Precautionary Savings*. Working Paper 2024-033. Federal Reserve Bank of St. Louis. DOI: [10.20955/wp.2024.033](#).
- Friedman, Milton (1957). *Theory of the Consumption Function*. Princeton university press.
- Gabaix, Xavier et al. (2016). “The Dynamics of Inequality”. In: *Econometrica* 84.6, pp. 2071–2111.
- Gazeaud, Jules, Eric Mvukiyehe, and Olivier Sterck (2023). “Cash Transfers and Migration: Theory and Evidence from a Randomized Controlled Trial”. In: *Review of Economics and Statistics* 105.1, pp. 143–157.
- Gollin, Douglas (2002). “Getting Income Shares Right”. In: *Journal of Political Economy* 110.2, pp. 458–474.

- Gollin, Douglas, Martina Kirchberger, and David Lagakos (2021). "Do Urban Wage Premia Reflect Lower Amenities? Evidence from Africa". In: *Journal of Urban Economics* 121. DOI: [10.1016/j.jue.2020.103301](https://doi.org/10.1016/j.jue.2020.103301).
- Gollin, Douglas, David Lagakos, and Michael E Waugh (2014). "The Agricultural Productivity Gap". In: *Quarterly Journal of Economics* 129.2, pp. 939–993.
- Gottschalk, Peter et al. (1994). "The Growth of Earnings Instability in the US Labor Market". In: *Brookings Papers on Economic Activity* 1994.2, pp. 217–272.
- Guvenen, Fatih (2009). "An Empirical Investigation of Labor Income Processes". In: *Review of Economic Dynamics* 12.1, pp. 58–79.
- Guvenen, Fatih and Anthony A Smith (2014). "Inferring Labor Income Risk and Partial Insurance from Economic Choices". In: *Econometrica* 82.6, pp. 2085–2129.
- Guvenen, Fatih et al. (2021). "What Do Data on Millions of US Workers Reveal about Lifecycle Earnings Dynamics?" In: *Econometrica* 89.5, pp. 2303–2339.
- Hall, Robert E (1971). "The Measurement of Quality Change from Vintage Price Data". In: *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press, pp. 240–272.
- Hamory, Joan et al. (Nov. 2020). "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata". In: *Journal of the European Economic Association* 19.3, pp. 1522–1555.
- Hansen, Gary D and Edward C Prescott (2002). "Malthus to Solow". In: *American Economic Review* 92.4, pp. 1205–1217.
- Harris, John R. and Michael P. Todaro (1970). "Migration, Unemployment and Development: A Two-Sector Analysis". In: *American Economic Review* 60.1, pp. 126–142.
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante (2010). "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006". In: *Review of Economic Dynamics* 13.1, pp. 15–51.
- Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante (2005). "Two Views of Inequality over the Life Cycle". In: *Journal of the European Economic Association* 3.2-3, pp. 765–775.
- Hendren, Nathaniel (2013). "Private Information and Insurance Rejections". In: *Econometrica* 81.5, pp. 1713–1762.
- Hendricks, Lutz and Todd Schoellman (2018). "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration". In: *Quarterly Journal of Economics* 133.2, pp. 665–700.
- Hirshleifer, Jack (1971). "The Private and Social Value of Information and the Reward to Inventive Activity". In: *American Economic Review* 61.4, pp. 561–574.
- Huggett, Mark, Gustavo Ventura, and Amir Yaron (2011). "Sources of Lifetime Inequality". In: *American Economic Review* 101.7, pp. 2923–2954.
- Jack, William and Tavneet Suri (2014). "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution". In: *American Economic Review* 104.1, pp. 183–223.
- Kaplan, Greg and Giovanni L Violante (2010). "How Much Consumption Insurance beyond Self-Insurance?" In: *American Economic Journal: Macroeconomics* 2.4, pp. 53–87.
- Karlan, Dean et al. (2014). "Agricultural Decisions After Relaxing Credit and Risk Constraints". In: *Quarterly Journal of Economics* 129.2, pp. 597–652.
- Kennan, John and James R Walker (2011). "The Effect of Expected Income on Individual Migration Decisions". In: *Econometrica* 79.1, pp. 211–251.

- Kinnan, Cynthia and Robert Townsend (2012). “Kinship and Financial Networks, Formal Financial Access, and Risk Reduction”. In: *American Economic Review* 102.3, pp. 289–293.
- Kleemans, Marieke (2015). *Migration Choice under Risk and Liquidity Constraints*. Tech. rep.
- Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh (2023). “The Welfare Effects of Encouraging Rural–Urban Migration”. In: *Econometrica* 91.3, pp. 803–837.
- Lagakos, David and Michael E Waugh (2013). “Selection, Agriculture, and Cross-Country Productivity Differences”. In: *American Economic Review* 103.2, pp. 948–980.
- Lagakos, David et al. (2020). “Migration Costs and Observational Returns to Migration in the Developing World”. In: *Journal of Monetary Economics* 113, pp. 138–154.
- Lewis, W. Arthur (1954). “Economic Development with Unlimited Supplies of Labour”. In: *The Manchester School* 22.2, pp. 139–191. DOI: [10.1111/j.1467-9957.1954.tb00021.x](https://doi.org/10.1111/j.1467-9957.1954.tb00021.x).
- Ligon, Ethan (1998). “Risk Sharing and Information in Village Economies”. In: *Review of Economic Studies* 65.4, pp. 847–864.
- (2025). “Risk Sharing Tests and Covariate Shocks: Drought, Floods, and Pests in Uganda”. In: *American Economic Review*. Forthcoming. DOI: [10.1257/aer.20240664](https://doi.org/10.1257/aer.20240664).
- Ligon, Ethan, Jonathan P Thomas, and Tim Worrall (2002). “Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies”. In: *Review of Economic Studies* 69.1, pp. 209–244.
- Low, Hamish, Costas Meghir, and Luigi Pistaferri (2010). “Wage Risk and Employment Risk over the Life Cycle”. In: *American Economic Review* 100.4, pp. 1432–1467.
- Lucas Jr, Robert E (2004). “Life Earnings and Rural-Urban Migration”. In: *Journal of Political Economy* 112.S1, S29–S59.
- Lucas, Robert E. Jr. (1987). *Models of Business Cycles*. Oxford: Basil Blackwell.
- Luxembourg Income Study (LIS) Database (2025). *Luxembourg Income Study (LIS) Database*. <http://www.lisdatacenter.org>. Luxembourg: LIS. Accessed <date of access>.
- MaCurdy, Thomas E (1982). “The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis”. In: *Journal of Econometrics* 18.1, pp. 83–114.
- Madhok, Raahil et al. (2025). *Internal Migration and the Spatial Reorganization of Agriculture*. Working Paper 33854. National Bureau of Economic Research.
- Mazzocco, Maurizio and Shiv Saini (2012). “Testing Efficient Risk Sharing with Heterogeneous Risk Preferences”. In: *American Economic Review* 102.1, pp. 428–468.
- McFadden, Daniel (1974). “Conditional Logit Analysis of Qualitative Choice Behavior”. In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York: Academic Press, pp. 105–142.
- Meghir, Costas et al. (2022). “Migration and Informal Insurance: Evidence from a Randomized Controlled Trial and a Structural Model”. In: *Review of Economic Studies* 89.1, pp. 452–480.
- Miner, Gwyneth (2025). “Reducing Migration Uncertainty: Conditional Unemployment Insurance for Kenyan Migrants”. Working paper.
- Mobarak, Ahmed Mushfiq and Maira Emy Reimão (2020). “Seasonal Poverty and Seasonal Migration in Asia”. In: *Asian Development Review* 37.1, pp. 1–42.
- Mongey, Simon and Michael E. Waugh (2024). *Discrete Choice, Complete Markets, and Equilibrium*. Working Paper 32135. National Bureau of Economic Research. DOI: [10.3386/w32135](https://doi.org/10.3386/w32135).
- Morduch, Jonathan (1995). “Income Smoothing and Consumption Smoothing”. In: *Journal of Economic Perspectives* 9.3, pp. 103–114.

- Morten, Melanie (2019). “Temporary Migration and Endogenous Risk Sharing in Village India”. In: *Journal of Political Economy* 127.1, pp. 1–46.
- Munshi, Kaivan and Mark Rosenzweig (2016). “Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap”. In: *American Economic Review* 106.01, pp. 46–98.
- Porcher, Charly (2025). “Migration with Costly Information”. Mimeo.
- Porcher, Charly, Eduardo Morales, and Thomas Fujiwara (2024). *Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach*. Working Paper 32413. Conditionally accepted, American Economic Journal: Microeconomics. National Bureau of Economic Research.
- Pulido, José and Tomasz Swiecki (2018). “Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps across Sectors in Indonesia”. Working paper, revision requested at American Economic Journal: Macroeconomics.
- Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu (2008). “Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis”. In: *Journal of Monetary Economics* 55.2, pp. 234–250.
- Samphantharak, Krislert and Robert M Townsend (2018). “Risk and Return in Village Economies”. In: *American Economic Journal: Microeconomics* 10.1, pp. 1–40.
- Santaaulalia-Llopis, Raül and Yu Zheng (2018). “The Price of Growth: Consumption Insurance in China 1989–2009”. In: *American Economic Journal: Macroeconomics* 10.4, pp. 1–35.
- Storesletten, Kjetil, Christopher I Telmer, and Amir Yaron (2004). “Consumption and Risk Sharing over the Life Cycle”. In: *Journal of Monetary Economics* 51.3, pp. 609–633.
- Townsend, Robert M. (1994). “Risk and Insurance in Village India”. In: *Econometrica* 62.3, pp. 539–591.
- Trouvain, Florian (2025). “Structural Change, Inequality, and Capital Flows”. Mimeo, University of Oxford.
- Udry, Christopher (1994). “Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria”. In: *Review of Economic Studies* 61.3, pp. 495–526.
- Working, Holbrook (1960). “Note on the Correlation of First Differences of Averages in a Random Chain”. In: *Econometrica* 28.4, pp. 916–918.
- World Bank (2021). *South Africa: Social Assistance Programs and Systems Review*. Report 164802. Washington, DC: World Bank. DOI: [10.1596/36514](https://doi.org/10.1596/36514).
- Yaari, Menahem E (1965). “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer”. In: *Review of Economic Studies* 32.2, pp. 137–150.
- Young, Alwyn (2013). “Inequality, the Urban-Rural Gap, and Migration”. In: *Quarterly Journal of Economics* 128.4, pp. 1727–1785.

A Data Appendix

A.1 Sample Selection and Weighting

All regressions in this section draw on three sample-selection indicators, constructed once at the household-wave level and then combined as needed for each specific analysis:

- *pp_in*: household income within the 3rd–99th percentile of the income distribution, computed separately by urban/rural sector and survey wave. This screens out implausible income levels (e.g. mis-reported near-zero or extreme values) without imposing a single economy-wide cutoff that would conflate sector- or period-specific differences in the income distribution with genuine outliers.

- *pp_in_con*: the analogous 3rd–99th percentile trim applied to household consumption, computed independently of *pp_in* (a household can fail one trim without failing the other) so that income-based and consumption-based analyses are not needlessly restricted by each other's outliers.
- *elig*: an eligibility indicator equal to one for individuals of working age (23–64) with active employment status, dropping labor-market inactive individuals and survey refusals. Employment status is used only as a sample restriction, never as a regression control, since it is itself an outcome of the migration decision under study.

For analyses that compare a household across two waves (e.g. growth in income or consumption around a migration event), both *pp_in/pp_in_con* and *elig* are required to hold in *both* the earlier and later wave — a trim on the current wave alone would not screen out a problematic lagged observation.

Weights. Two survey weights are used, matched to whether the analysis treats each wave as an independent cross-section or exploits within-household variation over time:

- *cs_weight*, a cross-sectional survey weight, is used for analyses comparing rural and urban households (or observations) within a single wave or pooled across waves without following individuals over time — e.g. the rural-urban income/consumption gap in levels, the selection-into-migration comparisons, and the income/consumption distribution histograms.
- *pn_weight*, a panel weight, is used wherever the specification follows the same household across waves — e.g. the growth-and-volatility regressions, the better/worse off regressions, and the fixed-effects version of the rural-urban gap.

The sample restriction and weight applied in each analysis are summarized below:

- **Descriptive summary statistics:** full sample (unrestricted); *cs_weight*.
- **Migration-flow shares:** full sample (unrestricted); *cs_weight*.
- **Mover characteristics** (age/education/assets vs. migrant status): *elig; pp_in/pp_in_con* added only for the income/consumption outcome; *cs_weight*.
- **"Better/worse off"** (income or consumption up vs. down, wave *t* vs. *t-1*): *pp_in/pp_in_con* and *elig*, both waves; *pn_weight*.
- **Growth and volatility around migration:** *pp_in/pp_in_con* and *elig*, both waves; growth annualized by years since the previous interview (NIDS wave spacing is uneven); *pn_weight*.
- **Migration probability by pre-move income decile:** rural pre-move sample, *elig, pp_in*; further restricted to households present in the survey's first (2008) wave, so the cross-sectional weight is drawn from one unambiguous snapshot even though income itself is averaged over all pre-move waves; *cs_weight* (2008 wave).
- **Income/consumption distribution histograms:** *pp_in/pp_in_con* and *elig*; residual (net of age, education, household size, gender, and year) additionally trimmed at its own 1st/99th percentile, symmetrically, since residualizing can introduce tail values that a level-based trim cannot catch; *cs_weight*.
- **Selection into migration** (pre-move income/consumption gap, rural stayers vs. future migrants): *elig, pp_in/pp_in_con*;

- **Variance of income/consumption residuals**, movers vs. stayers: `pp_in/pp_in_con,elig;cs_weight`.
- **Rural-urban gap, cross-section**: `pp_in/pp_in_con,elig`; age, education, and gender controls, but *not* household size, which is a plausible mediator of urban living rather than a pre-determined confounder; `cs_weight`.
- **Rural-urban gap, household fixed effects**: `pp_in/pp_in_con,elig;pn_weight`.

A.2 NIDS South Africa

The NIDS is a representative panel data set that collects information on income, consumption, wealth, and demographics for South African households and individuals roughly at biennial frequency from 2008 to 2017. A detailed codebook is available, which can be found clicking this [link](#), downloaded in August 2023.

Households. While the dataset is a panel dataset, there are no household identifiers that are valid across time. The only way to build a household panel is to combine different household units across waves by following specific individuals, which can be consistently tracked over time.

The researcher thus has to develop a method that defines a household head, which then is used to add different household survey together. I propose the following method.

- Drop top-up sample (for wave 5) and all TSMs.
- Calculate individual labor income.
- Calculate the cumulative individual labor income of the current and all later waves. (For example, if the current wave is wave 1, then take the sum of individual labor income of wave 1-5).
- Choose the highest earner among all members for each household.
- If there is no one chosen in the household, choose the eldest male member aged 15 to 64.
- If there are more than two chosen in the household, choose the elder male member. If there are still more than two, choose the member with smaller pid (personal id) to break the tie.
- Assume that household heads in wave X are still household heads in wave X+1. If the previous household heads are not present in later waves, apply the above procedure again.
- Only keep individuals identified as the household head for all waves. We use them to link households across different waves.
- Apply the same procedure to adult women if there is no male household head.

Because I have to define households heads without having an a-priori defined mapping from household units across time, there is a valid concern whether this procedure biases the composition towards having more urban households, and less households in small-scale farming. A reason could be that an individuals income is not well measured, or not measured at all, in subsistence agriculture. To investigate this concern, I analyze the composition of the full sample, and the composition of the selected sample with household observations that I could connect across waves.⁵² The upshot of the analysis is that there appears to be no major compositional shift.

⁵²See the stata code “`empirics/code/clean_data/south_africa/14_linking_sanity_check.do`” written by Qingyu Chen in the replication package.

Income. This variable is measured on a monthly basis, net of taxes, and includes imputed rental income from owner-occupied housing. Several imputation and correction procedures are applied to improve the quality of the data. Earned income or labor income includes wages, profits, and business income but drops other sources, especially government income, asset income, and other rental income including owner-occupied housing rent imputations.

I define labor income consistent with this variable, but instead of taking it directly, I use the imputed income measure `hhincome`, and I subtract individual income components. The imputation makes the whole thing a little bit more reliable so building on that variable and subtracting categories is preferable. If I get negative entries anywhere, I treat them as zeros.

Consumption. The baseline consumption expenditure measure includes food and non-food items measured over a month (30 days) including (imputed) rental expenditure.

I drop the following categories from the baseline consumption expenditure measure with the number of the category as outlined in the code book in parentheses: Jewelry (6), wedding ceremonies and funerals (13),⁵³ furniture bought with cash (31), lobola, which refers to a bride price (53), and income tax payments, which are already taken out of the household income measure and thus should not be included in consumption expenditure (54). I then use the baseline consumption expenditure measure, which includes some imputations, and subtract these specific categories from it to arrive at my consumption measure. This final measure is still highly correlated with the baseline consumption measure on the household level with a correlation coefficient of around .98 when measured in logs.

I also experimented with a more narrow measure of non-durable consumption. If one drops too many categories, there is hardly any consumption inequality in rural left, which makes sense. As more elastic consumption categories are dropped, subsistence requirements in basic items lead to very little variation across households. This observation has been made before, see especially BPP where food expenditure is used to impute overall non-durable expenditure to account for the low income elasticity in food, which otherwise generates artificially low passthrough coefficients.

Price deflators. I use the consumer price index from the South African Statistical office to deflate all income and expenditure variables in 2017-Rand.

Rural-Urban Status. A rural-urban status is provided, where rural includes both “traditional communities” as well as farming communities. I perform a robustness exercise later on to disentangle the role of artificial rural-urban reclassification without actual spatial moves.

A.2.1 Micro and Macro Data

I next assess the quality of the dataset by aggregating up income, and comparing it to aggregate statistics from the world bank.

⁵³Funerals are important events for rural communities that can cost several monthly wages for low-income rural households in South Africa, see Collins et al. (2009).

Figure 12. Micro and Macro Income Data – South Africa

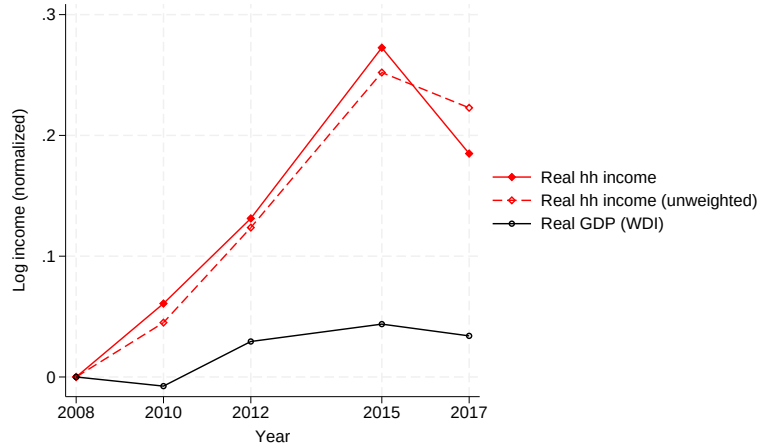


Figure 12 plots normalized log real income, i.e., I first compute sample averages, then divide each average by its value in the first year so I end up with an index of real income, and finally take the log. The real GDP measure comes again from the world bank measured at constant local currency units. Cross-sectional survey weights are used to compute the average, see the appendix for details.

Figure 13. Micro and Agg. Urban Share – South Africa

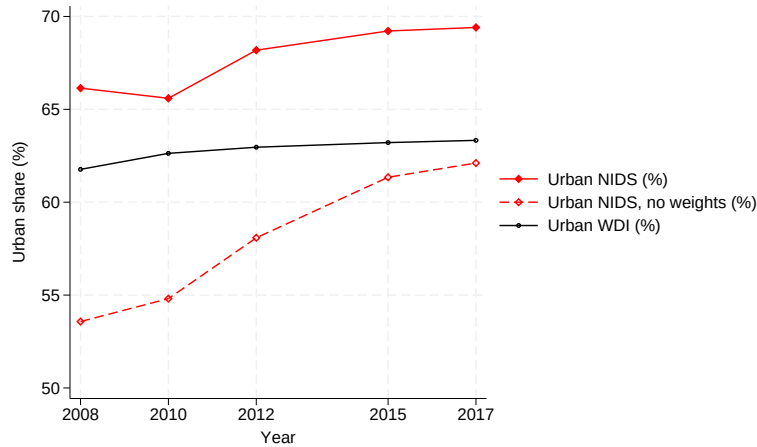


Figure 13 plots the share of urban households as reported by the World Bank Development Indicators, and as found in the micro data. I compute a weighted and an unweighted average using cross sectional weights as detailed in the appendix.

A.2.2 Sample Selection & Probability Weights

Sample selection. I select the sample aggressively to create a high-quality sub sample. I explain each restriction in detail, with a summary of how many observations are lost in each step provided in table 20.

I drop households that move across sectors. I use them to estimate rural-urban wage gaps and other parameters relevant to the model, but it is cleaner to estimate the income process for stayers only. I drop households where the month is missing.⁵⁴ Next, I impose an age limit $27 \leq age \leq 62$. The limit is slightly

⁵⁴This matters because I need to know the month because it tells me how far apart observations are in time, and I also control for aggregate shocks on the monthly level in the first stage regression, which seems important for agricultural households.

more generous compared to the Chinese application the reason being that I have difficulty to keep the sample size sufficient large. I drop households that are likely to split off, although I do not directly observe split-offs because there is no consistent household identifier over time. I drop households where the number of adults changes by more than 2, and I drop households that have more than 12 members. I now control for divorce in the first stage regression instead of dropping to keep the sample size higher. I restrict the sample to households with strong attachment to the labor market, which means I drop students, retirees, and heads that are out of the labor force. I do keep unemployed households. I drop ultra poor households that earn less than a dollar a day, which might be more reflective of measurement error.⁵⁵ I also drop the lowest 5% in each sector-year unit, but this turns out to have almost no bite because of the other restrictions. Lastly, I impose that income is missing at most once per household over the sample, based on advice from Daly, Hryshko, and Manovskii (2022).

Table 20. Sample Selection in South Africa

(a) Rural			(b) Urban		
Restrictions	Households	Observations	Restrictions	Households	Observations
All	2644	12597	All	3665	17340
No movers	2093	10002	No movers	3165	14904
Missing month	1722	8369	Missing month	2421	11444
Ages 27 to 62	1283	6200	Ages 27 to 62	1893	8881
Household splits	1179	5592	Household splits	1765	8376
Labor market attachment	884	4196	Labor market attachment	1537	7316
Income outliers	778	3671	Income outliers	1321	6357
Max one missing	331	1661	Max one missing	612	3051

NIDS from 2008 to 2017.

Probability weights. Every statistic is weighted by inverse probability weights. For cross-sectional objects, I use the weights *wgt*. For panel moments I use the weights *pweight*. I follow the advice in the codebook and trim the first wave of the panel weights at the 1st and 99th percentile.

Descriptive Statistics. Table 21 contains descriptive statistics for both the raw data, and the selected sample.

⁵⁵Because I take logs, low income realization can have an outsized influence on sample statistics.

Table 21. Summary statistics South Africa

(a) Full Sample				
	Rural		Urban	
	mean	sd	mean	sd
r_hhincome_lab_gov	4105.557	6285.731	8030.53	15066.41
Real expenditure	3642.57	5770.78	7728.415	12780.33
Years of schooling	8.856127	4.364013	10.72278	3.522302
Age	37.42919	14.55792	38.55696	13.18948
Male share	.5601422	.4963913	.6000816	.4898979
Household size	4.726216	3.173566	3.64497	2.358767
Observations	11531		14703	

(b) Selected Sample				
	Rural		Urban	
	mean	sd	mean	sd
r_hhincome_lab_gov	6835.937	8847.935	12318.69	19693.29
Real expenditure	4858.238	6779.317	9605.623	13432.56
Years of schooling	9.121613	4.582276	11.2527	3.197754
Age	42.15352	7.917871	42.72599	7.934546
Male share	.5586996	.4966919	.5798099	.4936702
Household size	3.798314	2.242986	3.581449	1.931491
Observations	1661		3051	

Variables refer to household head's demographics. In table 21a I do not select the sample other than dropping observations without urban-rural status. In table 21a I apply the same selection criteria defined earlier.

A.2.3 Residualizing the Data

I follow BPP and run a first-stage regression to purge the data from predictable life-cycle components. I run a standard OLS regression with the relevant log household income on the left hand side, and the following right hand side variables for each sector separately

- years of schooling
- age, age_square
- log household size, log number adults
- race, gender, marital status dummies
- urban-year fixed effects
- urban-month fixed effects.

This is a relatively parsimonious first stage regression similar to Aguiar, Bils, and Boar (2025). I run a regression of the form

$$y_{i,t} = \alpha + \gamma'_t \mathbf{x}_{i,t} + u_{i,t}$$

and then work with the residuals as income measure when computing autocovariances in the main part of the paper.

A.2.4 Additional Statistics

In this section I report additional details on rural-urban migration and rural-urban wage gaps. I impose less stringent sample selection criteria compared to when I estimate income risk. I consider households

between 24 and 64 years of age, and I drop households that are out of the labor force (i.e., households that do not want to work), and I drop the lowest 3% of observations for each wave-sector pair since the log transformation is sensitive to very small income levels.

Table 22. Migration and cross-sectional differences in South Africa (standardized)

(a) Rural stayers vs. rural movers		(b) Urban stayers vs. urban movers	
	(1) Rural-Urban Migrant		(1) Urban-Rural Migrant
age	-0.150***	age	0.005
gender	0.075**	gender	0.015
years of schooling	0.045	years of schooling	-0.071**
l_r_hhincome_lab_gov_yr_pa	-0.090***	log hh income	-0.013
children	0.034	children	-0.028
<i>N</i>	5666	<i>N</i>	11248
Standardized beta coefficients		Standardized beta coefficients	
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 22a is based only on rural stayers and rural-urban movers, while table 22b is based only on urban stayers and urban-rural movers. Standard errors are clustered at the household level and cross-sectional weights are applied. These tables report standardized beta coefficients from a multivariate linear probability model.

Rural-Urban Wage Gaps. Table 23 reports the cross-sectional wage gap.

Table 23. Cross-sectional rural-urban wage gap in South Africa

	urban	urban	urban
	0.727***	0.724***	0.450***
	(0.0428)	(0.0421)	(0.0367)
controls	No	No	Yes
time effects	No	Yes	Yes
<i>N</i>	15853	15853	15815
R^2	0.089	0.100	0.299

The outcome variable is log household labor income per adult broadly defined. Controls include age, years of school, sex, log household size, and year-sector fixed effects. Standard errors in parentheses are clustered at the household level. The second column includes only households who start in rural, and if migrated, never returned to rural. Standard errors in parentheses are clustered on the household level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 24. Within household rural-urban wage gap in South Africa

	all movers	rural-urban movers	urban-rural movers
urban	0.216***	0.340***	0.213*
	(0.0606)	(0.0830)	(0.113)
<i>N</i>	13142	6081	8041
R^2	0.806	0.751	0.800

The outcome variable is log household labor income per adult broadly defined. The second column includes only households who start in rural, and if migrated, never returned to rural. Standard errors in parentheses are clustered on the household level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls include year, and month fixed effects.

Consistent with the prior literature, the unadjusted wage gap is very large, and over 100%. After controlling for observables including age, education, and gender, the wage gap shrinks but remains substantial at around 30 log points.⁵⁶

⁵⁶The cross-sectional rural-urban wage gap in South Africa is somewhat smaller than reported in Lagakos et al. (2020). The discrepancy is entirely accounted for by the fact that I include race as a control variable.

Selection into Migration. Table 25 shows how rural-urban and urban-rural migrants are selected relative to the relevant population of stayers by running a simple bivariate linear regression before migration happens. Consistent with his findings, rural-urban movers tend to be more skilled. Somewhat surprisingly, selection on income is actually negative, and rural-urban migrants are much younger than rural stayers. This type of selection contrasts with urban-rural migrants, that are substantially older, and are negatively selected on assets but not on income. The massive age gaps reveals a potentially different migration motive across the two mover types, where retirement motives may loom larger for urban-rural migrants. Income-based migration stories appear more appropriate for young rural-urban migrants. This perspective is consistent with larger wage gains upon migration, and justifies the focus on rural-urban migration pursued in this paper.

Table 25. Selection into Migration in South Africa

	Age	Years of education	Log total net assets	Log income
Rural-Urban mover dummy	-6.025*** (0.642)	0.973*** (0.322)	-0.167 (0.302)	-0.276*** (0.0774)
Rural stayers baseline	38.76*** (0.308)	9.519*** (0.144)	10.71*** (0.0753)	9.880*** (0.0337)
<i>N</i>	6410	6409	1233	5667
<i>R</i> ²	0.034	0.005	0.000	0.007

	Age	Years of education	Log total net assets	Log income
Urban-Rural mover dummy	-0.523 (1.708)	-1.437*** (0.413)	-1.274*** (0.478)	-0.102 (0.131)
Urban stayers baseline	40.33*** (0.373)	11.48*** (0.102)	11.69*** (0.0898)	10.52*** (0.0400)
<i>N</i>	12922	12923	2961	11256
<i>R</i> ²	0.000	0.006	0.004	0.000

(a) Rural stayers vs. rural movers

(b) Urban stayers vs. urban movers

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 26a is based only on rural stayers and rural-urban movers, while table 26b is based only on urban stayers and urban-rural movers. Standard errors are clustered at the household level and cross-sectional weights are applied.

To further investigate the role of selection, I run a similar linear regression but now controlling for age for rural-urban migrants, which varies drastically across groups. Table reports the results 26. Strikingly, the positive link between education and migration completely disappears, while the negative selection on income remains large and highly statistically significant. The point here is that sorting is clearly important but casts doubt on the preconception that permanent rural-urban migration is predominantly driven by migration of more skilled individuals into the skill-intensive sectors, as argued in Young (2013).

Table 26. Selection into Migration in South Africa with Age Controls

	Years of education	Log total net assets	Log income
Rural-Urban mover dummy	0.115 (0.310)	-0.0770 (0.301)	-0.248*** (0.0795)
<i>N</i>	6409	1233	5667
<i>R</i> ²	0.120	0.006	0.010

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 26 is based only on rural stayers and rural-urban movers. I include age as a linear control variable. Standard errors are clustered at the household level and cross-sectional weights are applied.

Link between Rural and Urban Income among Migrants. Table 27 reports the results of a regression of urban on rural income among rural-urban movers. The first two columns report the estimates for movers, the two final columns report the estimates for rural stayers as placebo test. Consistent with the model, the IV estimate for movers of .58 is higher than the OLS estimate. This estimate is still substantially smaller than the IV estimate for stayers, which is around .82. It appears that income for households that stay in the same sector is more persistent than income for movers, and the IV strategy rules out simple measurement error as explanation.⁵⁷

⁵⁷Note that even the IV estimator for stayers is expected to be smaller than one due to imperfectly transitory shocks.

Table 27. Correlation across rural and urban income in South Africa

	OLS (movers)	IV (movers)	OLS (stayers)	IV (stayers)
rural labor income	0.279*** (0.0517)	0.583*** (0.186)	0.414*** (0.0301)	0.817*** (0.0626)
<i>N</i>	268	170	3011	2026
<i>R</i> ²	0.429	0.424	0.425	0.338

I regress income after the move on income before the move in column 1 and 2, where the second column uses lagged income as instrument. Column 3 and 4 run the same regression for stayers as placebo test. I implement the regression on the raw data to keep the sample size up, and control for the same observables used when residualizing the income data. Cross-sectional weights are applied, and standard errors in parentheses are clustered on the household level, * p<0.10, ** p<0.05, *** p<0.01.

Income Gains, Consumption Gains, and Volatility. Without controls, the probability of increasing one's income is about 50% after taking out time effects. The probability of an income gain increases substantially around the time of migration by 25%. Controlling for observables leads to an estimate whereby migration increases the probability of an income gain by about 20%. Similar results hold for consumption, see table 29. Thus, while rural-urban migration raises income on average, a substantial fraction of households – roughly 30%– experience income losses after migrating.

Table 28. Migration and Income Drop

	better	better
rural-urban move	0.254*** (0.0398)	0.208*** (0.0419)
constant	0.574*** (0.0121)	0.563*** (0.0169)
Time FE	Yes	Yes
Age controls	No	Yes
HH size	No	Yes
Observations	3480	3480
r ²	0.0172	0.0314

This table reports results from a linear probability model where the left hand side variable takes a value of one if households are better off, i.e., they improved their total per adult income relative to the last period. Standard errors are clustered on the household level, and cross-sectional panel weights are applied.

Table 29. Migration and Consumption Drop

	better_con	better_con
rural-urban move	0.302*** (0.0351)	0.237*** (0.0373)
constant	0.533*** (0.00998)	0.525*** (0.0124)
Time FE	No	Yes
Age controls	No	Yes
HH size	No	Yes
Observations	4084	4084
r ²	0.0339	0.0593

This table reports results from a linear probability model where the left hand side variable takes a value of one if households are better off, i.e., they improved their total per adult consumption expenditure relative to the last period. Standard errors are clustered on the household level, and cross-sectional panel weights are applied.

The first two columns in table 30 report the results on income growth and volatility. Reassuringly, mean

growth is positive, consistent with the sizable urban income premium estimated earlier.⁵⁸ Crucially, income volatility also spikes around the time of migration, consistent with risky rural-urban migration. The standard deviation of income growth relative to the mean is overall larger ($SD = \sqrt{0.165} \approx .4$ vs $mean \approx 0.3$). At the same time, income volatility is large even for rural households who don't migrate ($(SD = \sqrt{0.12} \approx .35)$). Importantly, the structural model will go beyond simple volatility measures and decompose volatility into different kinds of risk, i.e., transitory vs. persistent shocks. Transitory shocks, consistent with earlier results, are key drivers of the overall high degree of volatility. In the case of permanent rural-urban migration, persistent shocks will be the quantitatively dominant force that generates the differences in overall volatility. Consistent with this are differences in mean consumption and consumption volatility across groups, which are of a similar magnitude and a tell-tale sign of persistent shocks, see column 3 and 4 of 30.

Table 30. Migration and Income Volatility

	Income growth	Income volatility	Consumption growth	Consumption volatility
rural-urban move	0.323*** (0.0437)	0.0478* (0.0246)	0.278*** (0.0390)	0.0349 (0.0224)
constant	0.0314*** (0.00901)	0.120*** (0.00681)	0.00934 (0.00883)	0.125*** (0.00651)
Observations	3480	3480	4084	4084
R-squared	0.0751	0.0260	0.0983	0.0442

This table reports annualized log income and consumption growth regressed on a dummy for rural-urban migration, together with the analogous regressions using the squared residual from the growth regression as the dependent variable — a demeaned measure of growth volatility that isolates dispersion around each group's own mean from the mean-shift effect itself. All columns control for household size and a quadratic in age (both mean-centered) and include year fixed effects. Standard errors are clustered at the household level, and panel weights are applied. Table 35 in the appendix reports moments without controls, which are slightly larger. Estimates are annualized, i.e., I use $(y_{t+2} - y_t) / 2$ to compute annualized mean and volatility.

Rural-Urban Reclassification. Lastly, I condition on households moving into different counties. There is a concern to what extent the rural-urban differences are confounded by artificial reclassification. I now tackle this issue by reconsidering who is a rural-urban mover based on whether i) their rural-urban status is changing, and ii) whether the county of their household is changing. If only i) is the case but not ii), I assume that this household is misclassified and assume that the previous urban-rural status continues to apply.⁵⁹

⁵⁸ Annualized income growth using first-differences and a dummy that equals one only around the time of migration implies a larger urban premium than what the standard fixed effects estimator suggested, which wasn't annualized. This is consistent with negative selection on transitory shocks, which provides a boost around the time of migration that vanishes when households income gaps are based on long-term averages in each sector as done with the fixed effects estimator. Similarly, the volatility of stayers closely matches the volatility implied by the stochastic process estimated earlier. To see this, use the stochastic process to derive volatility of the growth rate, and plug in numbers close to the earlier estimates to illustrate that the implied income volatility estimate is surprisingly close, given differences in selection criteria and a simpler residualization procedure: $\mathbb{V}(\Delta y/2) = \sigma_p^2/2 + (1 - e^{-2\lambda\epsilon})\sigma_\epsilon^2/2 + \sigma_x^2/2 \approx 0 + (1 - e^{-1}) * 0.26/2 + .1/2 \approx .13$.

⁵⁹ Note that the county identifier is useful but unfortunately does not map one for one into rural-urban status in that there are many counties that have households classified as both urban and rural.

Table 31. County-to-county rural-urban migration flows in South Africa

(a) Full Sample		(b) Selected Sample	
	(1) Share		(1) Share
Rural stayer	0.264	Rural stayer	0.242
Urban stayer	0.626	Urban stayer	0.758
Rural-Urban mover	0.067	Rural-Urban mover	0.000
Urban-Rural mover	0.026	Urban-Rural mover	0.000
Two-way mover	0.018	Two-way mover	0.000
Observations	26245	Observations	4712

In table 31a I do not select the sample other than dropping observations without urban-rural status. In table 31b I apply the same selection criteria defined earlier.

Table 32. County-to-county migration and cross-sectional differences in South Africa

(a) Rural stayers vs. rural movers						(b) Urban stayers vs. urban movers					
	Age	Years of education	Log total net assets	Log income	Log asset-income ratio		Age	Years of education	Log total net assets	Log income	Log asset-income ratio
Rural-Urban mover dummy	-13.74*** (0.635)	2.112*** (0.223)	-0.0311 (0.250)	-0.648*** (0.0598)	0.315 (0.268)	Urban-Rural mover dummy	0.812 (2.082)	-1.614*** (0.449)	-1.682** (0.780)	-0.100 (0.170)	-1.525** (0.727)
Rural stayers baseline	38.27*** (0.347)	9.107*** (0.120)	10.69*** (0.0628)	9.617*** (0.0305)	-0.316*** (0.0573)	Urban stayers baseline	40.27*** (0.369)	11.47*** (0.101)	11.68*** (0.0892)	10.51*** (0.0395)	-0.167** (0.0667)
N	10932	10932	1856	9658	1856	N	13001	13002	2975	11329	2975
R ²	0.101	0.025	0.000	0.028	0.001	R ²	0.000	0.005	0.003	0.000	0.004

This table refers to the full sample but only considers households the first time they appear in the sample to zoom in on cross-sectional differences before migration. Table 32a is based only on rural stayers and rural-urban movers, while table 32b is based only on urban stayers and urban-rural movers.

Table 33. County-to-county migration and cross-sectional differences in South Africa (standardized)

(a) Rural stayers vs. rural movers		(b) Urban stayers vs. urban movers	
	(1) Rural-Urban Migrant		(1) Urban-Rural Migrant
age	-0.249***	age	0.025
gender	0.059**	gender	0.008
years of schooling	0.052**	years of schooling	-0.079**
log hh income	-0.130***	log hh income	0.016
children	0.050**	children	-0.023
N	9656	N	11321
Standardized beta coefficients		Standardized beta coefficients	
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 33a is based only on rural stayers and rural-urban movers, while table 33b is based only on urban stayers and urban-rural movers. Standard errors are clustered at the household level and cross-sectional weights are applied. These tables report standardized beta coefficients from a multivariate linear probability model.

Table 34 uses household fixed effects with and without additional controls in line with the previous specification. Note that the size of the coefficient increases.

Table 34. County-to-county within household rural-urban wage gap in South Africa

	all movers	rural-urban movers	urban-rural movers
urban	0.342*** (0.0609)	0.421*** (0.0815)	0.320** (0.133)
controls	Yes	Yes	Yes
<i>N</i>	22216	10719	11108
<i>R</i> ²	0.759	0.673	0.771

These results consider a household a mover only if their rural-urban switch coincides with a switch of the household's district council. The outcome variable is log household labor income broadly defined. The second column includes only households who start in rural, and if migrated, never returned to rural. Standard errors in parentheses are clustered on the household level, * p<0.10, ** p<0.05, *** p<0.01. Controls include year, and month fixed effects. In the appendix I report estimates controlling for a richer set of covariates including household size and age, leading to larger wage gaps.

Income and Consumption Volatility of Migrants without Controls. Table 35 reports income and consumption mean gains and volatility (as annualized variance) without controls, which complements the results in the main body of the paper.

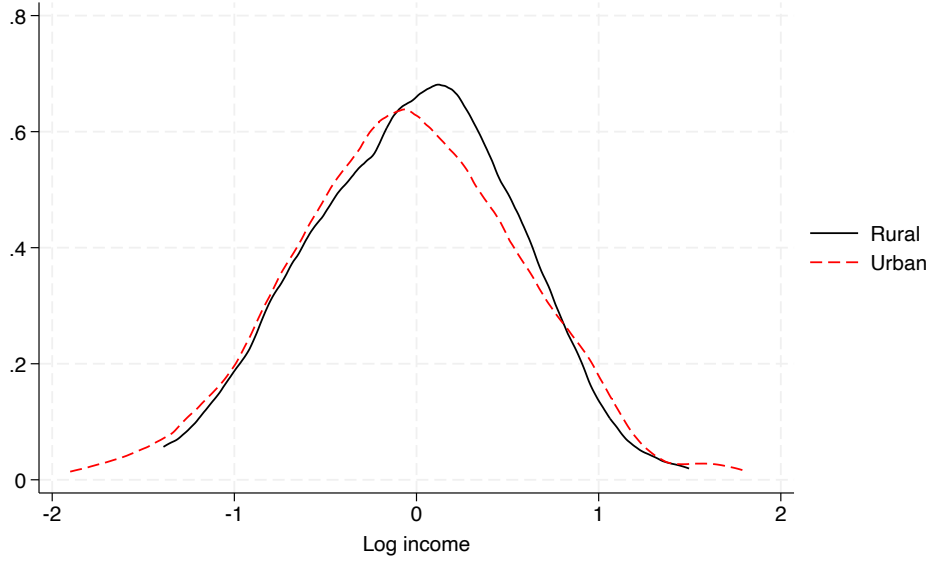
Table 35. Migration and Income Volatility (no controls)

	Income growth	Income volatility	Consumption growth	Consumption volatility
rural-urban move	0.367*** (0.0460)	0.0694*** (0.0265)	0.337*** (0.0373)	0.0491** (0.0243)
constant	0.0469*** (0.00687)	0.129*** (0.00565)	0.0202*** (0.00676)	0.130*** (0.00545)
Observations	3480	3480	4084	4084
R-squared	0.0534	0.0235	0.0644	0.0410

This table reports annualized log income and consumption growth regressed on a dummy for rural-urban migration, together with the analogous regressions using the squared residual from the growth regression as the dependent variable — a demeaned measure of growth volatility that isolates dispersion around each group's own mean from the mean-shift effect itself. All columns control for year fixed effects. Standard errors are clustered at the household level, and panel weights are applied.

Tails of the income distribution. The income process gives rise to a fat-tailed stationary income distribution where persistent shocks lead to a fanning out of the distribution while death shocks stabilize it. The tail-coefficient of the persistent income component p equals $\sqrt{\frac{2\phi}{\sigma_p^2}}$. To explore the fit of the model in this dimension, I plot the density of the average log residualized household income for rural and urban in figure 14.

Figure 14. Density of avg. log residualized household income South Africa



It is easy to see that the tail of the distribution of log income of urban households is fatter than the one of rural households, consistent with differences estimated in the volatility of persistent shocks.

I explore this more formally by running a regression of the log of the complementary CDF, which is a log linear function of income according to $Y^{-\zeta}$ where ζ is the tail coefficient. Table 36 reports the results of a regression that allows for intercept and slope coefficient to vary across sectors.

Table 36. Tail Coefficients

	South Africa	
	Estimate	SE
ζ_R	-3.60	0.115
ζ_U	-2.80	0.072

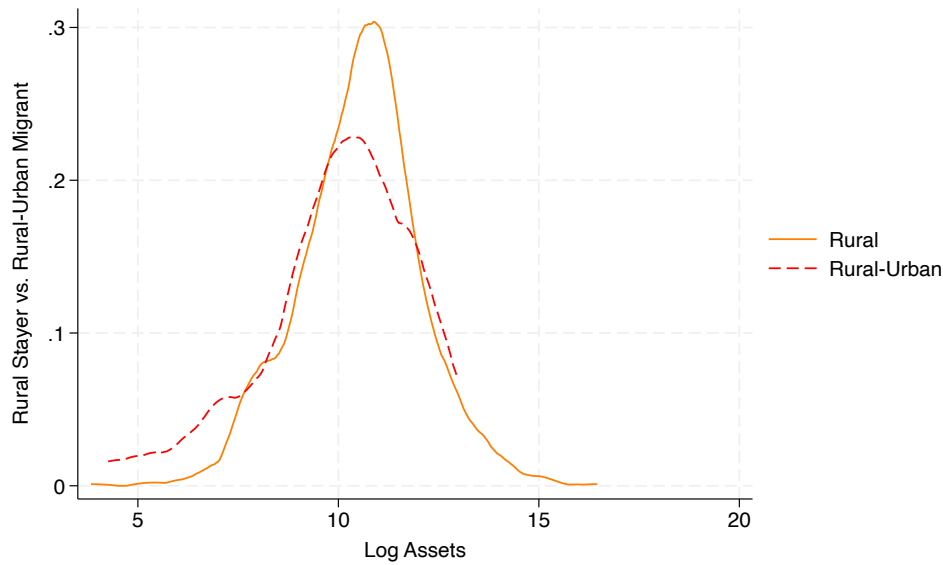
The income measure uses residualized log labor income broadly defined and averaged over different years. The bivariate regression is run separately for rural and urban stayers after dropping observations below the 80th percentile of income to only capture the right tail of the distribution.

The theory implied tail-coefficients that follow from the formula $\zeta = \sqrt{\frac{2\phi}{\sigma_p^2}}$ are largely in line with the results and consistent with the fatter tail in the urban sector, which is an untargeted moment.

Wealth and Migration. Before I study the nexus between wealth and migration, I stress that the data quality is somewhat weaker. Wealth is only measured irregularly, and at times in coarse intervals. Overall, the share of households with zero wealth seems inflated, also compared to other household level data in developing economies, see De Magalhaes and Santaaulalia-Llopis (2018) for relevant comparison.

Figure 15 plots a kernel density estimate of the asset distribution amongst households with non-zero assets. It focuses on rural households and compares rural stayers with future rural-urban movers. The plot shows that migrants have more mass on relatively low asset levels, consistent with the negative selection on wealth in the model.

Figure 15. Density of log assets in South Africa of rural stayers and migrants



When computing this Kernel density, I use average household assets across different waves, and use cross-sectional weights to obtain this average.

Additional Empirical Results. Perhaps unsurprisingly, earned income is more variable than disposable income, and especially so for rural households. Consistent with that, table 37 reports the share of remittances and government transfers in total disposable monthly household income, which helps account for these differences as both shares are higher for rural households. Quantitatively, government support is the more important factor constituting, on average, 13% of rural household income, in contrast to 7% for urban households. **Empirical Passthrough Income to Consumption.** Table 38 reports the empirical passthrough

Table 37. Income Shares in South Africa

	(1) Mean b
c.share_gov@0.urban	0.264
c.share_gov@1.urban	0.103
c.share_remit@0.urban	0.037
c.share_remit@1.urban	0.024
<i>N</i>	25934

Total disposable monthly household income is in the denominator. Government support includes pensions, but note that the sample is restricted for household heads between 25 and 60 years of age.

coefficients for South Africa.

For the Townsend-style regression I use year-month-sector fixed effects as well as a set of controls including education, a second order polynomial in age, log household size, and log number of adults. The first column uses a simple cross-sectional regression without household fixed effects so I include gender and race as additional controls. The second column adds household level fixed effects and is the one reported in the main text. Note that I use the total sample to implement this regression, which is complementary to the BPP results where I impose more stringent sample restrictions. Consistent with a large literature,

the passthrough coefficients from the simple regression are low. If I had village level outcomes and controlled for those, the estimates would probably be even lower. Note, however, that this does not mean that markets are complete in rural. Correlated shocks, like bad weather, together with measurement error and mean-reverting income processes would produce very low passthrough coefficients as I have shown when simulating the structural model.

The second table reports BPP passthrough coefficients for persistent shocks. This is a more interesting object than the passthrough for transitory shocks, which tends to be low in the data and in the structural model since transitory shocks are easy to insure. As usuals, measurement error and fully transitory shocks are not separately identified, so again it will be true that high measurement error would trivially lead to a low passthrough for transitory shocks. Note, however, that the BPP estimator for persistent shocks is robust to fully transitory shocks in the sense that these are partialled out. To implement the BPP estimator, I run an IV regression using the moment restrictions in the main text. I use the same restrictions as for the exercise when I measure income risk, and I residualize the data in the same way. I impose one additional restriction, which is that residualized income and consumption growth across waves is bounded by -100% and $+250\%$. While imposing additional restrictions on these growth rates is standard, I note that this matters and imposing no restrictions leads to nosy regressions with null results.

Table 38. Passthrough in South Africa

(a) β_{c-y}			(b) β_{c-p}^{BPP}		
	Log consumption	Log consumption		Rural	Urban
Log income	0.260*** (0.0141)	0.147*** (0.0144)	$\hat{\beta}$	0.224 (0.258)	0.621* (0.322)
Log income $\times$ 1 _{urban}	0.133*** (0.0212)	0.0534*** (0.0186)	N	385	610
Controls	Yes	Yes			
Year-month-sector fe	Yes	Yes			
Individual fe	No	Yes			
N	20878	20055			
R^2	0.658	0.815			

A.3 Estimating income risk

I build on the methodology developed in BPP. Consider the fix proposed in Crawley, Holm, and Tretvoll (2026) and discuss how things relate to level estimation and choice of weighting matrix, see also Heathcote, Perri, and Violante (2010) using levels. Note that a balanced sample is important according to Daly, Hryshko, and Manovskii (2022) since income volatility different for people who enter and exit out of ordinary frame.

I set up the model in continuous time, which is useful for a number of issues. First, it allows me to go back and forth between different time horizons, which comes in handy as the Chinese data is measured on an annual basis, i.e., income per year, but only reported every two years, while the South African data is measured monthly and the gap between households in different waves ranges from one to two and a half years. Second, it directly addresses issues of time aggregation as pointed out in the classic study of Working (1960). This matters, as recently pointed out by Crawley, Holm, and Tretvoll (2026), who show how to reconcile income risk estimates from first differenced vs. level specification.

A.3.1 Stochastic Income Process

I postulate an exogenous income process for residualized log income

$$y = p + \epsilon$$

where p is the persistent component of human capital and ϵ is a transitory component

$$\begin{aligned} dp &= \sigma_p dZ \\ d\epsilon &= dN (\epsilon' - \epsilon). \end{aligned}$$

The income process combines a random walk in p with a passing shock whereby transitory independent shocks ϵ hit the household at arrival rate λ_ϵ . Formally, Z is a Wiener process, dN is a Poisson process with intensity λ_ϵ , ϵ is drawn i.i.d. from some CDF $G(\epsilon)$. Additionally, I assume measured income y features classical measurement error x so

$$y^* = p + \epsilon + x.$$

Level moments. I focus on level moments, which are shown to be more robust according to Crawley, Holm, and Tretvoll (2026). There are also more level moments than differenced moments, which helps given the relatively short samples I work with. Denote residualized log income of household i as $\mathbf{y}_i^* = [y_{i,0}^* \ y_{i,1}^* \ \dots \ y_{i,T}^*]'$ and note that the Covariance matrix reads

$$V = \mathbb{E}_i [\mathbf{y}_i^* \mathbf{y}_i^{*'}]$$

since the means are zero by construction after residualizing income. I now derive the elements of V given the postulated income process focusing first on the variance term on the diagonal

$$\mathbb{E} [\hat{y}_0 \hat{y}_0] = \mathbb{V}(p_0) + \sigma_\epsilon^2 + \sigma_x^2$$

and more generally for $t \in [0, T]$

$$\mathbb{E} [\hat{y}_t \hat{y}_t] = \mathbb{V}(p_0) + t\sigma_p^2 + \sigma_\epsilon^2 + \sigma_x^2.$$

The covariance between time t and $t+k$ equals

$$\begin{aligned} \mathbb{E} [\hat{y}_t \hat{y}_{t+k}] &= Cov \left(p_t + \epsilon_t + x_t, p_t + \int_t^{t+k} \sigma_p dZ + \epsilon_{t+k} + x_{t+k} \right) \\ &= \mathbb{V}(p_t) + Cov(\epsilon_t, \epsilon_{t+k}) \\ &= \mathbb{V}(p_t) + \mathbb{P}(N_{t,t+k} = 0) \sigma_\epsilon^2 \\ &= \mathbb{V}(p_0) + t\sigma_p^2 + e^{-\lambda_\epsilon k} \sigma_\epsilon^2 \end{aligned}$$

where measurement error drops out conveniently for $k > 0$.

This continuous time approach is not fundamentally different from the discrete time version but it offers two additional benefits. First, it can handle data where time intervals between observations are non-constant and household specific. For example, for the South African data there are observations of monthly income in March 2010, and the same household's income shows up again in February 2012. For yet another

household, I observe their monthly income in August 2010 and June 2012. The continuous time setup can directly account for household specific time intervals between different waves of the panel. In the South African context with household specific time intervals across waves, I compute the theoretical covariance, for concreteness for a specific moment $m_{t,t+2} = \mathbb{E}_i [y_{i,t} y_{i,t+2}]$ as

$$m_{t,t+2}^{\text{theory}} = \frac{1}{N_{t,t+2}} \sum_i \mathbb{V}(p_{t_i}) + e^{-\lambda_\epsilon k_i} \sigma_\epsilon^2 \text{ for } t_i \in W_t \text{ and } t_i + k_i \in W_{t+2}$$

where W_t is a particular wave of the survey which appear at biennial frequency. For example, wave 1 appears in 2008 and wave 2 in 2010. The same household shows up in February in the first wave, and in August in the second, which I need to account for using the previous equation. Other theoretical moments are analogous. I emphasize that I treat θ as a theoretical moment because no information on income or consumption from the data is used. It does, however, use information from the data on when exactly households are surveyed. The asymptotics are thus best viewed as the limit of the average of repeated experiments where the time intervals on when households are surveyed are fixed across experiments.

GMM. I stack these moments into the vector $g(w_i, \theta)$ so that $\mathbb{E}[g(w_i, \theta)] = 0$ where θ is a vector of parameters that shape the household income process and w_i is random vector representing the data. I use the general method of moments (GMM) or the simulated method of moments (SMM) and estimate θ separately for urban and rural households. The standard quadratic form reads

$$\min_{\theta} \left[\sum_i g(w_i, \theta) \right]' \hat{\Omega} \left[\sum_i g(w_i, \theta) \right]$$

where Ω is the optimal weighting matrix, $\Omega = (\mathbb{E}[g(w_i, \theta) g(w_i, \theta)'])^{-1}$, and the hat indicates sample moments.

The first order condition reads

$$\left[\sum_i \nabla_{\theta} g(w_i, \hat{\theta}) \right]' \hat{\Omega} \left[\sum_i g(w_i, \theta) \right] = 0$$

and the asymptotic standard errors follow

$$\sqrt{N} (\hat{\theta} - \theta_0) \sim N \left(0, (\hat{G}' \hat{\Omega} \hat{G})^{-1} \right)$$

where $\hat{G} = \frac{1}{N} \sum_i \nabla_{\theta} g(w_i, \hat{\theta})$.

To ensure that variances are always positive, I have applied an exponential transformation, i.e., $\sigma_p^2 = e^{\tilde{\sigma}_p^2}$ etc. The exponential helps in that it ensure that my estimate for σ_p^2 is always weakly positive. Writing in general form, let $\tilde{\theta} = h(\theta)$ where $h : R^k \rightarrow R^k$ is an invertible function. The asymptotic variance of $\tilde{\theta}$ then follows from the delta method

$$AVar = N \left(0, h_{\theta} (\hat{G}' \hat{\Omega} \hat{G})^{-1} h'_{\theta} \right)$$

where h_{θ} is the Jacobian of h .

A.4 Deaton-Paxson Exercise

I next describe in more detail how I implemented the Deaton-Paxson exercise, which relates the rise in life-cycle inequality to the degree of persistent income shocks. I report results for several low and middle income countries in Latin America, Africa, and Asia, and the United States as benchmark, using harmonized data from the Luxembourg Center of Income studies (Luxembourg Income Study (LIS) Database (2025)).

Sample Selection. We impose the following aggregate selection criteria

1. Drop households with zero or negative income
2. Drop household with missing rural identifier – we use rural vs. non-rural as a distinct urban category is not always available
3. Drop waves of datasets where the number of observations in a sector is below 2200

We then further restrict the sample based on household head characteristics. Before we do so, we have to determine the household head. For some surveys the household head is provided, but often not highly correlated with income and prime age. I therefore use the following algorithm to determine the household head for all surveys:

1. Pick adult male. If there are several, pick the male closest to prime age of 45 years. Use: $|\text{age} - 45|$.
 - (a) If there are several with same distance to prime age, pick the one with higher education. If still tie, pick the one with lower individual id to break the tie.
 - (b) If there is no male, apply the same procedure to adult women

Having defined the head, I then further restrict the sample using the following characteristics

1. Age in $[26,60]$
2. Household size < 10
3. Available information on number of children (strictly younger than 18 years)
4. Keep households in the sample if the head is employed using the variable *emp*, which stands for employment and excludes households with a head that is unemployed, still in school, or retired. A measurement challenge could arise if employment is not well-defined in the rural sector due to the informality of work. However, after checking for employment rates across sectors, we did not find that this definition lead to artificially low employment in the rural sector.
5. Censor lowest 2% and highest 1% of income within each year and sector – I run this on the entire working sample, i.e., before I drop observations for other reasons.

I test whether the datasets provide sensible results regarding the returns to education, returns to experience/age, and cross-sectional rural-urban wage gaps, which is the case. See the code for more details.

Inequality over the life cycle. I next compute the log variance of income by sector-year-age bin.

I construct age bins using non-overlapping 5-year age intervals $[26,27,28,29,30]; [31,32,33,34,35]; \dots; [56,57,58,59,60]$ where the age in the middle is the age index of the age group as convention, i.e., the age index of the first group is 28. I next define the cohort of an age group as the current year minus the age index, i.e., in 2010 for the first age group the cohort would be $2010 - 28 = 1982$.

I use measures of real income comparable across time and countries to compute household equivalent income using a simple equivalent scheme, $Y^{\text{eqv}} = \frac{Y}{\text{adults} + .5 * \text{children}}$, i.e., children get a weight of half when computing an equivalent measure of income per capita where children are defined as household members below the age of 18.

I then construct year-age index-sector cells and compute the log variance of income for each cell using both raw income and adult-equivalent income

$$\sigma_{a,t,j}^2 = \frac{\sum_{i \in a,j,t} X_i^2}{N_{a,j,t}} - \left(\frac{\sum_{i \in a,j,t} X_i}{N_{a,j,t}} \right)^2, \quad (26)$$

where a is the age of the household head and t is calendar time and j is the sector which is rural or non rural. When doing this, we only consider cells for which we have at least 50 observations, $N_{a,t} \geq 50$,⁶⁰ and we use household weights to construct first and second moment.

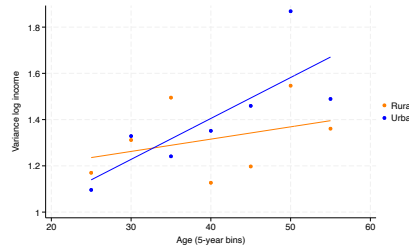
Time effects. I then run a linear regression of the form

$$\sigma_{a,t,j}^2 = \alpha + \gamma_t + \delta_a + u_{a,t,j}$$

where γ are time and age index fixed effects.⁶¹

Figure 16 illustrates the regression results graphically. The key takeaway is that urban income fans out at a significantly higher rate than rural income, consistent with the GMM estimates from the panel data. Since I did not residualize the data, inequality measured as the variance of log, and the rate at which inequality increases, are naturally larger. The GMM results thus constitute a conservative lower bound on the difference in permanent income risk across sectors.

Figure 16. Life Cycle Inequality in South Africa



The outcome variable is the variance of equivalent log household labor income broadly defined. To compute equivalent units, I assign children a weight of .5 while adults receive a weight of 1.0. I normalize the intercept by setting the age effect at 35 equal to the sample mean of cross-sectional inequality for this age group. I drop the lowest 3% of observations, and the largest .5%, and focus on employed households of age 25-59.

Table 39 reports the results for South Africa based on a linear regression model, i.e., $\delta_a = \delta \cdot a$.

⁶⁰It turns out that this restriction has no bite.

⁶¹As pointed out in Hall (1971), time, cohort, and age effects are not separately identified. Since age effects are the ultimate object of interest as they capture the rise in inequality over time, researchers face a somewhat arbitrary choice of whether to include cohort or time effects. I report results with year effects, which led to more sensible estimates, consistent with arguments in Heathcote, Storesletten, and Violante (2005).

Table 39. Within household rural-urban wage gap in South Africa

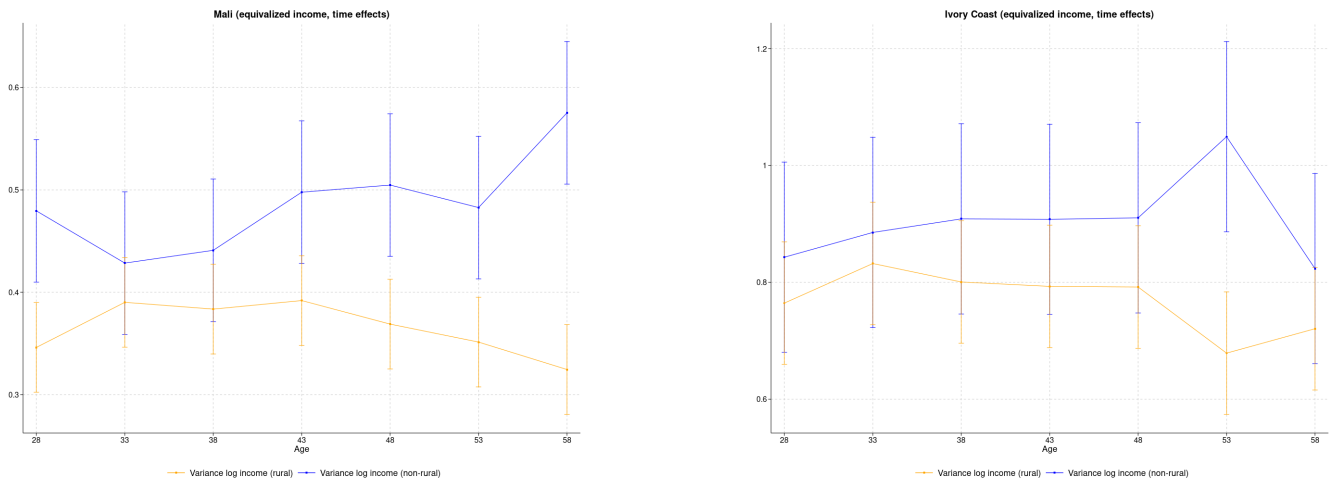
	variance log income
age	0.00532 (0.00406)
urban $\times$ age	0.0124* (0.00480)
time effects	Yes
<i>N</i>	70

The outcome variable is the variance of equivalent log household labor income broadly defined. To compute equivalent units, I assign children a weight of .5 while adults receive a weight of 1.0. I focus on households with working age heads that are in employment, and drop the lowest 3% and the highest .5%. Standard errors in parentheses are clustered on the age bin level, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Age bins are defined as non-overlapping 5-year intervals ([25,...,29],[30,...,34] and so on).

Relative to the GMM estimation, note that I impose virtually no sample restrictions and choose equivalized income, suggesting that my previous findings are not driven by sample selection criteria or using total household income. Nor is endogenous sample attrition a concern since the method does not rely on the panel structure of the data. Lastly, note that an additional benefit of the Deaton-Paxson approach is that it is readily implemented, which allows me to extend the analysis to Mali and Ivory Coast, two African developing economies with large rural sectors that have harmonized income measures provided by the Luxembourg Center for Income Studies (LIS). I show in the appendix A.4 that in both countries inequality rises faster in the urban sector, consistent with the findings from South Africa.

Figure 17 plots the evolution of life cycle inequality across age groups for Mali and Ivory Coast. In both countries, inequality in the rural sector fans out very little as households get older, in contrast to the urban sector, which technically is defined as non-rural in the LIS data. Focusing on prime-age households and comparing age group 33 and 53, we observe that cross-sectional inequality increases by about 10 log points in Mali and 20 log points in Ivory Coast, consistent with the dominant role of persistent shocks in the urban sector. Note that inequality is overall large here because I did not residualize the income data on observables. Additional details including robustness checks are deferred to the appendix, where I benchmark the results against life-cycle inequality in the US.

Figure 17. Life Cycle Inequality in Mali and Ivory Coast



95% confidence intervals are plotted around the age effects. To facilitate interpretation, I normalize the intercept by setting the age effect at 38 equal to the sample mean of cross-sectional inequality for this age group. Income is measured in adult equivalents.

Additional Results. I have also experimented with cohort effects instead of time effects, and I have extended the analysis to emerging markets in Asia and South America. Additional results are available upon request, which include exercises that use household income without adjusting for household size and number of children, and for a set of additional countries: South Africa (from LIS), India, China, Brazil, Colombia, Guatemala, Mexico, Paraguay, Peru, Vietnam. With the only clear exception being Vietnam, results are consistent with the ones presented in the paper in the sense that inequality increases relatively more in the urban sector as cohorts age.

A.5 BPP insurance coefficients

I consider the simplified model to derive the bias in the BPP estimator explicitly under the assumption that the income process combines a random walk and an imperfectly transitory shock. The simplified model in continuous time is summarized by the following equations

$$\begin{aligned} dc &= \beta_p dp + \beta_\epsilon d\epsilon \\ dp &= \sigma_p dZ \\ d\epsilon &= dN (\epsilon' - \epsilon) \end{aligned}$$

where the new transitory type ϵ' arrives at rate λ_ϵ , i.e., $\mathbb{E}[dN] = \lambda_\epsilon dt$.

Next, using the moment condition from BPP but adjusted for the biennial data (with 2-year intervals between observations so Δ also takes the difference over two years) I obtain

$$\begin{aligned} \frac{Cov(\Delta c_t, \Delta y_{t-2} + \Delta y_t + \Delta y_{t+2})}{Cov(\Delta y_t, \Delta y_{t-2} + \Delta y_t + \Delta y_{t+2})} &= \frac{Cov(\beta_p(p_t - p_{t-2}) + \beta_\epsilon(\epsilon_t - \epsilon_{t-2}), p_{t+2} - p_{t-4} + \epsilon_{t+2} - \epsilon_{t-4})}{Cov(p_t - p_{t-2} + \epsilon_t - \epsilon_{t-2}, p_{t+2} - p_{t-4} + \epsilon_{t+2} - \epsilon_{t-4})} \\ &= \frac{\beta_p(\mathbb{E}[\Delta p_t \Delta p_t]) + \beta_\epsilon(\mathbb{E}[\epsilon_t \epsilon_{t+2}] - \mathbb{E}[\epsilon_t \epsilon_{t-4}] - \mathbb{E}[\epsilon_{t-2} \epsilon_{t+2}] + \mathbb{E}[\epsilon_{t-2} \epsilon_{t-4}])}{\mathbb{E}[\Delta p_t \Delta p_t] + \mathbb{E}[\epsilon_t \epsilon_{t+2}] - \mathbb{E}[\epsilon_t \epsilon_{t-4}] - \mathbb{E}[\epsilon_{t-2} \epsilon_{t+2}] + \mathbb{E}[\epsilon_{t-2} \epsilon_{t-4}]} \\ &= \frac{\beta_p 2\sigma_p^2 + \beta_\epsilon 2\sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})}{2\sigma_p^2 + 2\sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})} \\ &= \beta_p \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})} + \beta_\epsilon \frac{\sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon} (1 - e^{-2\lambda_\epsilon})} \end{aligned}$$

where the second line uses the fact that the change in p is uncorrelated across non-overlapping intervals, e.g., $\mathbb{E}[\Delta p_t \Delta p_{t-2}] = 0$, and $\mathbb{E}[\epsilon_t \epsilon_{t+k}] = \mathbb{E}[\epsilon_t \epsilon_{t-k}] = \sigma_\epsilon^2 e^{-\lambda_\epsilon k}$. There is one final subtlety to note: for simplicity, I assumed in the derivation that $\mathbb{E}[\epsilon_{t+k} \epsilon_{t+k}] = \sigma_\epsilon^2 \forall k$. In the presence of migration, the statistical properties of realized income differ slightly from the postulated stochastic process because both the observed variance in the transitory component, and the correlation of transitory shocks over time, depend on selection into migration. Because migration is rare, the quantitative effect is negligible, and incorporated in the quantitative model when simulating the passthrough coefficients.

I have also explored versions of this bias with time aggregated data extending Crawley (2020)'s approach, which is very similar. Results available upon request.

B Computational Appendix

B.1 Solving the model

I build on the work of Achdou et al. (2022) and use a finite difference scheme to solve the model computationally. Two features are noteworthy. First, the setup is block-recursive in the sense that I can solve the urban problem without reference to the rural problem keeping the model tractable. Second, the rural problem features a non-linear discrete migration choice problem.

I focus on the fixed point problem for rural households, which is the key difference to the otherwise standard incomplete market model. It will use an iterative procedure to solve the household problem where n stands for a specific iteration and the algorithm stops when convergence up to some tolerance is achieved, i.e., $\|V_n - V_{n+1}\| < \epsilon$.

The system reads

$$(\rho + \phi + \lambda) V^{n+1} = u(c^n) + \lambda \tilde{V}_m^n + \mathcal{A}^n V^{n+1} + \frac{V^{n+1} - V^n}{\Delta} \quad (27)$$

$$u'(c^n) = \partial_b V^n \quad (28)$$

$$\pi^{\mathcal{I}} = \frac{(\mathbb{E}[V_U^{\mathcal{I}}] \cdot (-1))^{\frac{s}{1-\gamma}} \tau_{RU}^{-s}}{(V_R \cdot (-1))^{\frac{s}{1-\gamma}} + \mathbb{E}[V_U^{\mathcal{I}}] \cdot (-1)^{\frac{s}{1-\gamma}} \tau_{RU}^{-s}} \quad (29)$$

$$\tilde{V}_m^n = \zeta \tilde{V}_m^{pi,n} + (1 - \zeta) \tilde{V}_m^{fi,n} \quad (30)$$

where the value function depends on the triple of idiosyncratic states $x = (b, p, \epsilon)$. Note that the non-linear part $\tilde{V}_m^n$ has an n subscript, while the linear part has an $n + 1$ subscript.

I ensure $a_U \geq 0$ by imposing $V_U(a_U < 0, p_U, \epsilon_U) \rightarrow -\infty$, which is to say that households that cannot afford the monetary cost of migration will not migrate. One can show that $\hat{V} = V_R$ for households that cannot afford the cost of migration, and similarly the probability of migration converges to zero in that case, both of which are desirable implications of Frechet-distributed taste shocks.

To implement this computationally on a finite grid, I use a migrant's current normalized asset position $\tilde{a}$, and compute the normalized asset position in urban $\frac{b-f_{RU}}{\omega\Lambda}$. The reason this expression features the wage gap is that the actual urban assets of a migrant are $(b - f_{RU}) w_R$. When normalized by the effective urban wage facing a migrant, $w_U \Lambda$, this asset position equals $(b - f_{RU}) \frac{w_R}{w_U \Lambda} = \frac{(b-f_{RU})}{\omega\Lambda}$. Given the finite grid, it is likely that migrants' urban asset position is positioned between the rural grid points. In that case, I interpolate and assume that households are randomly allocated to adjacent asset grid points (in the level space, not log space).

To solve the model computationally, I use lemma 1 to write the system in normalized form, i.e., with wage normalized to unity on a normalized asset grid that is defined relative to the wage. In doing so I keep the rural-urban wage gap fixed, and iterate on it in the outer loop as described in the computational algorithm sketched out next.

Algorithm Bird's eye view.

1. Solve the urban household problem v^U . Note that urban wage is immediately implied by $w_U = (1 - \theta) \left(\frac{\theta}{r + \delta_k} \right)^{\frac{\theta}{1-\theta}} A_U$.
2. Since I have a solution for v_U and w_U I can solve the rural household problem v_R , which depends on the expected value of moving to urban. This, in turn, is a function of the rural-urban wage gap ω . Guess a wage gap ω , and solve the rural household problem inclusive of the migration choice.

3. The rural solution implies endogenous migration rates $\pi(x)$. To proceed, I solve for the stationary rural distribution g_R .
4. I can now check whether rural-urban migration given by

$$M = \frac{\Pi_{RU}}{\phi\delta_{UR} + \Pi_{RU}},$$

is consistent with labor market clearing using

$$\frac{w_U}{w_R} = \frac{(1-\theta) \left(\frac{\theta}{r+\delta_k} \right)^{\frac{\theta}{1-\theta}} A_U}{(1-\chi-\nu) \left(\frac{\nu}{r+\delta_k} \right)^{\frac{\nu}{1-\nu}} \left((1-M) \tilde{h}_R \right)^{-\frac{\chi}{1-\nu}}} \Rightarrow$$

$$1-M = \frac{1}{\tilde{h}_R} \left(\frac{\omega}{w_U} (1-\chi-\nu) \left(\frac{\nu}{r+\delta_k} \right)^{\frac{\nu}{1-\nu}} \right)^{\frac{1-\nu}{\chi}}.$$

When this is not the case, I update the wage gap ω . Specifically, I increase the wage gap whenever the migration-implied urban share is larger than the share implied by labor market clearing. I then return to step 2) and iterate up until I found a rural-urban wage gap that is consistent with labor market clearing and endogenous migration. I use Newton's method to find the equilibrium wage gap ω more quickly.

5. Having solved for the equilibrium wage gap, I combine the solution to the urban households problem together with the exogenous income process and the endogenous migration rates to compute the stationary urban income and wealth distribution. At rate ϕ , urban households die, and are replaced either by a newborn urban household with probability $1 - \delta_{UR}$ or by a migrant with probability δ_{UR} . To proceed, I have to compute the unconditional distribution of migrants over assets and income. I denote the conditional probability of a migrant with rural state-triple x_R to end up with state x_U in the urban sector as $\mathbb{P}(x_U|x_R)$. Let $\mathbf{P}^{RU}$ be a stochastic matrix of size $N^{\#a} \times N^{\#p} \times N^{\#e}$ where each row is associated with an initial rural state, and each column represents the conditional probability of ending up with a particular urban state. Taking into account of selection into migration, and the rural stationary distribution, I can compute the unconditional probability of a migrant ending up in a particular state denoted as $\kappa(x, \mathcal{I}) \in [0, 1]$, which depends on the information regime $\mathcal{I} \in \{pi, fi\}$ as selection changes when information changes,

$$\kappa(x_U, \mathcal{I}) = \frac{\int \mathbb{P}(x_U|x, \mathcal{I}) \pi_{RU}^{\mathcal{I}}(x) g_R(x) dx}{\int \mathbb{P}(x_U|x, \mathcal{I}) \pi_{RU}^{\mathcal{I}}(x) g_R(x) dx dx_U},$$

and the unconditional probability of a migrant arriving in state x_U equals

$$\kappa(x_U) = \zeta \kappa(x_U, pi) + (1 - \zeta) \kappa(x_U, fi).$$

Using the unconditional distribution of migrants over income and assets κ I can now construct the generator matrix that solves for the urban stationary distribution, which explicitly takes into account equilibrium selection into migration. Let $\mathcal{A}_x$ capture saving and income drift ignoring the death shock. Let $\mathcal{A}_{U, \text{death}}$ be a generator matrix capturing the evolution of income whenever the household is hit by a death shock and replaced by a new urban household – this happens with probability $1 - \delta_{UR}$

which is incorporated in $\mathcal{A}_{U,\text{death}}$. On the other hand, with probability δ_{UR} , the urban household is replaced by a rural migrant. This is captured using the generator matrix $\mathcal{A}_{\text{mig,RU}} = \phi\delta_{UR}(-\mathbb{I} + \iota\kappa')$ where $\mathbb{I}$ is the identity matrix, and $\iota = [1 \ 1 \dots 1]'$ is a row vector of ones. Putting the pieces together delivers the following generator matrix

$$\mathcal{A}_U = \mathcal{A}_{U,x} + \mathcal{A}_{U,\text{death}} + \mathcal{A}_{\text{mig,RU}}.$$

The stationary distribution of the urban sector g_U then solves

$$\mathbf{0} = \mathcal{A}_U^T g_U.$$

Focusing on the stationary distribution in the rural sector, note that migrant outflows are matched by urban-rural inflows in the steady state. This can be represented using

$$\mathcal{A}_{\text{mig,UR}} = \begin{bmatrix} [-\lambda_m \pi(x_1) + \lambda_m \pi(x_1)(g_{RU,0}(x_1))] & [\lambda_m \pi(x_1)(g_{RU,0}(x_2))] & \dots & [\lambda_m \pi(x_1)(g_{RU,0}(x_n))] \\ [\lambda_m \pi(x_2)(g_{RU,0}(x_1))] & [-\lambda_m \pi(x_2) + \lambda_m \pi(x_2)(g_{RU,0}(x_2))] & \dots & \dots \\ \dots & \dots & \dots & \dots \\ [\lambda_m \pi(x_n)(g_{RU,0}(x_1))] & \dots & \dots & [-\lambda_m \pi(x_n) + \lambda_m \pi(x_n)(g_{RU,0}(x_n))] \end{bmatrix}$$

where the diagonals capture the negative out-migration probability, and the in-migration uses the distribution $g_{RU,0}(x)$, which injects urban-rural migrants with zero assets and drawing their transitory and persistent type from the rural newborn income distribution. Note that the rows of this matrix sum to zero by construction.

A rural stationary distribution thus solves

$$\begin{aligned} \mathcal{A}_R &= \mathcal{A}_{R,x} + \mathcal{A}_{R,\text{death}} + \mathcal{A}_{\text{mig,UR}} \\ \mathbf{0} &= \mathcal{A}_R^T g_R. \end{aligned}$$

This complete the description of the algorithm.

B.2 Non-Linear Imaginary Grid

I build on the work of Achdou et al. (2022) to solve the model. Relative to their canonical finite-difference scheme applied to heterogeneous agent economies, I introduce a non-linear “imaginary grid”, which speeds up the solution algorithm and simplifies the KFE.

To set the stage, recall that the standard finite difference scheme iterates on the following discretized HJB equation,

$$(\rho + \phi) V_{n+1} = u(c_n) + A_n V_{n+1} + \frac{V_{n+1} - V_n}{\Delta} \quad (31)$$

up until convergence, $\|V_n - V_{n+1}\| < \epsilon$. An upwind scheme pins down the savings drift hidden in the matrix A_n , where the drift is implied by $u'(c) = V_a$ and a forward or backward approximation is chosen depending on whether savings are positive or negative. Given the concavity in V , the backward approximation implies a higher savings drift than the forward one, so for the algorithm to converge I take the drift that is closer to zero savings, and impose a zero drift when the two do not agree.

To solve for the distribution, take the transpose of the infinitesimal generator $A^* = A^T$, and solve the linear system $0 = A^*g$, which is only determined up to scale and requires one normalization.

Imaginary Grid. One simple but useful innovation in the paper is to build up the grid space in a clever way. First, I would like to use log-spaced asset grid points around the borrowing constraint so that relatively more grid points are places where the value function is most non-linear. In continuous time, however, this creates additional complications regarding the KFE along the transition path as the evolution of the distribution won't be mass-preserving anymore.⁶² Second, the same construction normalizes the grid space, which is desirable for applications with long-run growth where productivity growth shifts real income and assets upwards over time.

I next define a linear grid, which preserves the tractability of the standard finite difference scheme, also with respect to the KFE. I will show how the household will evaluate this linear grid in a non-linear fashion, which allows me to consider non-equidistant environments.

First, for concreteness, let $k \in \{0, 1, 2, \dots, 10\} = K$ be an index that runs from zero to ten with increments of unity (any equidistant spacing works just as well). And assume without loss of generality that $\tilde{a}_k = k$ so the normalized asset grid points are from zero to ten.

Next, define the non-linear mapping $R : f(\tilde{a}) \rightarrow a$, which follows the following functional form

$$f(\tilde{a}(k); k_n) = \begin{cases} e^k - e^0 & \text{if } k \leq k_n \\ \left(\frac{e^{k_n} - e^{k_n-1}}{k_n - k_{n-1}} \right) (k - k_n) + e^{k_n} - e^0 & \text{if } k > k_n \end{cases}$$

where k_n is one element in K and a parameter, k_{n-1} is the element below it, and k is the functional argument. This function ensures that for values of the index below k_n the grid is log-spaced. For values of the index above k_n the grid is linearly spaced where the jump to a higher index always equals $f(k_{n'+1}) - f(k_{n'}) = e^{k_n} - e^{k_{n-1}} \forall n' \geq n$.

Moreover, note that the derivative of f evaluated for some argument k reads

$$\frac{\partial f(\tilde{a}(k); k_n)}{\partial k} = \begin{cases} e^k & \text{if } k \leq k_n \\ \left(\frac{e^{k_n} - e^{k_n-1}}{k_n - k_{n-1}} \right) & \text{if } k > k_n \end{cases}$$

and is well-defined.

With this function f , I can now study the household consumption-saving problem where I focus only for notational simplicity on the risk-free case. Assume that the real asset position of a household is given by $a(k)$, and the standard budget constraint applies $\dot{a} = ra + y - c$ where r , y , and c are risk-free rate, income, and consumption. The HJB equation of the household problem reads

$$\rho V(a, y) = \max_{a, c} u(c) + V_a \dot{a} + V_y \dot{y}$$

with the usual first-order condition

$$V_a = u'(c). \quad (32)$$

The standard finite-difference solution algorithm approximates $V_a^F \approx \frac{V(a_{n+1}) - V(a_n)}{a_{n+1} - a_n}$ and $V_a^B \approx \frac{V(a_n) - V(a_{n-1})}{a_n - a_{n-1}}$ where F and B represent forward and backward difference. The backward difference is used for dissaving,

⁶²See the computational appendix of Achdou et al. (2022), which also proposes a solution. The solution proposed here is easier. The approach is potentially useful more broadly for researchers who want to place grid points over the state space in a non-equidistant fashion.

and the forward difference is applied for savings, and $V_a^0 = u'(c : \dot{a} = 0)$ is the approximation used when the household neither saves nor dis-saves.

In the usual finite-difference approximation, I solve for a fixed point in the K space, i.e.,

$$\rho V(k, y) = \max_{k, c} u(c) + V_k \dot{k} + V_y \dot{y} \quad (33)$$

where k is the fundamental grid on which the algorithm operates. Note that k is defined on an equidistant grid, which will become useful in a second.

Here is the key idea: the household doesn't mind whether they are solving $V(k)$ or $V(a(k))$ – either way the household simply maximizes utility. However, the first order condition that needs to hold is (32). This condition will pin down the optimal drift in the a -space.

The issue then becomes one of how to invert the first order condition implied by (33) to find the solution for (32). Clearly, these are going to be different, i.e., if you want to double your savings, you are not necessarily going to want to double the value of your index k , precisely because k is non-linear. Here is how to go from one to the other. First, note

$$\begin{aligned} V(a(k)) &= V(k) \\ \Rightarrow \\ V_a(a(k)) &= V_k(k) \frac{\partial k}{\partial a} = \frac{V_k(k)}{a'(k)} \end{aligned}$$

which is true by application of the chain rule, and noting $\partial k / \partial a = 1 / f'(k) = 1 / a'(k)$.

Note that $a'(k) > 0$ (almost everywhere except at zero, I will return to the boundary conditions below) so the sign of the derivative of the value function is always the same. Second, note that

$$\dot{a} = a'(k) \dot{k}$$

which offers a way to connect the optimal drift in the a -space with the drift in the K space. Note that $\dot{a}$ comes from the first order condition, and $a'(k)$ is known and depends on the function f , which by construction has a well-defined derivative.

Lastly, recall that the forward drift is used when the household saves, i.e., when $\dot{a} > 0$. Note, however, that whenever that is the case it is also going to be true that $\dot{k} > 0$, i.e., the forward drift also is the right one in the K -space. The reason is that f is a strictly increasing positive function of k , and the only way for a household to build up a higher asset position is to increase k . The approximation to $\dot{k}$ will then also depend on whether the forward or backward difference is used.

Lastly, what happens in the case of $\dot{a} = 0$? Instead of using forward or backward differences, I can directly evaluate the function $a'(k)$ at the value k , i.e., there is no approximation. At the boundaries I impose $V_a \geq u'(c)$, which in the K -space reads $V_k = V_a a'(k)$.

This routine delivers the value function in the K -space, and that value function is directly appropriate because I used $c(a(k))$ to compute it. No adjustments required.

What about the distribution? This will be conveniently defined over the K -space, which preserves the tractability implied by the linear grid-spacing. I solve the following linear system of equations, which solves

for the distribution of mass over the K space,

$$\mathbf{0} = A \begin{pmatrix} \dot{k} \end{pmatrix}^T p,$$

where $\dot{k}$ is the drift in k and not a . Given the solution p , which is a discretized probability distribution, the only thing left to do is to compute asset demand and update the interest rate. For that, note

$$B = \sum_k p_k a(k)$$

delivers aggregate asset demand, where the labor force is normalized to one. The probabilities live in the K -space, but they are multiplied with the asset value in the a -space.

Why is this convenient? I need fewer grid points and put more points where the curvature in the consumption policy function is. Second, the approach is measure preserving, also along the transition path, which is not the case with non-linear grids in Achdou et al. (2022). The reason is that the KFE never sees the non-linearity. The generator $A \begin{pmatrix} \dot{k} \end{pmatrix}$ is built on the equidistant K -space, where the standard finite-difference scheme conserves mass exactly, and the mapping f only enters when I translate the drift and read off asset demand. A non-linear grid applied directly in the a -space does not have this property, because the spacing enters the generator itself.

B.3 Grid Space and Simulation

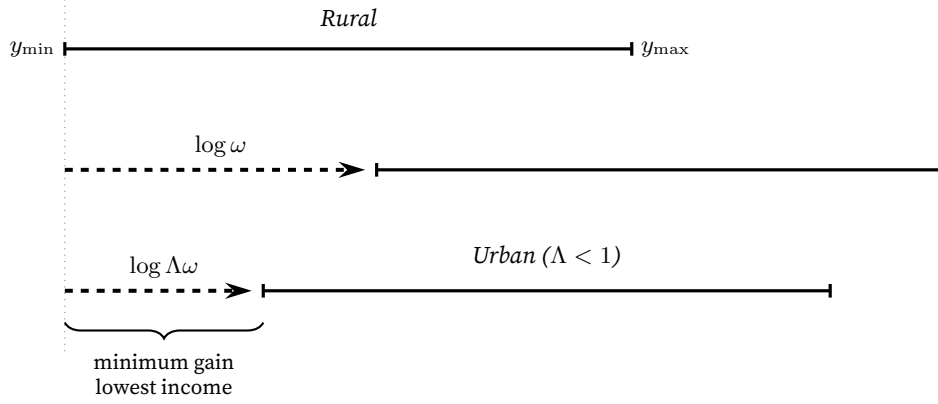
An important aspect of the solution routine is how the grid space is set up. Note that when households move from rural to urban, two things change. First, they end up on a different grid point that captures endowments in efficiency units in the new sector. And second, their income moves because efficiency units are multiplied by wages.

I propose a tractable structure that allows me to use the same normalization of the grid in both rural and urban, and then shift the grid using the rural-urban wage gap and an additional adjustment to make sure that the urban and rural income distribution display the same overlap that is found in the data. Getting this overlap right is key because the gains from migration for the low-income rural households depend crucially on the urban lower income bound. Using the same normalized grid space in rural and urban, the urban grid space in dollar terms is shifted by the rural-urban wage gap ω . This, however, leads to an overstatement of the income gains of rural-urban migrants at the bottom of the distribution, so I use the adjustment factor $\Lambda < 1$ to correct for this. I estimate Λ in the baseline and hold it fixed thereafter, including in counterfactuals where the endogenous rural-urban wage gap adjusts: Λ is a primitive, while the product $\Lambda\omega$ is an equilibrium object.⁶³ Figure 18 illustrates how ω and Λ shift the grid while using the same normalized space in each sector.

I use the finite difference scheme as developed in Achdou et al. 2022, which operates on the following grid space. Log income consists of two linearly spaced grids, $p_grid = \text{LinRange}(-2.5000, 2.5000, 21.0000)$, $\epsilon_grid = \text{LinRange}(-1.5000, 1.5000, 5.0000)$. The asset grid is initially linearly-spaced to obtain a good approximation when households save, in part, to pay the fixed cost of migration. I then switch to a log-linearly

⁶³ Λ and ω are separately identified because the former operates on income gains at the very bottom, while the latter is needed to match aggregate rural-urban migration. Λ and c_{RU} , by contrast, enter expected log urban income additively, so only their sum $c_{RU}^+ = c_{RU} + \log \Lambda$ — total skill downgrading in log efficiency units — is identified from households in the interior of the distribution. The two are separated only by behaviour at the lower bound, where the urban draw is censored and c_{RU} loses bite while Λ continues to move the floor.

Figure 18. Income Grid



All three grids cover the same normalized span $[y_{\min}, y_{\max}]$ but are positioned differently on the real log-income axis. The rural grid is the reference. Urban residents are shifted right by $\log \omega$ (the wage gap). With $\Lambda < 1$, migrants' effective income is shifted to the left relative to the right-shift implied by the rural-urban wage gap.

spaced grid, which again becomes linear at high levels of wealth. The grid uses 120.0000 points with a lower bound of zero.

B.4 Simulating the model

I simulate the income and consumption process of rural and urban households using the solution to the household problem. When I simulate the model, I do replicate the selection patterns induced by endogenous migration, and I compute moments on movers or stayers the same way I do in the empirical exercise.

Simulating Income and Consumption. I simulate the model using the KFE equation that comprises both endogenous savings drift and exogenous income drift. I only focus on surviving households by ignoring death shocks altogether, and I apply a small time step of $dt = 0.1000$, which lines up nicely with income and consumption data aggregated up over one month.⁶⁴ I describe how to draw the initial distribution below and focus on the evolution of an initial cohort hear. Given an initial cohort, I simulate forward the distribution

$$g_{t+dt} = (\mathbb{I} - \mathcal{A}_x^T dt)^{-1} g_t \quad (34)$$

where $\mathcal{A}_x^T$ is the adjoint of the savings and income drift captured in the matrix $\mathcal{A}_x$, and draw the next set of state variables using a uniform draw and the CDF $G_{t+dt}(x) = \int^x g_{t+dt}(u) du$.

In the rural sector, I incorporate endogenous rural-urban migration. Note that the migration opportunity over a time interval dt is approximately $\mathbb{P}(\text{mig. opportunity in } [t, t + dt]) = 1 - e^{-\lambda_m dt}$. The probability that the household is taking up this opportunity depends on the endogenous migration choice $\pi_{RU}(x)$, so the probability of migration over the interval $[t, t + dt]$ for a household of type x equals

$$\mathbb{P}(\text{migrate}|x) = 1 - e^{-\lambda_m \pi_{RU}(x) dt}. \quad (35)$$

Migrant households then draw their urban type consistent with the stochastic rural-urban transition matrix, which depends on the conditional distribution $\kappa(x_U|x)$. For these migrant households, I then apply

⁶⁴This delivers the same moments as allowing for death shocks and then dropping households that drop out of the sample due to death precisely because the death shock is completely random in this model of perpetual youth.

the appropriate urban generator to simulate the evolution of assets and income forward.

I use 50000.0000 observations to ensure the stability of the simulated moments, and I compute every moment 3.0000 to get rid of any sampling variability due to the random draws. Another important input into the simulation is the initial age of the average rural and urban household at the beginning of the sample. I explain next why this matters and how this informs the initial distribution from which the cohorts are drawn.

Drawing an initial cohort in each sector. To draw an initial cohort, note that the age distribution is key because as households age, their income is fanning out. While the increase in inequality over the sample period allows for identification of σ_p irrespective of the how much initial inequality there is, the parameter σ_0 depends on household age since it is a residual quantity that is needed to match overall inequality in the sample for any given estimate of σ_p , which depends on how old households in the sample are as the log variance of income is increasing linearly within each cohort.

Note, however, that migration confounds the link between inequality and household age because i) migrants raise the overall mean age in the urban sector, and ii) migrants draw their income from a distribution that is different from $N(c_{0,U}, \sigma_0)$. I deal with this by drawing from the stationary urban distribution, which incorporates the endogenous migrant inflow. I then use this distribution and drop older households, similar to how I proceed in the empirical data, up until the point where the average age of urban households coincides with the age of urban households in the first wave of the sample. To do this, I need to solve for the joint distribution including age, and then I need to draw from this joint distribution dropping older households, both of which is described in detail in the next paragraph.

Step 1. Solving for age distributions. I focus on the rural sector first. Let x denote the assets and income components as before, and let $a \in N_a$ denote age. The joint distribution $g(x, a)$ evolves according to a KFE with generator given by

$$\mathcal{A} = I_a \otimes \mathcal{A}_x + \mathcal{A}_{\text{age}}^{\text{drift}} \otimes I_x + \mathcal{A}_{\text{death}} + \mathcal{A}_{\text{mig}}, \quad (36)$$

where I_a is the identity matrix of size equal to the elements in N_a , $\mathcal{A}_x$ is the generator associated with income dynamics and savings drift, and $\mathcal{A}_{\text{age}}^{\text{drift}}$ captures the deterministic age drift.

Age evolves deterministically according to $\dot{a} = 1$. Discretizing the age grid yields the generator

$$\mathcal{A}_{\text{age}}^{\text{drift}} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}. \quad (37)$$

Households die at rate ϕ and are replaced by newborns who inherit the asset position but draw a new income state from distribution μ_y . Let I_b denote the identity matrix over assets and let μ_y be a column vector over income states. Define

$$P_y = ((\iota_y \mu_y') \otimes I_b). \quad (38)$$

where ι_y is a column vector of ones with dimension equal to the income grid.

Let $M_{a=0}$ denote the age-zero matrix of size $N_a \times N_a$ with ones in the first column and zero elsewhere.

The injection operator then equals

$$P_{\text{death}} = M_{a=0} \otimes P_y. \quad (39)$$

The associated generator is

$$\mathcal{A}_{\text{death}} = \phi(-I_{ax} + P_{\text{death}}), \quad (40)$$

where $I_{ax} = I_a \otimes I_x$.

Turning to migration, note that every rural-urban migrant in steady state is replaced by an urban-rural one that draws from the rural newborn income distribution and enters with zero assets. Let $e_{b=0}$ denote the selector for the zero-asset grid point and let ι_b be a vector of ones over the asset grid. Define the asset transition matrix

$$P_b^{\text{mig}} = \iota_b e'_{b=0}. \quad (41)$$

Define

$$P_{x,0}^{\text{mig}} = (\iota_y \mu'_y) \otimes P_b^{\text{mig}}. \quad (42)$$

I next account for the fact that migration is endogenous and use $\lambda_m \pi(x)$, which is a column vector, to obtain

$$P_{x,1}^{\text{mig}} = (\lambda_m \pi(x) \iota'_x) \cdot * P_{x,0}^{\text{mig}}$$

The full migration injection operator is given by

$$P_{\text{mig}} = M_{a=0} \otimes P_{x,1}^{\text{mig}}. \quad (43)$$

The associated generator is

$$\mathcal{A}_{\text{mig}} = -(I_a \otimes \text{Diag}(\lambda_m \pi(x))) + P_{\text{mig}}. \quad (44)$$

I can now compute the stationary distribution $g_{\text{R}}(x, a)$ by solving

$$\mathbf{0} = \mathcal{A}^T g_{\text{R}}.$$

Focusing on the urban sector, note that the only complication is that rural-urban migrants do not enter the urban sector at age zero, which is why migration raises the average age of urban households. The age distribution of rural-urban migrants thus partly determines the average age of urban households. Let x denote the assets and income components as before, and let $a \in N_a$ denote age. The joint distribution $g(x, a)$ evolves according to a KFE with generator given by

$$\mathcal{A} = I_a \otimes \mathcal{A}_x + \mathcal{A}_{\text{age}}^{\text{drift}} \otimes I_x + (1 - \delta_{\text{RU}}) \mathcal{A}_{\text{death}} + \delta_{\text{RU}} \mathcal{A}_{\text{RU}}, \quad (45)$$

where $\mathcal{A}_{\text{RU}}$ can be constructed analogous to before but now incorporates age as state variable, $\kappa(x, a)$.

Step 2. Draw from initial distribution

1. Solve for $g(x, a)$.

Solve for the age distribution using the KFE that now includes a drift term and Poisson death shock for age. In the presence of endogenous migration, there is no closed-form solution for the stationary age distribution. I therefore solve the model with age state variable computationally on a grid that includes all integers from zero to one hundred, $\text{age} a \in \{0, 1, \dots, 100\}$.

2. Find the age cutoff.

For each sector $j \in \{R, U\}$, find the smallest cutoff age $\bar{a}_s$ such that the mean of the truncated marginal age distribution equals $a_{j,\min}$:

$$\bar{a}_j = \min \left\{ c \in \mathbb{Z}_+ : \frac{\int_0^c a g(a) da}{\int_0^c g(a) da} \geq a_{j,\min} \right\}, \quad (46)$$

where $g(a) = \int g(\mathbf{x}, a) d\mathbf{x}$ is the marginal age density. This is found by bisection on $c \in \{0, \dots, 100\}$.

3. Form the initial distribution.

The initial joint distribution over $\mathbf{x}$ is

$$g_0(\mathbf{x}) = \frac{\int_0^{\bar{a}_s} g(\mathbf{x}, a) da}{\int \int_0^{\bar{a}_s} g(\mathbf{x}, a) da d\mathbf{x}}, \quad (47)$$

normalized to integrate to one. This is the $\mathbf{x}$ -marginal of the joint distribution conditioned on age $\leq \bar{a}_j$.

This distribution is then used to draw initial income when simulating the model.

Table 40 summarizes the details of the simulation.

Table 40. Details Simulation

	Simulation
Households (N)	50000.0000
Number of simulations (S)	3.0000
Waves	5.0000
Time step (dt)	0.1000
Rural age (first wave)	10.5000
Urban age (first wave)	10.5000

Comparison

Table 41 summarizes the two approaches. In practice, the age-init strategy yields a better-behaved income covariance matrix across waves: the diagonal of $\text{Cov}(\log y_t, \log y_{t+j})$ grows more slowly from wave 1 because young households in the initial cross-section have had less time for permanent shocks to accumulate. The legacy strategy tends to over-estimate variance at early waves and under-estimate the rate of fanning-out relative to the data.

The age-init approach is used as the preferred specification for the results reported in the main text. The legacy approach is retained in the code for comparison and robustness checks; it can be invoked via `simulate_model` while the age-init version is invoked via `simulate_model_age_init`.

Implementation note. Both approaches are implemented in `simulate_model.jl`. The initialization block in each sector's simulation function (`simulate_moments_rural`, `simulate_moments_urban`) accepts an optional `lg_int` argument. When `lg_int = nothing` (the default), the legacy path is taken. When `lg_int`

Table 41. Simulation initialization: comparison of approaches

	Legacy	Age-init
Correct asset-income joint	No (independence)	Yes
Mixed-age cross-section	No (single cohort)	Yes
Urban migrants embedded	No	Yes
Additional computation	None	One KFE solve
Requires calibrated $a_{\min}$	Yes	Yes

is supplied—as done in `simulate_model_age_init` `insolve_simulate_final.jl`—the age-truncated joint distribution is used directly. The two top-level wrappers return named tuples with identical field schemas, enabling field-by-field comparison of any moment.s../values/cov_y_sim_R_bl.tex

Table 42. Simulated Income Covariance Matrix

(a) Rural						(b) Urban					
	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}		y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.582					y_t	0.418				
y_{t+2}	0.289	0.604				y_{t+2}	0.229	0.532			
y_{t+4}	0.216	0.293	0.537			y_{t+4}	0.218	0.251	0.466		
y_{t+6}	0.188	0.22	0.301	0.543		y_{t+6}	0.212	0.239	0.271	0.405	
y_{t+8}	0.177	0.191	0.224	0.307	0.616	y_{t+8}	0.209	0.233	0.259	0.29	0.432

Simulations from the model. See the text for how to replicate selection on age and migration in the simulated data.

Table 43. Simulated Income Covariance Matrix Ignoring Selection

(a) Rural						(b) Urban					
	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}		y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.586					y_t	0.417				
y_{t+2}	0.288	0.606				y_{t+2}	0.229	0.531			
y_{t+4}	0.215	0.299	0.542			y_{t+4}	0.218	0.25	0.465		
y_{t+6}	0.189	0.229	0.314	0.548		y_{t+6}	0.212	0.238	0.271	0.404	
y_{t+8}	0.177	0.201	0.243	0.327	0.621	y_{t+8}	0.209	0.233	0.259	0.289	0.431

Simulations from the model for different selection biases. In the rural sector, imagine there was no migration amongst households in the simulated sample, holding everything else fixed. In the urban sector, I compute a different kind of bias by ignoring rural-urban migration and simply simulating the income process of a cohort that, at the beginning of the sample, matches the average age of urban households. In both cases, the statistical income process would be exactly correct as there is no selection into migration.

The next table simulates the consumption covariance matrices for each sector.

SMM – Estimation Algorithm. Objective function: Let $\theta \in \mathbb{R}^N$ denote the vector of $N = 13$ structural parameters and $m^d \in \mathbb{R}^M$ the vector of $M = 14$ empirical target moments. For a given θ , the model is solved and simulated to produce model-implied moments $m(\theta)$. The objective is a weighted mean absolute deviation:

$$Q(\theta) = \frac{1}{M} \sum_{j=1}^M w_j \left| \frac{m_j(\theta) - m_j^d}{m_j^d} \right|,$$

where $w_j \geq 0$ are pre-specified moment weights. Moments are normalized by their data counterpart so that all deviations are dimensionless and comparable across moments of different scales. Moments with $w_j = 0$ are excluded from the criterion but are still computed and reported.

Table 44. Simulated Income Covariance Matrix

(a) Rural						(b) Urban					
	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}		y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.582					y_t	0.418				
y_{t+2}	0.289	0.604				y_{t+2}	0.229	0.532			
y_{t+4}	0.216	0.293	0.537			y_{t+4}	0.218	0.251	0.466		
y_{t+6}	0.188	0.22	0.301	0.543		y_{t+6}	0.212	0.239	0.271	0.405	
y_{t+8}	0.177	0.191	0.224	0.307	0.616	y_{t+8}	0.209	0.233	0.259	0.29	0.432

Simulations from the model. See the text for how to replicate selection on age and migration in the simulated data.

Box constraints via logit transform: Each parameter has a known economic lower bound ℓ_i and upper bound u_i . To avoid projection steps, we reparametrize using the element-wise logit (log-odds) transform:

$$\varphi_i = \log\left(\frac{\theta_i - \ell_i}{u_i - \theta_i}\right), \quad \theta_i = \ell_i + \frac{u_i - \ell_i}{1 + e^{-\varphi_i}}.$$

The optimizer operates over the unconstrained vector $\varphi \in \mathbb{R}^N$; the inverse sigmoid maps any φ back to a θ that is strictly feasible. The starting point $\theta^{(0)}$ is clamped to $[\ell_i + \varepsilon(u_i - \ell_i), u_i - \varepsilon(u_i - \ell_i)]$ with $\varepsilon = 10^{-6}$ before applying the transform to ensure the initial $\varphi^{(0)}$ is finite.

Nelder-Mead minimization. We minimize $Q(\varphi) \equiv Q(\theta(\varphi))$ using the derivative-free Nelder-Mead simplex algorithm (via Optim.jl). The algorithm maintains a simplex of $N + 1$ points in $\mathbb{R}^N$ and iteratively contracts it toward the minimum through reflection, expansion, and contraction steps. No gradient or Hessian information is required, which speeds up the procedure. Convergence is declared when the relative change in the objective across iterations falls below 10^{-6} , with a hard cap of 2,000 iterations.

Warm-starting. Each evaluation of Q requires solving the model (an HJB-KFE system), which is computationally expensive. To reduce solve times during optimization, the solution from the most recently evaluated simplex vertex is passed as the starting guess for the next model solve. Because consecutive simplex vertices typically correspond to nearby parameter vectors, the warm-started solver converges in far fewer iterations than a cold start.

Identification analysis via the Jacobian. After convergence to $\hat{\theta}$, a numerical Jacobian $G \in \mathbb{R}^{M \times N}$ of the moment function is computed by forward finite differences:

$$G_{ji} = \frac{m_j(\hat{\theta} + h_i e_i) - m_j(\hat{\theta})}{h_i}, \quad h_i = \max(0.005 \cdot |\hat{\theta}_i|, 10^{-5}),$$

where e_i is the i -th unit vector. The relative step size of 0.5% is chosen to balance truncation error against numerical cancellation. The matrix G is reported as a table (rows = moments, columns = parameters) to make the moment-parameter mapping transparent. As a scalar summary of identification strength, we report $\kappa(G^\top G)$, the condition number of the information-matrix analogue: a large condition number signals near-collinearity among the columns of G , i.e., weak identification of some parameter combination.

B.5 Insurance Counterfactuals: Construction and Scenarios

This subsection describes the insurance counterfactuals of section 4.1 in detail, introduces several different ways to model insurance, and explains why the baseline approach was chosen in the main body of the paper.

In all of them the object handed to the migrant is the conditional mean of its urban labor income,

$$\bar{y}_t(p, \epsilon) := \mathbb{E}[(w_U \Lambda) e^{p_t + \epsilon_t} | p, \epsilon],$$

where time is indexed from the date the household joins the arrangement, and the expectation is taken over future income realizations conditional on the types at that date and on survival. I write y_t for realized income and $\bar{y}_t$ for its conditional mean. Types are observed throughout, which isolates the role of risk cleanly and abstracts away from moral hazard and adverse selection.

What differs across the three constructions below is how mortality risk is insured, and whether the borrowing constraint is in place.

B.5.1 Different Ways to Insure Risk

Income Pooling within Cohort. Households form cohorts to pool their income, so each household in a cohort receives the cohort's average income. This swaps a stochastic income stream for a deterministic one, and nothing else changes: households still earn r on assets, still discount at $\rho + \phi$, and the borrowing constraint is in place. Assets evolve as

$$\dot{b}_t = r b_t + \bar{y}_t - c_t, \quad b_t \geq 0,$$

and the urban value function solves the time-dependent problem

$$(\rho + \phi) v(b, t) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b v) \dot{b} + \partial_t v,$$

with the state constraint $\partial_b v(0, t) \geq \bar{y}_t^{-\gamma}$.

Every risk-averse household prefers the cohort average to its own draw at the time it joins, so participation is universal. Staying in requires commitment, which I assume. Note that the borrowing constraint stays in place, so households cannot borrow against their now safe future income. That is what separates this construction from the next one.

The Frictionless Benchmark. Discount the insured stream at r and let the household borrow freely against it,

$$\mathcal{H}(p, \epsilon) := \int_0^\infty e^{-rt} \bar{y}_t(p, \epsilon) dt,$$

which follows conveniently from the Feynman-Kac formula, whereby $\mathcal{H}$ solves the linear system

$$r\mathcal{H} = (w_U \Lambda) e^{p+\epsilon} + \mathcal{A}\mathcal{H},$$

with $\mathcal{A}$ the infinitesimal generator capturing the evolution of urban income among surviving households. In the absence of uncertainty, and with frictionless borrowing, consumption obeys the standard Euler equation, consumption is linear in wealth, and the value function takes a closed form

$$c_t = m \cdot (b_t + \mathcal{H}), \quad m := \frac{(\rho + \phi) + r(\gamma - 1)}{\gamma},$$

$$V_t = \frac{c_t^{1-\gamma}}{1-\gamma} \frac{1}{m}.$$

This is the optimal policy when households can borrow freely, and it is why the value function has a closed form. The problem is that no financial sector could fund it. The household is impatient, $\rho + \phi > r$, so it front-loads consumption and borrows against human wealth priced at r . Its own income can only support payments discounted at $r + \phi$, because both the income and the payments stop at death.

To see this, follow one cohort of insured migrants of measure one after migration. Cohort income and cohort consumption are

$$Y_t = \bar{y}_t e^{-\phi t}, \quad C_t = m \mathcal{H} e^{-(m-r)t} e^{-\phi t}.$$

For simplicity, consider the case with flat income and compare the present discounted value of income and consumption of a cohort without assets

$$\frac{\int_0^\infty e^{-rt} C_t dt}{\int_0^\infty e^{-rt} Y_t dt} = \frac{m(r + \phi)}{r(m + \phi)} = 1 + \frac{\phi(m - r)}{r(m + \phi)} > 1.$$

The excess term is mortality times impatience. Set either ϕ or $m - r$ to zero and it disappears. At my parameters the cohort consumes 19 percent more in present value than it ever earns with flat income, and 33 percent more at the urban variance of persistent shocks, which raises the level path of $\bar{y}_t$ and with it $\mathcal{H}$. What happens is that households die in debt and nobody bears the loss. One might think the small open economy assumption takes care of this. It does not. An open economy can run a deficit, but the resources still have to come from somewhere. The gap also grows in σ_p^2 , so it is tied to the parameter whose effect I want to measure. I report this case as an upper bound and not as a baseline.

The Complete Insurance Contract. The migrant hands over its urban labor income claim and its assets, and receives a consumption stream of its choosing, with both legs paid only while it is alive. This insurance breaks even by construction, and simultaneously insures mortality risk. It is thus the best deal a competitive financial sector could offer, and it also overcomes the borrowing constraint.

Here, human wealth is

$$\tilde{\mathcal{H}} := \int_0^\infty e^{-(r+\phi)t} \bar{y}_t dt, \quad (r + \phi) \tilde{\mathcal{H}} = (w_U \Lambda) e^{p+\epsilon} + \mathcal{A} \tilde{\mathcal{H}},$$

and the household's resource constraint is

$$\int_0^\infty e^{-(r+\phi)t} c_t dt = b + \tilde{\mathcal{H}}.$$

Both sides now discount at the same rate, so break-even holds as an identity, for any γ , ϕ and σ_p^2 . The optimal stream still satisfies $\dot{c}/c = (r - \rho)/\gamma$, the ϕ cancelling between preferences and prices, so the closed form survives with two new constants,

$$c_t = \tilde{m} \cdot (b_t + \tilde{\mathcal{H}}), \quad \tilde{m} := \frac{(\rho + \phi) + (r + \phi)(\gamma - 1)}{\gamma}.$$

The major issue with this approach is that by insuring mortality risk, it also offers households better terms on their wealth, which introduces an asymmetry across sectors unrelated to income risk.

In this setting, a household earns $r + \phi$ on its assets inside the contract and r on them outside it, so insurance is worth more the wealthier the household is. The premium over cohort pooling is under one

percent at zero assets, but it is roughly nine percent at five years of income and close to 30 percent at twenty. Selection on wealth is what I want to measure, so I do not want a construction that moves it for reasons outside the model.

The three constructions are nested. Drop the constraint $b_t \geq 0$ from cohort pooling and the solution is again $c_t = m(b_t + \mathcal{H})$, since $\int_0^\infty e^{-rt} \bar{y}_t dt$ is exactly $\mathcal{H}$. The frictionless benchmark is therefore cohort pooling without the borrowing constraint, and the difference between the two is the shadow value of that constraint and nothing else. The complete insurance contract lies between them. It gives the household the same income claim, better terms on its assets, and no borrowing constraint, so it is weakly better than cohort pooling everywhere, and it is weakly worse than the frictionless benchmark, which is not funded.

B.5.2 Why Income Pooling by Cohort is the Baseline

Income pooling leaves mortality risk and the borrowing constraint in place, exactly as in the baseline. Only the riskiness of income changes. That is what isolates the role of income risk, and it is why I use this construction in the main text. The cost is the closed form. Once the cohort is formed income is deterministic, so time since arrival is a sufficient statistic and (p, ϵ) drops out of the household's state vector. Insurance removes the income states from the problem. What is left is a one dimensional problem in assets, which I solve backward in calendar time with the same upwind scheme I use everywhere else.

B.5.3 Scenarios: Insuring One Margin or Both

Pooling can be offered against two sources of risk: the uncertainty about the initial urban draw, and the evolution of urban earnings after that. Insuring only the second margin, i.e., income risk after migration, means replacing the risky urban continuation value with the value delivered by cohort pooling, while leaving the household uncertain about which (p_U, ϵ_U) it will draw. Insuring both margins means forming the cohort before the draw is realized, so the payment is contingent on the rural type only,

$$\bar{y}_t^m(p_R) = \mathbb{E}[\bar{y}_t(p_U, \epsilon_U) | p_R],$$

where the expectation is computed against the distribution $p_U | p_R$ estimated earlier. Because $\bar{y}_t$ is linear in the level of arrival income, averaging the paths over the draw is the same as computing the path for the average level. Nothing new has to be solved.

Using these value functions together with the taste shocks delivers well-defined migration probabilities that incorporate the value of the insurance. Writing ni for the case in which only post-migration income risk is removed and $nm-ni$ for the case in which both margins are removed, the migration probabilities equal

$$\pi_{RU}^{ni}(x) = \frac{(\omega)^s \cdot \tau_{RU}^{-s} \left(-\mathbb{E} \left[v_U^{ni}(b, p_U, \epsilon_U) | x \right] \right)^{-\frac{s}{\gamma-1}}}{(-v_R)^{-\frac{s}{\gamma-1}} + (\omega)^s \cdot \tau_{RU}^{-s} \left(-\mathbb{E} \left[v_U^{ni}(b, p_U, \epsilon_U) | x \right] \right)^{-\frac{s}{\gamma-1}}},$$

$$\pi_{RU}^{nm-ni}(x) = \frac{(\omega)^s \cdot \tau_{RU}^{-s} \left(-v_U^{nm-ni}(b, p_R) \right)^{-\frac{s}{\gamma-1}}}{(-v_R)^{-\frac{s}{\gamma-1}} + (\omega)^s \cdot \tau_{RU}^{-s} \left(-v_U^{nm-ni}(b, p_R) \right)^{-\frac{s}{\gamma-1}}}.$$

One advantage of this approach relative to solving a full planner's problem is that it does not require taking a stance on the risk embedded in preference shocks.⁶⁵ The risk that is insured is about income alone, and

⁶⁵In principle, a full planner solution would also provide insurance against these preference shocks, see LMW and Donald, Fukui,

household preference shocks still generate heterogeneity in migration choices.

A clarification is needed regarding how to handle ex-ante information, captured by the share $1-\zeta$. When only after-migration income shocks are insured, insurance can be provided to all households regardless of the information regime. When insurance also covers the initial draw, two possibilities arise. First, one could restrict access to households without ex-ante information, which may be preferable given the potential for unraveling due to adverse selection among fully informed households. Second, if households must commit to the insurance scheme before the migration opportunity arrives, then – given risk-averse preferences – all households would opt in, and one can assume universal insurance. The second approach introduces some misallocation, as it ignores privately held information on the household side. I report results under both, labelled $nm-ni$ (A) and $nm-ni$ (B) respectively, and they turn out to be close.

Table 45 reports the ni counterfactual under all three constructions. The choice matters for both the aggregate response and for selection. Cohort pooling raises the urban share by eight percentage points, the complete contract by ten, and the frictionless benchmark by thirteen. Selection differs by more. The reason is how each construction treats wealth. Relaxing the borrowing constraint is worth most to the poor, since they are the ones it binds on, while paying $r + \phi$ on assets is worth most to the rich.

Table 45. Insurance counterfactual under alternative constructions

Model moments	baseline	insurance construction		
		cohort pooling	frictionless	complete contract
Urban population share (GE)	0.65	0.73	0.78	0.75
Urban population share (PE)	0.65	0.77	0.85	0.79
Log(migrant/stayer income in rural)	-0.16	-0.15	-0.19	-0.11
Slope migration rate on income decile	-0.0075	-0.0100	-0.0157	-0.0074

All three columns remove post-migration income risk and differ only in how it is delivered. Cohort pooling is the baseline: it leaves mortality risk and the borrowing constraint in place. The frictionless column lets households borrow freely against expected income priced at r , which no financial sector could fund, so it is reported as an upper bound. The complete contract prices both legs at $r + \phi$; it breaks even, but it also insures mortality risk on assets. PE holds the rural-urban wage gap at its baseline value, GE lets it adjust. The third row is the mean log income gap between future migrants and stayers while both are still in the rural sector, and the last row is the slope of the migration rate on the rural income decile. For both, a more negative entry means migrants are drawn from further down the rural income distribution.

Table 46 reports the full set of simulated moments across the four cells, together with the variance-reduction robustness column, and separates the partial from the general equilibrium response. The partial equilibrium column holds the rural-urban wage gap at its baseline value; the general equilibrium column lets it adjust.

B.6 Decomposing the Welfare Impact of Migration

This appendix collects the analytics behind the selection results in section 4.2. The question it answers is why insurance and information produce opposite selection patterns, and what a migration response does and does not reveal about the gains from migration. Throughout I work with the uninformed household in the derivations.

and Miyauchi (2026) and Mongey and Waugh (2024). Moreover, the planner's problem is somewhat awkward in this setting due to the overlapping generations structure and the fact that migrants do not internalize that the households replacing them draw, on average, from a superior urban human capital distribution, $c_{R,0} < c_{U,0}$. While the extent to which migrants internalize improved outcomes for their offspring is an interesting avenue for future research, I abstract from this issue to focus on the role of intra-generational income risk in rural-urban migration.

Table 46. Moments in counterfactual simulations

Model moments	baseline	counterfactual			robustness $\sigma_{p,U}^2 = .003$
		$\zeta = 0$	ni	$nm-ni$	
Urban population share (GE)	0.65	0.76	0.73	0.79	0.67
Urban population share (PE)	0.65	0.80	0.77	0.86	0.68
Rural-urban migration rate	0.02	0.03	0.03	0.04	0.02
Rural-urban wage gap ω	1.06	0.93	0.98	0.89	1.04
Log wage gap (cross-section)	0.45	0.46	0.34	0.21	0.43
Log wage gap (fixed effects)	0.35	0.55	0.19	0.02	0.32
Log(migrant/stayer income in rural)	-0.16	-0.11	-0.15	-0.17	-0.15
Slope migration rate on income decile	-0.0075	-0.0057	-0.0100	-0.0151	-0.0073
Prob. income improvement rel. to stayer	0.18	0.28	0.11	0.03	0.17
Variance of urban persistent type	0.21	0.12	0.12	0.12	0.05

The $nm-ni$ column reports scenario (A). PE holds the rural-urban wage gap at its baseline value; all remaining rows are general equilibrium. The final column lowers the variance of urban persistent shocks to its rural value as robustness check to see that insurance operates different from reducing volatility, which simultaneously reduces average income.

B.6.1 Three objects

It is convenient to write values in consumption units. For any value function $V < 0$ define the certainty equivalent

$$\mathcal{C}(V) := (-V)^{\frac{1}{1-\gamma}}, \quad (48)$$

which is the constant consumption stream delivering V , up to a factor common to all values that cancels in every ratio below. This is the same transformation used in the welfare measure λ_W of section 4.1: a proportional change in $\mathcal{C}$ is a consumption-equivalent change.

Write $\mathcal{C}_U(x) = \mathcal{C}(\bar{V}_U^{pi}(x))$ for the certainty equivalent of the urban prospect before the draw is known, and $\mathcal{C}_R(x) = \mathcal{C}(V_R(x))$ for the rural one. Both are in levels, not in the wage-normalized units of proposition 1, so the wage gap ω does not appear below.⁶⁶ Using $(-V)^{-\frac{s}{\gamma-1}} = \mathcal{C}(V)^s$, the migration probability (7) and the value of the migration opportunity become

$$\pi_{RU}(x) = \frac{\hat{\mathcal{C}}_U(x)^s}{\hat{\mathcal{C}}_U(x)^s + \mathcal{C}_R(x)^s}, \quad \hat{\mathcal{C}}_U := \frac{\mathcal{C}_U}{\tau_{RU}}, \quad (49)$$

$$\tilde{\mathcal{C}}(x) = \left[\hat{\mathcal{C}}_U(x)^s + \mathcal{C}_R(x)^s \right]^{\frac{1}{s}}, \quad (50)$$

where $\hat{\mathcal{C}}_U$ is the urban prospect net of the migration wedge. The two are symmetric,

$$\pi_{RU} = \left(\frac{\hat{\mathcal{C}}_U}{\tilde{\mathcal{C}}} \right)^s, \quad 1 - \pi_{RU} = \left(\frac{\mathcal{C}_R}{\tilde{\mathcal{C}}} \right)^s. \quad (51)$$

The household's certainty equivalent of the migration opportunity is a power sum of the two options, and s says how close it is to the maximum: as s grows the taste shocks vanish, $\tilde{\mathcal{C}}$ approaches the larger of the two and π_{RU} approaches zero or one, so the better option fully determines the expected value. For finite s the aggregate lies strictly above both options, $\tilde{\mathcal{C}} > \max\{\hat{\mathcal{C}}_U, \mathcal{C}_R\}$, and the difference is due to preference heterogeneity.

⁶⁶In levels the migrant's urban problem depends on the assets she carries across and on the urban wage. Since w_U is pinned down by A_U alone, (17), and migration is permanent, the urban prospect is the same object in partial and in general equilibrium. In normalized units ω enters through the asset argument, $\bar{v}_U(\hat{b}_R/\omega, p_R)$, and a household held at fixed $\hat{b}_R = b/w_R$ gains real assets when w_R rises. The code reports welfare at fixed $\hat{b}_R$, which adds a revaluation of order $(1 - \theta_L) \Delta \log w_R$ to the general-equilibrium gain of households with assets. It is zero for the newborn and for the poorest household, and small for the rest.

Dividing the two shares in (51) and taking logs, the log odds of migrating are

$$\ell(x) := \log \frac{\pi_{RU}(x)}{1 - \pi_{RU}(x)} = s \left[\log \hat{C}_U(x) - \log C_R(x) \right]. \quad (52)$$

Now consider an intervention that improves the urban prospect and leaves the primitives of the rural problem untouched, which is what insurance does. Define the *risk discount*

$$RD(x) := \log C_U^{ni}(x) - \log C_U(x), \quad (53)$$

the proportional increase in the certainty equivalent of the urban prospect when income shocks after arrival are pooled away. In words, $e^{RD} - 1$ is the permanent proportional increase in urban consumption that would make a household indifferent between the risky urban prospect and the insured one. It measures the value of insurance conditional on moving. It is not the total impact on rural utility since most households, most of the time, aren't moving, nor does it take account of GE effects etc.

B.6.2 The response and the gain

Proposition A.1. *Let an intervention change the urban problem so that $\log C_U(x)$ rises by $RD(x)$, and leave the primitives of the rural problem untouched. Let $\lambda_W(x)$ be the consumption-equivalent welfare gain of a rural household at x , in partial or in general equilibrium. Then, at every rural state and in both equilibria,*

$$\frac{\Delta \ell(x)}{s} + \log(1 + \lambda_W(x)) = RD(x). \quad (54)$$

Proof. From (52), $\Delta \ell = s [\Delta \log C_U - \Delta \log C_R]$, since τ_{RU} is fixed. The first term is RD : migration is permanent, so $\bar{V}_U$ does not contain V_R , and w_U does not depend on M , so nothing in general equilibrium reaches the urban problem. The second term is $\log(1 + \lambda_W)$ by definition of a consumption equivalent under CRRA utility, since V_R is the household's lifetime value and a proportional change in C_R is a proportional change in consumption. $\square$

The identity is exact rather than a first-order approximation. It says that the risk discount gets spent on two things at once: part of it shows up as movement, and the rest shows up as a higher value of staying rural. The migration response and the welfare gain are therefore not competing measures of the same thing, and given the risk discount, knowing one of them pins down the other.

Define the *welfare share* of the risk discount,

$$\psi(x) := \frac{\log(1 + \lambda_W(x))}{RD(x)}, \quad (55)$$

so that $1 - \psi$ is the share that shows up as movement. Then

$$\Delta \ell = s(1 - \psi) RD, \quad (\text{log odds}) \quad (56)$$

$$\Delta \log \pi_{RU} \approx s(1 - \pi_{RU})(1 - \psi) RD, \quad (\text{proportional response}) \quad (57)$$

$$\Delta \pi_{RU} \approx s \pi_{RU}(1 - \pi_{RU})(1 - \psi) RD, \quad (\text{level response}) \quad (58)$$

$$\log(1 + \lambda_W) = \psi RD, \quad (\text{welfare}) \quad (59)$$

all evaluated at x , where the two middle lines use $d\pi/d\ell = \pi(1 - \pi)$ and hold to first order in $\Delta\ell$. The case $\psi = 0$ is the textbook comparative static that holds the rural value fixed. It is what the odds ratios of section 4.2 report, and it is an upper bound on the response. Table 47 reports the responses with the rural value moving, so the identity closes on its rows.

One more object is useful. Holding the fallback fixed and differentiating $\tilde{C}^s = \hat{C}_U^s + C_R^s$, (51) gives directly

$$\frac{\partial \log \tilde{C}(x)}{\partial \log C_U(x)} = \pi_{RU}(x), \quad \Delta \log \tilde{C}(x) \approx \pi_{RU}(x) \text{RD}(x). \quad (60)$$

The elasticity of the value of a migration opportunity with respect to the urban prospect is exactly the migration probability, which is intuitive since an option is worth what it delivers times how often it gets taken. Note that this is the gain from a single opportunity holding the rural value fixed, and not the household's lifetime gain. The two are easy to confuse, and confusing them is what makes the incidence of the welfare gain look like the incidence of the migration response.

B.6.3 What determines the welfare share

All the incidence results now come down to $\psi(x)$. I discuss the three things that determine it in turn.

The option recurs. Write the rural problem so the migration option appears explicitly,

$$(\rho + \phi) V_R(x) = u(c) + \mathcal{A}V_R(x) + \lambda_m [\tilde{V}_m(x) - V_R(x)], \quad (61)$$

with $\mathcal{A}$ the generator of the rural state under the household's own policy. Rearranging as $(\rho + \phi + \lambda_m) V_R = u(c) + \mathcal{A}V_R + \lambda_m \tilde{V}_m$ shows that the household's value is a weighted average of its own flow and the value of receiving a migration opportunity, the latter carrying weight $\lambda_m / (\rho + \phi + \lambda_m)$. At the baseline calibration $\lambda_m = 0.2000$ against $\rho + \phi = 0.05$, so that weight is 0.80. Opportunities arrive four times faster than the household discounts and dies, so what a rural household's value largely is, is the prospect of future chances to leave rather than the current flow.

Suppose for the moment that the state never moved, so $\mathcal{A} = 0$, and write w for the weight in certainty-equivalent terms, $w := \partial \log C_R / \partial \log \tilde{C}$. This is $\lambda_m / (\rho + \phi + \lambda_m)$ times $\tilde{V}_m / V_R$, so 0.80 is an upper bound. The urban prospect enters $\tilde{C}$ with weight π_{RU} by (60) and the rural value with weight $1 - \pi_{RU}$, and the rural value itself moves by ψ . Hence $\psi = w [\pi_{RU} + (1 - \pi_{RU}) \psi]$, or

$$\psi(x) \approx \frac{w \pi_{RU}(x)}{1 - w + w \pi_{RU}(x)}. \quad (62)$$

Two implications follow. First, ψ is much larger than π_{RU} , by roughly a factor of three in the bottom income decile and considerably more at the top, which is why the lifetime gain is so much bigger than the gain from a single opportunity in (60). Second, ψ is concave in π_{RU} , so repetition compresses differences across households. A household that takes one opportunity in seven and a household that takes one in two both migrate eventually, and mainly differ in when.

The state moves. The term $\mathcal{A}V_R$ that I set to zero matters quantitatively. Rural income is dominated by imperfectly transitory shocks arriving at rate $\lambda_{\epsilon,R} = 0.4987$, so a rural household redraws its income roughly every two years and passes through much of the rural income distribution over a career. Its welfare share

is therefore an average of π_{RU} along its own path, not its value at the point where it happens to be observed,

$$\log(1 + \lambda_{\text{W}}(x_0)) = \mathbb{E}_0 \left[\int_0^\infty e^{-(\rho+\phi)t} \Omega(x_t) \pi_{\text{RU}}(x_t) \text{RD}(x_t) dt \right], \quad (63)$$

where Ω collects the discounting and arrival terms of (61). Because the rural income process mean-reverts quickly, the distribution of x_t conditional on x_0 converges within a few years. The current income decile is highly informative about $\pi_{\text{RU}}(x_0)$ and much less informative about the path average, so ψ is flatter across households than (62) suggests. The frozen-state formula is only a guide to the shape here, and it understates ψ badly at the top of the income distribution, where it implies a value several times smaller than the model delivers. The same mean reversion is why measured consumption insurance looks high in the rural sector in section 3.3, and why relaxing the borrowing constraint does so little in section 4.3.

The rural wage rises. In general equilibrium the counterfactual raises the rural wage by 0.0845 log points, because land becomes more abundant as households leave. This raises C_{R} for every rural household, migrant or not, and the identity (54) absorbs it: $\lambda_{\text{W}}^{\text{GE}}$ exceeds $\lambda_{\text{W}}^{\text{PE}}$, so $\psi^{\text{GE}} > \psi^{\text{PE}}$, and the response is damped further. Since the wage gain goes to every rural household and does not vary much with rural income, it also compresses whatever differences in ψ the first two forces leave, which is why the spread in ψ across deciles in table 47 is smaller in general equilibrium than in partial equilibrium.

Table 47 puts numbers on the split, since every row is on one basis and ψ can be read off it directly as $\log(1 + \lambda_{\text{W}})/\text{RD}$. In partial equilibrium most of the risk discount shows up as movement; in general equilibrium, where the rural wage rises for migrants and stayers alike, the shares are closer to even.

B.6.4 Information is not a level shift

Information changes the shape of the urban prospect rather than its level, and this is why it tilts selection the other way. An informed household observes p_{U} before deciding, so its migration probability is

$$\tilde{\pi}_{\text{RU}}^{fi}(x) = \int \sigma(\ell(x, p_{\text{U}})) dF(p_{\text{U}}|p_{\text{R}}), \quad \sigma(\ell) = \frac{1}{1 + e^{-\ell}}, \quad (64)$$

where $\ell(x, p_{\text{U}})$ is (52) evaluated at the known draw, while the uninformed household's probability is $\sigma(\bar{\ell}(x))$ with $\bar{\ell}$ built from the draw-averaged value $\bar{V}_{\text{U}}^{pi}$. The difference has two parts.

The first is a level effect. Averaging values before taking the certainty equivalent penalizes dispersion in the draw, so $\bar{\ell}(x) \leq \mathbb{E}[\ell(x, p_{\text{U}})]$. Removing that penalty raises the log odds, exactly as insurance does, and this part of information behaves like a level shift.

The second is an option value, and it has no counterpart under insurance. Since σ is convex wherever $\pi_{\text{RU}} < 1/2$, which holds for every rural household in the baseline, Jensen's inequality gives $\mathbb{E}[\sigma(\ell)] > \sigma(\mathbb{E}[\ell])$. To second order,

$$\Delta \log \tilde{\pi}_{\text{RU}}^{fi}(x) \approx \frac{1}{2} (1 - \pi_{\text{RU}}) (1 - 2\pi_{\text{RU}}) \mathbb{V}[\ell(x, p_{\text{U}}) | p_{\text{R}}], \quad (65)$$

so the value of learning the draw scales with the dispersion of the log odds across draws, $s^2 \mathbb{V}[\log C_{\text{U}} | p_{\text{R}}]$. That dispersion is larger for households whose consumption depends more on labor income, since the draw enters only through labor income; it is therefore larger for high-income rural households, whose human

wealth is large relative to their assets. This is the same labor share θ_L that governs the wealth gradient in appendix B.8.

Information therefore raises migration most where the option was least likely to be exercised, and insurance raises it most where it was most likely to be exercised. The two interventions tilt selection in opposite directions for reasons that have nothing to do with which households the risk costs most.

B.6.5 Quantitative results

Table 47 evaluates these objects in the baseline economy, by rural income decile. Every row comes from the same three solutions, at the same baseline states, weighted by the same baseline distribution, so the migration rates and the welfare gains are directly comparable and no row moves with composition.

Table 47. The Incidence of Risk and the Anatomy of the Response

rural income decile	1	2	3	4	5	6	7	8	9	10
RD, % of urban consumption	19.8	20.4	20.5	20.4	20.3	22.6	21.1	20.8	20.9	20.4
λ_W , PE, %	6.73	6.64	6.71	6.58	6.55	5.94	6.12	6.11	6.09	5.59
λ_W , GE, %	12.70	12.73	12.79	12.78	12.79	12.46	12.55	12.57	12.58	12.31
π_{RU} , baseline	0.109	0.070	0.061	0.053	0.048	0.030	0.034	0.030	0.024	0.018
$\Delta\pi_{RU}$, PE	0.1266	0.1019	0.0907	0.0801	0.0760	0.0672	0.0594	0.0526	0.0436	0.0327
$\Delta\pi_{RU}$, GE	0.0639	0.0516	0.0450	0.0404	0.0389	0.0336	0.0301	0.0269	0.0222	0.0170
$\Delta\ell$, PE	0.989	1.061	1.066	1.061	1.050	1.259	1.158	1.141	1.138	1.144
$\Delta\ell$, GE	0.582	0.649	0.645	0.651	0.650	0.796	0.731	0.724	0.720	0.736

Every row is on one convention: the full state space, weighted by the baseline rural distribution, binned by model income at baseline. Bin membership and weights never move with the counterfactual, so no row reflects a change in composition. π_{RU} is the probability of moving conditional on a migration opportunity arriving, at rate λ_m . RD is defined in (53) and reported as 100 ($e^{RD} - 1$); λ_W is the lifetime consumption-equivalent gain of section 4.1; $\Delta\ell$ is the change in log odds, averaged across states within the decile. By (54), $\Delta\ell/s$ plus $\log(1 + \lambda_W)$ equals RD column by column, in partial and in general equilibrium alike, once λ_W and RD are converted from percent to log points. Rural income depends on the persistent and the transitory component but not on wealth, so it takes finitely many values on the grid and its distribution has mass points, one of them larger than a decile. Where a decile boundary falls inside such a point it splits that income type by wealth, which is why the migration rate and the risk discount in decile six, and more mildly in decile eight, depart from the trend together. The pattern across deciles is unaffected.

Three facts carry the argument.

The risk discount rises with rural income, from 0.181 percent of urban consumption in the bottom decile to 0.185 percent in the top. Post-arrival income risk is not a poor household's problem in this model. If anything it costs the rich slightly more, because their consumption depends more on labor income.

The migration response runs the other way, and steeply. The baseline probability of moving falls from 0.109 to 0.018 across deciles, and the increase under insurance falls with it, from 0.0639 to 0.0170. This is the margin at work: a shift in the log odds moves more households where more households sit near the point of indifference.

The welfare gain follows neither. It is 12.70 percent in the bottom decile against 12.31 percent in the top, and nearly as flat in partial equilibrium. Insurance is worth about the same to everyone even though it moves the poor far more.

The reason is that the two questions have different horizons. A household with a low migration probability today is not one that never leaves; it is one that leaves later. Rural income mean-reverts quickly, so a household in the top decile will pass through the bottom, and the differences the response records at a point in time are largely differences in timing. The response measures who moves now. Welfare measures who leaves eventually, and almost everybody leaves eventually.

Information runs the other way. Revealing the draw raises migration by a factor of 2.6 in the bottom rural income decile and 15.1 in the top, because the option to move on a good draw is worth most to households that would not move on an average one.

Two checks confirm that insurance's selection tilt is the margin rather than the incidence of risk. Applying the average log-odds shift to the baseline margin, $\pi(1 - \pi)\overline{\Delta\ell}$, reproduces -0.0156 of the actual -0.0163 change in the decile slope of the migration rate. And a placebo: cutting the migration wedge τ_{RU} for every household by 17.2 percent, leaving all risk intact, raises the urban population share to 0.73, the same level insurance achieves, and produces a decile slope of -0.0103 against -0.0100 under insurance. A uniform reduction in the cost of moving reproduces insurance's selection pattern almost exactly.

B.6.6 What this implies for reading migration responses

Three qualifications are worth stating explicitly.

The odds ratios quoted in section 4.2 hold the rural value fixed and are therefore upper bounds. The actual response, reported in table 47, is smaller by the factor $1 - \psi$ of (56), and since ψ is largest where migration is most likely, the correction tilts the log-odds response further toward the rich. The selection argument in the text is conservative.

General equilibrium breaks the link between who moves and who gains, and the identity shows how. The counterfactual raises the rural wage, and that gain accrues to households that never migrate. It enters λ_W and therefore ψ , so the force that multiplies the welfare gain from insurance, comparing tables 14 and 13, is the same force that damps the migration response. Welfare rises and movement falls together, because they share one budget, RD. Information is not a level shift, so none of this applies to it.

Finally, the selection pattern of a migration response identifies the margin that was relaxed, not the incidence of the friction that was removed. A change in the level of the urban prospect tilts the response toward households that were already close to moving. A change in its shape tilts the response toward households whose option was far out of the money. Observing that the poor respond more to insurance and the rich respond more to information is therefore informative about what each policy does, and uninformative about which group the risk was hurting.

B.7 Jacobian

I next compute the Jacobians associated with the loss function and evaluated at the optimal parameter vector. I do this using both level changes and percent, where the former is the relevant object to compute standard errors, while the latter can be more intuitively mapped into how changes in parameters change moments. I do not attempt to compute standard errors, in part because the minimization routine takes moments from different datasets, including aggregates, which makes the computation of standard errors difficult.

B.8 Assets and Migration Rates – Analytical and Quantitative Results

This section uses an approximation that shows in closed form how assets play an ambiguous role in raising and reducing migration rates. The ambiguity hinges on two countervailing forces. On the one hand, a household with large asset holdings has relatively little to gain from migration. The reason is that the gains are tied to increases in labor income, which is a relatively unimportant source of income for high-wealth households. On the other hand, an increase in assets reduces consumption volatility associated with migration, which makes migration less risky.

To complement the quantitative findings in the main text, I provide a simple approximation here. Let $\tilde{\rho} := \rho + \phi$ denote the effective discount rate. Recall the value function given by

Table 48. Jacobian in levels

	c_0	ρ_{RU}	c_{RU}	σ_{RU}	ζ	s	τ_{RU}	Λ_{RU}	$\sigma_{0,R}^2$
$\hat{\beta}_{cs}$	-0.48	0.77	-0.54	0.52	-0.08	0.03	1.07	-1.11	-0.35
$\hat{\beta}_{FE}$	-0.12	-0.13	0.18	1.01	-0.83	0.05	1.15	0.01	0.68
$\hat{\rho}_{RU}^{IV}$	0.98	0.38	0.32	0.39	0.16	0.00	0.15	0.01	-0.78
$\Delta \tilde{y}^2$	0.03	-0.01	-0.04	0.13	0.07	-0.00	-0.04	0.00	-0.13
$\hat{\beta}_{R,RU}^d$	-0.01	0.41	0.06	0.18	-0.39	-0.01	-0.03	-0.01	-1.06
$\hat{\beta}_{dec}^\pi$	0.00	0.02	0.00	0.01	-0.02	-0.00	0.00	-0.00	-0.04
$\hat{\beta}_{R,RU}^{better}$	-0.07	-0.07	0.44	0.35	-0.46	0.02	0.52	0.01	0.29
$\ln(\Lambda\omega)$	0.20	0.84	-0.91	0.13	0.53	0.03	1.06	-0.00	0.00
$\text{Var}(y_{R,0})$	0.03	0.01	0.01	-0.01	0.00	0.00	0.01	-0.00	1.00

Table 49. Jacobian in percent

	c_0	ρ_{RU}	c_{RU}	σ_{RU}	ζ	s	τ_{RU}	Λ_{RU}	$\sigma_{0,R}^2$
$\hat{\beta}_{cs}$	-0.44	0.64	-0.08	0.26	-0.06	0.29	1.15	-1.00	-0.05
$\hat{\beta}_{FE}$	-0.11	-0.11	0.02	0.51	-0.65	0.40	1.24	0.01	0.10
$\hat{\rho}_{RU}^{IV}$	0.91	0.32	0.05	0.20	0.13	0.04	0.16	0.01	-0.11
$\Delta \tilde{y}^2$	0.03	-0.01	-0.00	0.07	0.06	-0.03	-0.04	0.00	-0.02
$\hat{\beta}_{R,RU}^d$	-0.01	0.34	0.01	0.09	-0.30	-0.06	-0.03	-0.01	-0.15
$\hat{\beta}_{dec}^\pi$	0.00	0.02	0.00	0.01	-0.02	-0.00	0.00	-0.00	-0.01
$\hat{\beta}_{R,RU}^{better}$	-0.06	-0.06	0.06	0.18	-0.35	0.21	0.57	0.01	0.04
$\ln(\Lambda\omega)$	0.19	0.71	-0.13	0.06	0.41	0.26	1.14	-0.00	0.00
$\text{Var}(y_{R,0})$	0.02	0.01	0.00	-0.01	0.00	0.01	0.02	-0.00	0.15

$$V_0 = \mathbb{E}_0 \left[\int_0^\infty e^{-\tilde{p}t} \frac{c_t^{1-\gamma}}{1-\gamma} dt \right]. \quad (66)$$

I make the simplifying assumption that households consume their flow income so

$$c_t = rb_t + we^{p_t} = y_t.$$

Next, recall that p_t follows a Brownian motion with zero drift and volatility σ_p , and I abstract away from imperfectly transitory shocks in this exercise.

I next show the ambiguous effect of asset holdings on migration. To this end, define the asset and labor income shares $\theta_{B,t} := rb_t/y_t$ and $\theta_{L,t} := we^{p_t}/y_t$ with $\theta_{B,t} + \theta_{L,t} = 1$. Log labor productivity equals $p_t = p_0 + \sigma_p W_t$, $W_t \sim \mathcal{N}(0, t)$. Future income can be written in terms of initial income, income shares, and the random evolution of labor income,

$$y_t = y_0 (\theta_{B,0} + \theta_{L,0} e^{\sigma_p W_t}).$$

I next use a first-order Taylor approximation around $h(p) := \log(\theta_{B,0} + \theta_{L,0} e^p)$ for $p = 0$,

$$\begin{aligned} h(p) &\approx h(0) + h'(0)p, \\ &= \theta_{L,0} p. \end{aligned}$$

Using this, income evolves approximately according to

$$y_t \approx y_0 e^{\theta_{L,0} \sigma_p W_t}. \quad (67)$$

Substituting (67) into the value function in (66) yields

$$\begin{aligned} V_0 &\approx \frac{y_0^{1-\gamma}}{1-\gamma} \int_0^\infty e^{-(\tilde{\rho} - (\theta_{L,0} \sigma_p (1-\gamma))^2 / 2)t} dt \\ &= \frac{y_0^{1-\gamma}}{1-\gamma} \frac{1}{\tilde{\rho} - \frac{(\theta_{L,0} \sigma_p (1-\gamma))^2}{2}}. \end{aligned}$$

The approximation assumes $\tilde{\rho} > \frac{(\theta_{L,0} \sigma_p (1-\gamma))^2}{2}$ and illustrates how risk reduces expected utility (recall the value function is negative so utility decreases when $\tilde{\rho} - \frac{(\theta_{L,0} \sigma_p (1-\gamma))^2}{2}$ becomes smaller, which is the case when risk goes up).

Next, I use this closed form solution to derive the gains from rural-urban migration. Recall that $p_{RU} = \rho_{RU} p_R - c_{RU}^+ + u$ where $u \sim N(0, \sigma_{RU}^2)$ and $c_{RU}^+ = c_{RU} + \log \Lambda$ is skill downgrading that incorporates the Lambda-grid-shift. Moreover, recall that assets can be transferred from rural to urban sector frictionlessly, while labor income further needs to account for the differences in the wages across sectors, i.e., w_U and w_R .

I proceed by assuming that income of rural stayers is given by the first order approximation derived before, where I ignore the migration arrival rates. Income of a rural-urban mover is random and equals

$$y_{0,U} = rb + w_U e^{\rho_{RU} p_R - c_{RU}^+ + u}.$$

The ratio of urban and rural income of a mover equals

$$\begin{aligned} \frac{y_{0,U} - y_{0,R}}{y_{0,R}} &= \frac{w_U e^{\rho_{RU} p_R - c_{RU}^+ + u} - w_R e^{p_R}}{rb + w_R e^{p_R}} \\ &= \theta_{L,0,R} \left(\omega e^{(\rho_{RU} - 1) p_R - c_{RU}^+ + u} - 1 \right). \end{aligned}$$

I focus on a household for which the expected income gains are close to zero, i.e., $e^{\log \omega + (\rho_{RU} - 1) p_R - c_{RU}^+ + u} - 1 \approx 0$. In that case,

$$\frac{y_{0,U}}{y_{0,R}} \approx 1 + \theta_{L,0,R} (\log \omega + (\rho_{RU} - 1) p_R - c_{RU}^+ + u).$$

Using this, the ratio of urban and rural expected utility, using the simplification that rural and urban labor share are identical, $\theta_{L,0,R} = \theta_{L,0,U} = \theta_{L,0}$ (ignoring the stochasticity of the urban labor share), equals

$$\begin{aligned} \frac{\mathbb{E}[V_{0,U}]}{V_{0,R}} &\approx \mathbb{E} \left[\left(1 + \theta_{L,0,R} (\log \omega + (\rho_{RU} - 1) p_R - c_{RU}^+ + u) \right)^{1-\gamma} \right] \frac{\tilde{\rho} - \frac{(\theta_{L,0} \sigma_{p,R} (1-\gamma))^2}{2}}{\tilde{\rho} - \frac{(\theta_{L,0} \sigma_{p,U} (1-\gamma))^2}{2}} \\ &= e^{-(\gamma-1)\theta_{L,0,R} \left[(\log \omega + (\rho_{RU} - 1) p_R - c_{RU}^+) - \frac{(\gamma-1)\theta_{L,0,R} \sigma_{RU}^2}{2} \right]} \frac{\tilde{\rho} - \frac{(\theta_{L,0} \sigma_{p,R} (1-\gamma))^2}{2}}{\tilde{\rho} - \frac{(\theta_{L,0} \sigma_{p,U} (1-\gamma))^2}{2}}. \end{aligned}$$

To finally arrive at the most simple version, let $\sigma_{p,R}^2 \approx 0$, $\gamma = 2$, roughly consistent with the quantitative application. Define the migration gain $d_{RU} = \log \omega + (\rho_{RU} - 1) p_R - c_{RU}^+$ to obtain an equivalent but simpler expression,

$$\frac{\mathbb{E}[V_{0,U}]}{V_{0,R}} \approx \frac{e^{-\theta_{L,0}[d_{RU} - \frac{1}{2}\theta_{L,0}\sigma_{RU}^2]}}{1 - \frac{1}{\tilde{\rho}}\frac{1}{2}(\theta_{L,0}\sigma_{p,U})^2}.$$

Finally, note that the migration share is given by

$$\pi_{RU} = \frac{1}{1 + \left(\tau_{RU} \times \frac{\mathbb{E}[V_{0,U}]}{V_{0,R}}\right)^s},$$

In general the wedge enters as τ_{RU}^s and the value ratio with exponent $\frac{s}{\gamma-1}$; with $\gamma = 2$ both exponents equal s , which is the case written here. The share is strictly falling in the ratio $\frac{\mathbb{E}[V_{0,U}]}{V_{0,R}}$.⁶⁷

One can now see exactly how the non-monotonicity of the link between migration and wealth comes about. Let $\theta_L := \theta_{L,0}$ and $a := \frac{1}{\tilde{\rho}}\frac{1}{2}\sigma_{p,U}^2$, and study the ratio R as a function of the labor share θ_L ,

$$R(\theta_L) := \frac{e^{-\theta_L[d_{RU} - \frac{1}{2}\theta_L\sigma_{RU}^2]}}{1 - a\theta_L^2} = \frac{\mathbb{E}[V_{0,U}]}{V_{0,R}}.$$

I next explore the derivative of this ratio with respect to the labor share. Since the asset share is the complement of the labor share, $\theta_B = 1 - \theta_L$, we can simply flip the sign to get the derivative of an increase in the asset share. I consider two cases that contain the intuition.

First, suppose $d_{RU} > \sigma_{RU}^2 + \frac{\sigma_{p,U}^2}{\tilde{\rho} - \frac{1}{2}\sigma_{p,U}^2}$. This is the case for the lowest-income rural households in my quantitative application. In this scenario, the derivative of the numerator is always negative. The higher the labor share, the greater the migration gains since the gains consist of wage gains times how important wage income is for total income. This importance is captured by the labor share. When the labor share is falling and the asset share rising, the gains from migration fall. This is true even though the denominator ($\frac{1}{1-a\theta_L^2}$) pushes in the opposite direction by reducing effective risk exposure in urban as the labor share falls. Consequently, for the lowest income type the relationship between assets and migration is monotonically falling once households can cover the fixed cost of migration.

Second, there are intermediate cases where the link between wealth and migration is non-monotone. The non-monotonicity derives from two countervailing forces. The by now familiar force is that a higher asset position discourages migration mechanically by reducing the labor share, which pins down how relevant migration gains are for household consumption, on average. On the other hand, more assets reduce consumption risk around migration, given by $\frac{1}{2}\theta_L^2\sigma_{RU}^2$, and assets also reduce consumption risk after migration captured in the term $1/(1 - a\theta_L^2)$. This positive impact of asset holdings on migration through reduced consumption volatility opens the door for non-monotone dynamics. A sufficient condition for non-monotone behavior in the migration probability to emerge is given by⁶⁸

$$0 < d_{RU} < \sigma_{RU}^2 + \frac{\sigma_{p,U}^2}{\tilde{\rho} - \frac{1}{2}\sigma_{p,U}^2},$$

which holds for the majority of households.

The decomposition in the full model. Taking logs of $\tau_{RU}R(\theta_L)$, using $\log(1-a\theta_L^2) \approx -a\theta_L^2$ and $\frac{1}{s} \log \frac{\pi_{RU}}{1-\pi_{RU}} =$

⁶⁷It is easy to get confused here: because the value functions are negative for $\gamma > 1$, a high ratio means migration is unattractive as the negativity causes a sign flip. In the case with $\gamma < 1$, it would be the other way around and high ratios would indicate high urban continuation values relative to rural.

⁶⁸This condition is most conveniently derived from taking logs and differentiating w.r.t. the labor share.

$-\log(\tau_{RU}R)$, gives

$$\frac{1}{s} \log \frac{\pi_{RU}}{1 - \pi_{RU}} \approx -\log \tau_{RU} + \theta_L d_{RU} - \frac{1}{2} \theta_L^2 \sigma_{RU}^2 - a \theta_L^2. \quad (68)$$

To check that its logic survives savings, the borrowing constraint, transitory shocks, and the option to wait for a better draw, I decompose the wealth gradient of the log odds in the full model into the same pieces, using the insurance benchmarks of appendix B.5. The gradient of the baseline log odds splits exactly into the gain term of equation (68), a risk term (baseline versus cohort pooling), a constraint term (cohort pooling versus the frictionless benchmark), and a rural term (risk and the constraint on the side of staying). The four sum to the model's own gradient by construction. Table 50 reports them by rural income quintile at the median wealth of each quintile. The entries are elasticities of the odds of migrating with respect to wealth, divided by s ; to compare with the derivative of (68) with respect to the asset share, divide by $\theta_L \theta_B$. In the middle quintile the wealth gradient is 0.0147, of which the gain term contributes -0.0017 and the risk term 0.0113. The constraint term is 0.0398 in the bottom quintile and 0.0338 in the middle.

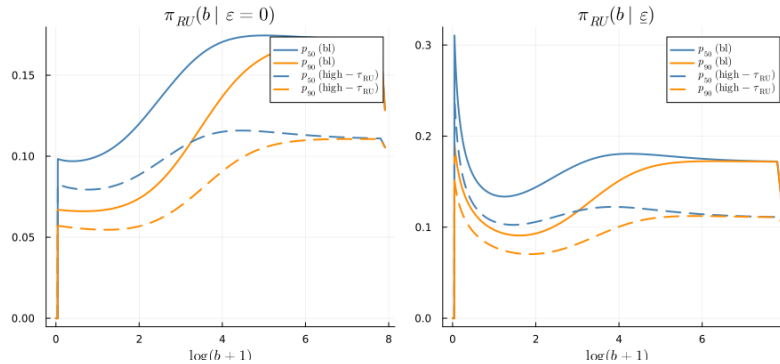
Table 50. Wealth Gradient of the Log Odds of Migrating, by Rural Income Quintile

quintile	1	2	3	4	5
total	0.0043	0.0120	0.0147	0.0180	0.0233
gain	-0.0044	-0.0024	-0.0017	-0.0011	-0.0003
risk	0.0105	0.0115	0.0113	0.0118	0.0128
constraint	0.0398	0.0367	0.0338	0.0328	0.0319
rural	-0.0417	-0.0338	-0.0287	-0.0255	-0.0211

Each entry is $\frac{1}{s} \partial \ell / \partial \log b$, the elasticity of the odds of migrating with respect to wealth divided by s , for an uninformed household with $\epsilon_R = 0$ at the median wealth of its rural income quintile. The gain term is $\theta_{L,R} - \theta_{L,U}$ from the frictionless benchmark with certain income on both sides, where $\theta_L = \mathcal{H}/(b + \mathcal{H})$ is the human-wealth share, the lifetime counterpart of the labor share in the text; risk is the gradient of the gap between the baseline and cohort pooling; constraint is the gradient of the gap between cohort pooling and the frictionless benchmark. These three rows vary the urban value only. The rural row is the gradient of the risk and constraint parts of the rural value. The rows sum to the total up to a residual of at most 0.000000.

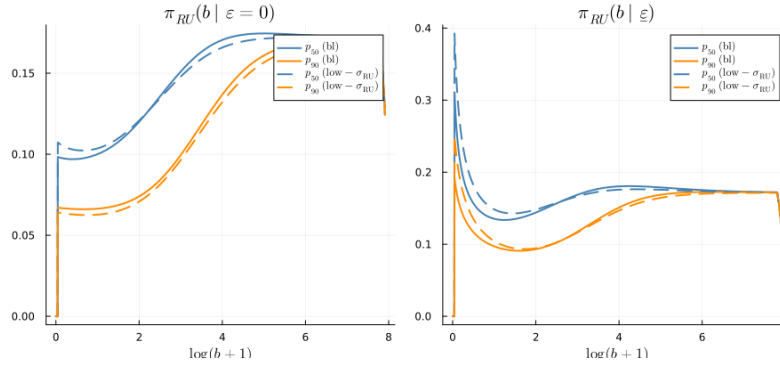
The figures below vary the wedge, the draw variance, the share of uninformed households, and the arrival rate of migration opportunities one at a time in the full quantitative model, which amplifies or reduces the channels above.

Figure 19. Selection into Migration – High τ_{RU}



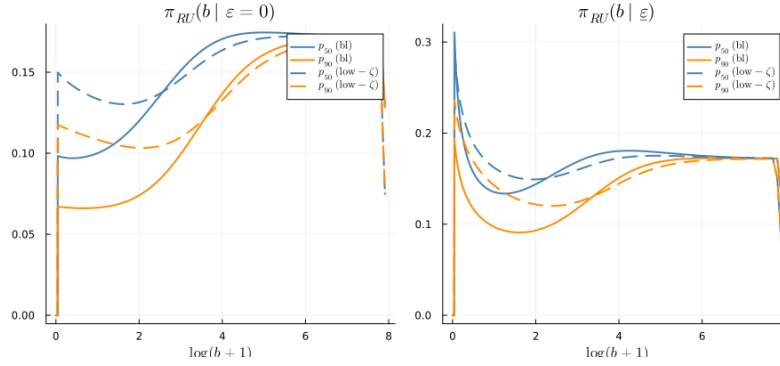
τ_{RU} (+10 log pts). This plot tests the pure utility-discount channel: an increase in the utility discount pushes down migration of wealthy households since they do not gain from wage gains (zero labor share) but they pay the utility cost, so their wealth is best consumed in rural. The drop of the very right tail is a grid issue. It is irrelevant, as the mass of agents in that region is virtually zero (<0.000001).

Figure 20. Selection into Migration – Low σ_{RU}



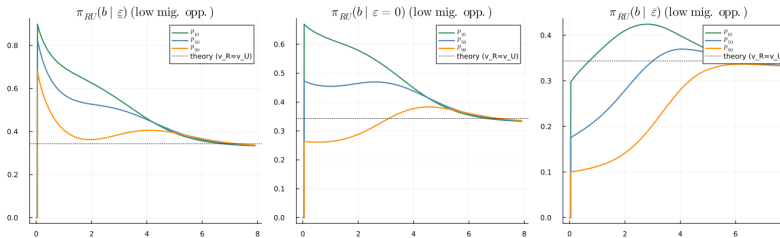
σ_{RU} (-30%). This plot tests whether a reduction in migration risk reduces the impact of assets on migration. This is the case as the slope flattens. The drop of the very right tail is a grid issue. It is irrelevant, as the mass of agents in that region is virtually zero (<0.000001).

Figure 21. Selection into Migration – Low ζ



ζ ($\times 0.5$). This plot tests whether reducing uncertainty around migration by providing more information also makes the impact of assets weaker. This seems to be the case as the dashed lines are flatter. The drop of the very right tail is a grid issue. It is irrelevant, as the mass of agents in that region is virtually zero (<0.000001).

Figure 22. Selection into Migration – Low λ_m



$\lambda_m = 0.005$. This plot is a sanity check for the wealthiest households, for whom a closed form solution is available: the urban value is the same as rural net of the migration wedge. The value functions in general do not agree due to the option value of rural-urban migration, but when migration arrival rates are small, this option value becomes small, and the closed form solution $\pi_{RU} = 1/(1+\tau_{RU}^S)$ becomes accurate for ultra-wealthy households. The sanity check is passed as the migration rates asymptote to the dashed line that computes the closed form solution. Consistent with this, note that labor income becomes eventually irrelevant as assets become ever-more important, so the different migration rates for different households converge as they get wealthier. The drop of the very right tail disappears here because the grid space issue disappears whenever the assets of rural-urban households live inside the urban grid, which is the case in this calibration with much higher rural-urban wage gap.

C Theoretical Appendix

C.1 Proofs

Migration probability. I derive the migration probability based on the random Frechet shocks. Note that the probability of migration is a function of the state variables captured in the the vector x ,

$$\begin{aligned}
\mathbb{P}(\text{migrate}, x) &= \mathbb{P}\left(V^U \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} > V_R \varphi_R^{1-\gamma}\right) \\
&= \mathbb{P}\left(\varphi_R^{\gamma-1} < \frac{V_R}{V_U} \left(\frac{\varphi_U}{\tau}\right)^{\gamma-1}\right) \\
&= \mathbb{E}_{\varphi_U} \left[\mathbb{P}\left(\varphi_R < \left(\frac{V_R}{V_U}\right)^{\frac{1}{\gamma-1}} \left(\frac{\varphi_U}{\tau}\right) \mid \varphi_U\right) \right] \\
&= \mathbb{E}_{\varphi_U} \left[\exp\left(-\left(\frac{V_R}{V_U}\right)^{\frac{-s}{\gamma-1}} \left(\frac{\varphi_U}{\tau}\right)^{-s}\right) \right] \\
&= \int_0^\infty \exp\left(-\left(\frac{V_R}{V_U}\right)^{\frac{-s}{\gamma-1}} \left(\frac{\varphi_U}{\tau}\right)^{-s}\right) s \varphi_U^{-s-1} e^{-\varphi_U^{-s}} d\varphi \\
&= \frac{1}{\left(\frac{V_R}{V_U}\right)^{\frac{-s}{\gamma-1}} \left(\frac{1}{\tau}\right)^{-s} + 1} \\
&= \frac{\tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}}}{(-V_R)^{\frac{-s}{\gamma-1}} + \tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}}}
\end{aligned}$$

where I assumed $\gamma > 1$.

Expected value of migration. I next derive the expected value of migration conditional on migration, i.e.,

$$\mathbb{E}\left[\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \mid \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U > (\varphi_R)^{1-\gamma} V_R\right].$$

To proceed, I need the distribution of $(\frac{\varphi_U}{\tau})^{1-\gamma} V_U | (\frac{\varphi_U}{\tau})^{1-\gamma} V_U > (\varphi_R)^{1-\gamma} V_R$, which can be derived as follows

$$\begin{aligned}
\mathbb{P}\left(\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \leq k \mid \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U > (\varphi_R)^{1-\gamma} V_R\right) &= \frac{\mathbb{P}\left(\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \leq k \cap \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U > (\varphi_R)^{1-\gamma} V_R\right)}{\mathbb{P}\left(\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U > (\varphi_R)^{1-\gamma} V_R\right)} \\
&= \frac{\mathbb{P}\left((\varphi_R)^{1-\gamma} V_R \leq \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \leq k\right)}{\pi_{RU}} \\
&= \frac{\int \mathbb{P}\left((\varphi_R)^{1-\gamma} > \left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} \frac{V_U}{V_R} \mid \varphi_U\right) dF(\varphi_U) \mathbb{1}_{\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \leq k}}{\pi_{RU}} \\
&= \frac{\int e^{-\left(\frac{\varphi_U}{\tau}\right)^{-s} \left(\frac{V_U}{V_R}\right)^{\frac{-s}{1-\gamma}}} dF(\varphi_U) \mathbb{1}_{\varphi_U \leq \tau \left(\frac{k}{V_U}\right)^{\frac{1}{1-\gamma}}}}{\pi_{RU}} \\
&= \frac{\int_0^{\tau \left(\frac{k}{V_U}\right)^{\frac{1}{1-\gamma}}} e^{-\left(\frac{\varphi_U}{\tau}\right)^{-s} \left(\frac{V_U}{V_R}\right)^{\frac{-s}{1-\gamma}}} s \varphi_U^{-s-1} e^{-(\varphi_U)^{-s}} d\varphi}{\pi_{RU}} \\
&= \frac{1}{1 + (\tau)^s \left(\frac{V_U}{V_R}\right)^{\frac{-s}{1-\gamma}}} \frac{\left[e^{-\left(\varphi_U\right)^{-s} \left(1 + (\tau)^s \left(\frac{V_U}{V_R}\right)^{\frac{-s}{1-\gamma}}\right)} \right]_0^{\tau \left(\frac{k}{V_U}\right)^{\frac{1}{1-\gamma}}}}{\pi_{RU}} \\
&= \frac{1}{1 + (\tau)^s \left(\frac{V_U}{V_R}\right)^{\frac{-s}{1-\gamma}}} \frac{\exp\left(-\left((-k)^{\frac{1}{1-\gamma}}\right)^{-s} \left((\tau)^{-s} (-V_U)^{\frac{-s}{1-\gamma}} + (-V_R)^{\frac{-s}{1-\gamma}}\right)\right)}{\pi_{RU}} \\
&= \frac{(\tau)^{-s} (-V_U)^{\frac{-s}{\gamma-1}}}{(\tau)^{-s} (-V_U)^{\frac{-s}{\gamma-1}} + (-V_R)^{\frac{-s}{\gamma-1}}} \frac{\exp\left(-\left((-k)^{\frac{s}{\gamma-1}}\right) \left(\tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}} + (-V_R)^{\frac{-s}{\gamma-1}}\right)\right)}{\pi_{RU}} \\
&= \exp\left(-\left((-k)^{\frac{s}{\gamma-1}}\right) \left(\tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}} + (-V_R)^{\frac{-s}{\gamma-1}}\right)\right)
\end{aligned}$$

where the domain of k is $(-\infty, 0]$. I can now compute the expected value using the distribution. Define $\tilde{\Phi} = \tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}} + (-V_R)^{\frac{-s}{\gamma-1}}$, and compute

$$\mathbb{E}\left[\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \mid \max\right] = \int_{-\infty}^0 k \frac{s}{\gamma-1} \tilde{\Phi} (-k)^{\frac{s}{\gamma-1}-1} \exp\left(-\tilde{\Phi} (-k)^{\frac{s}{\gamma-1}}\right) dk.$$

Next, use the change of variable $x := -k$, which flips both the limits and the sign,

$$\mathbb{E}\left[\left(\frac{\varphi_U}{\tau}\right)^{1-\gamma} V_U \mid \max\right] = - \int_0^{\infty} x \frac{s}{\gamma-1} \tilde{\Phi} x^{\frac{s}{\gamma-1}-1} \exp\left(-\tilde{\Phi} x^{\frac{s}{\gamma-1}}\right) dx.$$

Next, use another change of variable $y := \tilde{\Phi} x^{\frac{s}{\gamma-1}}$, so that $x = \left(\frac{y}{\tilde{\Phi}}\right)^{\frac{\gamma-1}{s}}$ and $\frac{s}{\gamma-1} \tilde{\Phi} x^{\frac{s}{\gamma-1}-1} dx = dy$,

$$\begin{aligned} \mathbb{E} \left[\left(\frac{\varphi_U}{\tau} \right)^{1-\gamma} V_U | \max \right] &= - \int_0^\infty \left(\frac{y}{\tilde{\Phi}} \right)^{\frac{\gamma-1}{s}} \exp(-y) dy \\ &= - \left(\frac{1}{\tilde{\Phi}} \right)^{\frac{\gamma-1}{s}} \int_0^\infty y^{\frac{\gamma-1}{s}} \exp(-y) dy \\ &= - \left(\frac{1}{\tilde{\Phi}} \right)^{\frac{\gamma-1}{s}} \Gamma \left(1 + \frac{\gamma-1}{s} \right). \end{aligned}$$

Since the distribution of utility in each location conditional on selection into the location are equalized, it follows that this is the expected utility of receiving a migration opportunity. Further, note that I shift the original Frechet distribution so that the term $\Gamma \left(1 + \frac{\gamma-1}{s} \right)$ drops out. This is achieved by drawing from the distribution $F(x) = e^{-\left(\frac{x}{T}\right)^{-s}}$ with $T = \Gamma \left(1 + \frac{\gamma-1}{s} \right)^{\frac{1}{\gamma-1}}$. After making this assumption, the expected value of migration $\hat{V}$ becomes

$$\begin{aligned} \hat{V} &= - \left(\tau^{-s} (-V_U)^{\frac{-s}{\gamma-1}} + (-V_R)^{\frac{-s}{\gamma-1}} \right)^{\frac{1-\gamma}{s}} \\ &= - \left(\tau^{-s} (-V_U)^{\frac{s}{1-\gamma}} + (-V_R)^{\frac{s}{1-\gamma}} \right)^{\frac{1-\gamma}{s}}. \end{aligned}$$

Defining $\Phi = -\tilde{\Phi}^{\frac{1-\gamma}{s}} = \hat{V}$ the coincides with the notation in the main text.

C.2 Lemma 1. (sketch proof)

To prove lemma 1, I first focus on the value function in the urban sector, where $y = e^{p+\epsilon}$, and show that $V_U(b, \xi y) = \xi^{1-\gamma} v_U(b_0, y)$ where v_U is the solution to the household problem when the wage is normalized to unity and $b_0 = \frac{b}{\xi}$. Here ξ is a generic scale factor on labor income: it equals the sectoral wage w_j for a household that stays put, and the effective wage $w_U \Lambda$ for a rural-urban migrant.

Step 1. Note that for $v_U(b_0, y)$ to solve the normalized household problem, it must be that

$$\begin{aligned} (\rho + \phi) v_U(b_0, y) &= \max \frac{c^{1-\gamma}}{1-\gamma} + (\partial_{b_0} v_U) \dot{b}_0 + (\partial_y v_U) \dot{y} \\ \text{s.t.: } \dot{b}_0 &= r b_0 + y - c_0, \\ \text{FOC: } \partial_{b_0} v_U &= c_0^{-\gamma}, \end{aligned}$$

where I study a simplified problem that only has a drift in income for expositional purposes only, and c_0 denotes consumption associated with the normalized problem. I omit the borrowing constraint, but note that whenever this constraint is not zero, it needs to be scaled with the wage as well for the argument to go through.

Step 2. Show that $V_U = \xi^{1-\gamma} v_U(b_0, y)$. This is the case whenever $\xi^{1-\gamma} v_U(b_0, y)$ solves

$$\begin{aligned} (\rho + \phi) V_U(b, y) &= \max \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_U) \dot{b} + (\partial_y V_U) \dot{y} \\ \text{s.t.: } \dot{b} &= r b + \xi y - c. \end{aligned}$$

Conjecture this is the case. Then, the first order condition implies

$$\begin{aligned}
\partial_b V_U &= c^{-\gamma} \\
\partial_b \xi^{1-\gamma} v_U(b_0, y) &= c^{-\gamma} \\
\xi^{1-\gamma} \frac{\partial_{b_0} v_U(b_0, y)}{\xi} &= c^{-\gamma} \\
\partial_{b_0} v_U(b_0, y) &= \left(\frac{c}{\xi}\right)^{-\gamma} \\
c_0^{-\gamma} &= \left(\frac{c}{\xi}\right)^{-\gamma},
\end{aligned}$$

where the third line uses the chain rule. It follows that $c = c_0 \xi$.

Next, verify whether this consumption policy maximizes utility using the HJB equation

$$\begin{aligned}
(\rho + \phi) V_U(b, y) &= \max \frac{\xi^{1-\gamma} c_0^{1-\gamma}}{1-\gamma} + (\partial_b V_U) \xi \dot{b}_0 + (\partial_y V_U) \dot{y} \\
&= \max \frac{\xi^{1-\gamma} c_0^{1-\gamma}}{1-\gamma} + \xi^{1-\gamma} (\partial_{b_0} v_U) \dot{b}_0 + \xi^{1-\gamma} (\partial_y v_U) \dot{y} \\
&= \xi^{1-\gamma} \max \frac{c_0^{1-\gamma}}{1-\gamma} + (\partial_{b_0} v_U) \dot{b}_0 + (\partial_y v_U) \dot{y}.
\end{aligned}$$

subject to the flow budget constraint $\dot{b} = rb + \xi y - \xi c_0$. At this point, note that $\xi^{1-\gamma}$ is a positive scalar that can be pulled out of the max operator, and the right hand side is maximized by the policy c_0 when the budget constraint equals $\dot{b}_0 = rb_0 + y - c_0$ by assumption (recall we assumed v_U is a solution to the normalized problem). Lastly, note that the same budget constraint applies, $\dot{b} = rb + \xi y - \xi c_0 \Leftrightarrow \dot{b}_0 = rb_0 + y - c_0$ since $\dot{b}_0 = \frac{\dot{b}}{\xi}$. This completes the first part of the proof.

Next, consider the rural household problem. This problem is complicated by the fact that rural households can move to the urban sector. I will show that a similar argument applies conditional on holding the rural-urban wage gap, ω , fixed.⁶⁹ Consider the rural value function with simplified income drift and for the case of partial information only (the more general case works exactly the same)

$$(\rho + \phi) V_R(b, \xi y) = \max \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_R) \dot{b} + (\partial_y V_R) \dot{y} + \lambda_m (\tilde{V}_{m,pi} - V_R)$$

⁶⁹Holding the wage gap fixed is equivalent to scaling both rural and urban wage. This is intuitive since the problem will not scale conveniently when only rural income is scaled but urban income is held fixed.

subject to the standard flow budget constraint. First, note

$$\begin{aligned}
\tilde{V}_{m,\text{pi}} &= - \left[\tau_{\text{RU}}^{-s} \left(\tilde{V}_{\text{U,pi}} (b_{\text{R}} - f_{\text{RU}} w_{\text{R}}, p_{\text{R}}) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + (V_{\text{R}} (b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) \cdot (-1))^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} \\
&= - \left[\tau_{\text{RU}}^{-s} \left((w_{\text{U}} \Lambda)^{1-\gamma} \tilde{v}_{\text{U,pi}} \left(\frac{b_{\text{R}} - f_{\text{RU}} w_{\text{R}}}{w_{\text{U}} \Lambda}, p_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + (V_{\text{R}} (b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) \cdot (-1))^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma} \left[\tau_{\text{RU}}^{-s} \left(\frac{(w_{\text{U}} \Lambda)^{1-\gamma}}{w_{\text{R}}^{1-\gamma}} \tilde{v}_{\text{U,pi}} \left(\frac{b_{\text{R}} - f_{\text{RU}} w_{\text{R}}}{w_{\text{U}} \Lambda}, p_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + \left(\frac{V_{\text{R}} (b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) \cdot (-1)}{w_{\text{R}}^{1-\gamma}} \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma} \left[\left(\frac{\tau_{\text{RU}}}{\omega \Lambda} \right)^{-s} \left(\tilde{v}_{\text{U,pi}} \left(\frac{b_{\text{R}} - f_{\text{RU}}}{\omega \Lambda}, p_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + \left(\frac{V_{\text{R}} (b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) \cdot (-1)}{w_{\text{R}}^{1-\gamma}} \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}}.
\end{aligned}$$

I conjecture that the same normalization applies to the rural value function, in which case this expression simplifies to

$$\begin{aligned}
\tilde{V}_{m,\text{pi}} &= -w_{\text{R}}^{1-\gamma} \left[\left(\frac{\tau_{\text{RU}}}{\omega \Lambda} \right)^{-s} \left(\tilde{v}_{\text{U,pi}} \left(\frac{b_{\text{R}} - f_{\text{RU}}}{\omega \Lambda}, p_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + \left(\frac{V_{\text{R}} (b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) \cdot (-1)}{w_{\text{R}}^{1-\gamma}} \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma} \left[\left(\frac{\tau_{\text{RU}}}{\omega \Lambda} \right)^{-s} \left(\tilde{v}_{\text{U,pi}} \left(\frac{b_{\text{R}} - f_{\text{RU}}}{\omega \Lambda}, p_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} + \left(v_{\text{R}} \left(\frac{b_{\text{R}}}{w_{\text{R}}}, p_{\text{R}}, \epsilon_{\text{R}} \right) \cdot (-1) \right)^{-\frac{s}{\gamma-1}} \right]^{-\frac{\gamma-1}{s}} \\
&= w_{\text{R}}^{1-\gamma} \tilde{v}_{m,\text{pi}} \left(\frac{b_{\text{R}}}{w_{\text{R}}}, p_{\text{R}}, \epsilon_{\text{R}}; \omega \right),
\end{aligned}$$

where I make explicit the dependence on the rural-urban wage gap. An analogous argument applies to the migration probability π_{RU} .

Now the same steps as before apply, i.e., verify that $V_{\text{R}}(b, \xi y; \omega) = \xi^{1-\gamma} v_{\text{R}}(b_0, y; \omega)$ with $b_0 = \frac{b}{\xi}$ solves the below, subject to the budget constraint and transversality condition,

$$(\rho + \phi) V_{\text{R}}(b, \xi y) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_{\text{R}}) \dot{b} + (\partial_y V_{\text{R}}) \dot{y} + \lambda_m (\tilde{V}_{m,\text{pi}} - V_{\text{R}}).$$

Scaling of the Stationary Distribution. Given that the household problem satisfies $\xi c_0(b_0, y) = c(\xi b_0, \xi y)$, the stationary density of the normalized problem and the stationary density of the problem with wage equal ξ satisfy

$$g(b) = \frac{g_0(b_0)}{\xi}$$

where $b = b_0 \cdot \xi$. Integrating over b implies $G(b) = G_0\left(\frac{b}{\xi}\right)$. An implication of this is that household savings, $\dot{b}$, and aggregate asset demand, scale linearly in the wage rate.

Proof sketch. Analyze the KFE

$$\dot{g} = -\partial_b [\mu(b, y) g(b, y)] + \mathcal{L}_y g(b, y)$$

where μ is the saving policy and $\mathcal{L}_y$ is the infinitesimal generator capturing the income transitions. Note that $g_0(b_0, y)$ solves the normalized problem. Note that $\mu(b, \xi y) = \xi \mu_0$ since $\xi c_0(b_0, y) = c(\xi b_0, \xi y)$. Fur-

ther, conjecture that $g = \frac{1}{\xi} g_0$ is the stationary solution for the problem with wage ξ . Note that the ξ cancels ($\xi \mu_0 \cdot \frac{1}{\xi} g_0 = \mu_0 g_0$), and the rest of the problem is unchanged. Since $\partial_{b_0} [\mu_0 g_0] = 0$, it must be that $\partial_b [\mu(b, y) g(b, y)] = 0$.

Computational Implications. The computational significance is that I can solve the model in normalized form for a guess of the wage gap. Once I have a solution, the rural wage is implied by the total amount of efficiency units of rural labor (due to land congestion). The urban wage is then set such that it is consistent with the rural wage, and the wage gap, $w_U = \omega w_R$, and the free parameter that ensures this is the case is the urban TFP level. This means that I can solve the model faster as a function of the wage gap alone, and in a final step simply determine the urban productivity such that the urban wage is consistent with this equilibrium. This works because the urban economy is constant returns to scale and the interest rate is fixed, the urban wage does not depend on migration or urban share.

Proof that full information dominates partial information. I want to show that full-information case dominates partial-information case. This amounts to showing that the function

$$g(x) = - \left((-x)^{-\kappa} + (-b)^{-\kappa} \right)^{-\frac{1}{\kappa}}$$

is convex, which means that $\mathbb{E}[g(x)] > g(\mathbb{E}[x])$. Here $\kappa = \frac{s}{\gamma-1}$, so that g is the expected value of migration $\hat{V}$ derived above, with the migration wedge normalized to one. In the derivation I assume that x and b are negative, which holds exactly when $\gamma > 1$. The variable x stands in for the random urban outcome, while b is the fixed rural continuation value. This matters because it is an easy way to prove that the model where information is revealed before migration leads to higher expected utility. The derivative with respect to x equals

$$g'(x) = (-x)^{-(1+\kappa)} [-g(x)]^{1+\kappa}.$$

Differentiating once more yields

$$\begin{aligned} g''(x) &= (1+\kappa) (-x)^{-(2+\kappa)} [-g(x)]^{1+\kappa} \\ &\quad - (1+\kappa) [-g(x)]^\kappa g'(x) (-x)^{-(1+\kappa)} \\ &= (1+\kappa) (-x)^{-(2+\kappa)} [-g(x)]^{1+\kappa} \\ &\quad - (1+\kappa) (-x)^{-2(1+\kappa)} [-g(x)]^{1+2\kappa} \\ &= (1+\kappa) (-x)^{-(2+\kappa)} [-g(x)]^{1+\kappa} \left(1 - (-x)^{-\kappa} [-g(x)]^\kappa \right) \\ &= (1+\kappa) (-x)^{-(2+\kappa)} [-g(x)]^{1+\kappa} \left(1 - \frac{(-x)^{-\kappa}}{(-x)^{-\kappa} + (-b)^{-\kappa}} \right). \end{aligned}$$

The last line makes clear that $g''(x) > 0$. The bracket $1 - \frac{(-x)^{-\kappa}}{(-x)^{-\kappa} + (-b)^{-\kappa}}$ is positive for any κ , and the remaining factors are positive because $\kappa > 0$ and $x < 0, b < 0$. Both hold whenever $\gamma > 1$.

C.3 Additional micro-foundation for stationary equilibrium

There are additional micro-foundations to obtain a stationary equilibrium. The first matches the baseline in levels and on impact. The second matches it exactly.

First, suppose that all newborns are moved to the rural or urban sector with probability P_R and P_U . Then

the share of urban households equals in steady state

$$M = \frac{\phi P_U + \Pi}{\phi + \Pi}.$$

Note that the problem of an incumbent household is unchanged. Moreover, the share of newborn urban households is identical since newborn inflows are the gap between death and rural inflows to ensure a stationary urban share,

$$\phi P_U = \phi M - (1 - M) \Pi.$$

In the original model, this was equal to $M\phi(1 - \delta_{UR})$, so I need to show that $M(1 - \delta_{UR}) = P_U$. And this is the case since the different versions of the model match the same steady-state urban share, i.e., see in the original model

$$\dot{M} = -\phi M + \phi(1 - \delta_{UR})M + \Pi(1 - M) \Rightarrow M = \frac{\phi(1 - \delta_{UR})M + \Pi}{\phi + \Pi}.$$

Since the age distributions are identical, and the household problems are unchanged, the steady state between the two models is equivalent.

The two models part ways once migration changes. In the baseline the urban share solves $M = \frac{\Pi}{\phi\delta_{UR} + \Pi}$, so that $\log \frac{M}{1-M} = \log \Pi - \log(\phi\delta_{UR})$ and the urban share responds to migration like a logit,

$$\frac{\partial M}{\partial \log \Pi} = M(1 - M).$$

With a fixed P_U instead, the same derivative equals $\frac{\Pi}{\phi + \Pi}(1 - M)$, which is smaller by the factor $\frac{\phi\delta_{UR} + \Pi}{\phi + \Pi}$. The reason is that the baseline injects urban newborns in proportion to the urban share, so a rise in migration feeds on itself, while ϕP_U always injects a constant mass of young households into the urban sector, no matter the urban share. What is the same across the two is the effect on impact, $\frac{\partial \dot{M}}{\partial \Pi} = 1 - M$, evaluated at the common steady state.

Another micro-foundation that avoids urban-rural migration altogether is to consider an environment with different fertility rates. The fertility rate in the rural sector needs to be higher so that out-migration is exactly matched by newborn households, on top of the death shock that is equal across rural and urban sector. Let rural fertility be $\phi + g_{n,R}$ and urban fertility is $g_{n,U}$. In that case, urban inflows equal

$$\begin{aligned} \dot{M} &= (g_{n,U} - \phi)M + \Pi(1 - M) \\ (1 - M) &= g_{n,R}(1 - M) - \phi(1 - M) - \Pi(1 - M). \end{aligned}$$

Now any $1 - M$ is consistent with a stationary equilibrium as long as $g_{n,R} = \phi + \Pi$. The urban share, in turn, equals

$$M = \frac{\Pi}{\phi - g_{n,U} + \Pi}, \quad \phi > g_{n,U}.$$

This is the baseline expression with $\phi - g_{n,U}$ in the place of $\phi\delta_{UR}$. The two therefore agree exactly, and not only around the steady state: the urban share and its response to migration, $\frac{\partial M}{\partial \log \Pi} = M(1 - M)$, coincide at every migration rate.

In this version of the model, we need $g_{n,R}$ to change one for one with the migration rate. This could be further micro-founded using a positive link between income and fertility, i.e., if migration picks up, the

rural-urban wage gap goes up, and so does fertility to ensure $g_{n,R} = \phi + \Pi$.