Risk and Rural-Urban Structural Change*

Florian Trouvain
University of Oxford

Current Version: August 29, 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 dynamic general equilibrium model 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 and find that these risks together hold the urban population share roughly twenty percentage points below where it would otherwise be. Providing information about initial urban outcomes closes more than half of the gap, encouraging migration among high-income rural households, who have the most to lose from a low draw. Insurance raises migration by a similar amount, but instead draws in the rural poor, who are particularly exposed to urban income shocks after migration. Wealth is no substitute for insurance. While wealth reduces urban consumption risk, it also lowers the share of labor income in consumption and, with it, the gains from moving to a higher-wage location. A model that ignores initial uncertainty and persistent income shocks overstates the residual migration wedge by a factor of two.

*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, 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 as an inherently dynamic and risky decision in the absence of complete insurance markets. As a result, equilibrium allocations may be inefficient, shaped by migrants' risk exposure.

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 crucial to match the data and tease out the risky nature of rural-urban migration in the South African context, which forms the main application.

Two main lessons emerge. The first is that both sources of risk retard rural-urban structural change substantially. Removing both raises the urban share from 65 to 84 percent. An immediate implication is that models that ignore initial uncertainty and persistent shocks substantially overstate the residual migration wedge. This wedge has been found to be large in the literature, and my framework suggests that it is overstated by a factor of two, once risk is properly accounted for.

The second lesson is that how risk is removed matters for who moves. Risk needs three ingredients—outcomes that differ, uncertainty about them, and the absence of complete insurance markets—and each can be taken away separately to undo the risky nature of migration. Information removes the uncertainty, while insurance removes the consequence of heterogeneous outcomes. Both raise migration by comparable amounts, yet the two strategies draw in opposite ends of the rural income distribution. Providing information encourages migration amongst higher-income rural households, who were deterred by the possibility of a low urban draw. Providing insurance encourages migration amongst low-income rural households, who have no bufferstock savings to absorb urban income shocks that follow arrival.

Paradoxically, wealth is no substitute for insurance even though the impact of income risk on consumption volatility does decline in wealth. With standard preferences, incentives to migrate are proportional to the relative increase in household income. This relative increase is falling in household wealth simply because asset income becomes the dominant source of household income. While wealth allows rural households to cope with urban labor market risk, it simultaneously reduces the gains from migration. Insurance escapes this paradox: it makes labor income safe without making it unimportant.

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 depart from the previous literature, in particular Lagakos, Mobarak, and Waugh (2023) (henceforth LMW), in three crucial ways.

First, I develop a model with household birth and death dynamics. This frees the analysis of two limitations inherent to the commonly used infinite horizon frameworks: the setup allows me to incorporate fully persistent income shocks, which is ruled out by construction in stationary infinite horizon economies. Moreover, death and birth dynamics allow for the possibility that newborn rural households are not necessarily in the sector where they are most productive.[1] I go on to show that both features have important quantitative implications by considering model variants that shut down permanent shocks where migration is driven by low transitory shock realizations, as in Kleemans (2015).

The second key innovation is to tractably allow for different degrees of prior knowledge about initial urban labor market outcomes when migration opportunities arise. I achieve this by incorporating an information friction that nests two limiting cases: a full and partial-information case. In the full-information case, households observe their urban outcome before migrating. In the partial-information case, households do not observe their urban outcome ex ante, and therefore have to form expectations based on the statistical relationship between rural and urban income. The partial-information case constitutes a departure from standard Roy models with perfect information, which is especially relevant for the case of permanent rural-urban migration where households enter a new economic environment for the first time.

The third departure from previous work is to propose a more tractable model of selection into migration. I assume that whenever migration opportunities arrive, households receive an urban draw that can be arbitrarily correlated with their rural type. This approach sidesteps the curse of dimensionality in stochastic dynamic Roy models, while still matching arbitrary selection patterns into migration. I leverage this tractability by embedding my migration model into a general equilibrium setup with endogenous rural-urban wage gap.[2]

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 (2022), 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 in my sample, it is important for the rural sector where imperfectly transitory shocks are the dominant source of income volatility. In contrast, fully persistent shocks are more important in the urban sector so long-term urban earnings potential is uncertain, even in the case where a household has full information about their initial urban persistent productivity draw. This insight requires allowing for fully persistent shocks, and I explore how the usual AR(1) income process employed in infinite horizon settings leads to an understatement of urban risk and an overstatement of rural risk.

I complement this finding by studying sectoral income dynamics in other developing countries, which are consistent with the findings from South Africa. Income risk fundamentally differs across sectors in developing economies, which represents a standalone empirical contribution that should be of interest to a growing literature that uses heterogeneous agent models to study households in 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. Intuitively, frequent post-migration income losses indicate limited ex-ante information. Such losses are possible even when migration is highly

[1] Contrast this with infinite horizon frameworks that feature a fixed long-term urban productivity. Since households live forever, they already have selected into whichever sector is best for them, and, ironically, the only migration ever observed in the steady state is done by agents that are largely indifferent between residing in either sector.

[2] I also develop a new technique to solve continuous-time incomplete market models on a non-linear grid that is particularly helpful for models with non-convexities, like fixed cost of migration, which my framework features as well. See appendix B.

profitable on average.

Using these empirical moments, I estimate the structural parameters of the model. Only 20% of potential migrants have full information about their initial persistent urban income type, which is the key determinant of success in the urban sector. This large degree of uncertainty generates negative selection on rural income, consistent with the South African data. Intuitively, if the initial urban income draw is unknown at the time of the migration decision, high-income rural households have much to lose despite large expected income gains, whereas for low-income rural households migration largely represents upside risk.

These findings contrast with previous work, especially Young (2013), which suggests rural-urban migrants are positively selected relative to rural stayers. While the model is flexible and can, in principle, accommodate positive selection, both the data and theory point in the opposite direction.

I also explore the implications of the model for consumption risk, since income risk only deters migration when it translates into consumption risk. I find a seemingly high degree of consumption insurance in rural South Africa, indicated by a low passthrough from income shocks to consumption, consistent with prior literature and the seminal work of Townsend (1994). Surprisingly, model-simulated passthrough coefficients match these low empirical estimates very well, even though the model contains no insurance beyond bufferstock savings. The reason is that passthrough estimates are extremely sensitive to the income process: when shocks are mean-reverting, as they are for rural households, bufferstock savings insure effectively. This simple insight is quantitatively important, and moot unless a dynamic model that differentiates between persistent and imperfectly transitory shocks is employed.

The framework is thus consistent with a worsening of insurance when households move from the rural to the urban sector, which has been an influential theme in the literature (Munshi and Rosenzweig, 2016). Yet, my framework suggests an important twist: the loss of insurance does not require diminished access to informal rural insurance schemes, and is simply the consequence of sectoral differences in the nature of income risk. In the model persistent shocks pass through to consumption almost fully in both sectors. Yet, rural households appear better insured against income shocks simply because they are less exposed to persistent shocks.

In the final step I use the model to quantify how risk impacts rural-urban migration. I start by investigating the complex interaction between risk and migration. Recall that risk requires heterogeneity, uncertainty, and market incompleteness all switched on. Shutting down any one of them eliminates risk, but their micro and macro implications differ substantially, i.e., who selects into migration, what are their welfare implications, and by how much does the aggregate urban population share increase.

First, consider providing information about initial urban persistent type to all potential rural-urban migrants. This raises migration rates, lifting the urban population share from 0.65 to 0.76 in general equilibrium, while substantially improving selection into migration. The intuition is that low rural types largely face upside risk, so they always switch sectors when given the opportunity. Informing a household in the middle or top of the distribution of a high urban draw, on the other hand, makes them want to move, which would have been too risky in the absence of information, given the possibility of a low urban draw.

Providing insurance works differently. Households receive an actuarially-fair safe urban income stream after migration, but learn nothing about their initial urban type, so only the income shocks that follow arrival are insured. This raises aggregate migration rates by a similar amount as full information about initial earnings. But it draws in different households: insurance induces more negative selection relative to the baseline. The reason is that wealthier rural households can already absorb urban income shocks with bufferstock savings, whereas low-income households cannot.

Whenever households accumulate large amounts of wealth that would allow them to cope with urban income risk effectively, the gains from migration become small. The reason is that migration gains are based on labor income gains, and the importance of labor income for household consumption is falling in household wealth. This is a straightforward explanation for the negative link between wealth and migration. LMW interpret this negative link as evidence against the role of risk in deterring migration. In contrast, my model reproduces it even though risk is a quantitatively important barrier.

Lastly, I use the model to quantify how much the literature overstates the residual migration wedge by ignoring the risky nature of rural-urban migration. At this point, the answer is unsurprising: by a lot. A researcher using the standard full-information, complete-markets model would have to inflate the migration wedge by 112.24 % to match observed migration flows. A large share of the residual cost of moving is the risk of moving.

A second, complementary approach reaches the same conclusion. Using a more standard model without initial uncertainty or persistent shocks and re-estimating inflates the migration wedge by 42.32 %.

Contribution to the Literature. The paper relates to a large literature at the intersection of macroeconomics and development. 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 constructs a heterogeneous agent model with endogenous migration to understand the impact of migration subsidies on consumption gains in the RCT of Bryan, Chowdhury, and Mobarak (2014). In contrast to LMW, where rural income risk leads to too much migration, I find that urban income risk reduces rural-urban migration. The findings are easily reconciled. In their setting, migration is mostly temporary and constitutes a coping strategy after a negative income shock; In my setting, migration is permanent, and uncertainty about long-run urban outcomes deters the move.[3]

The link between rural and urban productivity plays an important role in understanding agricultural productivity gaps, see Lagakos and Waugh (2013) and Alvarez-Cuadrado, Amodio, and Poschke (2023). This link is usually modeled in full-information environments. I stress how risk shapes selection into migration, and quantify its effect on the distribution of unobserved human capital across sectors using a dynamic equilibrium model.[4] I find that migrants are negatively selected, even compared to rural stayers. This contrasts with the influential work of Young (2013), which interprets the rural-urban wage gap as a product of positive sorting on unobserved skill.[5]

The role of risk in rural-urban migration has also been addressed in Munshi and Rosenzweig (2016) and Morten (2019), which focus on the interaction of migration and informal insurance using models of limited commitment, building on Ligon, Thomas, and Worrall (2002). I find that a simple model with bufferstock savings goes a long way in matching low passthrough estimates after allowing for more realistic income processes.[6] This is extremely desirable because models of limited commitment are difficult to solve and

[3]Treating 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). Section 4.2 develops the reconciliation.

[4]Imperfect 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.

[5]See 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.

[6]A large empirical literature has studied the degree of informal insurance in developing economies. See Alderman and Paxson

hard to implement, while benchmark heterogeneous agent models remain tractable even in general equilibrium.

Empirically, I build on a large literature in macroeconomics and labor economics to estimate persistent and transitory income shocks, especially Blundell, Pistaferri, and Preston (2008) (henceforth BPP).[7] I focus on income and consumption dynamics in developing economies, a relatively unexplored aspect with notable exceptions in De Magalhaes and Santaeulalia-Llopis (2018), De Magalhaes, Martorell, and Santaeulalia-Llopis (2024) and Carli et al. (2025). I contribute to this literature by showing that the nature of income risk systematically differs across sectors.

Experimental work by Baseler (2023), Barnett-Howell et al. (2025) and Miner (2025) studies whether information and insurance interventions induce rural-urban migration. In contemporaneous work, Baseler and Gordeev (2026) build a rich structural model of Kenyan rural-urban migration designed to reproduce these empirical studies in a unified framework. They use their model to rank migration frictions including urban unemployment, an information friction in the form of a systematic understatement of urban income, and a migration wedge. Notably, they find that urban unemployment and biased expectations about urban income account for relatively little, while the residual migration wedge dominates. My paper unpacks this residual wedge, and shows that much of it can be accounted for by risk, once a richer information structure and fully persistent income shocks are incorporated, consistent with the data but naturally ruled out by infinite horizon models. I show that ignoring these risks overstates the migration wedge by between 42.32 % and 112.24 %.[8]

Another important difference from this experimental literature is my focus on permanent rural-urban migration, which is less likely to respond to experimental intervention: both the gains and the costs at stake are large relative to the incentives usually provided through experimental variation. Consistent with this view is that the RCT of Bryan, Chowdhury, and Mobarak (2014) raised short-term migration but did not induce permanent rural-urban migration, see Mobarak and Reimão (2020). Lastly, I complement the experimental literature by studying the role of general-equilibrium feedback effects, which turn out to be quantitatively important and ensure that rising rural-urban migration rates translate to income gains for rural stayers. This arises because land is a factor in fixed supply so that land abundance caused by out-migration lifts wages for non-migrants, consistent with recent empirical work by 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.

(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), Santaeulalia-Llopis and Zheng (2018), Attanasio et al. (2022), Meghir et al. (2022), and Ligon (2025).

[7]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. (2021), and especially Crawley, Holm, and Tretvoll (2022) on which I build most directly.

[8]Note 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.

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 $j \in \{\text{R,U}\}$ equals

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

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:

$$dp = \sigma_{p,j}\, dZ, \tag{2}$$

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

where dZ is a Wiener process,[9] 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 are sometimes described as "passing shocks" as outcomes eventually revert back. The setup builds directly on Crawley, Holm, and Tretvoll (2022), and nests the canonical persistent-transitory model of MaCurdy (1982) when $\lambda_{\epsilon,j}$ becomes large.[10]

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.[11]

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

[9] The 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.

[10] The 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.

[11] The 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.

$$V_{\mathrm{U}}\left(b,p,\epsilon\right)=\max_{\{c_s\}}\mathbb{E}_t\left[\int_t^{\infty}e^{-(\rho+\phi)(s-t)}\frac{c_s^{1-\gamma}}{1-\gamma}ds\right] \tag{4}$$
$$\text{s.t.}\qquad \dot{b}=rb+W_{\mathrm{U}}\left(p,\epsilon\right)-c,\qquad b\geq 0,$$

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 $f_{\mathrm{RU}}\geq 0$ paid in rural labor such that urban assets after migration equal $b_{\mathrm{U}}=b_{\mathrm{R}}-f_{\mathrm{RU}}w_{\mathrm{R}}$.
- Migration includes an idiosyncratic taste shock φ, and a disutility wedge $\tau_{\mathrm{RU}}>0$, 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, $\tau_{\mathrm{RR}}=1$.
- 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.[12] Because the wedge admits this consumption-equivalent interpretation, the analysis is robust to mis-measured price gaps across sectors.[13]

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

$$V_m=\max_{\mathrm{R,U}}\left\{\mathbb{E}\left[V_{\mathrm{U}}\left(b_{\mathrm{U}},p_{\mathrm{U}},\epsilon_{\mathrm{U}}\right)|p_{\mathrm{R}},\epsilon_{\mathrm{R}}\right]\cdot\left(\frac{\varphi_{\mathrm{U}}}{\tau_{\mathrm{RU}}}\right)^{1-\gamma},V_{\mathrm{R}}\left(b_{\mathrm{R}},p_{\mathrm{R}},\epsilon_{\mathrm{R}}\right)\cdot\varphi_{\mathrm{R}}^{1-\gamma}\right\}, \tag{5}$$

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 $(p_{\mathrm{U}},\epsilon_{\mathrm{U}})$. 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

$$p_{\mathrm{U}}|p_{\mathrm{R}}\sim N\left(\rho_{\mathrm{RU}}\cdot p_{\mathrm{R}}+c_{\mathrm{RU}},\sigma_{\mathrm{RU}}\right), \tag{6}$$

[12] The 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.

[13] That 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 (2017), which document surprisingly small price gaps across sectors.

where the parameter $\rho_{\text{RU}} > 0$ induces a positive correlation between rural and urban type, σ_{RU} is the standard deviation around the conditional mean, and $c_{\text{RU}} \leq 0$ captures skill downgrading,[14] where $F(.|p_{\text{R}})$ denotes the CDF associated with (6).

In contrast, I assume that transitory types are independent across sectors, $\epsilon_{\text{U}} \perp \epsilon_{\text{R}}$, 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

$$\mathbb{E}\left[V_{\text{U}}\left(b_{\text{U}}, p_{\text{U}}, \epsilon_{\text{U}}\right)|p_{\text{R}}\right] = \int_{\epsilon'}\int_{p_{\text{U}}} V_{\text{U}}\left(b_{\text{U}}, p_{\text{U}}, \epsilon'\right) dF\left(p_{\text{U}}|p_{\text{R}}\right) dG_{\text{U}}\left(\epsilon'\right) := \bar{V}_{\text{U}}^{pi}\left(b_{\text{R}}, p_{\text{R}}\right). \quad \text{(partial information)}$$

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,

$$\mathbb{E}\left[V_{\text{U}}\left(b_{\text{U}}, p_{\text{U}}, \epsilon_{\text{U}}\right)|p_{\text{R}}\right] = \int_{\epsilon'} V_{\text{U}}\left(b_{\text{U}}, p_{\text{U}}, \epsilon'\right) dG_{\text{U}}\left(\epsilon'\right) := \bar{V}_{\text{U}}^{fi}\left(b_{\text{R}}, p_{\text{U}}\right). \quad \text{(full information)}$$

Even though the urban persistent type always evolves stochastically, its realization is revealed right 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, $\bar{V}_{\text{U}}^{fi}$, 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 $w_{\text{U}}\Lambda e^{p_{\text{U}}+\epsilon}$, and total downgrading in log efficiency units is $c_{\text{RU}}^{+} = c_{\text{RU}} + \log\Lambda$. 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.[15]

Migration Choice Probabilities. The probability of migrating for the imperfect and full information case

[14] This 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).

[15] Section 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.

takes the convenient logit form (McFadden, 1974)

$$\pi_{\mathrm{RU}}^{\mathcal{I}}\left(b_{\mathrm{R}}, p_{\mathrm{U}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right)=\frac{\tau_{\mathrm{RU}}^{-s}\left(-\bar{V}_{\mathrm{U}}^{\mathcal{I}}\left(b_{\mathrm{R}}, p_{j(\mathcal{I})}\right)\right)^{-\frac{s}{\gamma-1}}}{\left(-V_{\mathrm{R}}\left(b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right)\right)^{-\frac{s}{\gamma-1}}+\tau_{\mathrm{RU}}^{-s}\left(-\bar{V}_{\mathrm{U}}^{\mathcal{I}}\left(b_{\mathrm{R}}, p_{j(\mathcal{I})}\right)\right)^{-\frac{s}{\gamma-1}}} \text{ for } \mathcal{I} \in\{pi, fi\}, \tag{7}$$

where the minus term accounts for the fact that the value function is negative for $\gamma > 1$, and $p_{j(\mathcal{I})} = p_{\mathrm{U}}$ in the full information case and p_{R} otherwise.[16]

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

$$\mathbb{E}\left[V_m^{pi} | b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right]=-\left[\tau_{\mathrm{RU}}^{-s}\left(-\bar{V}_{\mathrm{U}}^{pi}\left(b_{\mathrm{R}}, p_{\mathrm{R}}\right)\right)^{-\frac{s}{\gamma-1}}+\left(-V_{\mathrm{R}}\left(b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right)\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}}:=\tilde{V}_m^{pi}\left(b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right), \tag{8}$$

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

$$\mathbb{E}\left[V_m^{fi} | p_{\mathrm{U}}, b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right]=-\left[\tau_{\mathrm{RU}}^{-s}\left(-\bar{V}_{\mathrm{U}}^{fi}\left(b_{\mathrm{R}}, p_{\mathrm{U}}\right)\right)^{-\frac{s}{\gamma-1}}+\left(-V_{\mathrm{R}}\left(b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right)\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}}:=\bar{V}_m^{fi}\left(p_{\mathrm{U}}, b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right), \tag{9}$$

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

$$\tilde{V}_m^{fi}\left(b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right)=\int_{p_{\mathrm{U}}} \bar{V}_m^{fi}\left(p_{\mathrm{U}}, b_{\mathrm{R}}, p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right) dF\left(p_{\mathrm{U}} | p_{\mathrm{R}}\right). \tag{10}$$

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.

Urban-Rural Reverse Migration. Since I focus on stationary equilibria, I need reverse urban-rural migration to balance rural outflows. I assume that a fraction δ_{UR} of newborn urban dynasties are forced to the rural sector, where they draw their persistent and transitory type from the distributions $p \sim N\left(c_{0,\mathrm{R}}, \sigma_{0,\mathrm{R}}\right)$, and $\epsilon \sim N\left(0, \sigma_{\epsilon, j}\right)$, and enter with zero assets.

Modeling reverse flows this way is helpful as the urban household problem can be solved without any reference to the rural household problem, and rural equilibrium outcomes such as migration and the rural wage. Obtaining such a block-recursive structure is particularly desirable in the presence of an endogenous rural-urban wage gap, which raises the computational burden by adding a non-linear fixed point problem. [17]

[16] I 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.

[17] Alternatively, one could generalize the framework and allow for two-way migration using the same machinery. My view is that the reasons for the less common urban-rural migration are different, for instance tied to retirement decisions or family planning. This would require an altogether different model beyond the scope of the paper. In appendix A, I confirm this conjecture by documenting

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

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

where $G_{\text{R}}(x)$ 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 $\tilde{\pi}^{fi}_{\text{RU}}(b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) = \int \pi^{fi}_{\text{RU}}(b_{\text{R}}, p_{\text{U}}, p_{\text{R}}, \epsilon_{\text{R}})\, dF(p_{\text{U}}|p_{\text{R}})$. In the partial information case, there are only unconditional probabilities as realized urban productivity p_{U} is unknown before migration so $\tilde{\pi}^{pi}_{\text{RU}}(b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) = \pi^{pi}_{\text{RU}}(b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}})$.

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

$$\dot{M} \quad = \quad \Pi_{\text{RU}}(1-M) - \phi\delta_{\text{UR}} M, \tag{11}$$

which depends on endogenous rural-urban migration, as well as the reverse flow $\phi\delta_{\text{UR}}M$. 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 $H_{\text{R}} = \int_{i\in\text{R}} e^{p_i+\epsilon_i} di$ and reproducible capital K_{R} using a constant-returns-to-scale Cobb-Douglas production function

$$Y_{\text{R}} = (T)^{\chi} (K_{\text{R}})^{\nu} (H_{\text{R}})^{1-\chi-\nu}\,, \tag{12}$$

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},

$$Y_{\text{U}} = (K_{\text{U}})^{\theta} (A_{\text{U}} H_{\text{U}})^{1-\theta}\,. \tag{13}$$

Summary of the Model. I conclude the description of the economic environment with a short discussion of the rural household problem using a standard HJB equation that highlights the information structure and reveals what makes the model so tractable. The recursive formulation reads

$$(\rho+\phi)\, V_{\text{R}}(b_{\text{R}}, p_{\text{R}}, \epsilon_{\text{R}}) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_{\text{R}})\, \dot{b} + \mathcal{A}_y V_{\text{R}} + \lambda_m \left(\zeta \tilde{V}^{pi}_m + (1-\zeta)\, \tilde{V}^{fi}_m - V_{\text{R}}\right), \tag{14}$$

where $\mathcal{A}_y$ is the infinitesimal generator of the income process, and $\zeta\tilde{V}^{pi}_m + (1-\zeta)\,\tilde{V}^{fi}_m$ is the expected value of a migration opportunity, which intuitively depends on the share ζ of partial information households.

The HJB equation in (14) also shows why my approach is so tractable: the persistent urban productivity type does not enter rural households' state space. Urban outcomes matter, but they are subsumed in the expectations defined in (8) and (10), which collapse into functions of rural productivity alone given the statistical link postulated in (6). Standard dynamic Roy models instead carry each sector's productivity as

how urban-rural migrants differ systematically from rural-urban ones in terms of income, assets, age, and education.

a separate state variable, which quickly becomes intractable. My framework accommodates an arbitrary number of sectors, since households only ever track the permanent and transitory type of their current sector, while still allowing for arbitrary selection into migration including sectoral comparative advantage. This is particularly true for fully-informed households, who select into migration based on rural and urban productivity draws.[18]

The cost of this approach is that whenever migration opportunities arise, households obtain a new urban draw, which must be independent of previous draws so that urban outcomes stay out of the rural state space. I view this as a sensible assumption whenever sectoral switches can be represented as rare opportunities, which appears to be the case in the context of rural-urban migration.

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 $\omega := \frac{w_{\mathrm{U}}}{w_{\mathrm{R}}}$ 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 $w_{\mathrm{U}}\Lambda e^{p+\epsilon}$, so the relevant scale factor is the effective urban wage $w_{\mathrm{U}}\Lambda$ 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.[19] Unsurprisingly, rural-urban migration is increasing in the rural-urban wage gap summarized in the following proposition.

[18] The full-information case can then also be interpreted as an environment where returning to the rural sector is costless right around the time of migration, without explicitly modeling computationally costly two-way migration. A common criticism raised toward models with income risk in the urban sector is that unsuccessful migrants could simply migrate back undermining the role of such risk. If this was the case, then the structural model would infer that the full information case is an appropriate description of the environment.

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

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

$$\pi_{RU}^{\mathcal{I}} = \frac{(\omega\Lambda)^{s} \cdot \tau_{RU}^{-s} \left(-\bar{v}_{U}^{\mathcal{I}}\right)^{-\frac{s}{\gamma-1}}}{\left(-v_{R}\right)^{-\frac{s}{\gamma-1}} + (\omega\Lambda)^{s} \cdot \tau_{RU}^{-s} \left(-\bar{v}_{U}^{\mathcal{I}}\right)^{-\frac{s}{\gamma-1}}} \text{ for } \mathcal{I} \in \{pi, fi\},$$

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 $\tilde{\pi}^{\mathcal{I}}$.[20]

The steady-state urban population share follows from setting $\dot{M}$ in (11) to zero,

$$M = \frac{\Pi_{\mathrm{RU}}}{\Pi_{\mathrm{RU}} + \phi\delta_{\mathrm{UR}}}. \tag{15}$$

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

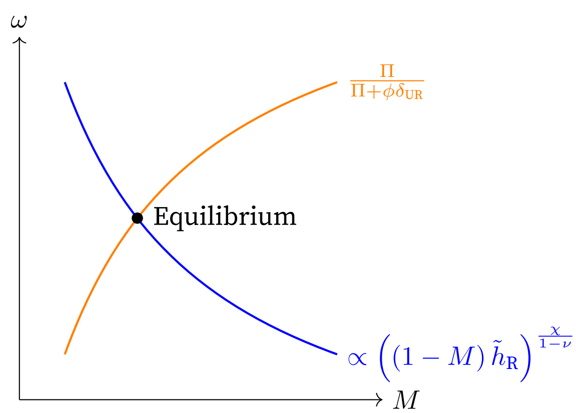
$$w_{\mathrm{R}} = (1-\chi-\nu)\left(\frac{\nu}{r+\delta_k}\right)^{\frac{\nu}{1-\nu}} \left(\left(1-M\right)\tilde{h}_{\mathrm{R}}\right)^{-\frac{\chi}{1-\nu}} \tag{16}$$

$$w_{\mathrm{U}} = (1-\theta)\left(\frac{\theta}{r+\delta_k}\right)^{\frac{\theta}{1-\theta}} A_{\mathrm{U}}, \tag{17}$$

where $\tilde{h}_{\mathrm{R}} = \int h_{\mathrm{R}}(x)\, dG_{\mathrm{R}}(x)$ 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.

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.

[20] In a slight abuse of notation, let $\bar{v}_{\mathrm{U}}^{pi} = \tilde{v}_{\mathrm{U}}^{pi}$, since there are only unconditional expectations in the partial information case.

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.[21]

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.[22]
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 is rare, justifying the focus of the paper on permanent migration.[23]

[21] I 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.

[22] Table 38 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)).

[23] I 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.

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 Blundell, Pistaferri, and Preston (2008) 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. I 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. Table 21 in the appendix contains a detailed breakup of sample selection by sector, and table 22 reports summary statistics for both the raw sample and the selected sample.

I first 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.

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

$$y^*_{i,t} = p_{i,t} + \epsilon_{i,t} + x_{i,t}, \tag{18}$$

The autocovariance moments implied by (18), and conditional on household survival, equal

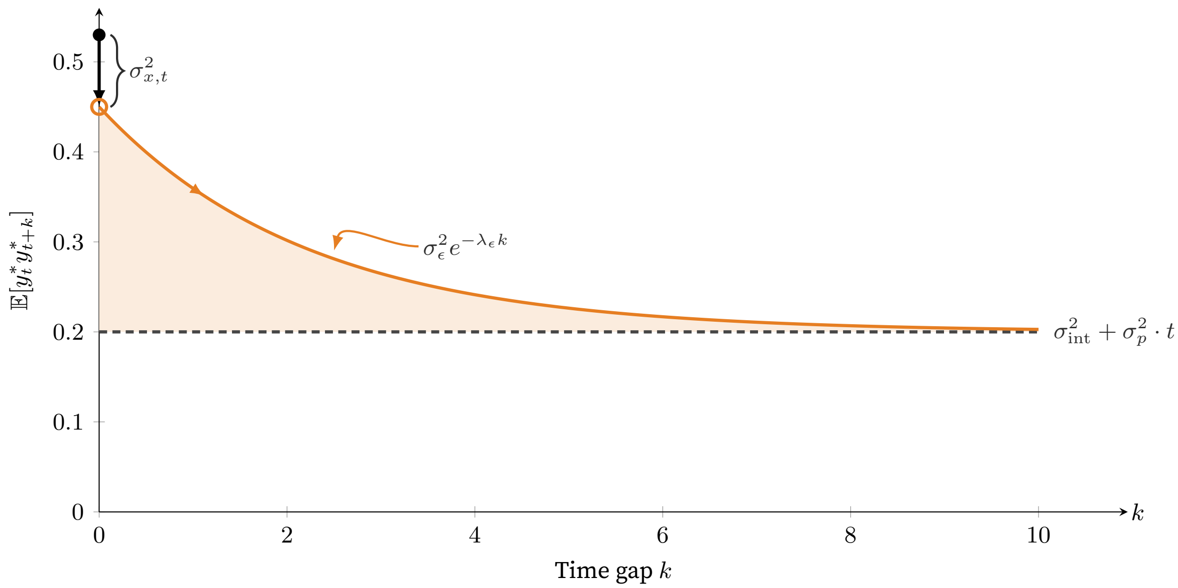
$$\mathbb{E}\left[y^*_{t_1} y^*_{t_2}\right] = \sigma^2_{\text{int}} + \sigma^2_p \cdot t_1 + e^{-(t_2-t_1)\lambda_\epsilon}\sigma^2_\epsilon + \mathbb{1}_{\{t_1=t_2\}}\sigma^2_{x,t_1} \tag{19}$$

where $\mathbb{1}$ is an indicator function accounting for classical measurement error that appears only on diagonal entries of the covariance matrix, and $t_1 \leq t_2$ are time indices relative to an initial period normalized to zero. The autocovariances are fully determined by the set of parameters $\theta = \{\sigma^2_{\text{int}}, \sigma^2_p, \lambda_\epsilon, \sigma^2_\epsilon, \{\sigma^2_{x,t}\}\}$.[24]

[24] The focus on level moments, in contrast to first differenced income growth rates as in BPP, offers three benefits. First, level moments are more robust according to Crawley, Holm, and Tretvoll (2022) in the presence of model misspecification. Second, there are more level moments than difference moments, which matters for the short sample I have to work with. Third, they identify initial dispersion for entering cohorts needed for the calibration.

Figure 2 plots this autocovariance against the time gap $k = t_2 - t_1$, and shows how each parameter is separately identified. Persistent shocks set the level. Because the random walk has no mean reversion, σ_p^2 raises the covariance between two waves regardless of how far apart they are, and therefore pins down the level the profile approaches at long gaps. This can be seen in how the covariance asymptotes to the dashed black line. Passing shocks account for the decline in the covariance: they contribute $e^{-k\lambda_\epsilon}\sigma_\epsilon^2$, so their dispersion σ_ϵ^2 fixes how far above that level the profile starts at short gaps, while λ_ϵ fixes how fast it falls back toward it. Classical measurement error appears only on the diagonal, $t_1 = t_2$, and shows up as a vertical drop between $k = 0$ and $k > 0$ that leaves the rest of the profile untouched.

Figure 2. Identification of Income Process



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

Before turning to the results, one caveat needs to be addressed. When estimating income risk, I focus on households that do not move across sectors, based on the idea that different economic environments entail different economic risks. Focusing on a sample of stayers, however, could lead to selection bias. It turns out that because migration is rare, endogenous selection induces a negligible bias regarding the last four parameters $\{\sigma_p^2, \lambda_\epsilon, \sigma_\epsilon^2, \{\sigma^2_{x,t}\}\}$, in the sense that simulated autocovariance moments based on the model with endogenous selection into migration differ by less than 1% from the simple GMM procedure used here that ignores this issue of selection. In contrast, selection is quantitatively important when decomposing initial inequality at the beginning of the sample, σ^2_{int}, into inequality at birth (σ_0^2) vs. income shocks as households age ($\sigma_p^2 \cdot age$).[25] This is why inequalities at birth, $\{\sigma_{0,\text{R}}, \sigma_{0,\text{U}}\}$, have to be inferred from the

[25] If there was no migration, persistent inequality at the beginning of the sample would simply be a linear function of inequality at birth and average age of the population ($\widetilde{age}$) times the volatility of persistent shocks,

$$\sigma^2_{\text{int}} = \sigma_0^2 + \sigma_p^2 \cdot \widetilde{age}.$$

However, this simple decomposition is no longer true in the presence of rural-urban migration. Note that by construction, rural-urban migrants are older than urban newborns, and when they draw their urban type they obtain a draw from a distribution that differs from the newborn urban distribution. Dropping migrants from the sample does not solve this problem because some urban households, that appear to be stayers, might have migrated to the urban sector before the sample starts in 2008.

minimum distance estimator that uses the full equilibrium model in the next section.

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

$$\hat{V} = \hat{\mathbb{E}}_i \left[\boldsymbol{y_i^*} \boldsymbol{y_i^*}' \right]$$

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. Turning to the results, table 2 reports the empirical autocovariance matrices. Reading down a column increases the time gap between the two waves; moving down a sub-diagonal holds the gap fixed and shifts both waves later in time. The two directions identify the two shocks separately.

First, reading down a column of the left rural panel, the autocovariances decline as log income is measured further apart in time. This decay is the signature of passing shocks: its magnitude identifies σ_ϵ^2 and its speed identifies λ_ϵ. Glancing at the right panel, the same decline is much weaker in urban, indicating a larger role for imperfectly transitory shocks in the rural sector.

Second, persistent shocks lead to a fanning out of inequality over time, since the covariance between two waves a fixed distance apart grows the later the pair is observed. Moving down the first sub-diagonal of the urban matrix, the autocovariances rise, while the same sequence in the rural matrix is flat. This fanning out identifies σ_p^2, and suggests that persistent income shocks are more important in urban than in rural South Africa.

Table 2. Income Covariance Matrix – South Africa

(a) Rural

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.52				
y_{t+2}	0.27	0.55			
y_{t+4}	0.23	0.27	0.49		
y_{t+6}	0.22	0.18	0.25	0.48	
y_{t+8}	0.18	0.16	0.23	0.34	0.57

(b) Urban

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.52				
y_{t+2}	0.27	0.63			
y_{t+4}	0.24	0.26	0.58		
y_{t+6}	0.26	0.25	0.32	0.53	
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.

The GMM routine infers the parameters of the income process by minimizing the distance between the empirical covariance matrices and the theoretical moments from (19). To implement the GMM routine, I use an optimal weighting scheme, and select relevant moments from the covariance matrix that are below the diagonal entries since the diagonal entries are confounded by time-varying measurement error.[26]

The point estimate of the variance of the random walk component in the rural sector is small and equals 0.003 . In contrast, the urban sector exhibits a volatility of around 0.0117 , comparable to estimates for the US, see Crawley, Holm, and Tretvoll (2022), and roughly four times as large as in rural South Africa.

In contrast, passing shocks are more important for rural households. A large variance of around $\sigma_\epsilon^2 = .26$

[26]Income is measured on a monthly basis for the South African sample so the error induced by time aggregation is negligible. The code also allows for time-aggregated data relevant whenever income or consumption are averaged over an entire year.

together with strong persistence of $\lambda_\epsilon = .5$,[27] imply that imperfectly transitory shocks are an important source of income volatility in rural South Africa. The role of imperfectly transitory shocks in the urban sample is much smaller, as indicated by the low variance of these shocks. More generally, the urban sample does not have much power to tell apart a high λ_ϵ from a low variance of shocks σ^2_ϵ since in either case passing shocks don't matter.

Lastly, I infer measurement error as a residual from the diagonal of the covariance matrix given estimates for the parameters of the income process, which appears overall large, and I stress that fully transitory shocks and measurement error are not separately identified (Friedman, 1957). Measurement error will be important later on when computing passthroughs from income shocks to consumption, which are sensitive to the degree of measurement error.

Table 3. Income Risk in South Africa

	Rural		Urban	
	Estimate	SE	Estimate	SE
σ^2_p	0.003	0.0065	0.0117	0.0045
λ_ϵ	0.4987	0.2611	0.3333	0.7473
σ^2_ϵ	0.2602	0.0987	0.05	0.0415
σ^2_{int}	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 $\{\sigma^2_{x,0}, \sigma^2_{x,2}, \sigma^2_{x,4}, \sigma^2_{x,6}, \sigma^2_{x,8}\}^{\text{R}} = \{$ 0.1088 , 0.1213 , 0.0515 , 0.0482 , 0.1161 $\}$, and $\{\sigma^2_{x,0}, \sigma^2_{x,2}, \sigma^2_{x,4}, \sigma^2_{x,6}, \sigma^2_{x,8}\}^{\text{U}} = \{$0.1678 , 0.2583 , 0.17 , 0.0902 , 0.0962 $\}$.

The difference in permanent volatility across the two sectors is an important input into the quantification, and I therefore do not want to rest it on relatively noisy estimates alone. First, I build on the approach of Deaton and Paxson (1994), which leverages the insight that persistent shocks generate a systematic increase in within-cohort inequality over the life cycle. The strength of this approach is that it can be implemented using cross-sectional data alone. Consistent with the panel evidence, I find that inequality within cohorts fans out at a higher rate in the urban sector in several other emerging economies, including Mali and Côte d'Ivoire, see appendix A.4. Second, I have explored additional panel evidence from another large emerging market, China, which yields very similar results to the South African case.[28]

A key assumption throughout is that income risk is a property of the sector, and not of the individual household. The small migrant sample makes it infeasible to estimate an income process for migrants alone, which would address the question directly. Two pieces of indirect evidence speak in its favor. First, the estimated income process for urban South Africa looks very similar to the one estimated for the United States, even though the share of recent rural-urban migrants is far larger in urban South Africa than it is in the US. If formerly rural households brought a different income process with them, we would expect the two to look quite different. Second, recent migrants in South Africa do not experience faster income growth than households that have always resided in the urban sector, so the two groups appear to be on similar mean income trajectories.

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

[27] This degree of persistence corresponds to a coefficient of autocorrelation of about $.6 = e^{-.5}$.
[28] Results available upon request.

Read this way, the classic dual-sector perspective of Lewis (1954) extends naturally to risk: the two sectors differ not only in the level of income they offer, but in the kind of risk that comes with it. And because that difference is a feature of the sector rather than of the household, a household that moves to the city takes on the urban income process along with the urban wage.

The analysis so far has only focused on income dynamics among households that remain in the same sector. I next turn to rural-urban migration and rural-urban income gaps, which provide additional key moments for the quantification.

3.1.3 Rural-Urban Wage Gaps and Selection into Migration

I first report rural-urban wage gaps, before I analyze how migrants are selected relative to rural stayers. I now use less strict sample restrictions to keep the sample size large given relatively few migration events.[29]

Rural-Urban Wage Gaps. Table 4 reports the cross-sectional rural-urban wage gaps. The raw rural-urban wage gap, which is omitted from the table, is around 70 log points. After controlling for observables, the wage gap falls to around 45 log points.[30] Since my model does not feature an endogenous schooling choice, this residual wage gap is the relevant object to target. The second row reports results based on a regression with household fixed effects, which further reduces the gap to about 30 log points, in line with the previous literature, see Lagakos et al. (2020).[31]

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)

An important theme of the paper is that average wage gaps do not contain sufficient information to understand the gains and costs associated with rural-urban migration. Figure 3 plots the entire empirical income distribution in each sector after controlling for age and education, and makes the following point: the two distributions overlap a great deal, so many rural households earn more than their urban counterparts. Migrants could therefore conceivably end up with an urban income below what they earned in the rural sector, despite the positive mean gap.

Selection into Migration. I explore how rural-urban migrants are selected relative to rural stayers by running a simple cross-sectional regression amongst all households in the rural sector

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

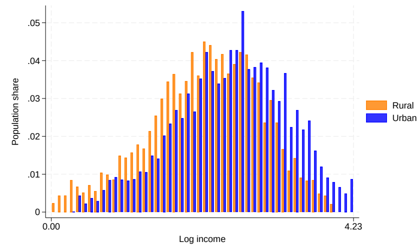
where $D\left(i \in \text{RU}\right)$ is a dummy variable that takes a value of one if the household migrates permanently to

[29] I 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.

[30] A 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.

[31] See 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.

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.

the 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 reports the results, and shows that selection into migration is negative in the sense that income of future migrants is 30% lower than the income of stayers when both groups are present in the rural sector. After controlling for observables, the coefficient remains large at around -15%. Using log consumption instead of log income leads to extremely similar results, suggesting an important role for permanent income differences.

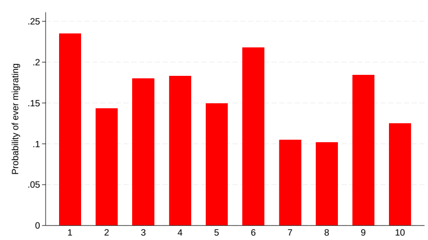
Table 5. Selection into Migration

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 illustrates how the likelihood of migration falls in household income when split into deciles.

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 .

The negative selection of permanent rural-urban migrants contrasts with the influential argument in Young (2013), suggesting that rural-urban wage gaps are explained by the sorting alone. A key piece of evidence in Young (2013) is that migrants are more educated than rural stayers. I replicate a virtually identical educational gap in the South African data: movers have roughly two more years of schooling than stayers. However, the gap is a cohort effect. Migrants are young, and younger South African cohorts are far better educated than older ones. The schooling difference between movers and stayers disappears fully once age is controlled for, see table 22b in the appendix. Conditional on age, rural-urban migrants are no more educated than rural stayers, yet they remain negatively selected on both income and consumption.

Rural-Urban Income Correlation. If urban income is partly unknown, the correlation between rural and urban income among movers becomes an important empirical object to discipline the model. I estimate the reduced-form link between rural and urban income using a simple linear regression of urban income on lagged rural income,

$$y_{i,\mathrm{U},t} = \alpha + \beta y_{i,\mathrm{R},t-2} + u_{i,t}. \tag{20}$$

Since measurement error and transitory shocks induce a downward bias, I estimate equation (20) using an IV approach with lagged income in $y_{i,t-4}$ as an instrument for $y_{i,t-2}$. Table 6 reports the results for both OLS and IV. The IV estimate is around .6, and twice as large as the OLS estimate. This points toward the sizable role of measurement error, and highlights the importance of differentiating between transitory and persistent shocks. The structural model will go further and show that even the IV estimate does not have a clean structural interpretation, i.e., is not exactly equal to ρ_{RU}. The IV, however, is essential to purge the empirics from time-varying measurement error.

Table 6. Rural-Urban Wage Gaps

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

Table 28 in the appendix reports additional details, and implements a placebo test for stayers. The IV estimate for stayers is substantially higher, at around .8.

Income Gains, Losses, and Volatility. I next document how likely it is that migrants improve their income after moving, and I compute the related moment of income volatility around the time of migration. While these moments are rarely reported, they are central to tease out how risky migration is together with the structural model.

Figure 5 compares household income and consumption pre and post migration, and only migrants on top of the 45-degree line have improved their income. Two findings stand out. First, not all households gain from migration: while most observations lie above the 45-degree line, a non-trivial share of migrants experience income and consumption losses after migration. Second, there is substantial dispersion in urban outcomes for rural households with initially similar income or consumption, as can be gauged by the low predictive power of initial income.

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,

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

where $D_{i,t}\left(\mathrm{U}\right)$ 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 in table 7 reports the result: the probability of income gains is about 20% higher at the time of moving. Taken together with the baseline rate of improving your income with

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.

a probability of 55%, this still leaves a substantial share of around 25% of households who don't improve their income upon migration.

Tightly related to the possibility of income losses is the overall increase in volatility associated with migration relative to rural stayers. To put this volatility into perspective, I compare it to annualized income growth associated with migration, using the following two regression models

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

where the first row estimates income growth around migration. The second row computes the variance of income growth after residualizing the data appropriately so that the variance isolates dispersion around each group's mean income growth. The final two columns in table 7 report the results on income growth and volatility. Reassuringly, mean growth is positive, consistent with the sizable urban income premium estimated earlier.

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 29 and 31 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 also spikes around the time of migration, consistent with risky rural-urban migration. At the same time, income volatility is large even for rural households who do not migrate ($SD = \sqrt{0.12} \approx .35$). Importantly, income volatility for stayers is mostly driven by passing shocks, which will be more easily insured using bufferstock savings.[32] This is precisely why the level of volatility is not by itself informative

[32] The volatility of stayers closely matches the volatility implied by the stochastic process estimated earlier. To see this, use the

about risk exposure. Whether the spike around migration is different in kind depends on whether it reflects persistent shocks. Table 31 in the appendix reports differences in mean consumption and consumption volatility across groups. Consumption volatility increases by a similar magnitude as income volatility around migration, which is what one would expect if the underlying shocks are persistent. I will return to the model's ability to match rural and urban consumption patterns in section 3.3, and use the model to decompose the increase in volatility into transitory and persistent shocks.

3.2 Model Calibration and Estimation

I discipline the model with a blend of calibration and indirect inference. A small set of parameters is fixed using external evidence or accounting normalizations. The remaining primitives are estimated by simulated method of moments (SMM). Below I list the parameters taken from external sources and the normalizations, describe the set of parameters estimated internally, and summarize which empirical moments most directly identify each parameter.

3.2.1 Calibrated and Externally Estimated Parameters

Preferences and Production. The interest rate of the risk-free bond is set to be $r = 2.7\%$.[33] The discount factor equals ρ =0.0200 , and the coefficient of relative risk aversion is set to γ =2.0000 , consistent with recent evidence in Choukhmane and Silva (2022).

In a stationary environment, rural outflows balance rural inflows. For a value of ϕ chosen below, the value of δ_{UR} consistent with a stationary equilibrium thus equals $\delta_{\mathrm{UR}} = \frac{\Pi_{\mathrm{RU}}}{\phi}\left(\frac{1-M}{M}\right)$, which only depends on known parameters or directly observable objects such as the aggregate rural-urban migration rate, Π_{RU}, as well as the rural-urban population ratio. Given a migration rate of about 2% together with an urban share of .65 implies δ_{UR} =0.3500 .

Urban production follows the standard neoclassical model with a reproducible capital share of 40%, a labor share of 60% and a depreciation rate of $\delta_k = 5\%$. Rural production follows Hansen and Prescott (2002) with land, capital, and labor share equal to 30%, 10%, and 60%, respectively. This calibration lines up well with Gollin (2002), which documents that the labor share is roughly stable across the development spectrum.

Income Process. I pick the parameters governing the income process consistent with the estimates from the empirical section. I chose a death shock ϕ =0.0300 consistent with an average career of 33 years as in Gabaix et al. (2016).[34] Further, I set average human capital amongst urban newborns equal to $c_{0,\mathrm{U}} = -.4$. This generates a realistic urban income distribution, appearing log-normal over most of its domain but with a thick right tail. Choosing $c_{0,\mathrm{U}}$ externally does not materially affect the analysis as the key parameter governing unobserved human capital differences across sectors is the gap between urban and rural initial average log human capital, $c_{0,\mathrm{U}} - c_{\mathrm{R},0}$. This gap is inferred in the structural estimation because $c_{\mathrm{R},0}$ is al-

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: $\mathbb{V}(\Delta y/2) = \sigma_p^2/2 + \left(1 - e^{-2\lambda_\epsilon}\right)\sigma_\epsilon^2/2 + \sigma_x^2/2 \approx 0 + \left(1 - e^{-1}\right) * 0.26/2 + .1/2 \approx .13.$

[33] Interest rates are important because they shape to what extent endogenous wealth inequality impacts consumption inequality. In related work I consider a general equilibrium version of the theory where land-intensive production sustains high interest rate, and explore their implications for consumption inequality and intergenerational mobility, see Trouvain (2025a).

[34] This stabilizes the income distribution and implies a tail-coefficient of log permanent income of $\sqrt{\frac{2\phi}{\sigma_p^2}}$. 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 37 in the appendix.

lowed to vary freely. Aside from this, the rural and urban income processes are disciplined by the estimates reported in the previous section.

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.[35] I calibrate these two parameters as follows. First, I choose an arrival rate of migration of $\lambda_m = 0.2000$. 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).[36] 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 $f_{\text{RU}} = 0.0460$ in my quantification. If the fixed cost is much higher, the transfer would not impact rural-urban migration relative to a control group. Similarly, if the fixed cost was much lower, non-treated households would also be able to migrate and again there would be no measurable gap between treatment and control group. 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.

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 (2022)
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		
$\sigma^2_{p,\text{R}}$	0.0030	own estimation
$\lambda_{\epsilon,\text{R}}$	0.4987	own estimation
$\sigma^2_{\epsilon,\text{R}}$	0.2602	own estimation
$\sigma^2_{p,\text{U}}$	0.0100	own estimation
$\lambda_{\epsilon,\text{U}}$	0.3333	own estimation
$\sigma^2_{\epsilon,\text{U}}$	0.0500	own estimation
$c_{0,\text{U}}$	-0.40	urban income distribution

[35] The 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).

[36] See also Bryan, Chowdhury, and Mobarak (2014) and Balboni et al. (2022).

3.2.2 Internally Estimated Parameters

I next select several empirical moments guided by the theoretical framework to estimate the remaining parameters using the simulated method of moments (SMM). Even though all parameters

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

are jointly identified in the SMM routine, certain moments are especially informative about particular parameters. I highlight those links to provide intuition and motivate the inclusion of the respective moments.

The cross-sectional rural-urban wage gap helps identify a newborn rural cohorts' average permanent type at entry, $c_{0,\mathrm{R}} < 0$. Intuitively, differences in average human capital are sensitive to newborn rural cohorts' average permanent type at entry.[37] The rural-urban wage gap, after controlling for household fixed effects, is particularly sensitive to the migration wedge τ_{RU}. A higher wedge requires higher real wage gain so households are willing to move, which implies a higher fixed effects estimator. Skill-downgrading c_{RU} is intimately linked to negative selection into migration, as skill-downgrading does not impact the least-skilled rural households due to a natural lower bound on the persistent productivity draw. That is, a higher skill loss upon migration only impacts sufficiently skilled migrants.

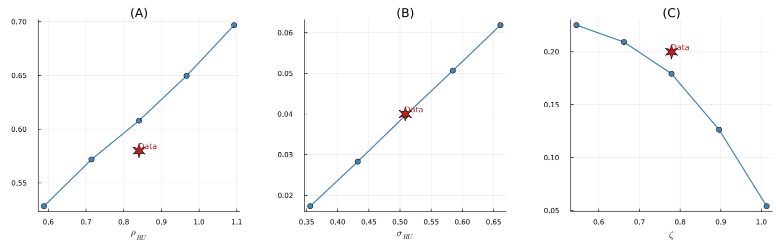
In contrast, the parameter Λ is identified by the worst urban outcome facing the lowest-income rural households. This is a moment I impose rather than measure: I require that such a household loses 5% of income upon moving. An adjustment of this kind is necessary because the rural-urban wage gap satisfies $\omega > 1$, so that without it even the worst possible urban draw would leave the poorest rural household better off than staying, and migration would be a guaranteed gain. Estimating this number directly is difficult, since extreme outliers are dropped from the analysis to begin with, so it is best read as a normalization that keeps some downside risk present at the bottom of the rural distribution.

Relative to the previous literature, there are three novel parameters: the average link between rural and urban productivity, ρ_{RU}, the extent to which this average is informative about actual urban outcomes parameterized by the volatility, σ_{RU}, and finally, the share of households that do not know their urban persistent draw ex-ante, ζ. To identify ρ_{RU}, I use the reduced-form link between rural and urban income based on the IV estimates of about .6. To identify the volatility of urban draws relative to the mean, I match the difference in income volatility of stayers relative to movers of about .04. And finally, to identify the share of households without ex-ante information of urban persistent type, I match the share of households that improve their income around the time of migration. Intuitively, if more households had ex-ante information, the share of households doing worse in the urban sector would fall.

Figure 6 confirms this logic by plotting the link between simulated moment and respective parameters while holding all other parameters fixed at their optimal value. At the same time, all parameter estimates are interrelated. For example, changes in volatility also impact the share of households that experience income losses in the urban sector in the sense that an increase in σ_{RU} makes good and bad tail outcomes more likely.

[37] One interpretation for $c_{0,\mathrm{R}} < c_{0,\mathrm{U}}$ is the superior quality of schooling in urban, recall I condition on total years of schooling.

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 concern preference heterogeneity s, and initial inequality $\{\sigma_{0,\text{R}}, \sigma_{0,\text{U}}\}$. To identify the degree of preference heterogeneity, note that for extreme levels of taste heterogeneity ($s \rightarrow 0$), migration choices would be unrelated to expected income gains, and identical across households that can afford the fixed cost. Conversely, for $s \rightarrow \infty$, migration choice probabilities feature strong selection and migration probabilities become polar in the sense that households of a certain type x always or never migrate, i.e., $\pi(x) \in \{0, 1\}$. A helpful moment to estimate the degree of preference heterogeneity is thus the rate at which migration rates decline in income as parameterized by the decline from lowest to highest income decile, which is about .0075. Intuitively, this slope asymptotes to zero when taste shocks become large. Lastly, initial inequality is inferred by finding an initial degree of dispersion that is consistent with measured inequality in each sector. This moment is quite sensitive to selection into migration, which is why it is included in the structural estimation routine.

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

$$\mathcal{L}(\theta) = \min_{\theta \in \Theta} \sum_k \frac{|\text{model}_k - \text{data}_k|}{\frac{1}{2}\left(|\text{model}_k| + |\text{data}_k|\right)} \tag{21}$$

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 collects all empirical moments, and compares them to the simulated model-based moments. The model replicates salient features of rural-urban migration in South Africa, and in particular novel moments related to the correlation of rural and urban income, and the likelihood of an income loss upon migration.

Table 10 shows the estimated parameter values that minimize the loss function.

Three parameters are of particular importance to understand the risky nature of migration, namely the statistical link between rural and urban income captured in ρ_{RU} and σ_{RU}, as well as the share of households that migrate without prior knowledge of their persistent urban type parameterized by the share ζ.

Turning to the statistical link between rural and urban income, ρ_{RU}, note that the parameter estimate is substantially higher than the reduced form coefficient from the IV regression. This follows from three aspects of the environment. First, mean-reverting passing shocks induce a downward-bias in the estimator similar to classic measurement error. Second, note that the effect of skill-downgrading, c_{RU}, is inherently asymmetric in its impact along the income distribution. Skill-downgrading represents a shift downward on the income grid for higher-income rural-urban migrants. This downward shift, however, disappears for the least fortunate households as they are already at the very bottom of the distribution. For these households,

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	$c_{0,\mathrm{R}}$	c_{RU}	Λ	ρ_{RU}	σ_{RU}	ζ	s	τ_{RU}	$\sigma_{\mathrm{R},0}$	$\sigma_{\mathrm{U},0}$
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.

migration largely represents upside risk, which weakens the link between rural and urban income. Third, a fraction of households $1-\zeta$ knows their urban type ex-ante. The ensuing selection weakens the link between rural and urban income as only households that actually have received a relatively high urban draw move to the urban sector.

The standard deviation around expected urban persistent type, σ_{RU}, is large and about $.5$. This represents substantial risk for households that don't observe their urban persistent type ex-ante. This finding may be surprising in light of the seemingly small interaction term of migration and volatility of only $.04$ in table 7 in the empirical subsection.

In the absence of a structural model, this would have been easily misinterpreted as migration being not particularly risky. The key to this result is the distinction between imperfectly transitory and persistent shocks. Imperfectly transitory shocks are larger in the rural sector, which, ceteris paribus, would lead to a fall in income volatility amongst movers. At the same time, a large increase in the volatility of the persistent income component raises volatility. Together, this leads to an increase in volatility of around .04. When I use the model to simulate the volatility of the persistent type p alone, it ends up substantially higher and equal to 0.06 . This number lines up nicely with the standard deviation of $.5$ since volatility is computed as variance in annualized form over two years, i.e., $\mathbb{V}(\Delta p/2) = .5^2/4 = 0.0625$.[38]

Regarding information about initial urban persistent type, I estimate ζ =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. In addition, note that rural-urban migrants that improve their income by less than 10% would have preferred

[38] This result is robust to allowing for a correlation structure between rural and urban transitory type. In fact, if rural and urban transitory type were perfectly correlated, then the share of income volatility which is due to transitory shocks amongst rural-urban movers would be even lower. This follows from noting that if the transitory type remains the same, it contributes nothing to the overall volatility based on first-differences. Consequently, an even higher share of the empirically measured volatility would have to accrue to the persistent component, making migration more risky.

to stay in the rural sector, since the migration wedge τ_{RU} is estimated to be about 10%, which further raises the share of households that wouldn't have moved had they had full information. In sum, sizable income gains of migration of 35% are consistent with large risks that manifest themselves in disparate outcomes leaving a substantial share of migrant households worse off.

The estimation also pins down the taste dispersion parameter s at around 8.6. This is larger than what is commonly found in spatial models, but consistent with the estimates in LMW. To match the strong degree of negative selection into migration, a low degree of preference heterogeneity is necessary, which maps into a large shape parameter s. Initial inequality at birth in rural equals 0.14 , which is sensible and lines up with the empirical estimate. In the urban sector, initial inequality at birth is estimated to be very low at 0.0100 . This low estimate is due to the impact of endogenous rural-urban migration on urban inequality. Since migration is risky and induces uneven outcomes, measured inequality in the urban sector is inflated by migration. Consequently, to match observed urban inequality, initial inequality at birth needs to be very low. Appendix 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.

3.4 Selection into Migration

Income. While the model matches negative selection on income by design, it can be used to ask whether that selection operates through the persistent or the transitory type. Figure 7 plots the migration choice probabilities π_{RU} as a function of the persistent and the transitory income type separately, i.e.,

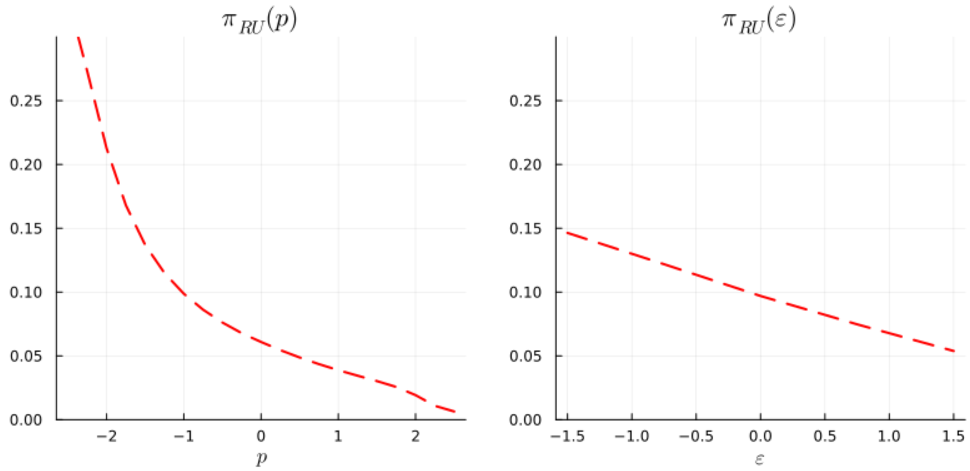
$$\pi_{\text{RU}}(p) := \frac{\int_\epsilon \int_b \pi_{\text{RU}}(b, p, \epsilon)\, dG_{\text{R}}(b, p, \epsilon)}{\int_\epsilon \int_b dG_{\text{R}}(b, p, \epsilon)},$$

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 is a natural implication of the lottery-like aspect of rural-urban migration, where low-income rural households have relatively more to gain than high-income rural households. It is this simple yet quantitatively powerful channel that generates the negative selection patterns found in the data.

The right panel plots the link between migration and transitory income states, which is also negative. Intuitively, since transitory shocks are uncorrelated across sectors, a household with a low rural transitory type has a higher incentive to move. At the same time, we see that the gradient at which migration probabilities fall in transitory income is much less steep compared to the persistent type. The reason is that a low transitory draw fades relatively quickly, in contrast with a low persistent type. Persistent shocks are the key determinant of long-run consumption, and hence the far more important driver of migration.

Figure 7. Selection into Migration – Income



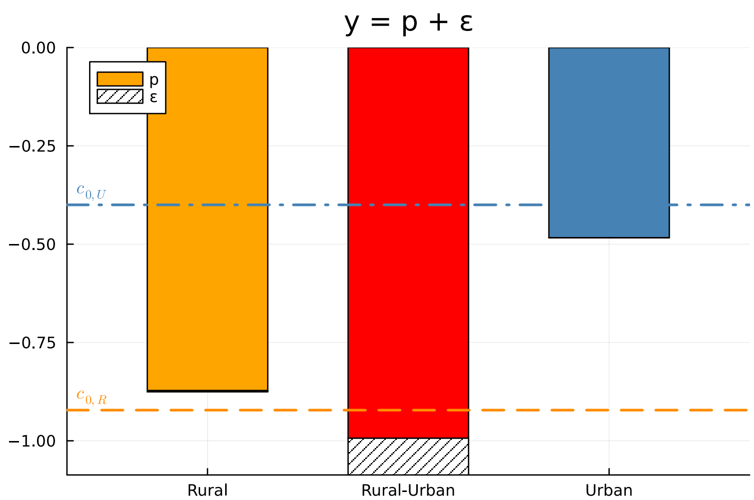
I compute unconditional migration probabilities using $\pi_{\mathrm{RU}}(p) = \int_{\epsilon}\int_{b} \pi_{\mathrm{RU}}(x)\, dG_{\mathrm{R}}(b,\epsilon)$, which gives the unconditional probability that a household of type p seizes the migration opportunity.

Taking account of the negative selection into migration is important to understand the human-capital gap across sectors. Figure 8 plots the average persistent and transitory type of rural households, rural-urban migrants while still in rural, and urban households. The plot illustrates that the cross-sectional gap between rural and urban households, i.e., the gap between the left and right bar, understates the gap in human capital between newborn rural and newborn urban households, $c_{0,\mathrm{U}} - c_{0,\mathrm{R}}$. This discrepancy is due to negative selection into migration, which makes the rural sector appear more productive, while relatively low-income migrants pull down average human capital in the urban sector. Adding the sectoral wage gap to the corrected human-capital gap gives the expected income difference between a newborn urban and a newborn rural household,

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

This gap is well above the measured cross-sectional wage gap of 0.45 and the fixed-effects gap of 0.35 . Consequently, my framework suggests that being born in the rural sector imposes an even larger burden than implied by simple regression-based approaches, which are confounded by endogenous migration.

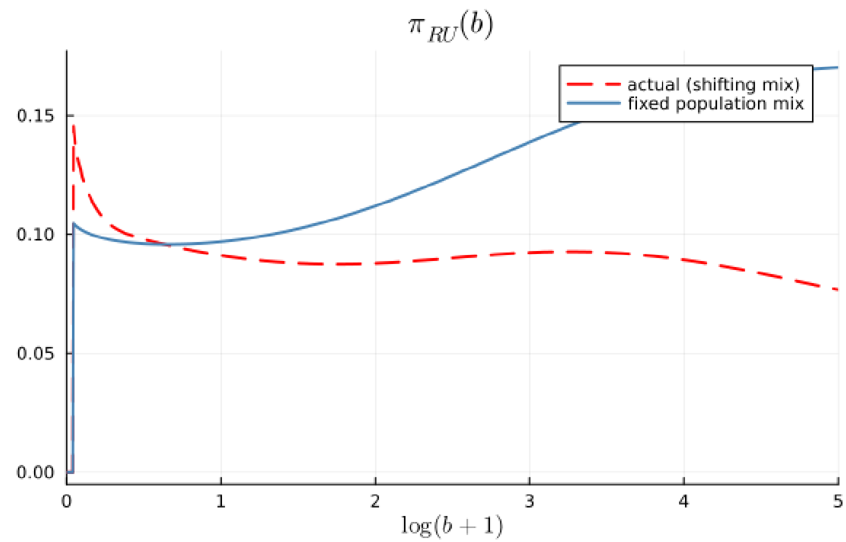
Figure 8. Unobserved log human capital



To compute the average amongst migrants, note that the distribution of rural-urban migrants follows from Bayes' law $dG_{\text{RU}}(x) = dG_{\text{R}}(x)\,\pi_{\text{RU}}(x) \,/ \int \pi_{\text{RU}}(x)\,dG_{\text{R}}(x)$. 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 in figure 9 represents the unconditional migration probability that a household with a certain amount of assets chooses to migrate, given the opportunity. We see a jump in the migration probability as soon as households are able to cover the fixed cost. From then on, the link between wealth and migration is non-monotone. The dynamics are somewhat hard to grasp as income types are changing with assets, so I compute migration probabilities in assets, holding the distribution of income fixed in the solid blue line. Again, an initial jump is followed by non-monotone patterns that are somewhat surprising, and point to a confluence of factors.

Figure 9. Selection into Migration – Wealth



I compute unconditional migration probabilities using $\pi_{\mathrm{RU}}(b) = \int_p \int_\epsilon \pi_{\mathrm{RU}}(x)\, dG_{\mathrm{R}}(p, \epsilon)$, 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., $g_{\mathrm{RU}}(x) = \pi_{\mathrm{RU}}(x)\, g_{\mathrm{R}}(x) \,/\, \int \pi_{\mathrm{RU}}(x)\, dG_{\mathrm{R}}(x)$.

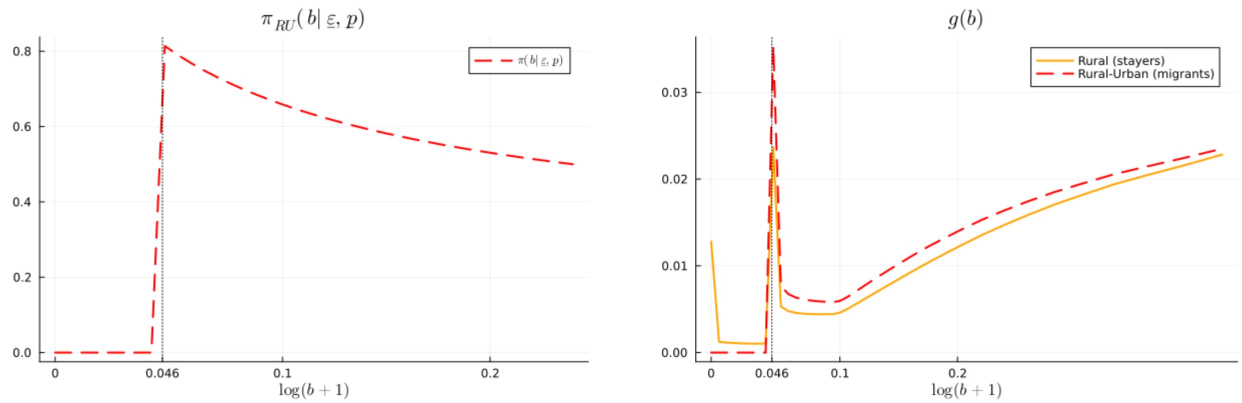
The non-monotonic link between wealth and migration reflects three channels. Two of them raise migration. First, wealth allows households to afford the fixed migration cost. Second, and consistent with the logic in Bryan, Chowdhury, and Mobarak (2014), migration is risky, and wealth reduces the impact of migration risk on consumption volatility. Appendix B.7 shows in closed form, using first and second-order approximations, how migration risk falls as wealth increases.

The third channel works the other way, especially so for households with low levels of labor income. The reason is that as wealth increases, the share of labor income falls mechanically. Yet, the gains from rural-urban migration are tied to increases in labor income. If labor income becomes an unimportant source of income, rural-urban migration becomes unattractive. The left panel of figure 10 is a case in point, and shows the migration probabilities of the lowest-income rural households as a function of wealth. The migration probability spikes once the household can cover the fixed cost, and then declines monotonically thereafter.

This labor-income-share channel depends on two assumptions. First, preferences are CRRA. 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.[39] Second, the migration wedge appears as an urban consumption disutility. This cost affects consumption financed by labor or asset income equally, yet the gains from migration for high wealth households are much smaller. Together, this explains the negative link between migration and wealth for rural households with low labor income.

[39] 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.

Figure 10. Migration and Wealth



LMW read the negative selection of migrants on wealth as evidence that risk is not an important barrier. This requires wealth to work only through risk-bearing ability. Yet, wealth also lowers the share of labor income in household consumption, and thus shrinks the gains from migration. The two channels pull in opposite directions, so the sign of the wealth gradient is uninformative about the importance of risk. In fact, I will show that risk is a first-order barrier to migration in my model. This is in spite of the negative link between wealth and migration, which the model reproduces, see the right panel of figure 9: there is substantially more mass in the left tail of the migrant wealth distribution than in that of stayers.[40]

The model also predicts that the effect of an unconditional transfer is non-monotone in initial wealth for low-income households: 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 labor share.

3.5 Consumption Insurance

I next turn to consumption risk. Income risk matters only when it translates into consumption risk. Whether risk deters migration therefore depends on households' ability to self-insure.

One approach, pioneered by Townsend (1994), computes the passthrough of idiosyncratic income shocks on household consumption, which is commonly used to infer the degree of consumption insurance. A related but more structural approach is pursued in BPP, which isolates the passthrough of persistent income shocks on consumption. The estimators are defined as follows

$$\beta^{\mathrm{T}}_{c\_y} = \frac{Cov(c_{i,t}, y_{i,t}|\alpha_i, \gamma_t, x_{it})}{\mathbb{V}\left(y_{i,t}|\alpha_i, \gamma_t, x_{it}\right)}, \tag{Townsend}$$

$$\beta^{\mathrm{BPP}}_{c\_p} = \frac{Cov\left(\Delta c_{i,t}, \Delta y_{i,t-2} + \Delta y_{i,t} + \Delta y_{i,t+2}\right)}{Cov\left(\Delta y_{i,t}, \Delta y_{i,t-2} + \Delta y_{i,t} + \Delta y_{i,t+2}\right)}, \tag{BPP}$$

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

[40]The 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.

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 coefficients for rural and urban South Africa. The Townsend-style passthrough regression suggests a low passthrough of income shocks to consumption in both rural and urban of around 0.15 and 0.20 . The BPP estimator reveals a large gap between the rural and urban estimated passthrough of persistent shocks of 0.22 in rural versus 0.62 in urban. The low passthrough from income shocks to consumption for rural households is in line with the large empirical literature,[41] while the urban estimate lines up well with results for US households from BPP.

Table 11. Empirical passthrough estimates

	South Africa	
	Rural	Urban
$\hat{\beta}^{\mathrm{T}}_{c\_y}$	0.15	0.20
$\hat{\beta}^{\mathrm{BPP}}_{c\_p}$	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.

I next compute the same passthrough coefficients on model-simulated data, reported in table 12. The model replicates the low passthrough from income shocks to consumption in the rural sector surprisingly well, even though there is no informal insurance at all. It also reproduces the large rural-urban gap in the BPP coefficient, and matches the urban estimate almost exactly. The reason the model can capture

Table 12. Simulations

	South Africa	
	Rural	Urban
$\beta^{\mathrm{T}}_{c\_y}$	0.19	0.13
$\beta^{\mathrm{BPP}}_{c\_p}$	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 low passthrough from income shocks to consumption for rural households is that I entertain a richer income process. First, and most obviously, classical measurement error (or fully transitory shocks with bufferstock savings) induces a downward bias in Townsend-style regression. This pedestrian point turns out to be quantitatively important because measurement error appears large in surveys, particularly so in developing economies.

Second, allowing for both persistent and imperfectly transitory shocks leads to a large downward bias in the BPP estimator in the rural sector where imperfectly transitory shocks are important. The BPP estimator essentially conflates transitory and persistent shocks, overstating the latter, which leads to a downward bias in the passthrough coefficient as the shock variance appears in the denominator.

To see the bias more easily, consider the following approximation of the log of consumption as a simple

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

linear function of permanent and transitory log income

$$dc = \beta_p dp + \beta_\epsilon d\epsilon, \tag{22}$$

with $\beta_p > \beta_\epsilon$ as the passthrough from permanent shocks is larger than from transitory shocks. In this case, the BPP estimator using (22) yields an estimate between β_p and β_ϵ given by

$$\beta_{c\_p}^{\text{BPP}} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon}\left(1 - e^{-2\lambda_\epsilon}\right)}\beta_p + \frac{\sigma_\epsilon^2 e^{-2\lambda_\epsilon}\left(1 - e^{-2\lambda_\epsilon}\right)}{\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon}\left(1 - e^{-2\lambda_\epsilon}\right)}\beta_\epsilon. \tag{23}$$

Since $\beta_\epsilon < \beta_p$, the passthrough coefficient $\beta_{c\_p}^{\text{BPP}}$ is downward-biased, and the strength of the bias depends on the ratio of transitory to persistent shocks.[42] The sense in which the naive BPP estimator overstates the volatility of persistent shocks is that it is meant to isolate the volatility of the persistent term, but with imperfectly transitory shocks, instead, one gets $\sigma_p^2 + \sigma_\epsilon^2 e^{-2\lambda_\epsilon}\left(1 - e^{-2\lambda_\epsilon}\right)$, which is strictly larger, and quantitatively important in the rural sector where σ_ϵ^2 is large and λ_ϵ is relatively small.

BPP are, of course, aware that their estimator fails when the income process is misspecified relative to the canonical persistent-transitory model. What is interesting here is that the misspecification is only severe in the rural sector. Failing to account for this leads one to wrongly conclude that informal insurance must be important when simple bufferstock savings are, in fact, sufficient to sustain smooth consumption in an environment where income is volatile but mean-reverting.

The higher measured passthrough in the urban sector is consistent with the worsening of insurance documented by Munshi and Rosenzweig (2016). In my framework, however, no informal insurance is lost upon migration. Bufferstock savings insure mean-reverting shocks well and persistent shocks poorly, and persistent shocks are simply more common in the urban sector. Indeed, the true passthrough of persistent shocks is nearly the same in both sectors, 0.80 against 0.89 , in contrast to the large gap in the measured coefficients.[43] In other words, migration is risky not because access to informal insurance is lost, but because more persistent income shocks are difficult to insure.

4 Risk and Rural-Urban Structural Change

This section quantifies by how much risk retards rural-urban structural change. Recall that risk needs three ingredients: outcomes that differ across households, initial uncertainty, and the absence of complete insurance markets. Shutting down any one of them eliminates risk, yet the analysis shows that who selects into migration and who gains differ sharply.

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. Finally, section 4.3 shows what a model without risk must wrongly attribute to the migration disamenity τ_{RU}.

[42] The bias here is distinct from the ones explored in Kaplan and Violante (2010), and in particular the case where the unit root in the income process is swapped for an AR(1) process with high persistence. This would be the case when $\sigma_p \to 0$. In that case the BPP coefficient in (23) does agree with the structural coefficient β_ϵ. The issue here is that there is both an AR(1) component and a random walk, in which case a severe bias emerges whenever passing shocks are important, which hinges on the arrival rate λ_ϵ and high variance σ_ϵ^2.

[43] These are standard estimates that common bufferstock models deliver, see Blundell, Low, and Preston (2013) and Kaplan and Violante (2010). Interestingly, in the rural-urban context, the urban coefficient is additionally pushed up by migration. Many migrants enter the urban sector with little wealth, and low-wealth hand-to-mouth households bias the BPP passthrough estimator upward.

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

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

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 household would be if it were born into the counterfactual economy.

Lastly, I define insurance as follows. An efficient financial sector pays migrants the present-discounted value of expected labor income. The present value of that stream, $H(p, \epsilon)$, solves the linear system

$$rH = (w_{\mathrm{U}}\Lambda)\, e^{p+\epsilon} + \mathcal{A}H, \tag{24}$$

where $\mathcal{A}$ is the infinitesimal generator of the urban income process. Equation (24) can readily be solved using the Feynman-Kac formula. Once H is pinned down, urban consumption, which is no longer risky and admits the usual closed-form solution as linear function of total wealth, $b + H$, given by $c = \mathrm{mpc} \cdot (b + H)$ with the marginal propensity to consume equals $\mathrm{mpc} = (\tilde{\rho} + r(\gamma - 1))/\gamma$. The value function inherits the closed form $V = \mathrm{mpc}^{-\gamma}(b + H)^{1-\gamma} / (1 - \gamma)$.

Insuring post-migration income shocks, i.e., insuring against adverse evolutions of persistent and transitory type after their initial realization, still leaves the initial draw uncertain. That is, rural households do not know which H they receive. I label this economy ni, for no income risk. Insuring both, which I denote nm-ni, additionally pays the uninformed households the conditional expectation $H_m(p_{\mathrm{R}}) = \mathbb{E}[H_{\mathrm{U}}|p_{\mathrm{R}}]$ before the draw is realized: the ni economy removes income risk after arrival, and nm additionally removes the risk in the migration draw itself.[44] The insured household still pays the fixed cost f_{RU} out of financial wealth, and I contrast the impact of insurance with that of providing liquidity alone by swapping the zero-borrowing constraint with a natural borrowing constraint as detailed below.

Two features of this construction are deliberate. First, the insurance is asymmetric: it covers urban income and leaves rural income risky. That isolates the object of interest, which is how risky migration is. Second, it is an accounting device and not a policy proposal. It answers how important risk is in deterring migration, and what models that ignore risk are missing. It is thus a necessary building block toward a

[44] Insurance and information about the initial draw are rivals rather than complements: H_m is payable only to households that contract before learning p_{U}. If the draw is publicly revealed before contracting, the ex-ante insurance opportunity disappears, see Hirshleifer (1971); if households privately learn it, insurance is subject to adverse selection and unravels, see Hendren (2013). Note also that the insurance leaves the taste draw untouched. Households continue to move partly for idiosyncratic reasons, and the exercise does not provide insurance for these idiosyncratic taste shocks, see Donald, Fukui, and Miyauchi (2023) for recent work exploring these issues. Appendix B.5 contains additional details.

richer optimal policy exercise.[45]

Results. The first counterfactual removes all uncertainty about the initial urban draw by setting the share of imperfectly informed households to zero, i.e., $\zeta = 0$. Recall that in the baseline model the estimated ζ =0.78 means that roughly four in five potential migrants move without knowing their long-run urban productivity, and the estimated σ_{RU} =0.51 means that they are exposed to large risks.

The second counterfactual studies the impact of insuring urban income risk. I distinguish between insuring income shocks after arrival in the urban sector, which means that initial types $(p_{\text{U}}, \epsilon_{\text{U}})$ have materialized. Conditional on these types, households are offered the present discounted value of their expected labor income so that an artificial financial intermediary would break even if it were to underwrite such contracts. Recall that this insurance is labeled ni. A third exercise combines it with full information ($\zeta = 0$). I can go one step further and also insure the initial urban draw for households with imperfect information using the baseline share ζ =0.78 , which is labeled nm-ni.[46]

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		
	GE	PE	$\Delta \log w_{\text{R}}$
Baseline	0.65	0.65	.
Information ($\zeta = 0$)	0.76	0.80	0.1373
Insurance, income risk (ni)	0.78	0.85	0.1585
Information and insurance ($\zeta = 0$, ni)	0.84	0.87	0.2615
Insurance, both (nm-ni)	0.84	0.90	0.2572

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 differ in how the initial draw is neutralized: in ($\zeta = 0$, ni) it is revealed before the decision, in nm-ni it 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, removing uncertainty around the initial persistent draw raises the urban share by eleven percentage points; insuring post-migration income raises the urban share by thirteen percentage points. Removing both raises it by nineteen percentage points, so each source on its own closes more than half of the gap. Both sources of risk thus represent a quantitatively important barrier to migration.

Second, comparing the first two rows with the third and fourth shows that resolving uncertainty about the initial draw is worth eleven percentage points when post-migration income is risky and only six when it is insured. The intuition is that both frictions operate through the same channel – each depresses the certainty equivalent of the urban option – so fixing one lowers the return to fixing the other. This has a direct implication for the experimental literature on information and insurance interventions: the effects of such interventions are not additive, and a bundled intervention will deliver less than the sum of its parts.

Third, rows four and five deliver identical impacts on aggregate rural-urban structural change, even though one scenario provides information while the other provides full insurance including the initial type. I will return to this fact and highlight that their welfare implications and who selects into migration differ

[45] Solving the planner problem is a very different exercise that hinges on three elements beyond risk. First, a planner would internalize the effect of a household's migration choice on their offspring. Second and related, a planner would have to take a stance on how to trade off utility across cohorts. Third, a planner would internalize pecuniary externalities, which matter here because the rural wage depends on the rural-urban migration rate.

[46] An alternative approach would be to lower income shocks. I pursue this in the appendix where I set $\sigma_{p,\text{U}}$ to the lower rural value. The reason this is not my baseline is that this simultaneously reduces average outcomes since I don't have a drift term correction. These results, together with additional robustness exercises and welfare decompositions, are relegated to appendix B.5.

drastically even though they lead to the same aggregate outcome.

Fourth, general equilibrium feedback is quantitatively important, reducing the migration response by roughly one-third, as can be seen by comparing the first two columns. The difference is accounted for by the endogenous rural wage response shown in the third column. How large this feedback is depends on the importance of land in rural production, which I turn to at the end of this section.

Table 14 reports the welfare effects associated with each counterfactual, focusing on newborn rural households. These households would give up between 17.80 % and 45.19 % of consumption for the removal of migration risk once the rural wage response is taken into account.

The average and the poorest rural household tell the same story as newborn households. The first averages over the stationary rural distribution, the second places a Dirac mass on the lowest-income, lowest-wealth household. In either case, welfare effects are sizable.

Comparing the ordering of low vs. average households in terms of welfare gains across the different policies is instructive: under information provision, the average household gains relatively more than the poorest household, while under insurance provision it is the other way around. I will deconstruct these differences below when I analyze the different migration patterns each policy produces.

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)	27.11	15.42	22.49	24.80
Information and insurance ($\zeta = 0$, ni)	45.19	22.36	42.07	38.38
Insurance, both (nm-ni)	40.92	21.44	36.17	37.06

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 recent empirical evidence in Madhok et al. (2025).

The role of land. The strength of this general-equilibrium channel is governed by the land share χ. The rural wage is inversely related to the rural population, $w_{\mathrm{R}} \propto \left((1-M)\tilde{h}_{\mathrm{R}}\right)^{-\frac{\chi}{1-\nu}}$. Holding the rural-urban wage gap ω and average rural skill $\tilde{h}_{\mathrm{R}}$ fixed,[47] a link between urban productivity growth $g_{A_{\mathrm{U}}}$ and rural-urban structural change emerges,

$$\frac{d\log\left(1-M\right)}{dt} = -\frac{1-\nu}{\chi}\, g_{A_{\mathrm{U}}},$$

which is reminiscent of Lucas (2004). A smaller land share therefore implies faster structural change: when land is less important, rural wages respond less to migrant outflows, so more households must leave to keep pace with urban productivity growth.

This mechanism disciplines the calibration of χ. I set the land share to approximately one-third, following Hansen and Prescott (2002).[48] The resulting pace of structural change is consistent with the historical

[47] In current work in progress, I study a richer model with transition dynamics where urban productivity growth endogenously generates a constant rural-urban wage gap and a constant average rural type $\tilde{h}_{\mathrm{R}}$ alongside a steadily declining rural population share.

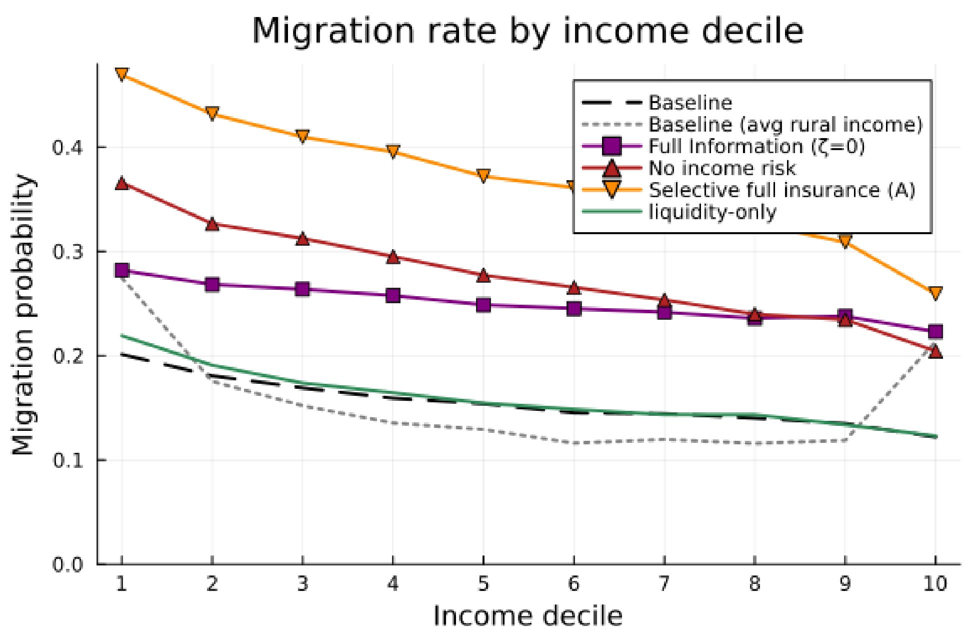
[48] This is larger than the land share of approximately ten percent implied by the village-level estimates of Akram, Chowdhury, and

experience of advanced economies and with the cross-country relationship between productivity growth and rural population decline; see, for example, Trouvain (2025b).

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 very 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.[49] 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. In contrast, for the lowest-income rural households, uncertainty about the initial persistent urban income type only represents upside risk so they migrate either way.

In contrast, post-migration income risk especially deters the rural poor, and insuring such risk leads to a strong migration response of low-income rural households as can be seen in the red line in the plot, which

Mobarak (2017). These estimates exploit village-level variation and need not identify the aggregate general-equilibrium elasticity relevant for a two-sector model; see LMW for the mapping from their estimates to the land share. Extrapolating this estimate to the aggregate economy would imply a much faster decline in the rural population than is observed over the long run.

[49] It 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 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.

is shifted up and clock-wise relative to the baseline, indicative of more negative selection into migration. Insurance of post-migration income shocks is particularly important for low-income rural households who tend to be asset poor and are thus more vulnerable to income shocks.

Rural-Urban Wage Gaps. Table 15 explores the implications more formally for common statistics related to rural-urban wage gaps and how negatively migrants are selected on income, complementing figure 11.

The first thing to note is that providing information helps little to close the cross-sectional rural-urban wage gap, and, to the contrary, increases the measured gap, despite the fact that the rural-urban efficiency wage gap falls substantially, see row two in table 13. The fixed effects wage gap increases even more by about 20 log points. Both are consequences of the more positive selection into migration visible in figure 11, where the negative link between rural income and migration flattens substantially. As households with high urban draws leave the rural sector, selection into rural-urban migration improves, raising the average urban type. At the same time, it reduces the average rural type since urban and rural type are correlated. This leads to a widening cross-sectional wage gap. The fixed-effects estimator increases even more since information provision means households with a low urban draw decide to stay, while households with a high draw move, amplifying the wage gains from migration.

Table 15. Risk and Selection into Migration

	log rural-urban wage gap		
	cross-section	fixed effects	log(migrant/stayer income)
Baseline	0.45	0.35	-0.16
Information ($\zeta = 0$)	0.46	0.55	-0.11
Insurance, income risk (ni)	0.24	0.09	-0.19
Information and insurance ($\zeta = 0$, ni)	0.30	0.32	-0.10
Insurance, both (nm-ni)	0.12	-0.08	-0.19

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

The opposite is true when providing insurance, see row three and five. Selection becomes substantially more negative as low-income low-wealth rural households were held back by urban income risk. In the scenario with insurance only for after-migration income shocks (ni), the cross-sectional wage gap is cut in half, and the within-household wage gap drops to only 9 log points. Risk was a major deterrent to migration, which is gone, so rising migration of both low and high types reduces the rural-urban wage gap.

Liquidity and Poverty Traps. Note that the insurance counterfactual simultaneously eases liquidity constraints as households implicitly borrow against future labor income, so one could be concerned that what matters is liquidity, and not insurance per se. To separate the two mechanisms, I solve a third counterfactual where the urban borrowing constraint is relaxed to the natural borrowing limit,

$$b_{\text{U}} \geq -\frac{w_{\text{U}} e^{\underline{p}+\underline{\epsilon}}}{r},$$

with income risk fully intact. Moreover, I allow rural households to pay the fixed cost of migration by borrowing from their urban borrowing limit. At this point, even the lowest-income rural households can pay the fixed cost of migration, so the mechanism usually described as a poverty trap is switched off.

The green line again plots migration rates against income in the counterfactual economy where the urban borrowing limit is equal to the natural borrowing limit, and migrants can borrow against it while still rural to pay the fixed cost. Perhaps surprisingly, the impact of liquidity is quantitatively small. As one would expect, there is some response among the lowest-income rural households, but overall the impact appears small relative to any other counterfactual I consider.

The main reason is that a more generous borrowing constraint is not an effective tool to cope with persistent shocks. At the same time, liquidity matters for the poorest who are actually constrained: relaxing the borrowing limit raises welfare for such households by 16.60 %. This is a large number, but it perhaps doesn't seem as transformative as one might think, especially compared to providing insurance. The reason is that rural income is dominated by passing shocks that arrive at rate $\lambda_{\epsilon,\mathrm{R}}$ =0.4987 . A rural household thus redraws its income every two years so few households remain ultra-poor for very long. The welfare gains are already much smaller for newborn rural households, at 7.73 %. These households also do not have any assets, yet they can finance the migration fixed cost relatively quickly so that the borrowing limit matters little. The gains for the average household are smaller yet, at 1.07 % consistent with the small aggregate effect on migration rates relative to baseline.

I make this point more directly by replicating the experiment of Banerjee, Duflo, and Sharma (2021) in my model. The fixed cost f_{RU} is calibrated in section 3 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 my economy raises labor income by only 0.01 , 0.02 and 0.01 log points after two, six and ten years. The share who have moved after ten years rises by 0.03 , from a base of 0.15 among the untreated. The reason the results are so small is that the untreated catch up on their own.

The same experiment in an economy with less mean-reverting income looks completely different. I halve the arrival rate of passing shocks, holding the variance of rural income fixed, so that a low income state lasts about four years instead of two. Labor income now rises by 0.25 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 welfare gain for the poorest rural household rises accordingly, from 16.60 % to 28.69 %. The low-churn economy is strikingly close to the empirical estimates in Banerjee, Duflo, and Sharma (2021), who find labor income gains of 27% after seven years. My model can generate their result when income at the bottom is less mean-reverting. The general insight here is that the strength of poverty traps is largely a question about the income process, and less so about whether fixed costs exist or not.

The fact that risk plays a much more important role for aggregate development than liquidity echoes the findings of Karlan et al. (2014), who show that risk constrains agricultural investment by more than liquidity. In the context of rural-urban migration, insurance is only one way to tackle risk, and information provision represents a different strategy. While they generate a similar aggregate migration response, who selects into migration and who gains the most differs drastically. While optimal (and real world) policy is beyond the scope of the paper, any such policy consideration will have to take account of the endogenous migration responses studied here.

In a final exercise, I study what share of the usually large migration wedge inferred in models with migration frictions,[50] is accounted for by risk. Again, this directly bears on follow-up policy questions. One common response to large migration wedges is a call for infrastructure investment to most literally lower the cost of migration. This would be ineffective if much of the wedge is due to risk.

4.3 Migration Wedges and Risk

Common approaches to modeling rural-urban migration ignore risk, both in terms of what is known about urban persistent income types before migration, and exposure to persistent income shocks after migra-

[50] See for instance Bryan and Morten (2019), Lagakos et al. (2020), LMW, or Baseler and Gordeev, 2026.

tion. Of course, given the many degrees of freedom, and in particular the migration wedge τ_{RU}, all of these alternative models proposed in the literature still match rural-urban migration rates.

I next quantify how much imperfect information and uninsurable income shocks are soaked up in the residual migration wedge, if ignored. I pursue this with two complementary approaches. The first approach treats my framework as ground truth, and asks by how much the migration wedge has to increase to match the baseline urban share of 65% if initial information was perfect, or income shocks after migration were insured, or both. This is what a researcher would infer about the migration wedge if they knew all structural parameters of the environment except for τ_{RU}, and treated urban income as known or insured.

At the same time, a researcher that ignores migration risk may also infer other structural parameters. This asks for a second approach where I simply shut down persistent risk around and after migration, and reestimate the model. Both approaches lead to the same conclusion: ignoring risk overstates the migration wedge substantially.

The Migration Wedge: Model as Ground Truth. Table 16 reports the results where I treat my model as ground truth, and ask: by how much does the migration wedge have to increase to match an urban population share of 65% in any of the counterfactual economies where I shut down risk. The bottom line is that the migration wedge has to be inflated substantially, up to a factor of 2, to match the baseline urban share.

A researcher who assumes households know their initial urban productivity before moving must inflate the wedge by 43.78 % to rationalize observed flows. A researcher who assumes that only expected income matters must inflate the wedge by 42.72 %. One who assumes both, initial information and safe income – must inflate it by 112.24 %. This scenario is even more extreme than the final counterfactual where insurance is extended to the initial draw among uninformed households. The reason is that it is extremely difficult to deter a household from moving to the urban sector when they know they have a high draw, which is what requires the large wedge in column three. The results are consistent with the findings in Porcher, Morales, and Fujiwara (2024), which also shows that ignoring information frictions overstates migration wedges, although they arrive at this result in a very different context and without bufferstock savings and fixed migration costs.

Table 16. Risk and the Inferred Migration Wedge

	baseline	$\zeta = 0$	ni	$\zeta = 0, ni$	nm-ni
$\Delta\tau_{\text{RU}}$ (%)	0.0	43.78	42.72	112.24	69.94

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 complete markets in the urban sector, 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.

Migration Wedge and Welfare. A corollary of the relatively small migration wedge I inferred in the baseline model is that welfare differences across sectors are much smaller than wage gaps. To see the centrality of the migration wedge for welfare gaps, consider a static model without heterogeneity in income types or preferences, where the migration wedge would be identical to the wage gap, i.e., $\log \omega = \log \tau_{\text{RU}}$, and thus identical to the consumption-equivalent welfare gap.

In table 17, I report the welfare gap across average rural and urban households, using the stationary distribution. Column one row one reports the consumption-equivalent welfare gap, which is only 12.30 %. This lines up well with the small migration wedge in the baseline model.

The table also highlights that in a richer model with income risk, assets, and preference shocks, the simple wage gap is no longer proportional to the welfare gap, which can be seen when comparing the first and second column, which report actual and naive wage-based welfare gap. Interestingly, the welfare gap is closer to the simple wage gap when focusing on newborn households, especially those without assets, see the second and third column. This is again explained by selection: newborn households tend to have much lower income than average rural households, as households with a successful rural career stay, while less successful households try to migrate to the urban sector. The difference between the last two rows is explained by the option to migrate and the ability to insure against income shocks, both of which require bufferstock savings.[51] In other words, while average cross-sectional wage gaps grossly overstate the welfare gap, they provide a reasonable approximation for the most vulnerable rural households, who can neither migrate nor insure against large income shocks.

Table 17. Income and Welfare Gaps

	actual gap (%)	naive prediction (%)
Average household	12.30	56.65
Newborn, stationary assets	46.71	76.63
Newborn, no assets	61.76	76.63

All three rows are the baseline economy and differ only in which households are compared. The first uses each sector's stationary distribution over types and assets. The second holds types at the newborn draw and gives each sector its stationary distribution of assets. The third is the newborn draw at zero assets. The naive prediction is what the income gap would be worth if consumption were simply labor income. Mean log income differs across sectors by $\Delta = \log\omega + (\bar{p}_{\text{U}} + \bar{\epsilon}_{\text{U}}) - (\bar{p}_{\text{R}} + \bar{\epsilon}_{\text{R}})$, where bars denote means over the relevant distribution, so a household with $c = y$ would need $e^{\Delta} - 1$ more consumption in the rural sector to be as well off as in the urban sector. It depends only on the type draw, ignoring assets altogether, so the last two rows of the second column are equal by construction. The actual gap is the consumption-equivalent difference computed from the value functions, which incorporates risk, savings, and the option to migrate.

The Migration Wedge: Reestimating the Model. In this exercise, I no longer treat my model as the ground truth. Instead, I use a more standard model where I shut down initial uncertainty and persistent shocks to report what migration wedge would be implied in this setup.[52]

I implement this as follows. I shut down persistent shocks in either sector, $\sigma_{p,j} = 0,\ j \in \{\text{R}, \text{U}\}$, and ask the imperfectly transitory shock to match the variance of the change in log income and its first-order autocorrelation.[53] I do this using income data net of measurement error. If measurement error was included, mean reversion would be even stronger and risk would be even less relevant, amplifying the result.

I set $\rho_{\text{RU}} = 1$, rid migratoin of uncertainty about urban persistent type by setting the variance to zero, $\sigma_{\text{RU}} = 0$, and impose $c_{\text{RU}} = \log\Lambda = 0$ so there is no skill downgrading. Consequently, the rural persistent type is perfectly predictive of the urban type. Note that rural-urban migrants still have a lower persistent type than urban natives as long as $c_{0,\text{R}} < c_{0,\text{U}}$. Moreover, I restrict heterogeneity in persistent type amongst households born in the rural sector to zero, i.e., $\sigma^2_{0,\text{R}} = 0$, while still allowing for differences in persistent urban productivity amongst urban natives, i.e., $\sigma^2_{0,\text{U}} > 0$, just as in LMW. Crucially, I hold f_{RU} and λ_m fixed across the different model variants, which is necessary to compare the implied migration wedge τ_{RU} sensibly across models, see footnote 35.

This leaves me with four parameters $\Theta = \{c_{0,\text{R}}, \sigma^2_{0,\text{U}}, \tau_{\text{RU}}, s\}$, which I identify using the following four moments: the cross-sectional and the fixed effects rural-urban wage gap, selection into migration, and overall urban inequality. The system is exactly identified. Urban productivity A_{U} again adjusts so that the urban population share matches its baseline value.

[51]I verified that this is not primarily driven by differences in initial dispersion. Rural newborns draw their persistent type from a much more dispersed distribution than urban ones, $\sigma^2_{0,\text{R}}$ =0.14 against 0.0100 . Giving urban newborns the rural dispersion lowers the urban value, but only modestly, leaving a gap of 41.19 %.

[52]The exercise is meant to bring the model closer to LMW or Kleemans (2015).

[53]This requires two moments for two parameters. Because the persistent type is assumed to be fixed, it drops out of any within-household difference.

Table 18 reports the parameters implied by the restricted income process without persistent shocks. The rural income process changes relatively little, consistent with the fact that persistent shocks were unimportant to begin with. In contrast, the urban income process changes more dramatically, with a doubling of the variance of the transitory shock. Intuitively, the omission of permanent shocks forces the restricted income process to attribute more variance and more persistence to the imperfectly transitory shock process.

Table 18. 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.

Table 19 reports the fit of the model, and 20 reports the implied parameter values.

Table 19. Moments Targeted, Restricted Model

Moments	data SA	model SA
Log rural-urban wage gap (cross-section)	0.45	0.44
Log rural-urban wage gap (fixed effects)	0.34	0.35
Log(migrant income in rural/stayer income in rural)	-0.14	-0.13
Initial inequality urban	0.18	0.152

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 20. Estimated Parameters, Restricted Model

Parameter	$c_{0,\mathrm{R}}$	$\sigma_{\mathrm{U},0}$	s	τ_{RU}
Estimate	-0.73	0.17	7.03	1.53

The remaining parameters of the full model are imposed by the restriction rather than estimated, as described above. Urban productivity adjusts to A_{U} =0.73 so that the urban population share matches its baseline value, and the implied rural-urban wage ratio is ω =1.26 .

The new model variant matches rural-urban wage gaps, both cross-sectional and within-households. In contrast to the richer model, the within-household wage gains are temporary, and rural-urban migration is a strategy to escape a negative rural income shock: future migrants earn -0.13 log points less than stayers before moving, due to negative temporary shocks. On arrival in the urban sector, they draw a new urban shock. The measured gain from migrating is therefore the equilibrium wage ratio, ω =1.26 , plus mean reversion.[54][55]

[54] This mechanism is exactly consistent with the large IV estimates in Bryan, Chowdhury, and Mobarak (2014), and explained in more detail in LMW.

[55] In this restricted economy, I can also ask: how much more migration do we get if we insure urban income shocks after migration. This raises the urban share from 0.65 to 0.71 , which is substantially lower than in the baseline economy with persistent shocks. This highlights that the strong response in the baseline economy isn't trivially implied, but depends on the large risk due to persistent shocks.

The migration wedge is substantially larger, rising by 42.32 % relative to baseline (1.08 vs 1.53). Recall that the migration wedge in my framework has a consumption-equivalent interpretation: the wedge acts as if a migrant's urban consumption stream was scaled down by $1/\tau_{\mathrm{RU}}$. In the baseline model net of preference shocks, a migrant would need eight percent more urban consumption to be indifferent. After ignoring initial uncertainty and persistent shocks, that number rises to 53 percent. The large discrepancy is a consequence of the sizable risk that permanent rural-urban migrants face.

Two complementary approaches thus arrive at the same conclusion: much of the large residual migration wedge is accounted for by risk.

5 Conclusion

Rural-urban income gaps are large, and a common interpretation is that they represent efficient sorting. This paper takes a different view. Permanent migration is a dynamic decision that is extremely risky for most households, due to initial uncertainty about where a household lands on the urban income distribution and persistent income shocks thereafter.

Risk constitutes a quantitatively important barrier to rural-urban structural change. Removing risk raises the urban population share by about 20 percentage points in the South African context. A researcher who ignores risk overstates the residual migration wedge by a factor of two.

Risk arises from the interplay of heterogeneous outcomes, uncertainty, and the absence of insurance markets. Removing any one of these elements eliminates risk. However, information and insurance generate fundamentally different selection into migration and welfare implications, which matters for any policy aimed at the market failures caused by the risky nature of rural-urban migration.

Wealth is not an effective tool for raising rural-urban migration, even though it enables households to cope with urban risk. The reason is that the gains from migration are tied to labor income, and these gains become ever smaller in relative terms as households get wealthier, rendering migration unattractive.

This paper is only a first step, and many open issues remain. For instance, how could insurance be effectively provided in the presence of moral hazard, which I have abstracted from throughout? Similarly, is it possible to provide more information, perhaps by improving search and matching in the labor market? Lastly, how do these results generalize along a transition path, and what would such a model of necessarily unbalanced growth reveal about the process of long-run development?

Despite these open issues, I want to establish one point. Moving to the urban sector is risky from a rural household's point of view, because initial urban human capital is uncertain, and subject to persistent shocks. This ought to matter for the process of rural-urban structural change.

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: *The review of economic studies* 89.1, pp. 45–86.

Aguiar, Mark, Mark Bils, and Corina Boar (2024). "Who are the Hand-to-Mouth?" In: *Review of Economic Studies*.

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 (2023). "Selection, absolute advantage, and the agricultural productivity gap". In.

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. (2022). *Growing Apart: Declining Within-and Across-Locality Insurance in Rural China*. Tech. rep. National Bureau of Economic Research.

Balboni, Clare et al. (2022). "Why do people stay poor?" In: *The 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 Weerdt, 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: *The 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: *The Economic Journal* 129.617, pp. 143–171.

Braxton, J Carter et al. (2021). *Changing income risk across the US skill distribution: Evidence from a generalized Kalman filter*. Tech. rep. National Bureau of Economic Research.

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 et al. (2025). "Social Norms and Transfer Progressivity in the Village". In: Work in progress.
Caselli, Francesco (2005). "Accounting for cross-country income differences". In: *Handbook of economic growth* 1, 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 (2022). *What drives investors' portfolio choices? separating risk preferences from frictions*. Tech. rep. Working paper.
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 Holm, and Hakon Tretvoll (2022). "A Parsimonious Model of Idiosyncratic Income". In.
Daly, Moira, Dmytro Hryshko, and Iourii Manovskii (2022). "Improving the measurement of earnings dynamics". In: *International Economic Review* 63.1, pp. 95–124.
De Magalhaes, Leandro, Enric Martorell, and Raül Santaeulalia-Llopis (2024). "Transfer Progressivity and Aggregate Development". In: Unpublished Manuscript.
De Magalhaes, Leandro and Raul Santaeulalia-Llopis (2018). "The Consumption, Income, and Wealth of the Poorest". In: *The Journal of Development Economics*.
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 (2023). "Unpacking Aggregate Welfare in a Spatial Economy". In.
— (2025). *Unpacking aggregate welfare in a spatial economy*. Tech. rep. National Bureau of Economic Research.
Dvorkin, Maximiliano and Brian Greaney (2025). *The geography of wealth: shocks, mobility, and precautionary savings*. Tech. rep.
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: *The 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 (2017). *In search of a spatial equilibrium in the developing world*. Tech. rep. National Bureau of Economic Research.
Gollin, Douglas, David Lagakos, and Michael E Waugh (2014). "The agricultural productivity gap". In: *The 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: *The 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: *The 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: *The 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: *The 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, William Arthur (1954). "Economic development with unlimited supplies of labour". In.

Ligon, Ethan (1998). "Risk sharing and information in village economies". In: *The Review of Economic Studies* 65.4, pp. 847–864.

— (2025). "Risk sharing tests and covariate shocks: Drought, Floods, and Pests in Uganda". In.

Ligon, Ethan, Jonathan P Thomas, and Tim Worrall (2002). "Informal insurance arrangements with limited commitment: Theory and evidence from village economies". In: *The 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*. Tech. rep. 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: *The Review of Economic Studies* 89.1, pp. 452–480.

Miner, Gwyneth (2025). "Overcoming Migration Barriers: The Impact of an Income Smoothing Program for Kenyan Migrants". In.

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*. Tech. rep. National Bureau of Economic Research.

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*. Tech. rep. 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 and locations in indonesia". In: *Unpublished Working Paper, Vancouver School of Economics*.

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.

Santaeulalia-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: journal of the Econometric Society*, pp. 539–591.

Trouvain, Florian (2025a). "A General Equilibrium Analysis of Consumption Risk and Inequality in Rural Economies". In: *Working Paper, University of Oxford*.

— (2025b). "Structural Change, Inequality, and Capital Flows". In: *Mimeo, University of Oxford*.

Udry, Christopher (1994). "Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria". In: *The 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: Journal of the Econometric Society*, pp. 916–918.

World Bank (2021). *South Africa: Social Assistance Programs and Systems Review*. Tech. rep. Washington, DC: World Bank.

Yaari, Menahem E (1965). "Uncertain lifetime, life insurance, and the theory of the consumer". In: *The Review of Economic Studies* 32.2, pp. 137–150.

Young, Alwyn (2013). "Inequality, the urban-rural gap, and migration". In: *The 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. misreported 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*;c
- **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.[56] The upshot of the analysis is that there appears to be no major compositional shift.

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

[56]See the stata code “empirics/code/clean_data/south_africa/14_linking_sanity_check.do" written by Qingyu Chen in the replication package.

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),[57] 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.

[57]Funerals 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

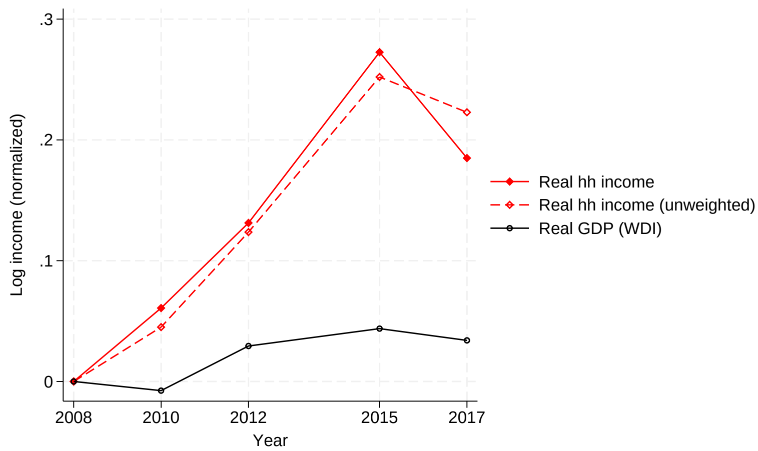


Figure 12 plots normalized log real income, i.e., I first compute sample averages, then divide each average by its value in the first year so I end up with an index of real income, and finally take the log. The real GDP measure comes again from the world bank measured at constant local currency units.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

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

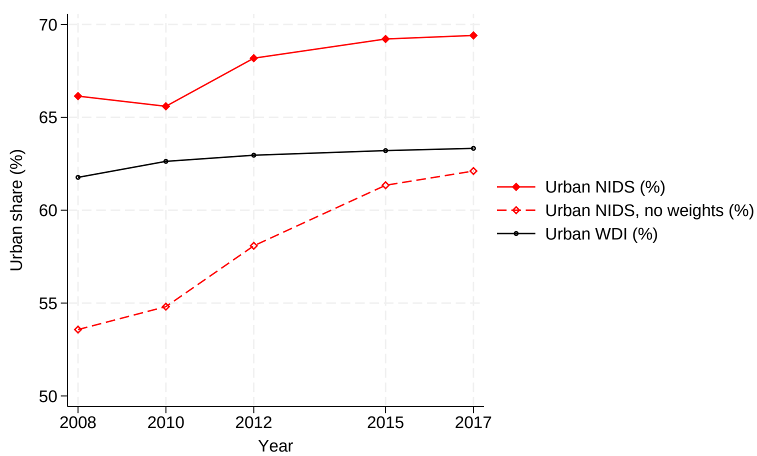


Figure 13 plots the share of urban households as reported by the World Bank Development Indicators, and as found in the micro data. I compute a weighted and an unweighted average using cross sectional weights as detailed in the appendix.

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 21.

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.s[58] Next, I impose an age limit $27 \leq age \leq 62$. The limit is slightly

[58] This 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 fist 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.[59] 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 21. Sample Selection in South Africa

(a) Rural

Restrictions	Households	Observations
All	2644	12597
No movers	2093	10002
Missing month	1722	8369
Ages 27 to 62	1283	6200
Household splits	1179	5592
Labor market attachment	884	4196
Income outliers	778	3671
Max one missing	331	1661

(b) Urban

Restrictions	Households	Observations
All	3665	17340
No movers	3165	14904
Missing month	2421	11444
Ages 27 to 62	1893	8881
Household splits	1765	8376
Labor market attachment	1537	7316
Income outliers	1321	6357
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 22 contains descriptive statistics for both the raw data, and the selected sample.

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

Table 22. Summary statistics South Africa

(a) Full Sample

	Rural mean	sd	Urban 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 mean	sd	Urban 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 22a I do not select the sample other than dropping observations without urban-rural status. In table 22a 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_sqare
- 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 (2024). 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 23. Migration and cross-sectional differences in South Africa (standardized)

(a) Rural stayers vs. rural movers

	(1)
	Rural-Urban Migrant
age	-0.150***
gender	0.075**
years of schooling	0.045
l_r_hhincome_lab_gov_yr_pa	-0.090***
children	0.034
N	5666

Standardized beta coefficients
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Urban stayers vs. urban movers

	(1)
	Urban-Rural Migrant
age	0.005
gender	0.015
years of schooling	-0.071**
log hh income	-0.013
children	-0.028
N	11248

Standardized beta coefficients
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 23a is based only on rural stayers and rural-urban movers, while table 23b 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 24 reports the cross-sectional wage gap.

Table 24. Cross-sectional rural-urban wage gap in South Africa

urban	0.727***	0.724***	0.450***
	(0.0428)	(0.0421)	(0.0367)
controls	No	No	Yes
time effects	No	Yes	Yes
N	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 25. 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)
N	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.[60]

[60]The 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 26 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 26. Selection into Migration in South Africa

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

(a) Rural stayers vs. rural movers

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

(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 27a is based only on rural stayers and rural-urban movers, while table 27b 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 27. 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 27. Selection into Migration in South Africa with Age Controls

	Years of education	Log total net assets	Log income
Rural-Urban mover dummy	0.115	-0.0770	-0.248***
	(0.310)	(0.301)	(0.0795)
N	6409	1233	5667
R^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 27 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 28 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.[61]

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

Table 28. Correlation across rural and urban income in South Africa

	OLS (movers)	IV (movers)	OLS (stayers)	IV (stayers)
rural labor income	0.279^{***}	0.583^{***}	0.414^{***}	0.817^{***}
	(0.0517)	(0.186)	(0.0301)	(0.0626)
N	268	170	3011	2026
R^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 30. Thus, while rural-urban migration raises income on average, a substantial fraction of households – roughly 30%– experience income losses after migrating.

Table 29. Migration and Income Drop

	better	better
rural-urban move	0.254^{***}	0.208^{***}
	(0.0398)	(0.0419)
constant	0.574^{***}	0.563^{***}
	(0.0121)	(0.0169)
Time FE	Yes	Yes
Age controls	No	Yes
HH size	No	Yes
Observations	3480	3480
r2	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 30. Migration and Consumption Drop

	better_con	better_con
rural-urban move	0.302^{***}	0.237^{***}
	(0.0351)	(0.0373)
constant	0.533^{***}	0.525^{***}
	(0.00998)	(0.0124)
Time FE	No	Yes
Age controls	No	Yes
HH size	No	Yes
Observations	4084	4084
r2	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 31 report the results on income growth and volatility. Reassuringly, mean growth is positive, consistent with the sizable urban income premium estimated earlier.[62] 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 31.

Table 31. Migration and Income Volatility

	Income growth	Income volatility	Consumption growth	Consumption volatility
rural-urban move	0.323***	0.0478*	0.278***	0.0349
	(0.0437)	(0.0246)	(0.0390)	(0.0224)
constant	0.0314***	0.120***	0.00934	0.125***
	(0.00901)	(0.00681)	(0.00883)	(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 36 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.[63]

[62] 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: $\mathbb{V}(\Delta y/2) = \sigma_p^2/2 + \left(1 - e^{-2\lambda_\epsilon}\right)\sigma_\epsilon^2/2 + \sigma_x^2/2 \approx 0 + \left(1 - e^{-1}\right) * 0.26/2 + .1/2 \approx .13$.

[63] 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 32. County-to-county rural-urban migration flows in South Africa

(a) Full Sample

	(1) Share
Rural stayer	0.264
Urban stayer	0.626
Rural-Urban mover	0.067
Urban-Rural mover	0.026
Two-way mover	0.018
Observations	26245

(b) Selected Sample

	(1) Share
Rural stayer	0.242
Urban stayer	0.758
Rural-Urban mover	0.000
Urban-Rural mover	0.000
Two-way mover	0.000
Observations	4712

In table 32a I do not select the sample other than dropping observations without urban-rural status. In table 32b I apply the same selection criteria defined earlier.

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

(a) Rural stayers vs. rural movers

	Age	Years of education	Log total net assets	Log income	Log asset-income ratio
Rural-Urban mover dummy	-13.74***	2.112***	-0.0311	-0.648***	0.315
	(0.635)	(0.223)	(0.250)	(0.0598)	(0.268)
Rural stayers baseline	38.27***	9.107***	10.69***	9.617***	-0.316***
	(0.347)	(0.120)	(0.0628)	(0.0305)	(0.0573)
N	10932	10932	1856	9658	1856
R^2	0.101	0.025	0.000	0.028	0.001

(b) Urban stayers vs. urban movers

	Age	Years of education	Log total net assets	Log income	Log asset-income ratio
Urban-Rural mover dummy	0.812	-1.614***	-1.682**	-0.100	-1.525**
	(2.082)	(0.449)	(0.780)	(0.170)	(0.727)
Urban stayers baseline	40.27***	11.47***	11.68***	10.51***	-0.167**
	(0.369)	(0.101)	(0.0892)	(0.0395)	(0.0667)
N	13001	13002	2975	11329	2975
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 33a is based only on rural stayers and rural-urban movers, while table 33b is based only on urban stayers and urban-rural movers.

Table 34. County-to-county migration and cross-sectional differences in South Africa (standardized)

(a) Rural stayers vs. rural movers

	(1) Rural-Urban Migrant
age	-0.249***
gender	0.059**
years of schooling	0.052**
log hh income	-0.130***
children	0.050**
N	9656

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Urban stayers vs. urban movers

	(1) Urban-Rural Migrant
age	0.025
gender	0.008
years of schooling	-0.079**
log hh income	0.016
children	-0.023
N	11321

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table refers to the full sample but focuses on cross-sectional differences in household observables before migration. Table 34a is based only on rural stayers and rural-urban movers, while table 34b 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 35 uses household fixed effects with and without additional controls in line with the previous specification. Note that the size of the coefficient increases.

Table 35. County-to-county within household rural-urban wage gap in South Africa

	all movers	rural-urban movers	urban-rural movers
urban	0.342***	0.421***	0.320**
	(0.0609)	(0.0815)	(0.133)
controls	Yes	Yes	Yes
N	22216	10719	11108
R^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 36 reports income and consumption mean gains and volatility (as annualized variance) without controls, which complements the results in the main body of the paper.

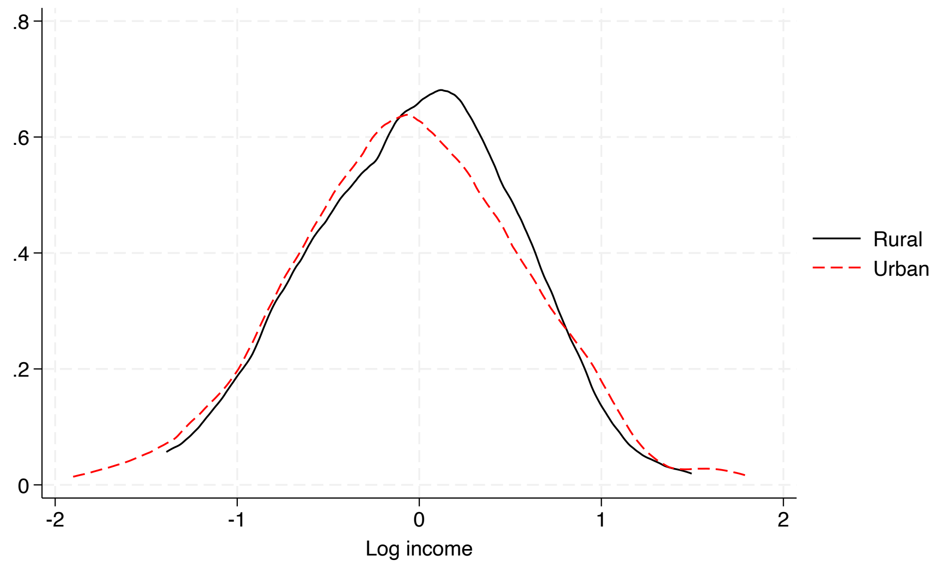
Table 36. Migration and Income Volatility (no controls)

	Income growth	Income volatility	Consumption growth	Consumption volatility
rural-urban move	0.367***	0.0694***	0.337***	0.0491**
	(0.0460)	(0.0265)	(0.0373)	(0.0243)
constant	0.0469***	0.129***	0.0202***	0.130***
	(0.00687)	(0.00565)	(0.00676)	(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 37 reports the results of a regression that allows for intercept and slope coefficient to vary across sectors.

Table 37. Tail Coefficients

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

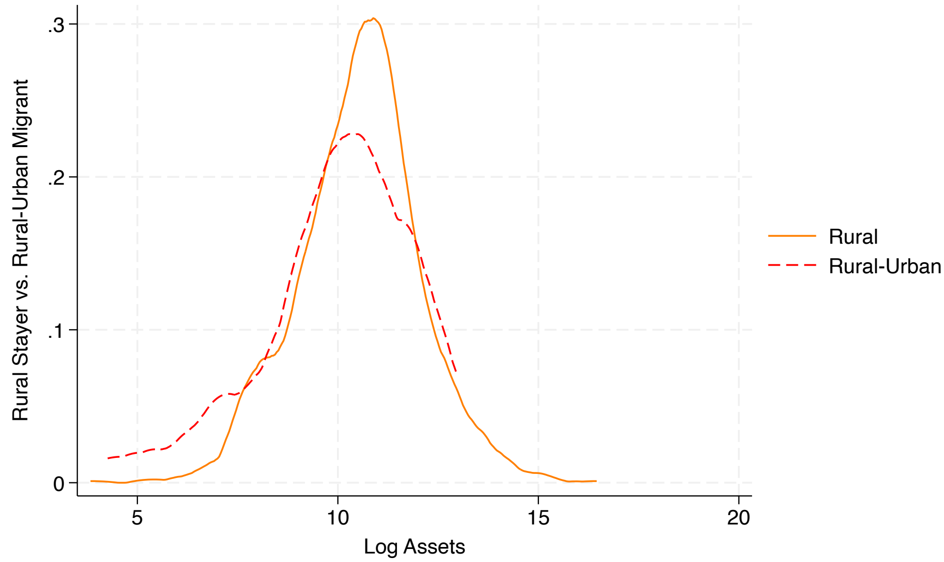
The income measure uses residualized log labor income broadly defined and averaged over different years. The bivariate regression is run separately for rural and urban stayers after dropping observations below the 80th percentile of income to only capture the right tail of the distribution.

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 Santaeulalia-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 38 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 39 reports the empirical passthrough

Table 38. 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
N	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 39. Passthrough in South Africa

(a) β_{c_y}

	Log consumption	Log consumption
Log income	0.260***	0.147***
	(0.0141)	(0.0144)
Log income $\times$ $\mathbb{1}\_urban$	0.133***	0.0534***
	(0.0212)	(0.0186)
Controls	Yes	Yes
Year-month-sector fe	Yes	Yes
Individual fe	No	Yes
N	20878	20055
R^2	0.658	0.815

(b) $\beta_{c\_p}^{\text{BPP}}$

	Rural	Urban
$\hat{\beta}$	0.224	0.621*
	(0.258)	(0.322)
N	385	610

A.3 Estimating income risk

I build on the methodology developed in BPP. Consider the fix proposed in Crawley, Holm, and Tretvoll (2022) 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 (2022), 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

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

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 (2022). There are also more level moments that differenced moments, which helps given the relatively short samples I work with. Denote residualized log income of household i as $\boldsymbol{y_i^*} = [y_{i,0}^* \; y_{i,1}^* \; ... \; y_{i,T}^*]'$ and note that the Covariance matrix reads

$$V = \mathbb{E}_i \left[\boldsymbol{y_i^*} \boldsymbol{y_i^*}\right]'$$

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

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

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

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

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

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

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 $m_{t,t+2} = \mathbb{E}_i \left[y_{i,t} y_{i,t+2}\right]$ as

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

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

$$\min_{\boldsymbol{\theta}} \left[\sum_i \boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)\right]' \widehat{\Omega} \left[\sum_i \boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)\right]$$

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

The first order condition reads

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

and the asymptotic standard errors follow

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

where $\hat{G} = \frac{1}{N} \sum_i \nabla_\theta \boldsymbol{g}\left(\boldsymbol{w}_i, \hat{\boldsymbol{\theta}}\right)$.

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

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

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: |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];...;[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, $Y^{\text{eqv}} = \frac{Y}{adults+.5*children}$, 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

$$\sigma^2_{a,t,j} = \frac{\sum_{i\in a,j,t} X_i^2}{N_{a,j,t}} - \left(\frac{\sum_{i\in a,j,t} X_i}{N_{a,j,t}}\right)^2, \tag{25}$$

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, $N_{a,t} \geq 50$,[64] and we use household weights to construct first and second moment.

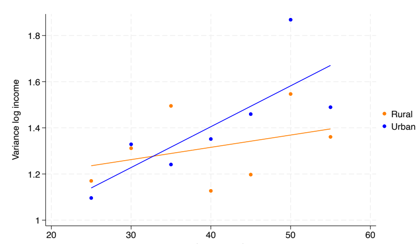
Time effects. I then run a linear regression of the form

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

where γ are time and age index fixed effects.[65]

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

[64] It turns out that this restriction has no bite.

[65] As 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 abitrary 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 40. Within household rural-urban wage gap in South Africa

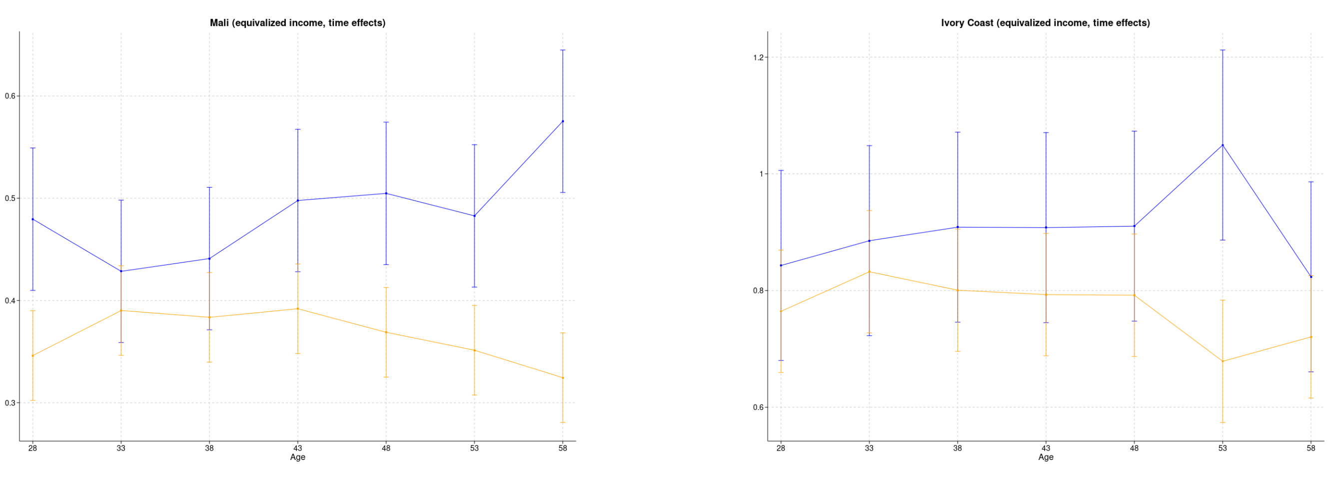
	variance log income
age	0.00532
	(0.00406)
urban × age	0.0124*
	(0.00480)
time effects	Yes
N	70

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

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

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

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

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

where the second line uses the fact that the change in p is uncorrelated across non-overlapping intervals, e.g., $\mathbb{E}\left[\Delta p_t \Delta p_{t-2}\right] = 0$, and $\mathbb{E}\left[\epsilon_t\epsilon_{t+k}\right] = \mathbb{E}\left[\epsilon_t\epsilon_{t-k}\right] = \sigma_\epsilon^2 e^{-\lambda_\epsilon k}$. There is one final subtlety to note: for simplicity, I assumed in the derivation that $\mathbb{E}\left[\epsilon_{t+k}\epsilon_{t+k}\right] = \sigma_\epsilon^2\, \forall k$. 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

$$(\rho + \phi + \lambda) V^{n+1} = u(c^n) + \lambda \tilde{V}_m^n + \mathcal{A}^n V^{n+1} + \frac{V^{n+1} - V^n}{\Delta} \tag{26}$$

$$u'(c^n) = \partial_b V^n \tag{27}$$

$$\pi^{\mathcal{I}} = \frac{\left(\mathbb{E}\left[V_{\text{U}}^{\mathcal{I}}\right] \cdot (-1)\right)^{\frac{s}{1-\gamma}} \tau_{\text{RU}}^{-s}}{(V_{\text{R}} \cdot (-1))^{\frac{s}{1-\gamma}} + \mathbb{E}\left[V_{\text{U}}^{\mathcal{I}}\right] \cdot (-1)^{\frac{s}{1-\gamma}} \tau_{\text{RU}}^{-s}} \tag{28}$$

$$\tilde{V}_m^n = \zeta \tilde{V}_m^{pi,n} + (1-\zeta) \tilde{V}_m^{fi,n} \tag{29}$$

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 $a_{\text{U}} \geq 0$ by imposing $V_{\text{U}}(a_{\text{U}} < 0, p_{\text{U}}, \epsilon_{\text{U}}) \to -\infty$, which is to say that households that cannot afford the monetary cost of migration will not migrate. One can show that $\hat{V} = V_{\text{R}}$ 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 $\frac{b - f_{\text{RU}}}{\omega\Lambda}$. The reason this expression features the wage gap is that the actual urban assets of a migrant are $(b - f_{\text{RU}}) w_{\text{R}}$. When normalized by the effective urban wage facing a migrant, $w_{\text{U}}\Lambda$, this asset position equals $(b - f_{\text{RU}}) \frac{w_{\text{R}}}{w_{\text{U}}\Lambda} = \frac{(b - f_{\text{RU}})}{\omega\Lambda}$. 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 $w_{\text{U}} = (1-\theta)\left(\frac{\theta}{r+\delta_k}\right)^{\frac{\theta}{1-\theta}} A_{\text{U}}$.

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

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

is consistent with labor market clearing using

$$\frac{w_{\text{U}}}{w_{\text{R}}} = \frac{(1-\theta)\left(\frac{\theta}{r+\delta_k}\right)^{\frac{\theta}{1-\theta}} A_{\text{U}}}{(1-\chi-\nu)\left(\frac{\nu}{r+\delta_k}\right)^{\frac{\nu}{1-\nu}} \left((1-M)\tilde{h}_{\text{R}}\right)^{-\frac{\chi}{1-\nu}}} \Rightarrow$$

$$1 - M = \frac{1}{\tilde{h}_{\text{R}}} \left(\frac{\omega}{w_{\text{U}}} (1-\chi-\nu) \left(\frac{\nu}{r+\delta_k}\right)^{\frac{\nu}{1-\nu}} \right)^{\frac{1-\nu}{\chi}}.$$

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 $1-\delta_{\text{UR}}$ 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 $\mathbb{P}(x_{\text{U}}|x_{\text{R}})$. Let $\mathbf{P}^{\text{RU}}$ be a stochastic matrix of size $N^{\#a} \times N^{\#p} \times N^{\#\epsilon}$ 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,

$$\kappa(x_{\text{U}}, \mathcal{I}) = \frac{\int \mathbb{P}(x_{\text{U}}|x,, \mathcal{I}) \pi_{\text{RU}}^{\mathcal{I}}(x) g_{\text{R}}(x) dx}{\int \mathbb{P}(x_{\text{U}}|x,, \mathcal{I}) \pi_{\text{RU}}^{\mathcal{I}}(x) g_{\text{R}}(x) dx dx_{\text{U}}},$$

and the unconditional probability of a migrant arriving in state x_{U} equals

$$\kappa(x_{\text{U}}) = \zeta \kappa(x_{\text{U}}, pi) + (1-\zeta)\kappa(x_{\text{U}}, fi).$$

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 $\mathcal{A}_{\text{U,death}}$ 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 $1-\delta_{\text{UR}}$

which is incorporated in $\mathcal{A}_{\text{U,death}}$. On the other hand, with probability δ_{UR}, the urban household is replaced by a rural migrant. This is captured using the generator matrix $\mathcal{A}_{\text{mig,RU}} = \phi\delta_{\text{UR}}\left(-\mathbb{I} + \iota\kappa'\right)$ where $\mathbb{I}$ is the identity matrix, and $\iota = [1\,1\ldots1]'$ is a row vector of ones. Putting the pieces together delivers the following generator matrix

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

The stationary distribution of the urban sector g_{U} then solves

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

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

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

where the diagonals capture the negative out-migration probability, and the in-migration uses the distribution $g_{\text{RU},0}\left(x\right)$, 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

$$\mathcal{A}_{\text{R}} = \mathcal{A}_{\text{R},x} + \mathcal{A}_{\text{R,death}} + \mathcal{A}_{\text{mig,UR}}$$
$$\mathbf{0} = \mathcal{A}_{\text{R}}^{T} g_{\text{R}}.$$

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,

$$\left(\rho + \phi\right) V_{n+1} = u\left(c_n\right) + A_n V_{n+1} + \frac{V_{n+1} - V_n}{\Delta} \tag{30}$$

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 $u'\left(c\right) = V_a$ 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 $\mathbf{0} = A^* g$, 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.[66] 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 $k \in \{0, 1, 2, ..., 10\} = K$ 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 $R : f(\tilde{a}) \to a$, 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

$$V_a = u'(c). \tag{31}$$

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,

[66] See 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.,

$$\rho V(k, y) = \max_{k,c} u(c) + V_k \dot{k} + V_y \dot{y} \tag{32}$$

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 (31). 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 (32) to find the solution for (31). 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,

$$\mathbf{0} = A\left(\dot{k}\right)^{T} p,$$

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

$$B = \sum_{k} p_k a\left(k\right)$$

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 $A\left(\dot{k}\right)$ 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

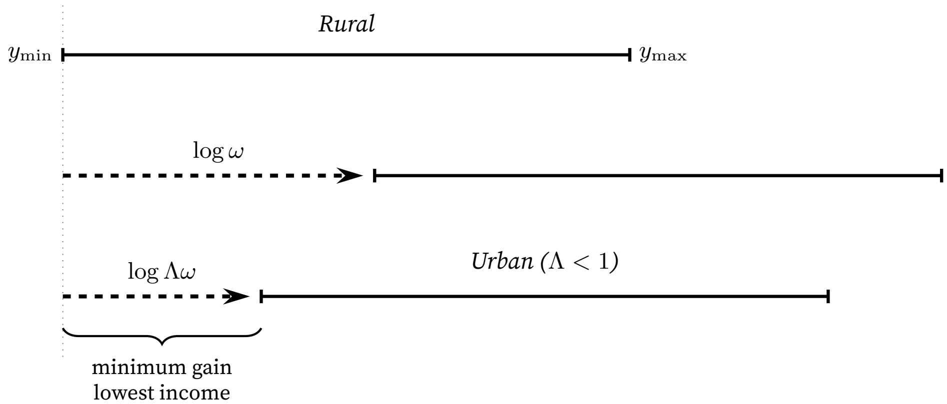
An important aspect of the solution routine is how the grid space is set up. Note that when households move from rural to urban, two things change. First, they end up on a different grid point that captures endowments in efficiency units in the new sector. And second, their income moves because efficiency 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.[67] 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 = LinRange(-2.5000 ,2.5000 ,21.0000), ϵ_grid = 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-

[67] 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 $c^{+}_{\text{RU}} = c_{\text{RU}} + \log\Lambda$ — 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.

linearly 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 .[68] 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

$$g_{t+dt} = \left(\mathbb{I} - \mathcal{A}_x^T dt\right)^{-1} g_t \tag{33}$$

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 $\pi_{\text{RU}}(x)$, so the probability of migration over the interval $[t, t+dt]$ for a household of type x equals

$$\mathbb{P}(migrate|x) = 1 - e^{-\lambda_{\text{m}}\pi_{\text{RU}}(x)dt}. \tag{34}$$

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

[68] This 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 $N(c_{0,\mathrm{U}}, \sigma_0)$. 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

$$\mathcal{A} = I_a \otimes \mathcal{A}_x + \mathcal{A}_{\text{age}}^{\text{drift}} \otimes I_x + \mathcal{A}_{\text{death}} + \mathcal{A}_{\text{mig}}, \tag{35}$$

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

$$\mathcal{A}_{\text{age}}^{\text{drift}} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}. \tag{36}$$

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

$$P_y = \left(\left(\iota_y \mu_y' \right) \otimes I_b \right). \tag{37}$$

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

$$P_{\text{death}} = M_{a=0} \otimes P_y. \tag{38}$$

The associated generator is

$$\mathcal{A}_{\text{death}} = \phi \left(-I_{ax} + P_{\text{death}}\right), \tag{39}$$

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

$$P_b^{\text{mig}} = \iota_b e'_{b=0}. \tag{40}$$

Define

$$P_{x,0}^{\text{mig}} = \left(\iota_y \mu'_y\right) \otimes P_b^{\text{mig}}. \tag{41}$$

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

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

The full migration injection operator is given by

$$P_{\text{mig}} = M_{a=0} \otimes P_{x,1}^{\text{mig}}. \tag{42}$$

The associated generator is

$$\mathcal{A}_{\text{mig}} = \left(-\left(I_a \otimes Diag\left(\lambda_m \pi(x)\right)\right) + P_{\text{mig}}\right). \tag{43}$$

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

$$\mathcal{A} = I_a \otimes \mathcal{A}_x + \mathcal{A}_{\text{age}}^{\text{drift}} \otimes I_x + \left(1 - \delta_{\text{RU}}\right) \mathcal{A}_{\text{death}} + \delta_{\text{RU}} \mathcal{A}_{\text{RU}}, \tag{44}$$

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, age$a \in \{0, 1, \ldots, 100\}$.

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

$$\bar{a}_j = \min\left\{ c \in \mathbb{Z}_+ : \frac{\int_0^c a g(a) da}{\int_0^c g(a) da} \geq a_{j,\min} \right\}, \tag{45}$$

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

3. Form the initial distribution.
   The initial joint distribution over$\mathbf{x}$ is

$$g_0(\mathbf{x}) = \frac{\int_0^{\bar{a}_s} g(\mathbf{x}, a) da}{\int \int_0^{\bar{a}_s} g(\mathbf{x}, a) da d\mathbf{x}}, \tag{46}$$

   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 41 summarizes the details of the simulation.

Table 41. 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

Table42 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`g_int` argument. When`g_int = nothing` (the default), the legacy path is taken. When`g_int`

Table 42. 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` in`solve_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 43. Simulated Income Covariance Matrix

(a) Rural

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.582				
y_{t+2}	0.289	0.604			
y_{t+4}	0.216	0.293	0.537		
y_{t+6}	0.188	0.22	0.301	0.543	
y_{t+8}	0.177	0.191	0.224	0.307	0.616

(b) Urban

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.418				
y_{t+2}	0.229	0.532			
y_{t+4}	0.218	0.251	0.466		
y_{t+6}	0.212	0.239	0.271	0.405	
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 44. Simulated Income Covariance Matrix Ignoring Selection

(a) Rural

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.586				
y_{t+2}	0.288	0.606			
y_{t+4}	0.215	0.299	0.542		
y_{t+6}	0.189	0.229	0.314	0.548	
y_{t+8}	0.177	0.201	0.243	0.327	0.621

(b) Urban

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.417				
y_{t+2}	0.229	0.531			
y_{t+4}	0.218	0.25	0.465		
y_{t+6}	0.212	0.238	0.271	0.404	
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.

Box constraints via logit transform: Each parameter has a known economic lower bound ℓ_i and upper

Table 45. Simulated Income Covariance Matrix

(a) Rural

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.582				
y_{t+2}	0.289	0.604			
y_{t+4}	0.216	0.293	0.537		
y_{t+6}	0.188	0.22	0.301	0.543	
y_{t+8}	0.177	0.191	0.224	0.307	0.616

(b) Urban

	y_t	y_{t+2}	y_{t+4}	y_{t+6}	y_{t+8}
y_t	0.418				
y_{t+2}	0.229	0.532			
y_{t+4}	0.218	0.251	0.466		
y_{t+6}	0.212	0.239	0.271	0.405	
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.

bound u_i. To avoid projection steps, we reparametrize using the element-wise logit (log-odds) transform:

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

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:

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

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 in detail 4.1.

The insurance construction. Suppose there exists an efficient financial sector that can fully insure migrants, both against the initial uncertainty surrounding migration outcomes and against subsequent long-term earnings risk. I assume the financial sector observes households' permanent and transitory types, which isolates the role of risk cleanly, and abstracts away from issues of moral hazard and adverse selection.

An efficient financial sector can commit to paying migrants a deterministic, life-contingent income stream equal to their expected labor income at each point in time,

$$y_s\left(p_t, \epsilon_t\right) := \mathbb{E}\left[\left(w_{\mathrm{U}} \Lambda\right) e^{p_s+\epsilon_s} | p_t, \epsilon_t\right],$$

where the expectation is taken over future income realizations conditional on current types and survival. This payment constitutes an actuarially fair annuity paid as long as the household is alive. Its present discounted value equals

$$H_{\mathrm{U}, t}\left(p_t, \epsilon_t\right) := \int_t^{\infty} e^{-r(s-t)} y_s\left(p_t, \epsilon_t\right) d s,$$

which follows conveniently from the Feynman-Kac formula, whereby H_{U} solves the linear system

$$r H_{\mathrm{U}}=\left(w_{\mathrm{U}} \Lambda\right) e^{p_t+\epsilon_t}+\mathcal{A} H_{\mathrm{U}},$$

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

$$\begin{aligned} c_t & =\frac{(\rho+\phi)+r(\gamma-1)}{\gamma} \cdot\left(a_t+H_{\mathrm{U}, t}\right), \\ V_t & =\frac{c_t^{1-\gamma}}{1-\gamma} \frac{1}{\frac{(\rho+\phi)+r(\gamma-1)}{\gamma}} . \end{aligned}$$

Insuring only the second margin, i.e., income risk after migration, means replacing the risky urban continuation value with this insured value while leaving the household uncertain about which $\left(p_{\mathrm{U}}, \epsilon_{\mathrm{U}}\right)$ they will draw. Insuring both margins means additionally offering a payout contingent on the rural type prior to migration,

$$H_m\left(p_{\mathrm{R}, t}\right)=\mathbb{E}\left[H_{\mathrm{U}}\left(p_t, \epsilon_t\right) | p_{\mathrm{R}}, \epsilon_{\mathrm{R}}\right],$$

where the expectation is computed against the distribution $p_{\mathrm{U}} | p_{\mathrm{R}}$ estimated earlier.

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 $nm\text{-}ni$ for the case in which both margins are removed, the migration probabilities equal

$$\begin{aligned} \pi_{\mathrm{RU}}^{ni}(x) & =\frac{(\omega)^s \cdot \tau_{\mathrm{RU}}^{-s}\left(-\mathbb{E}\left[v_{\mathrm{U}}^{ni}\left(H_{\mathrm{U}}, b\right) | x\right]\right)^{-\frac{s}{\gamma-1}}}{\left(-v_{\mathrm{R}}\right)^{-\frac{s}{\gamma-1}}+(\omega)^s \cdot \tau_{\mathrm{RU}}^{-s}\left(-\mathbb{E}\left[v_{\mathrm{U}}^{ni}\left(H_{\mathrm{U}}, b\right) | x\right]\right)^{-\frac{s}{\gamma-1}}}, \\ \pi_{\mathrm{RU}}^{nm\text{-}ni}(x) & =\frac{(\omega)^s \cdot \tau_{\mathrm{RU}}^{-s}\left(-v_{\mathrm{U}}^{nm\text{-}ni}\left(H_m, b\right)\right)^{-\frac{s}{\gamma-1}}}{\left(-v_{\mathrm{R}}\right)^{-\frac{s}{\gamma-1}}+(\omega)^s \cdot \tau_{\mathrm{RU}}^{-s}\left(-v_{\mathrm{U}}^{nm\text{-}ni}\left(H_m, b\right)\right)^{-\frac{s}{\gamma-1}}} . \end{aligned}$$

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.[69] The risk the financial sector insures is about income

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

alone, and 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 $nm\text{-}ni$ (A) and $nm\text{-}ni$ (B) respectively, and they turn out to be close.

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.

Table 46. Moments in counterfactual simulations

		insurance			robustness
Model moments	baseline	$\zeta = 0$	ni	$nm\text{-}ni$	$\sigma^2_{p,\mathrm{U}} = .003$
Urban population share (GE)	0.65	0.76	0.78	0.84	0.67
Urban population share (PE)	0.65	0.80	0.85	0.90	0.68
Rural-urban migration rate	0.02	0.03	0.04	0.05	0.02
Rural-urban wage gap ω	1.06	0.93	0.91	0.82	1.04
Log wage gap (cross-section)	0.45	0.46	0.24	0.12	0.43
Log wage gap (fixed effects)	0.35	0.55	0.09	-0.08	0.32
Log(migrant/stayer income in rural)	-0.16	-0.11	-0.19	-0.19	-0.15
Slope migration rate on income decile	-0.0075	-0.0057	-0.0157	-0.0203	-0.0073
Prob. income improvement rel. to stayer	0.18	0.28	0.06	-0.01	0.17
Variance of urban persistent type	0.21	0.12	0.12	0.12	0.05

The $nm\text{-}ni$ 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 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.7 Assets and Migration Rates – Analytical and Quantitative Results

This section uses an approximation that show 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,

and Miyauchi (2025) 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, $c_{\mathrm{R},0} < c_{\mathrm{U},0}$. 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 47. Jacobian in levels

	c_0	ρ_{RU}	c_{RU}	σ_{RU}	ζ	s	τ_{RU}	Λ_{RU}	$\sigma^2_{0,R}$
$\hat{\beta}_{cs}$	-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
$\hat{\rho}^{IV}_{RU}$	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
$\hat{\beta}^d_{R,RU}$	-0.01	0.41	0.06	0.18	-0.39	-0.01	-0.03	-0.01	-1.06
$\hat{\beta}^{\pi}_{dec}$	0.00	0.02	0.00	0.01	-0.02	-0.00	0.00	-0.00	-0.04
$\hat{\beta}^{\text{better}}_{R,RU}$	-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
Var$(y_{R,0})$	0.03	0.01	0.01	-0.01	0.00	0.00	0.01	-0.00	1.00

Table 48. Jacobian in percent

	c_0	ρ_{RU}	c_{RU}	σ_{RU}	ζ	s	τ_{RU}	Λ_{RU}	$\sigma^2_{0,R}$
$\hat{\beta}_{cs}$	-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
$\hat{\rho}^{IV}_{RU}$	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
$\hat{\beta}^d_{R,RU}$	-0.01	0.34	0.01	0.09	-0.30	-0.06	-0.03	-0.01	-0.15
$\hat{\beta}^{\pi}_{dec}$	0.00	0.02	0.00	0.01	-0.02	-0.00	0.00	-0.00	-0.01
$\hat{\beta}^{\text{better}}_{R,RU}$	-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
Var$(y_{R,0})$	0.02	0.01	0.00	-0.01	0.00	0.01	0.02	-0.00	0.15

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 make migration less risky.

To complement the quantitative findings in the main text, I provide a simple approximation here. Recall the value function given by

$$V_0 = \mathbb{E}_0 \left[\int_0^\infty e^{-\tilde{\rho}t} \frac{c_t^{1-\gamma}}{1-\gamma} dt \right]. \tag{47}$$

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,

$$y_t = y_0 \left(\theta_{B,0} + \theta_{L,0} e^{\sigma_p W_t} \right).$$

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

$$y_t \approx y_0 e^{\theta_{L,0}\sigma_p W_t}. \tag{48}$$

Substituting (48) into the value function in (47) yields

$$\begin{aligned} V_0 &\approx \frac{y_0^{1-\gamma}}{1-\gamma}\int_0^{\infty} e^{-\left(\tilde{\rho}-(\theta_{L,0}\sigma_p(1-\gamma))^2/2\right)t}dt \\ &= \frac{y_0^{1-\gamma}}{1-\gamma}\frac{1}{\tilde{\rho}-\frac{(\theta_{L,0}\sigma_p(1-\gamma))^2}{2}}. \end{aligned}$$

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 $p_{\mathrm{RU}} = \rho_{\mathrm{RU}}p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u$ where $u \sim N(0, \sigma^2_{\mathrm{RU}})$ and $c^{+}_{\mathrm{RU}} = c_{\mathrm{RU}} + \log\Lambda$ 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

$$y_{0,\mathrm{U}} = rb + w_{\mathrm{U}}e^{\rho_{\mathrm{RU}}p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u}.$$

The ratio of urban and rural income of a mover equals

$$\begin{aligned} \frac{y_{0,\mathrm{U}} - y_{0,\mathrm{R}}}{y_{0,\mathrm{R}}} &= \frac{w_{\mathrm{U}}e^{\rho_{\mathrm{RU}}p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u} - w_{\mathrm{R}}e^{p_{\mathrm{R}}}}{rb + w_{\mathrm{R}}e^{p_{\mathrm{R}}}} \\ &= \theta_{L,0,\mathrm{R}}\left(\omega e^{(\rho_{\mathrm{RU}}-1)p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u} - 1\right). \end{aligned}$$

I focus on a household for which the expected income gains are close to zero, i.e., $e^{\log\omega + (\rho_{\mathrm{RU}}-1)p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u} - 1 \approx 0$. In that case,

$$\frac{y_{0,\mathrm{U}}}{y_{0,\mathrm{R}}} \approx 1 + \theta_{L,\mathrm{R},0}\left(\log\omega + (\rho_{\mathrm{RU}} - 1)\,p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u\right).$$

Using this, the ratio of urban and rural expected utility, using the simplification that rural and urban labor share are identical, $\theta_{L,0,\mathrm{R}} = \theta_{L,0,\mathrm{U}} = \theta_{L,0}$ (ignoring the stochasticity of the urban labor share), equals

$$\frac{\mathbb{E}\left[V_{0,\mathrm{U}}\right]}{V_{0,\mathrm{R}}} \approx \mathbb{E}\left[\left(1+\theta_{L,0,\mathrm{R}}\left(\log\omega + \left(\rho_{\mathrm{RU}}-1\right)p_{\mathrm{R}} - c^{+}_{\mathrm{RU}} + u\right)\right)^{1-\gamma}\right] \frac{\tilde{\rho} - \frac{\left(\theta_{L,0}\sigma_{p,\mathrm{R}}(1-\gamma)\right)^2}{2}}{\tilde{\rho} - \frac{\left(\theta_{L,0}\sigma_{p,U}(1-\gamma)\right)^2}{2}}$$
$$= e^{-(\gamma-1)\theta_{L,0,\mathrm{R}}\left[\left(\log\omega + (\rho_{\mathrm{RU}}-1)p_{\mathrm{R}} - c^{+}_{\mathrm{RU}}\right) - \frac{(\gamma-1)\theta_{L,0,\mathrm{R}}\sigma^2_{\mathrm{RU}}}{2}\right]} \frac{\tilde{\rho} - \frac{\left(\theta_{L,0}\sigma_{p,\mathrm{R}}(1-\gamma)\right)^2}{2}}{\tilde{\rho} - \frac{\left(\theta_{L,0}\sigma_{p,\mathrm{U}}(1-\gamma)\right)^2}{2}}.$$

To finally arrive at the most simple version, let $\sigma^2_{p,\mathrm{R}} \approx 0$, $\gamma = 2$, roughly consistent with the quantitative application. Define the migration gain $d_{\mathrm{RU}} = \log\omega + \left(\rho_{\mathrm{RU}}-1\right)p_{\mathrm{R}} - c^{+}_{\mathrm{RU}}$ to obtain an equivalent but simpler expression,

$$\frac{\mathbb{E}\left[V_{0,\mathrm{U}}\right]}{V_{0,\mathrm{R}}} \approx \frac{e^{-\theta_{L,0}\left[d_{\mathrm{RU}} - \frac{1}{2}\theta_{L,0}\sigma^2_{\mathrm{RU}}\right]}}{1 - \frac{1}{\tilde{\rho}}\frac{1}{2}\left(\theta_{L,0}\sigma_{p,\mathrm{U}}\right)^2}.$$

Finally, note that the migration share is given by

$$\pi_{\mathrm{RU}} = \frac{1}{1 + \left(\tau_{\mathrm{RU}} \times \frac{\mathbb{E}\left[V_{0,\mathrm{U}}\right]}{V_{0,\mathrm{R}}}\right)^{s}},$$

where the exponent is $\frac{s}{\gamma-1}$ in general and equals s here because $\gamma = 2$. The share is strictly falling in the ratio $\frac{\mathbb{E}[V_{0,\mathrm{U}}]}{V_{0,\mathrm{R}}}$.[70]

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 $a := \frac{1}{\tilde{\rho}}\frac{1}{2}\sigma^2_{p,\mathrm{U}}$, and study the ratio R as a function of the labor share θ_L,

$$R\left(\theta_L\right) := \frac{e^{-\theta_L\left[d_{RU} - \frac{1}{2}\theta_L\sigma^2_{\mathrm{RU}}\right]}}{1 - a\theta_L^2} = \frac{\mathbb{E}\left[V_{0,\mathrm{U}}\right]}{V_{0,\mathrm{R}}}.$$

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 $d_{\mathrm{RU}} > \sigma^2_{\mathrm{RU}} + \frac{\sigma^2_{p,\mathrm{U}}}{\tilde{\rho} - \frac{1}{2}\sigma^2_{p,\mathrm{U}}}$. 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 $\left(\frac{1}{1-a\theta_L^2}\right)$ 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

[70] It 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.

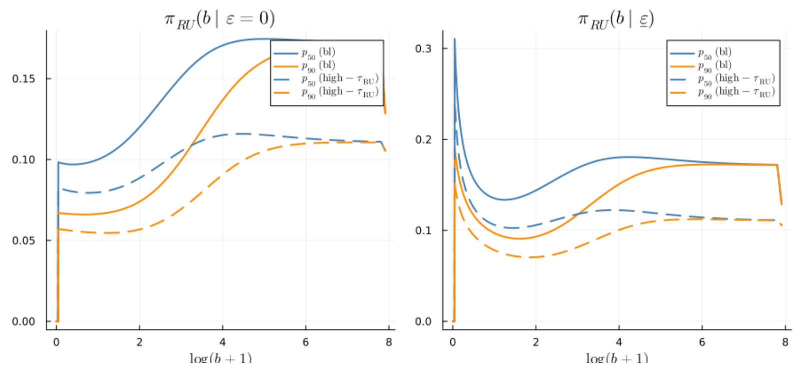
consumption risk around migration, given by $\frac{1}{2}\theta_L^2\sigma_{\text{RU}}^2$, and assets also reduce consumption risk after migration captured in the term $1/(1-a\theta_L^2)$. This positive impact of asset holdings on migration through reduced consumption volatility opens the door for non-monotone dynamics. A sufficient condition for non-monotone behavior in the migration probability to emerge is given by[71]

$$0 < d_{\text{RU}} < \sigma_{\text{RU}}^2 + \frac{\sigma_{p,\text{U}}^2}{\tilde{\rho} - \frac{1}{2}\sigma_{p,\text{U}}^2},$$

which holds for the majority of households.

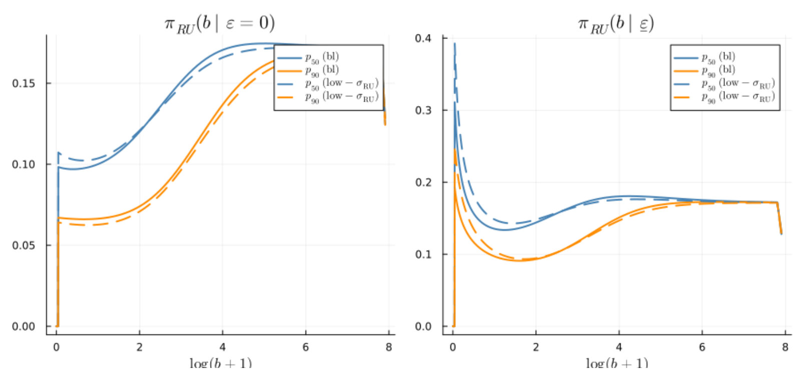
The below figures illustrate these forces in the full quantitative model by changing migration gains or risk, which then amplifies or reduces the aforementioned channels.

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 don't 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 space issue, that is irrelevant as the mass of agents in that region is virtually zero (<0.000001).

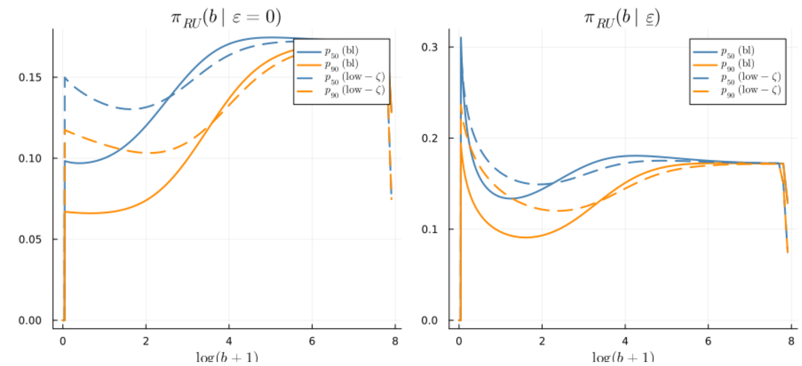
Figure 20. Selection into Migration – Low σ_{RU}



$\boldsymbol{\sigma_{\text{RU}}}$ (−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 space issue, that is irrelevant as the mass of agents in that region is virtually zero (<0.000001).

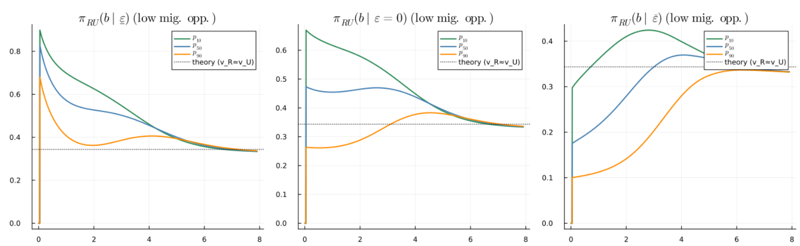
[71]This condition is most conveniently derived from taking logs and differentiating w.r.t. the labor share.

Figure 21. Selection into Migration – Low ζ



ζ **(×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 space issue, that 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 don't 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 $\pi_{\mathrm{RU}} = 1/(1+\tau^s_{\mathrm{RU}})$ 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,

$$\begin{aligned}
\mathbb{P}\left(\text{migrate}, x\right) &= \mathbb{P}\left(V^{\mathrm{U}}\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} > V_{\mathrm{R}}\varphi_{\mathrm{R}}^{1-\gamma}\right) \\
&= \mathbb{P}\left(\varphi_{\mathrm{R}}^{\gamma-1} < \frac{V_{\mathrm{R}}}{V_{\mathrm{U}}}\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{\gamma-1}\right) \\
&= \mathbb{E}_{\varphi_{\mathrm{U}}}\left[\mathbb{P}\left(\varphi_{\mathrm{R}} < \left(\frac{V_{\mathrm{R}}}{V_{\mathrm{U}}}\right)^{\frac{1}{\gamma-1}}\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)|\varphi_{\mathrm{U}}\right)\right] \\
&= \mathbb{E}_{\varphi_{\mathrm{U}}}\left[\exp\left(-\left(\frac{V_{\mathrm{R}}}{V_{\mathrm{U}}}\right)^{\frac{-s}{\gamma-1}}\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{-s}\right)\right] \\
&= \int_0^{\infty}\exp\left(-\left(\frac{V_{\mathrm{R}}}{V_{\mathrm{U}}}\right)^{\frac{-s}{\gamma-1}}\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{-s}\right)s\varphi_{\mathrm{U}}^{-s-1}e^{-\varphi_{\mathrm{U}}^{-s}}d\varphi \\
&= \frac{1}{\left(\frac{V_{\mathrm{R}}}{V_{\mathrm{U}}}\right)^{\frac{-s}{\gamma-1}}\left(\frac{1}{\tau}\right)^{-s}+1} \\
&= \frac{\tau^{-s}\left(-V_{\mathrm{U}}\right)^{\frac{-s}{\gamma-1}}}{\left(-V_{\mathrm{R}}\right)^{\frac{-s}{\gamma-1}}+\tau^{-s}\left(-V_{\mathrm{U}}\right)^{\frac{-s}{\gamma-1}}}
\end{aligned}$$

where I assumed $\gamma > 1$.

Expected value of migration. I next derive the expected value of migration conditional on migration, i.e.,

$$\mathbb{E}\left[\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma}V_{\mathrm{U}}|\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma}V_{\mathrm{U}} > \left(\varphi_{\mathrm{R}}\right)^{1-\gamma}V_{\mathrm{R}}\right].$$

To proceed, I need the distribution of $\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} | \left(\frac{\varphi}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} > (\varphi_{\mathrm{R}})^{1-\gamma} V_{\mathrm{R}}$, which can be derived as follows

$$
\begin{aligned}
\mathbb{P}\left(\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} \leq k | \left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} > (\varphi_{\mathrm{R}})^{1-\gamma} V_{\mathrm{R}}\right) &= \frac{\mathbb{P}\left(\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} \leq k \cap \left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} > (\varphi_{\mathrm{R}})^{1-\gamma} V_{\mathrm{R}}\right)}{\mathbb{P}\left(\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} > (\varphi_{\mathrm{R}})^{1-\gamma} V_{\mathrm{R}}\right)} \\
&= \frac{\mathbb{P}\left((\varphi_{\mathrm{R}})^{1-\gamma} V_{\mathrm{R}} \leq \left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} \leq k\right)}{\pi_{\mathrm{RU}}} \\
&= \frac{\int \mathbb{P}\left((\varphi_{\mathrm{R}})^{1-\gamma} > \left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} \frac{V_{\mathrm{U}}}{V_{\mathrm{R}}} | \varphi_{\mathrm{U}}\right) dF(\varphi_{\mathrm{U}}) \mathbb{1}_{\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} \leq k}}{\pi_{\mathrm{RU}}} \\
&= \frac{\int e^{-\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{-s} \left(\frac{V_{\mathrm{U}}}{V_{\mathrm{R}}}\right)^{\frac{-s}{1-\gamma}}} dF(\varphi_{\mathrm{U}}) \mathbb{1}_{\varphi_{\mathrm{U}} \leq \tau \left(\frac{k}{V_{\mathrm{U}}}\right)^{\frac{1}{1-\gamma}}}}{\pi_{\mathrm{RU}}} \\
&= \frac{\int_0^{\tau\left(\frac{k}{V^{\mathrm{U}}}\right)^{\frac{1}{1-\gamma}}} e^{-\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{-s} \left(\frac{V_{\mathrm{U}}}{V_{\mathrm{R}}}\right)^{\frac{-s}{1-\gamma}}} s\varphi_{\mathrm{U}}^{-s-1} e^{-(\varphi_{\mathrm{U}})^{-s}} d\varphi}{\pi_{\mathrm{RU}}} \\
&= \frac{1}{1+(\tau)^s \left(\frac{V_{\mathrm{U}}}{V_{\mathrm{R}}}\right)^{\frac{-s}{1-\gamma}}} \frac{\left[e^{-(\varphi_{\mathrm{U}})^{-s}\left(1+(\tau)^s \left(\frac{V_{\mathrm{U}}}{V_{\mathrm{R}}}\right)^{\frac{-s}{1-\gamma}}\right)} \right]_0^{\tau\left(\frac{k}{V_{\mathrm{U}}}\right)^{\frac{1}{1-\gamma}}}}{\pi_{\mathrm{RU}}} \\
&= \frac{1}{1+(\tau)^s \left(\frac{V_{\mathrm{U}}}{V_{\mathrm{R}}}\right)^{\frac{-s}{1-\gamma}}} \frac{\exp\left(-\left((-k)^{\frac{1}{1-\gamma}}\right)^{-s} \left((\tau)^{-s} (-V_{\mathrm{U}})^{\frac{s}{1-\gamma}} + (-V_{\mathrm{R}})^{\frac{s}{1-\gamma}}\right)\right)}{\pi_{\mathrm{RU}}} \\
&= \frac{(\tau)^{-s} (-V_{\mathrm{U}})^{\frac{-s}{\gamma-1}}}{(\tau)^{-s} (-V_{\mathrm{U}})^{\frac{-s}{\gamma-1}} + (-V_{\mathrm{R}})^{\frac{-s}{\gamma-1}}} \frac{\exp\Big(-\left((-k)^{\frac{s}{\gamma-1}}\right) \Big(\tau^{-s} (-V_{\mathrm{U}})^{\frac{-s}{\gamma-1}} + (-V_{\mathrm{R}})}{\pi_{\mathrm{RU}}} \\
&= \exp\left(-\left((-k)^{\frac{s}{\gamma-1}}\right) \left(\tau^{-s} (-V_{\mathrm{U}})^{\frac{-s}{\gamma-1}} + (-V_{\mathrm{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 $\tilde{\Phi} = \tau^{-s} (-V_{\mathrm{U}})^{\frac{-s}{\gamma-1}} + (-V_{\mathrm{R}})^{\frac{-s}{\gamma-1}}$, and compute

$$
\mathbb{E}\left[\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} | \max\right] = \int_{-\infty}^{0} k \frac{s}{\gamma-1} \tilde{\Phi} (-k)^{\frac{s}{\gamma-1}-1} \exp\left(-\tilde{\Phi} (-k)^{\frac{s}{\gamma-1}}\right) dk.
$$

Next, use the change of variable $x := -k$, which flips both the limits and the sign,

$$
\mathbb{E}\left[\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}} | \max\right] = -\int_{0}^{\infty} x \frac{s}{\gamma-1} \tilde{\Phi} x^{\frac{s}{\gamma-1}-1} \exp\left(-\tilde{\Phi} x^{\frac{s}{\gamma-1}}\right) dx.
$$

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 $\frac{s}{\gamma-1}\tilde{\Phi}x^{\frac{s}{\gamma-1}-1}dx = dy$,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\varphi_{\mathrm{U}}}{\tau}\right)^{1-\gamma} V_{\mathrm{U}}|\max\right] &= -\int_0^\infty \left(\frac{y}{\tilde{\Phi}}\right)^{\frac{\gamma-1}{s}} \exp\left(-y\right) dy \\
&= -\left(\frac{1}{\tilde{\Phi}}\right)^{\frac{\gamma-1}{s}} \int_0^\infty y^{\frac{\gamma-1}{s}} \exp\left(-y\right) dy \\
&= -\left(\frac{1}{\tilde{\Phi}}\right)^{\frac{\gamma-1}{s}} \Gamma\left(1 + \frac{\gamma-1}{s}\right).
\end{aligned}$$

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 $F\left(x\right) = e^{-\left(\frac{x}{T}\right)^{-s}}$ 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

$$\begin{aligned}
\hat{V} &= -\left(\tau^{-s}\left(-V_{\mathrm{U}}\right)^{\frac{-s}{\gamma-1}} + \left(-V_{\mathrm{R}}\right)^{\frac{-s}{\gamma-1}}\right)^{\frac{1-\gamma}{s}} \\
&= -\left(\tau^{-s}\left(-V_{\mathrm{U}}\right)^{\frac{s}{1-\gamma}} + \left(-V_{\mathrm{R}}\right)^{\frac{s}{1-\gamma}}\right)^{\frac{1-\gamma}{s}}.
\end{aligned}$$

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 $V_{\mathrm{U}}\left(b, \xi y\right) = \xi^{1-\gamma} v_{\mathrm{U}}\left(b_0, y\right)$ 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 $w_{\mathrm{U}}\Lambda$ for a rural-urban migrant.

Step 1. Note that for $v_{\mathrm{U}}\left(b_0, y\right)$ to solve the normalized household problem, it must be that

$$\begin{aligned}
\left(\rho+\phi\right) v_{\mathrm{U}}\left(b_0, y\right) &= \max \frac{c^{1-\gamma}}{1-\gamma} + \left(\partial_{b_0} v_{\mathrm{U}}\right)\dot{b}_0 + \left(\partial_y v_{\mathrm{U}}\right)\dot{y} \\
\text{s.t.:} \quad \dot{b}_0 &= rb_0 + y - c_0, \\
\text{FOC:} \quad \partial_{b_0} v_{\mathrm{U}} &= c_0^{-\gamma},
\end{aligned}$$

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 $V_{\mathrm{U}} = \xi^{1-\gamma} v_{\mathrm{U}}\left(b_0, y\right)$. This is the case whenever $\xi^{1-\gamma} v_{\mathrm{U}}\left(b_0, y\right)$ solves

$$\begin{aligned}
\left(\rho+\phi\right) V_{\mathrm{U}}\left(b, y\right) &= \max \frac{c^{1-\gamma}}{1-\gamma} + \left(\partial_b V_{\mathrm{U}}\right)\dot{b} + \left(\partial_y V_{\mathrm{U}}\right)\dot{y} \\
\text{s.t.:} \quad \dot{b} &= rb + \xi y - c.
\end{aligned}$$

Conjecture this is the case. Then, the first order condition implies

$$\begin{aligned}
\partial_b V_{\mathrm{U}} &= c^{-\gamma} \\
\partial_b \xi^{1-\gamma} v_{\mathrm{U}}(b_0, y) &= c^{-\gamma} \\
\xi^{1-\gamma} \frac{\partial_{b_0} v_{\mathrm{U}}(b_0, y)}{\xi} &= c^{-\gamma} \\
\partial_{b_0} v_{\mathrm{U}}(b_0, y) &= \left(\frac{c}{\xi}\right)^{-\gamma} \\
c_0^{-\gamma} &= \left(\frac{c}{\xi}\right)^{-\gamma},
\end{aligned}$$

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

$$\begin{aligned}
(\rho + \phi) V_{\mathrm{U}}(b, y) &= \max \frac{\xi^{1-\gamma} c_0^{1-\gamma}}{1-\gamma} + (\partial_b V_{\mathrm{U}}) \xi \dot{b}_0 + (\partial_y V_{\mathrm{U}}) \dot{y} \\
&= \max \frac{\xi^{1-\gamma} c_0^{1-\gamma}}{1-\gamma} + \xi^{1-\gamma} (\partial_{b_0} v_{\mathrm{U}}) \dot{b}_0 + \xi^{1-\gamma} (\partial_y v_{\mathrm{U}}) \dot{y} \\
&= \xi^{1-\gamma} \max \frac{c_0^{1-\gamma}}{1-\gamma} + (\partial_{b_0} v_{\mathrm{U}}) \dot{b}_0 + (\partial_y v_{\mathrm{U}}) \dot{y}.
\end{aligned}$$

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.[72] 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)

$$(\rho + \phi) V_{\mathrm{R}}(b, \xi y) = \max \frac{c^{1-\gamma}}{1-\gamma} + (\partial_b V_{\mathrm{R}}) \dot{b} + (\partial_y V_{\mathrm{R}}) \dot{y} + \lambda_m \left(\tilde{V}_{m,\mathrm{pi}} - V_{\mathrm{R}}\right)$$

[72] Holding 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

$$
\begin{aligned}
\tilde{V}_{m,\text{pi}} &= -\left[\tau_{\text{RU}}^{-s}\left(\tilde{V}_{\text{U,pi}}\left(b_{\text{R}}-f_{\text{RU}}w_{\text{R}},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(V_{\text{R}}\left(b_{\text{R}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}} \\
&= -\left[\tau_{\text{RU}}^{-s}\left(\left(w_{\text{U}}\Lambda\right)^{1-\gamma}\tilde{v}_{\text{U,pi}}\left(\frac{b_{\text{R}}-f_{\text{RU}}w_{\text{R}}}{w_{\text{U}}\Lambda},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(V_{\text{R}}\left(b_{\text{R}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma}\left[\tau_{\text{RU}}^{-s}\left(\frac{\left(w_{\text{U}}\Lambda\right)^{1-\gamma}}{w_{\text{R}}^{1-\gamma}}\tilde{v}_{\text{U,pi}}\left(\frac{b_{\text{R}}-f_{\text{RU}}w_{\text{R}}}{w_{\text{U}}\Lambda},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(\frac{V_{\text{R}}\left(b_{\text{R}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)}{w_{\text{R}}^{1-\gamma}}\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma}\left[\left(\frac{\tau_{\text{RU}}}{\omega\Lambda}\right)^{-s}\left(\tilde{v}_{\text{U,pi}}\left(\frac{\frac{b_{\text{R}}}{w_{\text{R}}}-f_{\text{RU}}}{\omega\Lambda},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(\frac{V_{\text{R}}\left(b_{\text{R}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)}{w_{\text{R}}^{1-\gamma}}\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}}.
\end{aligned}
$$

I conjecture that the same normalization applies to the rural value function, in which case this expression simplifies to

$$
\begin{aligned}
\tilde{V}_{m,\text{pi}} &= -w_{\text{R}}^{1-\gamma}\left[\left(\frac{\tau_{\text{RU}}}{\omega\Lambda}\right)^{-s}\left(\tilde{v}_{\text{U,pi}}\left(\frac{\frac{b_{\text{R}}}{w_{\text{R}}}-f_{\text{RU}}}{\omega\Lambda},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(\frac{V_{\text{R}}\left(b_{\text{R}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)}{w_{\text{R}}^{1-\gamma}}\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}} \\
&= -w_{\text{R}}^{1-\gamma}\left[\left(\frac{\tau_{\text{RU}}}{\omega\Lambda}\right)^{-s}\left(\tilde{v}_{\text{U,pi}}\left(\frac{\frac{b_{\text{R}}}{w_{\text{R}}}-f_{\text{RU}}}{\omega\Lambda},p_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}+\left(v_{\text{R}}\left(\frac{b_{\text{R}}}{w_{\text{R}}},p_{\text{R}},\epsilon_{\text{R}}\right)\cdot(-1)\right)^{-\frac{s}{\gamma-1}}\right]^{-\frac{\gamma-1}{s}} \\
&= w_{\text{R}}^{1-\gamma}\tilde{v}_{m,\text{pi}}\left(\frac{b_{\text{R}}}{w_{\text{R}}},p_{\text{R}},\epsilon_{\text{R}};\omega\right),
\end{aligned}
$$

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,

$$
(\rho+\phi)V_{\text{R}}(b,\xi y)=\max_c\frac{c^{1-\gamma}}{1-\gamma}+\left(\partial_b V_{\text{R}}\right)\dot{b}+\left(\partial_y V_{\text{R}}\right)\dot{y}+\lambda_m\left(\tilde{V}_{m,\text{pi}}-V_{\text{R}}\right).
$$

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

$$
\dot{g}=-\partial_b\left[\mu(b,y)\,g(b,y)\right]+\mathcal{L}_y g(b,y)
$$

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 $\xi c_0(b_0,y)=c(\xi b_0,\xi y)$. Further, 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, $w_{\mathrm{U}} = \omega w_{\mathrm{R}}$, 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

$$M = \frac{\phi P_{\text{U}} + \Pi}{\phi + \Pi}.$$

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,

$$\phi P_{\text{U}} = \phi M - (1 - M)\,\Pi.$$

In the original model, this was equal to $M\phi\,(1-\delta_{\text{UR}})$, so I need to show that $M\,(1-\delta_{\text{UR}}) = P_{\text{U}}$. 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

$$\dot{M} = -\phi M + \phi\,(1-\delta_{\text{UR}})\,M + \Pi\,(1-M) \Rightarrow M = \frac{\phi\,(1-\delta_{\text{UR}})\,M + \Pi}{\phi + \Pi}.$$

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 $M = \frac{\Pi}{\phi\delta_{\text{UR}} + \Pi}$, so that $\log\frac{M}{1-M} = \log\Pi - \log\left(\phi\delta_{\text{UR}}\right)$ 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 $\frac{\phi\delta_{\text{UR}}+\Pi}{\phi+\Pi}$. 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

$$\begin{aligned}\dot{M} &= \left(g_{n,U} - \phi\right) M + \Pi\,(1-M)\\ (1 - \dot{M}) &= g_{n,R}\,(1-M) - \phi\,(1-M) - \Pi\,(1-M)\,.\end{aligned}$$

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 $\phi\delta_{\text{UR}}$. 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 $g_{n,R} = \phi + \Pi$.