Selection and Heterogeneity in the Returns to Migration

Eduardo Cenci[1], Marieke Kleemans[2], and Emilia Tjernström[3]

[1] *University of Wisconsin, Madison*
[2] *University of Illinois, Urbana-Champaign*
[3] *University of Sydney*

April 2020

Latest version here.

Abstract

There is considerable debate on the returns to rural-urban migration in developing countries and magnitudes differ sharply depending on the method used. We aim to reconcile these divergent estimates by explicitly accounting for the role of heterogeneity in the returns to migration. We begin by using machine learning methods to examine key features of the distribution of conditional average returns. We exploit rich panel data from over 58,000 individuals in Indonesia whom we observe over a 28-year time period. The estimates from this exercise inform a multi-period Roy model, which allows for worker heterogeneity in both absolute and comparative advantage. We then explore several econometric models that allow us to estimate returns under both types of heterogeneity. The first of these is an instrumental variable approach to estimate marginal treatment effects using weather shocks interacted with historical migration patterns. The second is a correlated random coefficient model that gets identification from a parametric assumption on the relationship between comparative advantage and absolute advantage. This model lets us extrapolate the returns identified from switcher sub-populations to stayers—a group of particular interest to policy makers deciding whether to encourage migration as a development strategy. We then draw on recent developments in the literature on non-parametric panel data identification and employ a group random coefficient model that allows us to explicitly test the parametric assumptions that identify the returns to non-movers. We develop alternative identifying restrictions that allow us to estimate returns to the important non-mover population even when these parametric restrictions are rejected.

1 Introduction

As economies undergo structural transformation, labor tends to migrate out of rural areas and agriculture and into higher-productivity sectors in cities. The expected returns to migrating are an important factor in individuals' migration decisions, yet there is considerable debate on the magnitudes of these returns. On average in the developing world, urban residents tend to earn 2 to 3 times the amount of their rural counterparts and large wage differentials remain even when researchers control for a rich set of individual characteristics (Gollin *et al.*, 2014; Young, 2013).

Most economists would hesitate to interpret these cross-sectional wage gaps as returns to migration, since migrants may differ from non-migrants in many ways, some of which are unobservable to the econometrician. Two recent studies aim to address this issue by using panel data methods to account for time-invariant individual characteristics. Hicks *et al.* (2018), and Alvarez (2020) find that individual fixed effects regressions reduce the earnings gap by over 80 percent in data from Indonesia, Kenya and Brazil. Experimental estimates from Bangladesh documented in Bryan *et al.* (2014), which randomly incentivized people to migrate in the agricultural off season, are slightly larger at roughly 30 percent.

A potential way to reconcile these divergent estimates is to explicitly account for the role of heterogeneity. Different empirical methods rely on specific populations for identification, and if some groups have higher returns than others, estimates are bound to differ. If average income gaps are due to efficient sorting—as posited by Lagakos and Waugh (2013), Young (2013), and others—then existing earnings gaps would be poor predictors of the returns to migration for rural residents weighing the migration decision. Observational returns—such as those recorded by Hicks *et al.* (2018) and Alvarez (2020)—apply only to those who voluntarily move. Without further assumptions, it is hard to argue that these returns should necessarily extend to those who do not voluntarily move. In fact, returns to those who stayed behind may be smaller precisely because any higher returns have already been 'arbitraged away' by the early-movers.

Stayers could also have high expected returns if external constraints or costs keep them from migrating in the status quo—an argument that Lagakos *et al.* (2020) (henceforth LMMVW) carefully lay out. Experimental studies, such as Bryan *et al.* (2014), estimate the returns for individuals who are induced to move by a particular intervention. Again, we have little basis for extrapolating such estimates to the broader population that did not respond to the treatment. While these separate estimates are all informative, each approach provides but a piece of the puzzle of the returns to migration.

Our paper is closely related to LMMVW, who build a multi-region Roy (1951) model with heterogeneous workers who face different migration costs and test the model implications with detailed panel data from six developing countries. In line with LMMVW, we focus on heterogeneity in returns to migration, but also evaluate assumptions under which we may be able to extrapolate returns identified from switcher sub-populations to stayers. Our efforts are motivated by the fact that never-movers may be a particularly important group for policy makers deciding whether to encourage migration as a development strategy.

We begin by applying machine learning methods to explore both the determinants of migration and predictors of heterogeneity in returns to migration. Our data come from detailed panel surveys of over 58,000 individuals in Indonesia for whom we observe employment and location choices over a 28-year time period. We then develop a model of migration that builds on the two-period model in LMMVW. We extend their model to allow for back-and forth migration between rural and urban areas that we frequently observe in the data, and to include additional time periods to match the often-temporary nature of migration.

Starting with a framework that resembles the model in Lagakos and Waugh (2013), we illustrate the parallels of this model to the form of correlated random coefficient (CRC) model developed in Lemieux (1998) and Suri (2011), which allows for more than one type of unobserved heterogeneity. In particular, her model admits heterogeneity in both absolute and comparative advantage by imposing parametric restrictions on the relationship between the reduced form and the structural parameters. These restrictions allow us to use the estimated returns for switchers to extrapolate to stayers.

Drawing on recent developments in the literature on non-parametric panel data identification, we use the Group Random Coefficient (GRC) method developed by Ghanem *et al.* (2020) to estimate a more flexible version of the CRC model. The GRC model allows us to explicitly test the parametric identifying assumptions of the Suri (2011) model. In addition, the GRC reduced form allows us to examine some forms of heterogeneity without imposing the parametric restriction. Early results reveal substantial heterogeneity in returns across sub-populations. The specific form of heterogeneity generally does not conform to Suri's (2011) restriction, which highlights the importance of testing the restrictions and of exploring other identification approaches.

We then extend the model to accommodate temporary and permanent migration costs and discuss ways in which our estimation strategy can capture the heterogeneity in returns to migration within this extended version of the model. Work in progress develops alternative identifying restrictions that allow us to estimate returns to the important non-mover population even when these parametric restrictions are rejected.

2 Model

This section presents the model used in this paper in a few steps to parallel our different estimation strategies. Our most general model resembles Lagakos *et al.* (2019) with a few exceptions. We allow for back-and forth migration between rural and urban areas and consider a larger number of time periods to match the often temporary nature of migration.

We begin with a simple framework similar to that in Lagakos and Waugh (2013). The consumption of worker i in labor market l, which we restrict to either rural or urban, depends on two factors: her relative skill in the rural versus the urban labor market, and a place-specific productivity factor. For now, we assume that migration costs are zero.

In Section 4.3 we link this comparative-advantage driven model of migration choice to a correlated random coefficient (CRC) model that achieves identification using a parametric relationship between absolute and comparative advantage. We also use a more flexible generalized method of moments (GMM) estimation strategy that identifies heterogeneity in returns both in an unrestricted and a restricted model. The restricted model allows us—if the restriction holds—to estimate the returns for stayers.

We then extend the model to accommodate migration costs—both temporary and permanent—and discuss the extent to which our estimation strategy can capture the heterogeneity in returns to migration predicted by this extended versions of the model.

2.1 Modeling comparative advantage in rural-urban migration

The environment

The economy is composed of a unit measure of individuals indexed by i, who live for $T \geq 2$ periods. These individuals can work in one of two closed competitive labor markets: rural (r) and one urban (u). Labor markets are indexed by $l = \{r, u\}$. At time $t = 0$, a share π of individuals is born (or appear in the data) in the urban market.

Individuals make their migration decisions in period $t-1$ for the next period (t), and they can choose to stay where they are or migrate to the other market. In other words, migration occurs both from rural to urban and from urban to rural.

Endowments and preferences

Individuals are endowed with one unit of time which they supply inelastically to labor in each period. They are also endowed with a bivariate skill vector $z_i = \{z_i^r, z_i^u\}$ that determines how many efficiency units of labor they supply in each labor market.

We assume that each skill vector has a cumulative distribution function $F_l(z^l)$, and probability density function $f_l(z^l)$, with support on the positive reals. We assume these to be "regular" distributions, i.e., that their moments exist and that they are continuous and differentiable. $G(z^r, z^u)$, also a "regular" distribution, denotes the joint distribution of skills.

Technology, market clearing and equilibrium

A representative firm in each market uses labor as the only input to produce a single consumption good, Y. Production is linear in labor everywhere: $Y_l = A^l L_l$, where A^l is the productivity effect of labor market l area. We dub this term the "place factor".

Let Ω_l denote the set of workers employed in labor market l and N_l be the total number of workers employed there. Then

$$\begin{aligned} L_l &= \int_{\Omega_l} z^l \mathrm{dF} \\ N_l &= \int_{\Omega_l} \mathrm{dF} \end{aligned} \tag{1}$$

Finally, let C_t be the aggregate consumption in the economy in period t and assume that markets clear in each period:

$$\begin{aligned} \sum_{l \in \{r,u\}} N_{lt} &= 1, \\ \sum_{l \in \{r,u\}} Y_{lt} &= C_t \end{aligned} \tag{2}$$

We normalize the price of the consumption good to one. Thus, in equilibrium, the wages per efficiency unit of labor are denoted by $w^l = A^l,\ \forall l \in \{r, u\}$, which implies $w^u = A^u$ and $w^r = A^r$.[1]

[1]Notice that we could normalize the place factor in the rural labor market to 1 and get the following expression for the comparative advantage of individual i, $z_i^u/z_i^r = 1/A^u$, which is similar to the expression for comparative advantage in Lagakos and Waugh (2013) (equation 3): $z_n^i/z_a^i = p_a$.

Migration decision

We assume that individual i's consumption in each period t and market l is subject to a random shock $e^{u^l_{it}}$. Thus, the log of Mr i's consumption is

$$\begin{cases} c_{ilt} = \log z_i^r + \log A^r + u_{it}^r & \text{if } l = \text{ rural} \\ c_{ilt} = \log z_i^u + \log A^u + u_{it}^u & \text{if } l = \text{ urban} \end{cases} \tag{3}$$

Recall that migration decisions are made in period $t-1$ for the next period (t), thus, they are made in expectation. We assume that individuals "know their type" and have complete information on place factors A^l. Thus, their expected log consumption in labor market l and period t is

$$\log z_i^l + \log A^l + E\left[u_{it}^l\right]$$

Now we define some variables that represent the differences between urban and rural: $u_{it} \equiv u_{it}^u - u_{it}^r$, $d_i^z \equiv \log z_i^u - \log z_i^r$, and $d^A \equiv \log A^u - \log A^r$. Then we can write out the decision problem that a rural worker at time $t-1$ considers when deciding whether or not to migrate to the urban labor market in time t. She will migrate if

$$\begin{aligned} \log z_i^u + \log A^u + E\left[u_{it}^u\right] &\geq \log z_i^r + \log A^r + E\left[u_{it}^r\right] \\ \Rightarrow d_i^z + d^A &\geq E\left[u_{it}\right] \end{aligned} \tag{4}$$

To examine the decision to migrate from urban to rural, we just reverse the inequality to $<$.

Note that individual i does not base her migration decision on her initial labor market. In the absence of differences in the expected random shocks across labor markets, we would not expect to see additional movement after $t = 2$. Variation in expected random shocks $E\left[u_{it}\right] = E\left[u_{it}^r - u_{it}^u\right]$ will determine changes in migration status for periods $t \geq 3$.[2]

The result is intuitive: individuals migrate if their returns to migration, determined by the difference in skills d_i^z and the difference in place factors d^A, is larger than the expected realization of their random shocks.

[2]Without these random variations, individuals would choose a labor market for period $t = 2$ and stay there for all subsequent periods. For example, if $E\left[u_{it}\right] = 0$ for all i and t, an individual will choose urban for period $t = 2$ if $d_i^z + d^A \geq 0$ and remain urban for all next periods since neither d_i^z nor d^A will change for $t \geq 3$.

2.2 Adding a cost of being away from home (the "home factor")

In this extension, individuals are more productive at home, which we assume to be their starting labor market. This "home factor" is denoted by $\psi > 1$. Such an assumption can be justified by the existence of location-specific knowledge as in Bazzi *et al.* (2016) or a home-premium as in Kennan and Walker (2011).

Similar to Lagakos *et al.* (2020), who assume that an "inexperienced migrant" (i.e., a rural migrant who undergoes a disutility for being in the urban environment) can become "experienced" with some probability every period, we assume that the current labor market can become "home" for a migrant i with probability α_{it}. Like we did with the unobserved skills z_i^l, we assume that this probability follows a "regular" distribution $F_\alpha(\alpha)$ with support on the interval $[0, 1]$.

Therefore, we write the home factor as

$$\psi_{it} = \begin{cases} \psi & \text{if } l_{it} = l_{i1} \\ \psi^{\alpha_{it}} & \text{if } l_{it} \neq l_{i1} \end{cases}$$

where $l_{it} = r, u$ is the labor market of worker i in period t, and l_{i1} determines "home" (the labor market in the initial period).

We assume that individuals can still move in any direction and that there are no "one-time" migration costs (e.g., bus fares).

Migration decision

Define $\varphi \equiv \log \psi$ and consider an individual starting in the rural market ($l_{i1} = r$).

She migrates to (or remains at) the urban market in any period $t > 1$ if

$$\log z_i^u + \log A^u + E\left[\alpha_{it}\right] \log \psi + E\left[u_{it}^u\right] \geq \log z_i^r + \log A^r + \log \psi + E\left[u_{it}^r\right]$$

$$\Rightarrow d_i^z + d^A - (1 - E\left[\alpha_{it}\right])\varphi \geq E\left[u_{it}\right]$$

And she migrates to (or remains at) the rural market when the sign in the inequality changes to $<$.

Now, imagine that this individual had started in the urban market instead. In this case, she would migrate to (or remain at) the urban market in any period $t > 1$ if

$$\log z_i^u + \log A^u + \log \psi + E\left[u_{it}^u\right] \geq \log z_i^r + \log A^r + E\left[\alpha_{it}\right] \log \psi + E\left[u_{it}^r\right]$$

$$\Rightarrow d_i^z + d^A + (1 - E[\alpha_{it}])\varphi \geq E[u_{it}]$$

And she would migrate to (or remain at) the rural market when the sign in the inequality changes to $<$.

In summary, the decision to "go urban" is:

$$\begin{cases} d_i^z + d^A - (1 - E[\alpha_{it}])\varphi \geq E[u_{it}] & \text{if } l_{i1} = r \text{ (} i \text{ was born rural)} \\ d_i^z + d^A + (1 - E[\alpha_{it}])\varphi \geq E[u_{it}] & \text{if } l_{i1} = u \text{ (} i \text{ was born urban)} \end{cases}$$

There are important differences between this condition and the one we had before ($d_i^z + d^A \geq E[u_{it}]$).

1. The migration decision now depends on the initial labor market of individual i: l_{i1}.
2. Maintaining urban status is generally costlier for workers starting rural, than it is for workers starting urban since $E[\alpha_{it}] \in [0, 1]$ and $\varphi > 0$ will imply $(1 - E[\alpha_{it}])\varphi \geq 0$.
3. Changes in urban status over time (e.g., returned and back-and-forth migration) may occur regardless of variation in $E[u_{it}]$: variations in $E[\alpha_{it}]$ will suffice.

2.3 Allowing for permanent and temporary (one-time) migration costs

Assume individuals who chose to migrate out of labor market l' incur in costs equal to a share $(1 - s_l^m)$ of their consumption in the next period. In other words, they retain a share $s_l^m < 1$ of the consumption in their first period in the new market l. We assume this one-time migration cost to be uniform across destinations, individuals, and time ($s_l^m = s^m$ for all $l = \{r, u\}$).[3]

[3] We can, however, allow increasing flexibility:

1. Let the migration cost vary by destination l, which, in our case with two labor markets mean it varies with the direction of the move: $s_u^m \neq s_r^m$ (with s_u^m representing the one-time cost a $r \to u$ move and s_r^m representing that of a $u \to r$ move).
2. Let the migration cost vary by person (s_{il}^m): someone with access to migrant networks may have a lower cost than someone without that access.
3. Let the migration cost vary by person and time (s_{ilt}^m): this could explicitly account for cost decreasing with the amount of times someone has migrated in the past (with possibly different effect for each direction).

With we extend the model to allow several markets of each type—a number R of rural markets and a number U of urban markets—the term could grow in flexibility, differing for each origin-destination pair. For example, Before the difference between a certain home factor $\log \psi$ and the expectation of the home factor when away from home $E\left[\alpha_{it}\right]$ represented a permanent cost of migration. Now, on top of that permanent cost, migrants pay a one-time cost equal to a share $(1 - s_l^m)$ of their consumption in the next period every time they move. If before the addition of the permanent cost made keeping an urban status costlier for rural born individuals, now the addition of the temporary migration cost will discourage status switching (back-and-forth migration).

Migration decision

Define $m \equiv \log s^m$ and notice that $m < 0$ because $s^m \in (0,1)$.

Consider an individual who started showing in the data (was "born") in the the rural market ($l_{i1} = r$) and is currently at the rural market ($l_{it} = r$). She migrates to the urban market in period $t+1$ if

$$\log z_i^u + \log A^u + E\left[\alpha_{i,t+1}\right] \log \psi + \log s^m + E\left[u_{i,t+1}^u\right] \geq \log z_i^r + \log A^r + \log \psi + E\left[u_{i,t+1}^r\right]$$

$$\Rightarrow d_i^z + d^A - (1 - E\left[\alpha_{i,t+1}\right])\varphi + m \geq E\left[u_{i,t+1}\right]$$

And she remains at the urban market in the next period $(t+2)$ if

$$\log z_i^u + \log A^u + E\left[\alpha_{i,t+2}\right] \log \psi + E\left[u_{i,t+2}^u\right] \geq \log z_i^r + \log A^r + \log \psi + \log s^m + E\left[u_{i,t+2}^r\right]$$

$$\Rightarrow d_i^z + d^A - (1 - E\left[\alpha_{i,t+2}\right])\varphi - m \geq E\left[u_{i,t+2}\right]$$

Similarly, an individual "born" in the rural market in ($l_{i1} = r$) but currently at the urban market ($l_{it} = u$) will remain there in period $t+1$ if

$$d_i^z + d^A - (1 - E\left[\alpha_{i,t+1}\right])\varphi - m \geq E\left[u_{i,t+1}\right]$$

And she will still remain at the urban market in $t+2$ if

$$d_i^z + d^A - (1 - E\left[\alpha_{i,t+2}\right])\varphi - m \geq E\left[u_{i,t+2}\right]$$

---

s_{ilt}^m can be a function of distance between any two markets: $s_{it}^m = s_{it}^m(l_o, l_d)$, where l_o denotes origin, and l_d the destination.

In summary, the decision to "go urban" in $t+1$ will be

$$\begin{cases} d_i^z + d^A - (1 - E[\alpha_{i,t+1}])\varphi + m \geq E[u_{i,t+1}] & \text{if } l_{it} = r \ (i \text{ starts rural}) \\ d_i^z + d^A - (1 - E[\alpha_{i,t+1}])\varphi - m \geq E[u_{i,t+1}] & \text{if } l_{it} = u \ (i \text{ starts urban}) \end{cases}$$

And the decision to "remain urban" in $t+2$ will be[4]

$$\begin{cases} d_i^z + d^A - (1 - E[\alpha_{i,t+1}])\varphi - m \geq E[u_{i,t+2}] & \text{if } l_{it} = r \ (i \text{ starts rural}) \\ d_i^z + d^A - (1 - E[\alpha_{i,t+1}])\varphi - m \geq E[u_{i,t+2}] & \text{if } l_{it} = u \ (i \text{ starts urban}) \end{cases}$$

There is one important difference between this condition and the one we had before with the introduction of the permanent cost (home factor) in Section 2.2: the inclusion of one-time migration costs adds inertia. It is generally costlier to migrate than it is to remain. In other words, it is generally costlier to "switch" ($h_{it} = 0 \rightarrow 1$ and vice-versa) in the next period than it is to remain.

3 Data

We use the Indonesia Family Life Survey (IFLS) to test the predictions of the model described in the previous section. The IFLS is a panel dataset that is known for relatively low rates of attrition and representative of about 83% of the Indonesian population (Strauss et al., 2004)[5]. The analyses are based on all five waves of the IFLS, collected between 1993 and 2015 (Strauss, Witoelar and Sikoki, 2016).

The original 1993 sample consisted of 22,347 individuals, but with the inclusion of additional members from split-off households in subsequent rounds a sample of 58,337 unique individuals is obtained. Detailed data on current and previous employment and location were collected during each IFLS survey round, allowing us to create up to a 28-year annual individual employment and migration panel, from 1988 to 2015[6].

[4]The decisions for an individual "born" in the urban market are analogous, with the only difference being the sign of $(1 - E[\alpha_{i,t+1}])\varphi$, which turns positive.

[5]Attrition is often high in panel data, but with intensive focus on respondent tracking, the IFLS is well-suited to study migration. Re-contact rates between any two rounds are above 90%, and 87% of original households were contacted in all five rounds (Strauss et al. 2016). Thomas et al. (2012) contains a detailed discussion of tracking and attrition in the IFLS.

[6]Please refer to Kleemans (2015) and Kleemans and Magruder (2018) for more details on the IFLS panel dataset.

4 Empirical Strategy

In line with the model, we present three complementary approaches to study heterogeneous returns to migration. We use the insights of the model to understand and compare the empirical estimates and the sub-populations they apply to. We then use these estimates to inform subsequent extensions to the model.

First, we employ the machine learning techniques developed by Chernozhukov et al. (2018)—henceforth CDDF— to examine key features of the distribution of treatment effect heterogeneity. The CDDF method estimates these features for a binary, exogenous treatment effect. Selection into migration is of course not random, so we employ the approach in Deryugina *et al.* (2019), which adapts CDDF to instrumental variables.

The instrumental variable for this analysis follows Kleemans and Magruder (2018) who employ rainfall shocks interacted with historical migrant networks and show this is a good predictor of internal migration in Indonesia. To the extent that individuals respond to this instrument, we will be able to assign these individuals a propensity score, namely the probability that they migrate to an urban area.

Second, we leverage the above propensity score in the MTE approach redefined by Zhou and Xie (2019) to to estimate marginal treatment effects of the urban migration "treatment", allowing us to recover several policy-relevant parameters (Heckman and Vytlacil, 1998, 2007; Cornelissen *et al.*, 2016). The parameters that we can recover using this MTE/IV approach are particularly valuable: they can reveal an important degree of heterogeneity in returns to migration and they allow us to predict returns to urban migration for those who never migrate—using different assumptions than the CRC model assumptions below. The MTE/IV strategy also allows us to compare our results with experimental estimates reported in Bryan *et al.* (2014), Lagakos *et al.* (2019), and Lagakos *et al.* (2020).

Third, we take a different approach to selection bias by estimating a Correlated Random Coefficient (CRC) model. CRC models remove bias in average treatment effect estimates while allowing for more flexible selection patterns and heterogeneity in returns than fixed effects models. A large literature on CRC models explores different ways of estimating such models: Wooldridge (1997) and Heckman and Vytlacil (1998) use instrumental variables; Lemieux (1998) and Suri (2011) rely on a parametric restriction on the relationship between comparative advantage and absolute advantage.

We estimate a model similar to that in Lemieux (1998) and Suri (2011), which allows for more than one type of unobserved heterogeneity. In particular, the model admits heterogeneity in both absolute and comparative advantage by imposing parametric restrictions on the relationship between the reduced form and the structural parameters. These restrictions allow us to extrapolate from the estimated returns for switchers to those for the non-movers—but of course, the extrapolation is only valid if the parametric restriction holds.

We also present results from a group correlated coefficient (GRC) approach developed by Ghanem *et al.* (2020), which is a more flexible approach to estimating the CRC model proposed by Lemieux (1998) and Suri (2011). The GRC model has a reduced form that estimates average earnings/consumption in the rural labor market as well as average returns to urban migration for different groups of individuals—where groups are defined based on their migration trajectories. By imposing further restrictions, the model can recover the same parameter estimates as Suri (2011) and identify the returns to migration for non-migrants. Importantly, the GMM estimation allows us to test the parametric identification restriciton in a straightforward way. Furthermore, we can extend the GMM estimation method by imposing alternative restrictions that are less rigid than Suri's specific restriction on the form of comparative advantage.

4.1 Prediction with Machine Learning Techniques

This discussion draws on the discussion in Chernozhukov *et al.* (2019).[7] A key challenge with machine learning tools in high-dimensional settings is that they typically require strong assumptions to produce consistent estimators of conditional average treatment effects (CATE). The new method developed in **?** sacrifices some generalizability, but allow researchers to instead rely on fewer assumptions.

In particular, instead of trying to make inference on the full CATE function, the method focuses on making inference on key *features* of the CATE. These features are (1) the Best Linear Predictor (BLP) of the CATE function, (2) Sorted Group Average Treatment Effects (GATES), reporting predicted treatment effects at different deciles of the predicted treatment effect distribution, and (3) Classification Analysis (CLAN). The CLAN results show how covariates of interest differ between the units that we predict will be the most and least affected, with these most-affected and least-affected groups defined by the highest and lowest deciles of the predicted treatment effect distribution. Below, we provide some more intuition for the BLP.

Briefly, the method splits the data into an auxiliary subset, separate from the main data (the data is split into main and auxiliary many times, as is standard with ML techniques). Letting Y^0 and Y^1 denote potential outcomes under control and treatment, respectively,

[7] Section 6.2 in their paper describes the implementation algorithm in detail.

we can write out two key functions: $b_0(Z) := E[Y^0|Z]$, which is the baseline conditional average, and $s_0(Z) := E[Y^1|Z] - E[Y^0|Z]$. Given a randomly assigned treatment variable D, a known propensity score $p(Z)$, and a few more assumptions on the propensity score, the observed outcome can be written as a regression function (here, conditional on D, Z): $Y = b_0(Z) + Ds_0(Z) + U$, where $E[U|Z, D] = 0$.

Since our study does not have a randomized treatment, we rely on our instrument (weather shocks interacted with historical migration patterns) to predict migration for each individual-time observation and use the predicted probability of migration, $\hat{p}(Z)$, as in Deryugina *et al.* (2019).

We then proceed by using each auxiliary sample to train an ML estimator and obtain ML estimates of the baseline and treatment effects, called proxy scores. We will refer to the estimated proxies of $b_0(Z)$ and $s_0(Z)$ as $B(Z)$ and $S(Z)$, respectively. Note that we can then use these predicted proxies in the main sample to estimate the BLP of the conditional average treatment effect. Essentially, we regress the observed outcome on the treatment variable minus the (predicted) propensity score to estimate the average treatment effect and on the treatment variable minus the propensity score *interacted with* deviations of the $S(Z)$ that we estimated in the auxiliary data from the expected value of $S(Z)$ in the main sample. The coefficient on this second interaction term is what provides information about treatment effect heterogeneity. More specifically, we obtain the BLP parameters by estimating the following relationship in the main sample, using weighted OLS:

$$Y_i = \hat{\alpha}' X_{1,i} + \hat{\beta}_1(D_i - p(Z_i)) + \hat{\beta}_2(D_i - p(Z_i))(S_i - \mathbb{E}_N S_i) + \hat{\varepsilon}_i \tag{5}$$

where $S(Z)$ is written as S for simplicity, $\mathbb{E}_N[w(Z_i)\hat{\varepsilon}_i X_i] = 0$ with $w(Z_i) = \{p(Z_i)(1 - p(Z_i))\}^{-1}$. Further, $X_{i,1}$ is constructed as $X_i = [X'_{1,i}, (D_i - p(Z_i), (D_i - p(Z_i)(S_i - \mathbb{E}_N S_i)]'$, and $X_{1,i}$ includes a constant, $B(Z_i)$, and $S(Z_i)$. In the above regression, the estimated β_1 is the ATE, and β_2 is best linear predictor of the existing heterogeneity. If what we estimate in the auxiliary sample ($S(Z)$) is a perfect proxy for the true heterogeneity, $s_0(Z)$, then $\beta_2 = 1$. If there is no heterogeneity, and the estimates from the ML are pure noise, then $\beta_2 = 0$.

We consider a range of possible baseline variables that could plausibly be correlated with the returns to migration. This includes socio-demographic variables such as age, gender, education, marital status, household structure, assets, health indicators and information on formal and informal insurance; employment data, such as sector of employment and hours worked; and migration data, including location of birth, current location and migration history.

4.2 Instrumental Variables and Marginal Treatment Effects

4.3 The Group Correlated Coefficient Model

This section borrows notation from Ghanem *et al.* (2020) and briefly explains the approach presented there to estimate heterogeneity in returns to a binary choice variable: the group random coefficient (GRC). For clarity, we start with the simplest model presented in Section 2.1. After having established the notation and the rationale that connects that simple model and the estimation strategy, we repeat the process for the model extended in Sections 2.2 and 2.3.

Define an urban indicator $h_{it} \equiv 1\{l = \text{urban in time } t\}$, so that we can rewrite log consumption as

$$c_{it} = \log z_i^r + (\log z_i^u - \log z_i^r)h_{it} + \log A^r + (\log A^u - \log A^r)h_{it} + u_{it}^r + (u_{it}^u - u_{it}^r)h_{it}$$

With the definitions used before ($u_{it} \equiv u_{it}^r - u_{it}^u$, $d_i^z \equiv \log z_i^u - \log z_i^r$, and $d^A \equiv \log A^u - \log A^r$), this simplifies to

$$c_{it} = \log z_i^r + \log A^r + (d_i^z + d^A)h_{it} + u_{it}^r - u_{it}h_{it}$$

Now, let's normalize $A^r = 1$ so that $\log A^r = 0$ and $d^A = \log A^u$. Also, define $\varepsilon_{it} \equiv u_{it}^r + (u_{it}^u - u_{it}^r)h_{it} = u_{it}^r - u_{it}h_{it}$. The expression further simplifies to

$$c_{it} = \log z_i^r + (d_i^z + d^A)h_{it} + \varepsilon_{it}$$

Let a_i be a collection of all individual characteristics: the observable ones, denoted by X_i and omitted here for simplicity, and the unobservable ones like $z_i = \{z_i^r, z_i^u\}$. Then, the log consumption of individual i in period t can be modeled as a generic function of the individual's characteristics a_i and the urban indicator h_{it} (the adoption indicator) plus an additive error term ε_{it}.[8]

$$c_{it} = f(h_{it}, a_i) + \varepsilon_{it}$$

Define $\mu_i \equiv f(0, a_i)$ and $\Delta_i \equiv f(1, a_i) - f(0, a_i)$. Then, the log consumption of individual

[8] It is trivial to include also an additive time effect f_t, common to all individuals, which we do in our estimations. However, to keep the exposition of the model simple here, we omit these time fixed-effects as we do with the vector of time-invariant observable characteristics X_i.

i in period t becomes

$$c_{it} = \mu_i + \Delta_i h_{it} + \varepsilon_{it}$$

with $\mu_i = \log z_i^r$ and $\Delta_i = d_i^z + d^A$.

4.3.1 The unrestricted GRC model

Ghanem *et al.* (2020) follow Suri (2011) in assuming strict exogeneity: $E[\varepsilon_{it}|h_i, a_i] = 0$, where $h_i = (h_{i1}, ..., h_{iT})$ is the complete history of labor market choices of individual i.[9]

We can use the unrestricted GRC to estimate average returns to adoption for different groups of individuals in our sample. Such groups are determined by the different migration trajectories (h_i) we observe.

For example, with two periods ($T = 2$), the possible trajectories are as follows:

- The never-urban group, h_n, begin in the rural labor market and stay there;
- Joiners, h_{01}, are people who start in the rural labor market and move to the urban market;
- Leavers, h_{10}, move from the urban labor market to the rural market;
- Always-urban h_a start in the urban labor market and stay there;

Without further assumptions, we cannot identify the return to migration for non-switchers (we will do so in the next subsection). However, even the simplest version of our model and this unrestricted version of the GRC approach already allows us to move one step further than traditional panel fixed-effect models (FE). In these models, with the inclusion of an adoption dummy and a time fixed-effect dummy, the return to migration is identified out of switchers only (like here) and collapsed in a weighted average of those returns (unlike here) (Gibbons *et al.*, 2018). In the GRC model, we identify non-urban earnings and the returns to migration (the differential between consumption in urban relative to the rural markets) for every possible group of switcher in our data.

To the extent that these groups of migration trajectories are predictive of individuals' skills in urban and rural markets, the identification of average returns for groups defined by trajectories significantly improves our ability to identify and discuss heterogeneity in returns to migration. Generally, we can always identify at least 2^T groups based on the migration

[9] For example, with $T = 3$, an individual whose history is $h_i = (0, 1, 0)$ starts in the rural market in $t = 1$, migrates to the urban in $t = 2$ and returns to the rural market in $t = 3$.

histories (or trajectories) of individuals. Therefore, we can will be able to identify different returns to migration for $2^T - 2$ groups. With 5 survey waves available in our data, this means up to 30 different estimates for the return to migration.

The table below summarizes what we can estimate using the unrestricted version of the GRC model when the number of periods is $T = 2$ and $T = 3$ (it is trivial to extend the table for any $T \geq 2$). Notice that we can identify baseline consumption (μ_h) for all switchers and also for the "never-urban" group. We can identify the returns to urban migration (Δ_h), however, for the switchers only. For the last group, the "always urban" we identify only the sum of these two estimate $\kappa_h = \mu_h + \Delta_h$ (we deliberately substitute the subscript i for h to make it clear that our estimates refer not to an individual, but for a group of individuals who share a common migration history).

Table 1: Returns to urban migration by trajectory ($T = 2$ and $T = 3$, simple model)

Trajectory	Baseline log cons when rural (μ_h)	Return to urban migration (Δ_h)
	$T = 2$ **periods**	
$h_n = (0,0)$	$\log z_n^r$	?
$h_2 = (0,1)$	$\log z_{h_2}^r$	$d_{h_2}^z + d^A$
$h_3 = (1,0)$	$\log z_{h_3}^r$	$d_{h_3}^z + d^A$
$h_a = (1,1)$	$\kappa = \mu_a + \Delta_a = \log z_a^u + d^A$	
	$T = 3$ **periods**	
$h_n = (0,0,0)$	$\log z_n^r$	?
$h_2 = (0,0,1)$	$\log z_{h_2}^r$	$d_{h_2}^z + d^A$
$h_3 = (0,1,0)$	$\log z_{h_3}^r$	$d_{h_3}^z + d^A$
$h_4 = (0,1,1)$	$\log z_{h_4}^r$	$d_{h_4}^z + d^A$
$h_5 = (1,0,0)$	$\log z_{h_5}^r$	$d_{h_5}^z + d^A$
$h_6 = (1,0,1)$	$\log z_{h_6}^r$	$d_{h_6}^z + d^A$
$h_7 = (1,1,0)$	$\log z_{h_7}^r$	$d_{h_7}^z + d^A$
$h_a = (1,1,1)$	$\kappa = \mu_a + \Delta_a = \log z_a^u + d^A$	

If we are willing to impose the linear restrictions from Lemieux (1998) and Suri (2011), we can additionally identify the returns to migration for never-movers. By identifying (under some assumptions) the returns to urban migration for those who are never observed leaving the rural markets and vice-versa, these estimates answer relevant policy questions. Further our context, the non-switchers are the largest group in our data and are therefore relevant by sheer strength of numbers.

4.3.2 The restricted GRC model

The restricted version of the GRC imposes the parametric assumptions in Suri (2011) within a GMM framework. First, we assume a linear structure for the log consumption that allows for unobserved comparative and absolute advantage while also mapping well onto the terms of our migration choice model:

$$c_{it} = \log z_i^r + \left(d_i^z + d^A\right) h_{it} + \varepsilon_{it} = \mu_i + \Delta_i h_{it} + \varepsilon_{it} = \tau_i + \theta_i + \left(\phi\theta_i + \beta\right) h_{it} + \varepsilon_{it}$$

Thus

$$\mu_i = \tau_i + \theta_i \quad \text{and} \quad \Delta_i = \phi\theta_i + \beta$$

and

$$\log z_i^r = \tau_i + \theta_i, \quad \log z_i^u = \tau_i + (1 + \theta_i), \text{ and } \quad \beta = d^A = \log A^u$$

Notice that this structure implies a linear relationship between the abilities of groups of individuals—as defined by migration trajectories—in the rural market and their relative abilities the urban market. This relationship allows us to identify the returns to non-switchers from relationships between the returns of switchers.[10] While these may seem like strong assumptions, there is a trade-off: for example, unlike control function approaches, this method does not require assumptions on the parametric form of the distribution of log unobserved skills.

Re-phrasing the assumptions in Suri (2011) within this framework, we also assume the following:

$$\begin{aligned} &E[\theta_i] = 0 \\ &E[\tau_i|h_i] = E[\tau_i] \\ &E[\varepsilon_{it}|h_i, \tau_i, \theta_i] = 0 \end{aligned} \tag{6}$$

With these assumptions, and defining $\theta_h = E[\theta_i|h_i = h]$, we get the following results for any $h, h' \in \mathcal{H} = \{0,1\}^T$ (Proposition 1 in Ghanem *et al.* (2020)):

[10] Refer to Verdier (2019) for an illustration of this linear extrapolation process involved in a related model.

(i) $\Delta_h = \beta + \phi\theta_h$,

(ii) $\mu_h - \mu_{h'} = \theta_h - \theta_{h'}$,

(iii) $\Delta_h - \Delta_{h'} = \phi\left(\theta_h - \theta_{h'}\right)$, for $h \neq h'$.

Both the restricted and the unrestricted GRC model can be estimated using efficient generalized method of moments (GMM).

4.3.3 The GRC approach applied to the extended models

Define an "away from home" indicator $\lambda_{it} \equiv 1\left\{l_{it} \neq l_{i1}\right\}$, and a "just moved" indicator $\eta_{it} \equiv 1\left\{l_{i,t} \neq l_{i,t-1}\right\}$. then we can re-write log consumption as

$$c_{it} = \log z_i^r + (d_i^z + d^A)h_{it} + \varphi\left(1 + \left(\alpha_{it} - 1\right)\lambda_{it}\right) + (m)\eta_{it} + \varepsilon_{it}$$

We can also re-write the log consumption of individual i in period t in the general format we used before, explicitly accounting for the two new indicators:

$$c_{it} = f\left(h_{it}, \lambda_{it}, \eta_{it}, a_i\right) + \varepsilon_{it}$$

Notice that the new indicators λ_{it} and η_{it} are fully determined by an individual's characteristic a_i (their initial market, or home) and the individual's migration history h_i, which tells how many times the individual moved and when. Therefore, to estimate the model exposed in Sections 2.2 and 2.3 we do not need a new identifying assumption: the strict exogeneity assumed before ($E\left[\varepsilon_{it}|h_i, a_i\right] = 0$) will still work.

5 Preliminary results

The regressions for $T = 2$ use data from the last two periods available in our dataset (waves 4 and 5), while the regressions for $T = 3$ use data from the last three periods (waves 3, 4, and 5)

Table 2: Estimates for the return to urban migration using OLS, Panel FE, and GRC ($T = 2$)

Dependent variable: **log of real consumption**

Periods		**T = 2**		
Models	Pooled OLS	Panel (ind. FE)	unrestricted GRC	restricted GRC
urban dummy	0.1827	0.1826		
	(0.0145)	(0.0317)		
returns to migration (Δ_h)				
type $h_n = (0, 0)$: never urban				0.1781
				(0.0432)
type $h_2 = (0, 1)$: adopter			0.3360	0.3360
			(0.0320)	(0.0320)
type $h_3 = (1, 0)$: disadopter			-0.5230	-0.5230
			(0.0653)	(0.0653)
type $h_a = (1, 1)$: always urban				0.1951
				(0.0653)
baseline returns (μ_h)				
type $h_n = (0, 0)$: never urban			12.2180	12.2180
			(0.0132)	(0.0132)
type $h_2 = (0, 1)$: adopter			12.1224	12.1224
			(0.0302)	(0.0302)
type $h_3 = (1, 0)$: disadopter			12.6429	12.6429
			(0.0558)	(0.0557)
type $h_a = (1, 1)$: always urban				12.2077
				(0.0607)
N (obs.)	10,495	10,495	10,495	10,495
Controls	No	No	No	No
Time FE	No	No	No	No

Standard errors in parenthesis. Mean of dependent variable: 12.32 (s.d. 0.74)

Table 3: Estimates for the return to urban migration using OLS, Panel FE, and GRC ($T = 3$)

Dependent variable: **log of real consumption**

Periods			**T = 3**	
Models	Pooled OLS	Panel (ind. FE)	unrestricted GRC	restricted GRC
urban dummy	0.2370	0.2309		
	(0.0119	(0.0204)		
returns to migration (Δ_h)				
type $h_n = (0, 0, 0)$: never urban				0.1740
				(0.0325)
type $h_2 = (0, 0, 1)$: late adopter			0.4733	0.4543
			(0.0326)	(0.0317)
type $h_3 = (0, 1, 0)$: mixed disadopter			-0.1902	-0.2040
			(0.0904)	(0.0854)
type $h_4 = (0, 1, 1)$: early adopter			0.4117	0.4083
			(0.0284)	(0.0283)
type $h_5 = (1, 0, 0)$: early disadopter			-0.4816	-0.4227
			(0.0552)	(0.0551)
type $h_6 = (1, 0, 1)$: mixed adopter			0.0288	0.0333
			(0.0632)	(0.0633)
type $h_7 = (1, 1, 0)$: late disadopter			-0.5594	-0.5703
			(0.0755)	(0.0750)
type $h_a = (1, 1, 1)$: always urban				0.3418
				(0.0508)
baseline returns (μ_h)				
type $h_n = (0, 0, 0)$: never urban			12.0661	12.0661
			(0.0125)	(0.0124)
type $h_2 = (0, 0, 1)$: late adopter			11.9333	11.8978
			(0.0275)	(0.0231)
type $h_3 = (0, 1, 0)$: mixed disadopter			12.3124	12.2931
			(0.0678)	(0.0533)
type $h_4 = (0, 1, 1)$: early adopter			11.8719	11.9255
			(0.0304)	(0.0246)
type $h_5 = (1, 0, 0)$: early disadopter			12.2964	12.4244
			(0.0493)	(0.0416)
type $h_6 = (1, 0, 1)$: mixed adopter			12.2687	12.1506
			(0.0633)	(0.0416)
type $h_7 = (1, 1, 0)$: late disadopter			12.6609	12.5130
			(0.0713)	(0.0568)
type $h_a = (1, 1, 1)$: always urban				11.9654
				(0.0442)
N (obs.)	15,785	15,785	15,785	15,785
Controls	No	No	No	No
Time FE	No	No	No	No

Standard errors in parenthesis. Mean of dependent variable: 12.18 (s.d. 0.76)

References

Alvarez, J. A. (2020). The agricultural wage gap: Evidence from Brazilian micro-data. *American Economic Journal: Macroeconomics*, **12** (1), 153–73.

Bazzi, S., Gaduh, A., Rothenberg, A. D. and Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in indonesia. *American Economic Review*, **106** (9), 2658–98.

Bryan, G., Chowdhury, S. and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, **82** (5), 1671–1748.

Chernozhukov, V., Demirer, M., Duflo, E. and FernÃąndez-Val, I. (2019). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. *arXiv:1712.04802 [econ, math, stat]*, arXiv: 1712.04802.

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, **41**, 47–60.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. and Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, **109** (12), 4178–4219.

Ghanem, D., Tjernström, E., Cabanillas, O. B., Lybbert, T., Michler, J. and Michuda, A. (2020). A group random coefficient approach to modeling heterogeneity in technology adoption. *Working paper*.

Gibbons, C. E., Serrato, J. C. S. and Urbancic, M. B. (2018). Broken or fixed effects? *Journal of Econometric Methods*, **8** (1).

Gollin, D., Lagakos, D. and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, **104** (5), 165–70.

Heckman, J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *The Journal of Human Resources*, **33** (4), 974–987.

HECKMAN, J. J. and VYTLACIL, E. J. (2007). Chapter 70 econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. Handbook of Econometrics, Elsevier.

HICKS, J. H., KLEEMANS, M., LI, N. Y. and MIGUEL, E. (2018). *Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata*. Working Paper 23253, National Bureau of Economic Research.

KENNAN, J. and WALKER, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, **79** (1), 211–251.

KLEEMANS, M. (2015). *Migration choice under risk and liquidity constraints*. Working paper.

— and MAGRUDER, J. (2018). Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*, **128** (613), 2032–2065.

LAGAKOS, D., MARSHALL, S., MOBARAK, A. M., VERNOT, C. and WAUGH, M. E. (2019). Migration policy and observational returns to rural-urban migration in the developing world. *Working paper*.

—, MOBARAK, A. M. and WAUGH, M. E. (2020). *The welfare effects of encouraging rural-urban migration*. Working paper.

— and WAUGH, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, **103** (2), 948–80.

LEMIEUX, T. (1998). Estimating the effects of unions on wage inequality in a panel data model with comparative advantage and nonrandom selection. *Journal of Labor Economics*, **16** (2), 261–291.

SURI, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, **79** (1), 159–209.

VERDIER, V. (2019). Average treatment effects for stayers with correlated random coefficient models of panel data. *Working paper*.

YOUNG, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, **128** (4), 1727–1785, publisher: MIT Press.

ZHOU, X. and XIE, Y. (2019). Marginal treatment effects from a propensity score perspective. *Journal of Political Economy*, **127** (6), 3070–3084.