Risk Management and Aggregate Migration Flows*

Matthew Schwartzman†

Yale University

Abstract: I study whether the need to manage uninsured risk influences the destination choice of migrants in low-income settings. I develop a parsimonious model of migration as a risk management strategy. Risk-averse households have an incentive to migrate between origin-destination pairs whose economic shocks have low covariance, because these location pairs are better hedges for one another. Under common preference specifications, the model generates a risk-augmented gravity equation that I estimate using migration data from the Philippines. Empirical results confirm that migration is substantially higher between origin-destination pairs with low covariance of shocks.

*I am grateful to Nate Barker, Kevin Donovan, Rosie Li, Mushfiq Mobarak, Michael Peters, Mark Rosenzweig, Nick Ryan, Sam Solomon, and participants at Yale's development breakfast and development lunch for their helpful comments and suggestions.

†Email: matthew.schwartzman@yale.edu

1 Introduction

Economic development is tightly linked with migration. Development does not just change the technology or the sectoral composition of an economy; it also changes the spatial distribution of its population. As some locations grow and others fall in relative productivity, simply moving workers to the locations where they are most productive can raise living standards. When workers fail to respond to price signals and do not migrate to more productive locations, the result is misallocation across space, which typically implies misallocation across firms and sectors.

Misallocation has well-documented potential aggregate consequences (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008), and individuals can reap large benefits from migrating somewhere productive (Bryan et al., 2014). It is therefore a puzzle that workers do not always choose locations that offer lucrative compensation. Several credible explanations exist for this puzzle, including barriers to migration and the sorting of workers into locations based on their skills (Young, 2013; Bryan and Morten, 2019).

My focus in this paper is on a different distortion to workers' migration decisions: their desire to manage risk and smooth consumption. This desire influences migrants to choose destinations whose economic shocks have low covariance with the migrants' origin. Alongside well-known migration barriers like distance, migrants' risk-aversion-induced preference for low covariance origin-destination pairs may divert them away from the output-maximizing choice of location.

Households in developing countries frequently turn to non-market solutions in response to uninsured risk (Besley, 1995). One risk management strategy is for families to diversify risk across households whose income is subject to heterogeneous shocks. It is intuitive that these schemes are more effective the less households' income shocks are related to one another. To understand their effect on location decisions, I develop a simple model of migration as an *ex ante* risk management strategy: households from a common origin share location-specific aggregate risk between the origin and a destination. The migrant chooses a destination in anticipation of shocks and takes the variance of consumption into account.

I then integrate the model into a gravity equation framework and show that it has a clear implication: higher covariance of shocks between an origin and destination reduces migration between the two. Origin-destination pairs with low covariance are attractive because cross-location risk sharing means they reduce the variance of consumption. The implied gravity equation is simple to estimate and provides an empirically testable prediction of the model. I confirm this implication of the model with data on inter-provincial migration in the Philippines. Using rainfall data to measure economic shocks and long-form census data to measure migration, I find that origin-destination pairs with high covariance of shocks do tend to have lower levels of migration, all else equal.

Not only does the negative effect of origin-destination covariance exist, its estimated magnitude is significant. As expected, rainfall shocks matter more between pairs of provinces that are highly reliant on agriculture. My estimates imply that two high agriculture locations with typical covariance of rainfall shocks have 34% lower migration than a counterfactual pair with uncorrelated shocks.

From these empirical results I conclude that workers' risk management concerns shape aggregate migration patterns. Information only on the first moment of each location's productivity, such as which locations are most productive on average, is therefore not sufficient to explain and predict spatial patterns of development. The second moment of the joint distribution of location-level productivity plays a critical role precisely because of workers' risk aversion. The importance of cross-location covariance for migration illustrates that information on this second moment is necessary to understand the spatial allocation of labor.

This paper contributes to a few related strands of the economic literature. One is the large body of work on migration in developing countries that has followed the seminal paper by Harris and Todaro (1970) on rural to urban migration. An equally important literature in development economics concerns the risk coping strategies of the poor (Rosenzweig and Wolpin, 1993; Townsend, 1994, 1995; Kinnan and Townsend, 2012). Several recent studies, such as Meghir et al. (2019), Rosenzweig and Udry (2014), Munshi and Rosenzweig (2016), and Morten (2019), straddle these two topics and examine the relationship between risk management and migration. Such papers study both the role of informal insurance networks and the partial insurance offered by migration itself. In my model, I examine the interplay between informal insurance networks and migration.

Research on risk and migration tends to focus on the binary decision of whether or not to migrate. A notable exception is Rosenzweig and Stark (1989)'s study of marital migration in rural India, which shows that brides' destinations can be rationalized as part of a scheme that diversifies risk over distant villages. Similarly, the work I present here does not focus on the decision of whether to migrate; instead I ask if a risk management motive shapes the choice of destination. Data on every pair of provinces in the Philippines allow me to investigate this question at the level of the entire economy.

Studying migration between every origin-destination pair in the economy invites the use of techniques from the large spatial economics literature. Specifically, I model taste shocks over locations as following a Fréchet distribution in order to express migrant flows with a gravity equation. Redding and Rossi-Hansberg

(2017) provide a review of related spatial models. A recent example studying migration in a developing country is the work of Bryan and Morten (2019), who instead model skills as following a Fréchet distribution and use gravity equations to motivate a model of spatial sorting in Indonesia. To my knowledge, this paper is the first to integrate a risk management scheme into a gravity equation and use it to test the risk diversification motive of migration.

Finally, this paper contributes to the literature on risk-coping strategies and migration in the Philippines specifically. Fafchamps and Lund (2003) find evidence for kinship-based risk-sharing networks among rural Filipino households. A follow-up study by Fafchamps and Gubert (2007) finds little evidence of purposeful risk diversification in risk-sharing networks, but this result applies to within-village risk-sharing rather than risk-sharing between distant households.

On the reasons behind migration, Quisumbing and McNiven (2005) find in a survey of rural out-migrants that a large percentage of moves are motivated by either job search or marriage. Quisumbing and McNiven (2010) find that internal remittances in the Philippines influence the expenditures of households at the origin, and Quisumbing et al. (2012) find that migrant networks are important for risk management. 80% of households in their sample have at least one migrant child, 61% receive remittances, and the probability of receiving remittances is greater for households that suffered more economic shocks. These results provide grounding for the models I develop.

The remainder of the paper is as follows. In Section 2 I propose a model of migration as an *ex ante* risk management strategy. I derive the model's main implication and show how to test it empirically by estimating a gravity equation. Section 3 describes the application of the model to the Philippine context. I give an overview of the data and methods used to test the models' predictions.

Section 4 presents the main results, which confirm the central implication of the model. Section 5 concludes.

2 Migration as *ex ante* risk management

This section provides a simple spatial model where families share risk across locations, which each experience shocks to agricultural productivity. I show that families have an incentive to diversify risk across locations with less correlated shocks even within this simple model.

2.1 Local economies and within-location allocations

The economy features a set J of locations indexed by j. Each location has an agricultural sector, denoted by a, and a non-agricultural sector, denoted by n. For a given location j in period t, let A_{jt}^{a} and A_{j}^{n} be exogenous sectoral productivities, N_{jt}^{a} and N_{jt}^{n} be labor in each sector, and $N_{jt} = N_{jt}^{a} + N_{jt}^{n}$ be j's aggregate labor supply. Then output in each sector is as follows:

$$Y_{jt}^{a} = A_{jt}^{a}\left(N_{jt}^{a}\right)^{\alpha} N_{jt}^{1-\alpha} \tag{1}$$

$$Y_{jt}^{n} = A_{jt}^{n}\left(N_{jt}^{n}\right)^{\alpha} N_{jt}^{1-\alpha} \tag{2}$$

Here $\alpha \in (0,1)$ paramaterizes the decreasing marginal product of sector-specific labor. However, holding the share of labor in each sector fixed, production exhibits constant returns to scale with respect to aggregate labor supply. The agricultural and non-agricultural goods are perfect substitutes[1], so that aggregate output $Y_{jt} = Y_{jt}^{a} + Y_{jt}^{n}$.

[1] An equivalent assumption is that both goods are perfectly tradable.

The labor market is frictionless, so that the marginal product of labor is equal across sectors. Let ϑ^a_{jt} denote the share of labor in agriculture. Then ϑ^a_{jt}, wages w_{jt}, and per capita output y_{jt} can all be written as functions of A^a_{jt} and A^n_j:

$$A^a_{jt}\left(\vartheta^a_{jt}\right)^{\alpha-1} = A^n_j\left(1-\vartheta^a_{jt}\right)^{\alpha-1} \tag{3}$$

$$\vartheta^a_{jt} = \frac{\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}}{\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}+\left(A^n_j\right)^{\frac{1}{1-\alpha}}} \quad w_{jt} = \alpha\left[\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}+\left(A^n_j\right)^{\frac{1}{1-\alpha}}\right]^{1-\alpha} \quad y_{jt} = \left[\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}+\left(A^n_j\right)^{\frac{1}{1-\alpha}}\right]^{1-\alpha} \tag{4}$$

2.2 Location-specific shocks

Each location experiences shocks to agricultural productivity – in the empirical section of the paper, these shocks map to rainfall. Shocks are i.i.d. over time, but the aggregate agricultural shock in each period is correlated across locations with covariance matrix Σ^a. Within each location, the process for agricultural productivity follows:

$$A^a_{jt} = \overline{A}^a_j + z^a_{jt} \tag{5}$$

Here $\overline{A}^a_j$ is the long run average agricultural productivity in j and z^a_{jt} is the shock, with $\mathbb{E}[z^a_{jt}] = 0$. Taking the derivative of per capita output y_{jt} with respect to z^a_{jt}, one can write a first order approximation of y_{jt} in each period:

$$\frac{\partial y_{jt}}{\partial z^a_{jt}} = \left[\frac{\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}}{\left(A^a_{jt}\right)^{\frac{1}{1-\alpha}}+\left(A^n_j\right)^{\frac{1}{1-\alpha}}}\right]^{\alpha} = \left(\vartheta^a_{jt}\right)^{\alpha} \tag{6}$$

$$\implies \quad y_{jt} \approx \overline{y}_j + \left(\overline{\vartheta}^a_j\right)^{\alpha} z^a_{jt}, \tag{7}$$

where $\overline{y}_j$ and $\overline{\vartheta}^a_j$ are per capita output and agricultural share when $A^a_{jt} = \overline{A}^a_j$. The shock to per capita output in each period can be approximated as $z^y_{jt} =$

$y_{jt} - \overline{y}_j \approx \left(\overline{\vartheta}_j^a\right)^{\alpha} z_{jt}^a$. The output shock is therefore approximately proportional to the agricultural productivity shock z_{jt}^a, and the impact of z_{jt}^a increases in the long run agricultural share $\overline{\vartheta}_j^a$. This structure of shocks generates a covariance matrix of output shocks Σ^y, with entries σ_{jk}^y representing the covariance of per capita output between two locations j and k. σ_{jk}^y relates to σ_{jk}^a, the covariance of agricultural shocks, through agricultural employment shares in each location:

$$\sigma_{jk}^y \approx \left(\overline{\vartheta}_j \overline{\vartheta}_k\right)^{\alpha} \sigma_{jk}^a \tag{8}$$

2.3 Risk sharing across locations

In the model of migration as an *ex ante* risk management strategy, households diversify risk across locations in order to reduce the variance of consumption. These households have indirect utility over the mean and standard deviation of consumption $V(\sigma,\mu)$. I assume that $V_\mu > 0$ and $V_\sigma \leq 0$. If $V_\sigma < 0$ then households are risk averse. There is no savings technology, so consumption is equal to earnings plus transfers.

A family originates in origin $o \in J$. Some part of the family stays in o and another part either chooses a migration destination $d \in J$ or does not migrate at all. Those in the origin earn income averaging $\overline{y}_o$ and are subject to aggregate shocks z_{ot}^y that have mean zero and variance $(\sigma_o^y)^2$. Similarly, migrants to the destination earn $\overline{y}_d$ on average and experience aggregate shocks z_{dt}^y with mean zero and variance $(\sigma_d^y)^2$. Earnings shocks z_{ot}^y and z_{dt}^y have covariance σ_{do}^y. Because the focus of this paper is on aggregate shocks correlated across locations, I omit idiosyncratic risk from the model as a simplification.

When part of the family creates a new household and migrates to d, the two households (stayers, indexed oo, and migrants, indexed do) agree to a risk-sharing scheme. Under this scheme, each household keeps the mean of earnings in each location, but the two households use transfers to divide up earnings shocks z_{ot} and z_{dt} according to a cross-location risk-sharing parameter $\phi \in [0,1]$. Thus total income for each household for a given realization of shocks is as follows:

$$y_{oot} = \overline{y}_o + (1-\phi) z^y_{ot} + \phi z^y_{dt} \tag{9}$$

$$y_{dot} = \overline{y}_d + \phi z^y_{ot} + (1-\phi) z^y_{dt} \tag{10}$$

The case where $\phi = 0$ represents the absence of risk-sharing and the complete separation of the two households' incomes. When $\phi = \frac{1}{2}$, the households evenly divide up risk. For any given ϕ, the means of the two households' incomes are still $\overline{y}_o$ and $\overline{y}_d$, but the variances are as follows:

$$\text{Var}(y_{oo}) = (1-\phi)^2 (\sigma^y_o)^2 + \phi^2 (\sigma^y_d)^2 + 2\phi(1-\phi)\sigma^y_{do} \tag{11}$$

$$\text{Var}(y_{do}) = \phi^2 (\sigma^y_o)^2 + (1-\phi)^2 (\sigma^y_d)^2 + 2\phi(1-\phi)\sigma^y_{do} \tag{12}$$

As a simplification, I assume that ϕ is exogenous and identical across all families (for example, it is given by common norms). It is also possible to endogenize ϕ through a bargaining process. Under CARA preferences, assumed below, families with equal bargaining weights will always choose $\phi = \frac{1}{2}$, although there may be an overall transfer from one household to the other.

The family decides upon a migration destination taking the risk-sharing scheme outlined above as given. If households are risk averse, then given two destinations d and d' with equal means $\overline{y}_d$ and $\overline{y}_{d'}$, the migrant will tend to choose d over d' if $\text{Var}(y_{do}) < \text{Var}(y_{d'o})$. Equation (12) above makes it clear that d will be more desirable than d' if earnings there exhibit lower variance $(\sigma^y_d)^2$ and lower covariance σ^y_{do} with earnings in o. Intuitively d is more attractive because it provides a better hedge against earnings in o. Note that variance $(\sigma^y_d)^2$ influences how attractive destination d is to migrants from all origins, while covariance σ_{do} influences its attractiveness only to migrants from origin o.

2.4 Risk sharing meets gravity

I now turn to the migrant's choice of destination from the full set J of locations. For concision I will write the variance of consumption for a migrant from o to d as Var$(y_{do})=\nu^2_{do}$. The full indirect utility for a migrant i to move from o to d is as follows:

$$u^i_{do}=V(\nu_{do},\overline{y}_d)+a_d-c_{do}+\log(\xi^i_d) \tag{13}$$

Here $V(\nu_{do},\overline{y}_d)$ is as defined above, a_d is the exogenous amenity value of living in d, c_{do} represents migration costs of moving from o to d, and ξ^i_d is an i.i.d. individual-specific taste shock for location d. The migrant i draws a vector $\vec{\xi}^i$ of independent taste shocks for all locations in J from a Fréchet distribution with shape parameter θ. The migrant's choice of the destination that maximizes u^i_{do} is equivalent to choosing a destination to maximize $U^i_{do}=e^{u^i_{do}}$:

$$U^i_{do}=\exp(V(\nu_{do},\overline{y}_d)+a_d-c_{do})\xi^i_d=\tilde{U}_{do}\xi^i_d \tag{14}$$

The existing literature using gravity equations to model migration typically defines U^i_{do} with wage w_d (or real wage $\frac{w_d}{P_d}$) in place of $\exp(V(\nu_{do},\overline{y}_d))$. If w_d is taken to represent the time average of earnings in d then this commonly used specification corresponds to $V(\nu_{do},\overline{y}_d)=\log(\overline{y}_d)$ in my model, implying that the variance of consumption does not enter households' decisions.

Since ξ_d^i follows a Fréchet distribution, migration from d to o can be expressed by the following gravity equation:

$$m_{do} = \frac{\tilde{U}_{do}^{\theta}}{\sum_{k \in J} \tilde{U}_{ko}^{\theta}} L_o \tag{15}$$

Here m_{do} is the migrant flow from o to d and L_o is the total population originating from o. It is possible to re-write Equation (15) with m_{do} expressed as an exponent.

$$m_{do} = \exp\left(\log(L_o) - \log\left(\sum_{k \in J} \tilde{U}_{ko}^{\theta}\right) + \theta a_d + \theta V(\nu_{do}, \overline{y}_d) - \theta c_{do}\right) \tag{16}$$

The first two terms inside the exponent, $\log(L_o) - \log\left(\sum_{k \in J} \tilde{U}_{ko}^{\theta}\right)$, can be thought of as an origin fixed effect γ_o: for a given origin o, these terms are identical for all destinations d. The next three terms in the exponent, $\theta a_d + \theta V(\nu_{do}, \mu_d) - \theta c_{do}$, are equal to $\theta \log\left(\tilde{U}_{do}\right)$.

2.5 Empirical implications of the *ex ante* model

To understand how risk management affects migration, it is necessary to observe c_{do} separately from $V(\nu_{do}, \overline{y}_d)$. To do so, I assume that $c_{do} = \frac{c(\text{dist}_{do}) - \varepsilon_{do}}{\theta}$. Here dist_{do} represents the distance from origin o to destination d and ε_{do} is an idiosyncratic cost advantage of migration from o to d. Dividing by θ is a normalization. Next, to isolate the effect of cross-location covariance, I assume the following functional form of $V(\nu_{do}, \overline{y}_d)$:

$$V(\nu_{do}, \overline{y}_d) = \overline{y}_d - \frac{\lambda}{2} \nu_{do}^2 \tag{17}$$

This indirect utility function can be derived from constant absolute risk aversion (CARA) preferences over consumption, where $\lambda \geq 0$ is the risk aversion parameter. While the assumption of a specific functional form for $V(\cdot,\cdot)$ may seem strong, similar predictions can be obtained from weaker functional form assumptions such as additive separability. After plugging in this form of $V(\nu_{do},\overline{y}_d)$ to the expression for m_{do}, and decomposing ν_{do}^2 into its separate components as in (12), it is possible to see the effect of cross-location covariance σ_{do}^y.

$$m_{do} = \exp\left(\gamma_o - \theta\frac{\lambda}{2}\phi^2(\sigma_o^y)^2 + \theta a_d + \theta\overline{y}_d - \theta\frac{\lambda}{2}(1-\phi)^2(\sigma_d^y)^2 - \theta\lambda\phi(1-\phi)\sigma_{do}^y - c(\text{dist}_{do}) + \varepsilon_{do}\right) \tag{18}$$

The expression for m_{do} simplifies further once one collects $\gamma_o - \theta\frac{\lambda}{2}\phi^2(\sigma_o^y)^2$ into an origin fixed effect δ_o and $\theta a_d + \theta\overline{y}_d - \theta\frac{\lambda}{2}(1-\phi)^2(\sigma_d^y)^2$ into a destination fixed effect δ_d.

$$m_{do} = \exp(\delta_o + \delta_d - \theta\lambda\phi(1-\phi)\sigma_{do}^y - c(\text{dist}_{do}) + \varepsilon_{do}) \tag{19}$$

Equations (18) and (19) make the model's central prediction clear: when agents are risk averse, so that $\lambda > 0$, and engage in cross-location risk sharing, so that $\phi \in (0,1)$, migrant flows decrease in origin-destination output shock covariance σ_{do}^y. The intuition behind the model's prediction is that locations with less related shocks are better hedges for one another. It is possible to test this prediction empirically with data on migrant flows, the distance between origins and destinations, and a proxy for the origin-destination covariance σ_{do}^y. The empirical application of this paper uses rainfall shocks as the shock z_{jt}^a to agricultural productivity. Recalling that $\sigma_{do}^y = \left(\overline{\vartheta}_d\overline{\vartheta}_o\right)^\alpha \sigma_{do}^a$, it is clear that the covariance of rainfall shocks σ_{do}^a matters more between provinces with high agricultural shares of employment ϑ^a.

3 Application to the Philippines

3.1 Data and summary statistics

I use data from the Philippines to test whether origin-destination covariance of economic shocks reduces migrant flows, as predicted by the model of migration as an *ex ante* risk management strategy. Data on migration come from the long form of the 2000 and the 2010 Philippine Population and Housing Census, made available by IPUMS International (Minnesota Population Center, 2019). For each person in the sample, the census includes information on the province where the individual currently lives and the province where they lived 5 years ago. While the Philippines technically sub-divided some existing provinces into new ones in the 1990s, IPUMS provides a classification of 76 provinces that are consistent for the period from 1990 to 2010. I use this classification throughout my analysis. In 2000, these provinces had an average population of 1.03 million and an average land area of 3,947 km^2.

For each origin-destination pair, I construct the 5-year migrant flow m_{do} by adding together all the individuals currently in destination d who lived in origin o 5 years ago. When summing up migrants, I use the person weights provided by IPUMS for representativeness. The Philippine census also includes information on the industry of each worker in the sample, which I use to construct agricultural employment shares.

I use rainfall shocks as a proxy for economic shocks. The rainfall data comes from NASA's Global Precipitation Measurement (GPM) database. This data reports total precipitation for each $.1° \times .1°$ grid cell around the globe for the period from 2000 to 2020, measured over either hourly, daily, or monthly intervals.

To construct shocks, I proceed as follows. First, I use the provincial shapefiles described below to spatially aggregate the gridded monthly data over each province, which produces the spatially-weighted average of total rainfall for each province in each month of each year. I next temporally aggregate the province-month-year data over three successive months to produce rainfall data at the province-quarter-year level, which roughly captures seasons. Finally, I standardize the raw province-quarter-year rainfall by subtracting the province-quarter mean over 2000-2020 and dividing by the province-quarter standard deviation. In mathematical terms:

$$\text{Rainfall shock}_{pqy} = \frac{\text{Rainfall}_{pqy} - \overline{\text{Rainfall}}_{pq}}{SD_{pq}(\text{Rainfall})} \tag{20}$$

In other words, the rainfall shock measures how many standard deviations away from the province-quarter mean a given realization of rainfall is. It therefore captures a good season or bad season relative to expected rainfall in that season.

Once I have constructed the shocks, I compute their covariance across provinces. For example, for a given pair of provinces d and o, the covariance of rainfall shocks is:

$$\text{Cov}_{do}(\text{Rainfall shock}) = \frac{1}{83} \sum_{y=2000}^{2020} \sum_{q=1}^{4} \text{Rainfall shock}_{dqy} \times \text{Rainfall shock}_{oqy} \tag{21}$$

Repeating this procedure for every origin-destination pair generates a full covariance matrix of rainfall shocks that can be used to proxy for the covariance of agricultural productivity shocks (σ^a_{do} in the model).

For geography, I use the Philippine province shapefile from IPUMS whose boundaries match the time-consistent classification of provinces (Minnesota Population Center, 2019). I measure the distance between provinces as the straight line distance between each province's geographic centroid. I use these same shapefiles for the spatial aggregation of rainfall data described above.

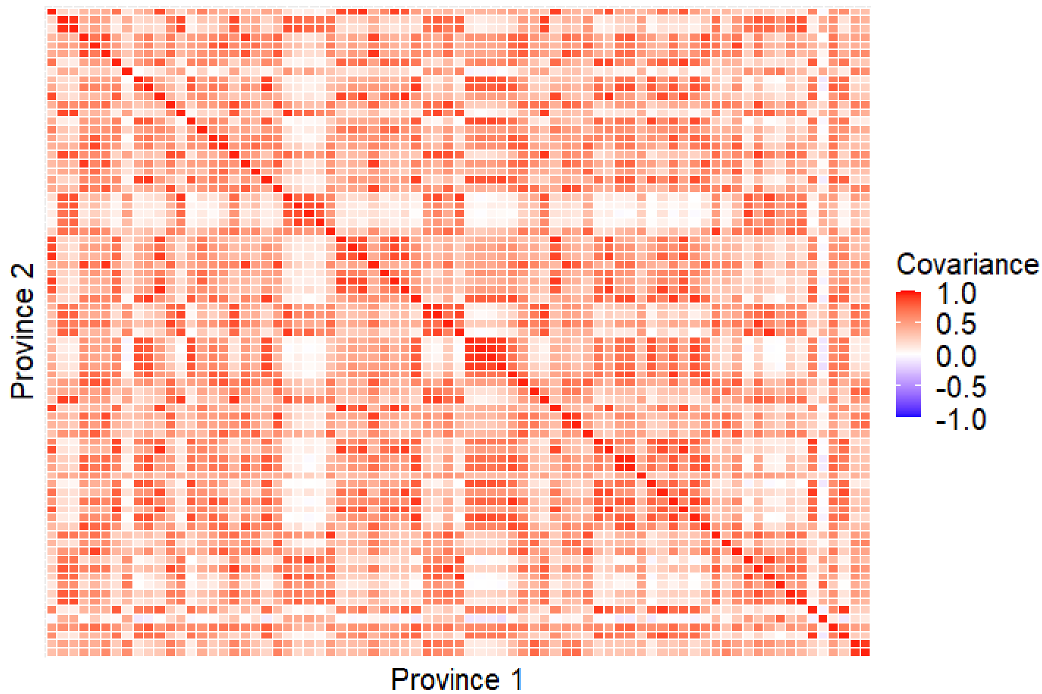
Since the Philippines are an archipelago, I supplement the straight line distance with driving time between each pair of provincial capitals as estimated by the Google Maps Distance Matrix API. The advantage of using driving time is that it accounts for physical barriers like water. For example, to drive between provinces on different islands requires a ferry trip, and since ferry travel is generally slower than road travel, the driving time penalizes origin-destination pairs that are not on the same island. A downside of driving time is that it is not available between all pairs of provinces. In particular, three provinces are inaccessible to the other provinces by car, so all origin-destination pairs involving these provinces are excluded from specifications that use driving time.

To summarize, for each origin-destination pair, I have migrant flows m_{do} and agricultural shares $\vartheta_d^a, \vartheta_o^a$ from the census, rainfall shock covariance σ_{do}^a from NASA's GPM database, distance $dist_{do}$ from geographic shapefiles, and driving time from Google Maps Distance Matrix API. Thus I have the required data to test the empirical implications outlined in Section 2.

I use a tile plot to visualize the inter-provincial rainfall shock covariance matrix in Figure 1. Same-province pairs appear along the diagonal. Since shocks are already standardized, the covariance is equivalent to the correlation and can only vary between -1 and 1. Essentially all origin-destination pairs have shocks with positive covariance, but the magnitude of this covariance varies widely.

I provide summary statistics for each origin-destination variable in Table 1. Since I am interested in modeling the choice of destination conditional on migration, I

Figure 1: Inter-provincial rainfall shock covariance matrix



Note: Each tile's color visualizes the covariance between the corresponding provinces according to the legend on the right hand side.

exclude the 76 origin-destination pairs for which the same province is the origin and destination. This leaves 5,700 pairs of provinces. The summary statistics show that migrant flows are right-skewed, so it is appropriate to model them as an exponential function of other variables. They also reveal that there are zeroes in the matrix of migrant flows. The model does not technically allow for zeroes, but in reality they could arise from the indivisibility of humans and from the long form census's coverage of only a sub-sample of the population. The methods I use, described in more detail in the next subsection, are designed in part to accommodate these zeroes.

Table 1: Summary statistics for origin-destination pairs

Statistic	N	Mean	St. Dev.	Min	Median	Max
Migrants (5 year)	11,400	332.23	1,592.72	0.00	38.00	63,226.00
Male migrants (5 year)	11,400	157.63	773.09	0.00	19.00	31,413.00
Female migrants (5 year)	11,400	174.47	821.74	0.00	19.00	31,813.00
Cov_{do}(rainfall shock)	5,700	0.42	0.23	-0.14	0.41	0.95
Ag. employment share	76	0.14	0.07	0.00	0.15	0.31
Distance (km)	5,700	539.84	324.84	9.27	496.86	1,440.72
Drive time (hours)	5,256	19.77	11.82	0.44	18.48	65.52

3.2 Methods

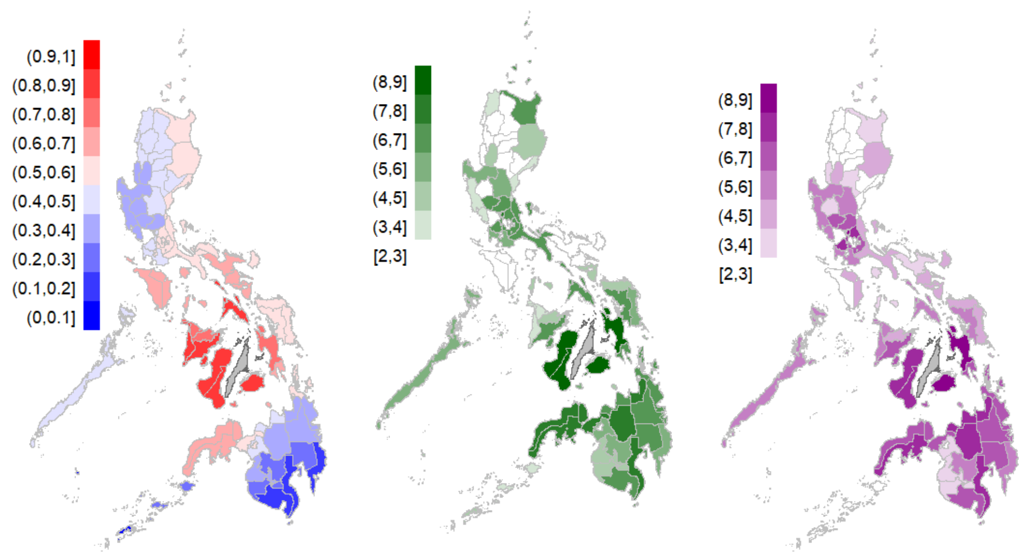
Figure 2 illustrates the main variation I use in the empirical section of the paper. I map the relationship between Cebu, the most populous Philippine province outside the Manila area, and each other province. Cebu is the province that appears shaded in grey with heavy boundaries in the lower center of each map.

Panel 2a displays the covariance of rainfall shocks in each province with shocks in Cebu. The typical province's shocks have a moderate positive covariance with those of Cebu, but the strength of this relationship varies substantially throughout the Philippines. This covariance tends to decrease in distance, though certainly not monotonically.

Panels 2b and 2c respectively present the flow of migrants into and out of Cebu over a 5 year horizon. Because migration is heavily right-skewed, I use the natural logarithm of migrant flows. The maps show that Cebu faces substantial variation in migrant flows to and from other provinces, and it appears that there is more intense migration between Cebu and provinces that are located a shorter distance away. These migrant flows do not account for things like population, so the amount of information one can glean from visual inspection is limited.

Figure 2: Provinces' relation to Cebu

(a) Rainfall covariance with Cebu **(b)** Log in-migrants to Cebu **(c)** Log out-migrants from Cebu



The central idea of the empirical analysis is to take the variation in log migrant flows displayed in Panels 2b and 2c, subtract out the effect of distance as well as fixed effects for each province as an origin and as a destination, and compare the remaining unexplained migrant flows against the variation in shock covariance displayed in Panel 2a. Of course, my analysis includes every origin-destination pair of provinces and not just those involving Cebu.

I use two methods to estimate the relationship between migrant flows and origin-destination covariance. The first is linear regression, which I estimate using ordinary least squares (OLS). The relevant regression equation comes from taking the log of the expression in (19) for migrant flows m_{do}:

$$\log(m_{dot}) = \beta_1 \mathrm{Cov}_{do}(\text{Rainfall shock}) + f(dist_{do}) + \delta_d + \delta_o + \tau_t + \varepsilon_{dot} \qquad (22)$$

The unit of observation is an origin-destination-year, since I have migration data for two census years (2000 and 2010). As suggested by the model, I include a cost function of distance as well as origin and destination fixed effects δ_o and δ_d. I also include period fixed effects τ_t to account for time-varying differences in country-wide migration.

The second method I use is Poisson regression, where the dependent variable is explicitly modeled as an exponential function of the covariates. The regression equation closely resembles (19), the expression generated by the model for migrant flows:

$$m_{dot} = \exp(\beta_1 \text{Cov}_{do}(\text{Rainfall shock}) + f(\text{dist}_{do}) + \delta_o + \delta_d + \tau_t + \varepsilon_{dot}) \tag{23}$$

A Poisson regression can handle zeroes, while for the OLS regressions I use $\log(m_{do}+1)$ as the dependent variable. The central prediction of the model is that when households are risk averse and engage in cross-location risk sharing, the point estimate for β_1 will be negative. The reason is that $\beta_1 = -\theta\lambda\phi(1-\phi)$ in the model, when $\text{Cov}_{do}(\text{Rainfall shock})$ is taken to represent σ^y_{do}.

The specifications above in (22) and (23) are straightforward, but a closer test of the model would account for agricultural employment shares. In the model, the covariance of output shocks $\sigma^y_{do} = (\vartheta^a_d \vartheta^a_o)^\alpha \sigma^a_{do}$: the covariance of agricultural (rainfall) shocks multiplied by agricultural shares in the origin and destination raised to α. Using this exact expression in empirical tests is tricky, since it requires taking a stance on the value of α. Nonetheless, the model is clear that the covariance of rainfall shocks has heterogeneous effects: it matters most for pairs of provinces that *both* have high agricultural shares.

To test this heterogeneity, I modify the specifications above to include an interaction between the covariance of rainfall shocks and a "high agriculture" dummy variable. This dummy variable is equal to 1 when both the origin and destination have above-median agricultural shares.

$$\log(m_{do}+1) = \beta_1 \mathbf{1}\{\text{High ag.}\}\text{Cov}_{do}(\text{Rainfall shock}) + \beta_2 \text{Cov}_{do}(\text{Rainfall shock}) \\ + f(dist_{do}) + \delta_d + \delta_o + \tau_t + \varepsilon_{dot} \tag{24}$$

$$m_{do} = \exp\big(\beta_1 \mathbf{1}\{\text{High ag.}\}\text{Cov}_{do}(\text{Rainfall shock}) + \beta_2 \text{Cov}_{do}(\text{Rainfall shock}) \\ + f(dist_{do}) + \delta_d + \delta_o + \tau_t + \varepsilon_{dot}\big) \tag{25}$$

The central identifying assumption when estimating these gravity equations is that the error term ε_{dot} is uncorrelated with the covariance of rainfall shocks. ε_{dot} represents idiosyncratic advantages of moving from o to d in period t not captured by other variables. The biggest threat to identification is that, as seen in Figure 2, $\text{Cov}_{do}(\text{rainfall shocks})$ is negatively correlated with the distance between d and o. As a result, a mis-specified cost of distance will bias the estimate of β_1. For example, if I were to control for log distance, as is common in the gravity equation literature, and the cost of distance was less concave than the natural logarithm, some residual distance cost would be left in the error term and β_1 would be biased. As a solution to this problem, I control for a fifth-degree polynomial of distance to allow for more flexibility.[2] For robustness, I add flexible controls for driving time (also as a polynomial) to some specifications.

[2] I use a fifth-degree polynomial because the first five coefficients of higher-degree distance polynomials are all significant. However, estimates of β_1 are similar when controlling for any distance polynomial that is at least quadratic.

4 Results

4.1 Baseline results

I present estimates of β_1 from the baseline specifications without heterogeneity (equations (22) and (23)) in Table 2. These results are mixed. Under OLS, there appears to be a significant negative effect of rainfall shock covariance on migrant flows. Point estimates from Poisson regressions are also negative, but they are smaller and statistically insignificant. Controlling for a polynomial of driving time between the origin and destination makes little difference, and in fact tends to increase the magnitude of the point estimates. These results are somewhat consistent with the model laid out above, in that one would expect an effect only for origin-destination pairs with high agriculture. It is therefore not surprising that the effect of the covariance of rainfall shocks over all origin-destination pairs is insignificant under some specifications.

4.2 Heterogeneity by agricultural shares

Table 3 displays estimates from the specifications in equations (24) and (25). These specifications are closer to the migration gravity equation generated by the model because they feature heterogeneity by agricultural shares. In each regression, I interact the covariance of rainfall shocks with a dummy variable that equals 1 if both the origin and the destination have an above-median agricultural employment share.

These results bear out the central prediction of the model: between origin-destination pairs with large agricultural sectors, migration decreases in the covariance of rainfall shocks, i.e. agricultural productivity shocks. For these pairs of provinces, the effect rainfall covariance is large, negative, and statistically

Table 2: Migration regressions on covariance of rainfall shocks

	Log migrants		Migrants	
	(1)	(2)	(3)	(4)
	OLS	OLS	Poisson	Poisson
Cov_{do}(rainfall shocks)	-0.8549**	-1.095***	-0.2581	-0.4150
	(0.3653)	(0.4111)	(0.5045)	(0.5176)
Distance control	Poly.	Poly.	Poly.	Poly.
Driving time control	None	Poly.	None	Poly.
Observations	11,400	10,512	11,400	10,512
Destination fixed effects	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes
Period fixed effects	Yes	Yes	Yes	Yes

Notes: $* \, p < 0.1$, $** \, p < 0.05$, $*** \, p < 0.01$
Two way (origin and destination) cluster robust standard errors displayed in parentheses. In OLS regressions I take the log of migrants plus one to include zeroes. When controlling for polynomials of distance and travel duration I use fifth-order polynomials.

significant: an increase in covariance by 0.1 decreases migration by 9.9-13.5%. Among the high agriculture subset of pairs, a one standard deviation (0.23) increase in rainfall covariance decreases migration by 20-27%. And a high agriculture origin-destination pair with the mean level of covariance 0.42 has 34-43% lower aggregate migration than if the pair were uncorrelated. These numbers are large. Rainfall covariance appears to play an economically important role among high agriculture pairs of provinces.

By contrast, the covariance of rainfall shocks has no statistically significant effect among the remaining origin-destination pairs. In the Poisson regressions, the

Table 3: Migration regressions on covariance of rainfall shocks, with heterogeneity by ag. shares

	Log migrants		Migrants	
	(1)	(2)	(3)	(4)
	OLS	OLS	Poisson	Poisson
$\mathbf{1}\{\text{High ag.}\}\times\text{Cov}_{do}(\text{rainfall shocks})$	-1.262**	-1.350**	-0.9903***	-1.007***
	(0.5277)	(0.5689)	(0.3716)	(0.3801)
$(1-\mathbf{1}\{\text{High ag.}\})\times\text{Cov}_{do}(\text{rainfall shocks})$	-0.1760	-0.4149	0.0013	-0.1168
	(0.4821)	(0.5071)	(0.5119)	(0.5199)
Distance control	Poly.	Poly.	Poly.	Poly.
Driving time control	None	Poly.	None	Poly.
Observations	11,400	10,512	11,400	10,512
Dest.$\times\mathbf{1}\{\text{High ag. } o\}$ fixed effects	Yes	Yes	Yes	Yes
Origin$\times\mathbf{1}\{\text{High ag. } d\}$ fixed effects	Yes	Yes	Yes	Yes
Period fixed effects	Yes	Yes	Yes	Yes

Notes: $*\ p<0.1$, $**\ p<0.05$, $***\ p<0.01$
Two way (origin and destination) cluster robust standard errors displayed in parentheses. In OLS regressions I take the log of migrants plus one to include zeroes. When controlling for polynomials of distance and driving time I use fifth-order polynomials. Distance and driving time polynomials are interacted with the high-agriculture dummy.

coefficient on rainfall covariance outside high agriculture pairs is close to zero. Again this finding is consistent with the theory: these are pairs where at least one province has a small agricultural sector, and the model is clear in equation (19) that both provinces in a pair must have a large agricultural sector for rainfall covariance to have a meaningful effect.

As in the regressions without heterogeneity, controlling for a polynomial of driving time makes little difference for the point estimates of interest. In Table

A1 in the appendix, I present similar regressions with a different coefficient for every possible pair of high agriculture and low agriculture origins/destinations (four categories in total). The results are consistent with those of Table 3: the effect of rainfall covariance is concentrated in high agriculture pairs.

In equation (19), the log of migration depends negatively on $\theta\lambda\phi(1-\phi)\left(\overline{\vartheta}_d^a\overline{\vartheta}_o^a\right)^\alpha\sigma_{do}^a$. Agricultural shock covariance σ_{do}^a corresponds to the covariance of rainfall shocks, and high-agriculture pairs are ones where $\left(\overline{\vartheta}_d^a\overline{\vartheta}_o^a\right)^\alpha$ is large. Rejecting that the coefficient on rainfall covariance is zero for these pairs means rejecting that $\lambda\phi(1-\phi)=0$. λ is the coefficient of absolute risk aversion and ϕ is the risk sharing parameter. In the context of the model, the regression results indicate that agents are risk averse ($\lambda>0$) and engage in risk sharing across locations ($\phi\in(0,1)$).

5 Conclusion

This paper documents a novel theoretical and empirical relationship: origin-destination pairs with more related economic shocks have lower migrant flows. I derive this relationship from a model of migration as an *ex ante* risk management strategy. I then estimate gravity equations using rainfall and census data from the Philippines to empirically validate this negative relationship. I find that origin-destination covariance of rainfall shocks has a sizable effect on migration: a pair of high agriculture provinces with average rainfall covariance has 34-43% lower migration than if the pair were uncorrelated. This relationship is robust to using OLS or Poisson regression, and to flexibly controlling for driving time between the origin and destination.

The relationship between origin-destination covariance and migration has implications for spatial patterns of economic development. My results demonstrate that risk management strategies are important enough to shape the allocation of labor across space. Economic activity will not naturally gravitate to the most productive locations if migrating to those locations leaves workers overly exposed to risk. It is already well-known that exposure to uninsured risk can influence workers' decision of *whether* to migrate. My contribution is to show that uninsured risk distorts the decision of *where* to migrate.

To my knowledge, this is the first paper to integrate a risk management scheme into a gravity equation for migration. Estimating similar gravity equations with more detailed data on local economic conditions, for example on wages or GDP per capita, would provide an additional test of the predictions derived from the model. Richer data may also allow for more relaxed assumptions about the functional form of the indirect utility function. Alternatively, future work could use dynamic discrete choice methods similar to Kennan and Walker (2011) on panel data for a deeper understanding of how risk management affects migrants' choice of destination.

Knowledge about the risk management strategies of workers in this context helps to inform a wide range of public policies, such as social insurance or inter-provincial transfer payments. My paper suggests that if such policies alter the inter-provincial covariance of shocks, they could shape the spatial distribution of economic activity.

References

Besley, Timothy, "Nonmarket institutions for credit and risk sharing in low-income countries," *Journal of Economic Perspectives*, 1995, *9* (3), 115–127.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

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

Fafchamps, Marcel and Flore Gubert, "The formation of risk sharing networks," *Journal of Development Economics*, 2007, *83* (2), 326–350.

— **and Susan Lund**, "Risk-sharing networks in rural Philippines," *Journal of Development Economics*, 2003, *71* (2), 261–287.

Harris, John R and Michael P Todaro, "Migration, unemployment and development: a two-sector analysis," *The American economic review*, 1970, *60* (1), 126–142.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kinnan, Cynthia and Robert Townsend, "Kinship and financial networks, formal financial access, and risk reduction," *American Economic Review*, 2012, *102* (3), 289–93.

Meghir, Costas, Ahmed Mushfiq Mobarak, Corina D Mommaerts, and Melanie Morten, "Migration and informal insurance," Technical Report, National Bureau of Economic Research 2019.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 7.2 [Philippines 2000]," 2019.

Morten, Melanie, "Temporary migration and endogenous risk sharing in village india," *Journal of Political Economy*, 2019, *127* (1), 1–46.

Munshi, Kaivan and Mark Rosenzweig, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (1), 46–98.

Quisumbing, Agnes and Scott McNiven, "Moving forward, looking back: The impact of migration and remittances on assets, consumption, and credit constraints in the rural Philippines," *The Journal of Development Studies*, 2010, *46* (1), 91–113.

Quisumbing, Agnes R and Scott McNiven, "Migration and the rural-urban continuum: Evidence from the Rural Philippines," Technical Report 2005.

_, _, and Marie Godquin, "Shocks, groups, and networks in Bukidnon, the Philippines," *Collective Action and Property Rights for Poverty Reduction: Insights from Africa and Asia*, 2012, pp. 79–109.

Redding, Stephen J and Esteban Rossi-Hansberg, "Quantitative spatial economics," *Annual Review of Economics*, 2017, *9*, 21–58.

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 2008, *11* (4), 707–720.

Rosenzweig, Mark R and Christopher Udry, "Rainfall forecasts, weather, and wages over the agricultural production cycle," *American Economic Review*, 2014, *104* (5), 278–83.

— **and Kenneth I Wolpin**, "Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India," *Journal of political economy*, 1993, *101* (2), 223–244.

— **and Oded Stark**, "Consumption smoothing, migration, and marriage: Evidence from rural India," *Journal of political Economy*, 1989, *97* (4), 905–926.

Townsend, Robert M, "Risk and insurance in village India," *Econometrica: Journal of the Econometric Society*, 1994, pp. 539–591.

— , "Financial systems in Northern Thai villages," *The Quarterly Journal of Economics*, 1995, *110* (4), 1011–1046.

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

A Additional tables and figures

Table A1: Migration regressions on covariance of rainfall shocks, with heterogeneity

	Log migrants		Migrants	
	(1)	(2)	(3)	(4)
	OLS	OLS	Poisson	Poisson
$\mathbf{1}\{\text{Low ag. } d\} \times \mathbf{1}\{\text{Low ag. } o\} \times \text{Cov}_{do}(\text{rainfall shocks})$	0.8238	-0.0650	0.0817	-0.4117
	(0.8302)	(0.7100)	(0.6679)	(0.5159)
$\mathbf{1}\{\text{Low ag. } d\} \times \mathbf{1}\{\text{High ag. } o\} \times \text{Cov}_{do}(\text{rainfall shocks})$	-0.3589	-0.5808	-0.3979	-0.6293
	(0.6167)	(0.6129)	(0.7038)	(0.6398)
$\mathbf{1}\{\text{High ag. } d\} \times \mathbf{1}\{\text{Low ag. } o\} \times \text{Cov}_{do}(\text{rainfall shocks})$	-0.3507	-0.6089	-0.4679	-0.6091
	(0.5933)	(0.6085)	(0.7558)	(0.6337)
$\mathbf{1}\{\text{High ag. } d\} \times \mathbf{1}\{\text{High ag. } o\} \times \text{Cov}_{do}(\text{rainfall shocks})$	-1.518***	-1.293**	-1.495**	-1.122*
	(0.4777)	(0.5277)	(0.6233)	(0.6638)
Distance control	Poly.	Poly.	Poly.	Poly.
Driving time control	None	Poly.	None	Poly.
Observations	11,400	10,512	11,400	10,512
Destination fixed effects	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes
Period fixed effects	Yes	Yes	Yes	Yes

Notes: $*\ p<0.1$, $**\ p<0.05$, $***\ p<0.01$
Two way (origin and destination) cluster robust standard errors displayed in parentheses. In OLS regressions I take the log of migrants plus one to include zeroes. When controlling for polynomials of distance and driving time I use fifth-order polynomials. Distance and driving time polynomials are interacted with the heterogeneous agricultural productivity indicators.