Online Appendix

Wealth Heterogeneity and the Income Elasticity of Migration

Samuel Bazzi

Contents

A Theoretical Results

This section provides proofs for the results in Section 2. Equations (3) and (4) can be obtained by integrating over landholdings R_{iv} in inequality (2) with (i) $\tau_{vj} > 0$ and $R_L \geq \underline{R}$ or (ii) $R_L < \underline{R}$ ($\Longleftarrow \tau_{vj} = 0$). First, consider the following expressions for the thresholds within which migration is both feasible and profitable in period t (from the perspective of $t-1$ decision-makers required to pay fixed upfront costs in that period)

$$R_{L,t-1} = \left(\frac{\tau_{vj} C_{vjt}}{p_{v,t-1}(\overline{\sigma}_v + a_{v,t-1})K_v^\theta}\right)^{1/\beta}; \quad R_{U,t-1} = \left(\frac{W_{vjt} - C_{vjt}}{\alpha_v p_{v,t-1}\overline{\sigma}_v \chi K_v^\theta}\right)^{1/\beta}, \tag{A.1}$$

where $\mathbb{E}_{t-1}[p_{vt}\sigma_{vt}] = \alpha_v p_{v,t-1}\overline{\sigma}_v$, which hinges on $\mathrm{cov}_{t-1}(p_{vt}, \sigma_{vt}) = 0$, i.e. households cannot forecast the relationship between rainfall and prices next period. This does not imply that past rainfall has no effect on contemporaneous prices. Rather, $a_{v,t-k}$ for $k > 0$ are elements of the error term $\sum_{s=0}^{q} \upsilon_s e_{v,t-s}$ in the $ARMA(1,Q)$ expression for rice prices. Thus, past output has a direct effect on current prices.[1]

If CIA constraints are binding, then the stock migration rate in period t is derived by integrating over all landholdings $R_{iv} \in [R_{L,t-1}, R_{U,t-1}]$ in village v (maintaining the innocuous normalization $\underline{R} = 1$ ha)

$$\mathbb{P}(R_{L,t-1} \leq R_{iv} \leq R_{U,t-1}) = \frac{M_{vt}}{N_{vt}} = \int_{R_{L,t-1}}^{R_{U,t-1}} \lambda_v R_{iv}^{-\lambda_v - 1} \mathrm{d}R_{iv} = R_{L,t-1}^{-\lambda_v} - R_{U,t-1}^{-\lambda_v}. \tag{A.2}$$

Replacing the expressions for R_L and R_U with those in equation (A.1) and taking the difference in logs between $t+1$ and t, we obtain equation (3).

On the other hand, if CIA constraints are not binding and $R_L < \underline{R}$ ($\Longleftarrow \tau_{vj} = 0$), then

$$\mathbb{P}(1 \leq R_{iv} \leq R_{U,t-1}) = \frac{M_{vt}}{N_{vt}} = \int_{1}^{R_{U,t-1}} \lambda_v R_{iv}^{-\lambda_v - 1} \mathrm{d}R_{iv} = 1 - R_{U,t-1}^{-\lambda_v}. \tag{A.3}$$

Similarly substituting for $R_{U,t-1}$ and taking differences in logs implies equation (4). Recall that, by definition, the expressions for the intensive margin in (A.2) and (A.3) must be greater than zero.

Proposition 1

The proofs in the presence of CIA constraints follow immediately from differentiation of equation (3). Letting $\Delta \ln(M_{v,t+1}/N_{v,t+1}) \equiv \Delta\widehat{M}_{v,t+1}$,

$$\frac{\partial \Delta \widehat{M}_{v,t+1}}{\partial \Delta \ln p_{vt}} = \frac{\lambda_v}{\beta} > 0; \quad \frac{\partial \Delta \widehat{M}_{v,t+1}}{\partial a_{vt}} = \frac{\frac{\lambda_v}{\beta}(\overline{\sigma}_v + a_{vt})^{\lambda_v/\beta - 1}(\tau_{vj}C_{vj,t+1})^{-\lambda_v/\beta}}{\left(\frac{\overline{\sigma}_v + a_{vt}}{\tau_{vj}C_{vj,t+1}}\right)^{\lambda_v/\beta} - \left(\frac{\alpha_v\overline{\sigma}_v}{W_{vj,t+1} - C_{vj,t+1}}\right)^{\lambda_v/\beta}} > 0. \tag{A.4}$$

The derivative with respect to rainfall last period, $a_{v,t-1}$, is identical to $\partial\Delta\widehat{M}_{v,t+1}/\partial a_{vt}$ with a leading negative sign and shifting all t subscripts back to $t-1$. The proof that rainfall shocks have no effect in the absence of CIA constraints is trivial since a_{vt} and $a_{v,t-1}$ do not enter equation (4). The positive effect of price shocks on $\widehat{M}_{v,t+1}$ in the presence of CIA constraints follows immediately from the fact that $\lambda_v/\beta > 0$. The proof that price shocks have a negative effect on the change in migration rates in the absence of CIA constraints proceeds by checking that the

[1] The expressions are more complicated if prices (i) follow a higher-order autoregressive process or (ii) have a forecastable nonzero drift term, and/or (iii) households do not have rational expectations over the high frequency seasonality in prices. Nevertheless, the assumptions here are largely consistent with the time series properties of rainfall and rice prices in Indonesia (and presumably elsewhere). Moreover, the first-order price formulation is sufficiently general to comprise more higher-order Markov processes (see Chambers and Bailey, 1996).

following expression satisfies increasing differences (over time) in (H_{vs}, p_{vs}),

$$\ln\left[1-(H_{vs}p_{vs})^{\frac{\lambda_v}{\beta}}\right],$$

where $H_{vs} = \alpha_v\overline{\sigma}_v\chi K_v^\theta/(W_{vj,t+1} - C_{vj,t+1})$. This condition holds so long as migration costs are non-increasing, $C_{vj,t+1} \leq C_{vjt}$, which seems plausible in most settings. Of course, taking the derivative with respect to the price *level*, we find

$$\frac{\partial\Delta\widehat{M}_{v,t+1}}{\partial p_{vt}} = \frac{-\frac{\lambda_v}{\beta}p_{vt}^{\lambda_v/\beta-1}\left(\frac{\alpha_v\overline{\sigma}_v\chi K_v^\theta}{W_{vj,t+1}-C_{vj,t+1}}\right)^{\lambda_v/\beta}}{1-\left(\frac{\alpha_v p_{vt}\overline{\sigma}_v\chi K_v^\theta}{W_{vj,t+1}-C_{vj,t+1}}\right)^{\lambda_v/\beta}} < 0. \tag{A.5}$$

Proposition 2

The fact that λ_v has an ambiguous effect on the intensive margin follows immediately from differentiating equations (3) or (4) and recognizing that the terms inside brackets $[\cdot]$ within the logarithm are less than one. That $\frac{\partial\Delta\widehat{M}_{v,t+1}}{\partial\Delta\ln p_{vt}\partial\lambda_v} = 1/\beta > 0$ in the presence of CIA constraints is immediate from equation (3). To show that $\partial^2\Delta\widehat{M}_{v,t+1}/\partial a_{vt}\partial\lambda_v > 0$, simply rearrange and differentiate equation (A.4) with respect to λ_v

$$\frac{\partial^2\Delta\widehat{M}_{v,t+1}}{\partial a_{vt}\partial\lambda_v} = \frac{\frac{1}{\beta}(\overline{\sigma}_v+a_{vt})^{-1}}{1-\left(\frac{\alpha_v\overline{\sigma}_v\tau_{vj}C_{vj,t+1}}{(\overline{\sigma}_v+a_{vt})(W_{vj,t+1}-C_{vj,t+1})}\right)^{\frac{\lambda_v}{\beta}}} + \frac{\frac{\lambda_v}{\beta^2}(\overline{\sigma}_v+a_{vt})^{-1}\left(\frac{\alpha_v\overline{\sigma}_v\tau_{vj}C_{vj,t+1}}{(\overline{\sigma}_v+a_{vt})(W_{vj,t+1}-C_{vj,t+1})}\right)^{\frac{\lambda_v}{\beta}}\ln\left(\frac{\alpha_v\overline{\sigma}_v\tau_{vj}C_{vj,t+1}}{(\overline{\sigma}_v+a_{vt})(W_{vj,t+1}-C_{vj,t+1})}\right)}{\left(1-\left(\frac{\alpha_v\overline{\sigma}_v\tau_{vj}C_{vj,t+1}}{(\overline{\sigma}_v+a_{vt})(W_{vj,t+1}-C_{vj,t+1})}\right)^{\frac{\lambda_v}{\beta}}\right)^2}. \tag{A.6}$$

Letting $x_v := \alpha_v\overline{\sigma}_v\tau_{vj}C_{vj,t+1}$ and $y_v := (\overline{\sigma}_v + a_{vt})(W_{vj,t+1} - C_{vj,t+1})$, recognizing that $x_v < y_v$ (for those migrating, i.e. $R_{iv} \in [R_L, R_U]$), and noting that $(y_v/x_v)^{\lambda_v/\beta} + (\lambda_v/\beta)\ln(x_v/y_v) > 1$, it can be shown that equation (A.6) is positive. In the absence of CIA constraints, a similar calculation on equation (A.5) shows that $\partial^2\Delta\widehat{M}_{v,t+1}/\partial p_{vt}\partial\lambda_v < 0$.

Multiple Labor Units. There are two ways to think about the household income maximization problem above in the context of allocating multiple units of household labor. In either approach, there is no tradeoff between holding on to one's land and migrating as in Jayachandran (2006). Moreover, the key insight in inequality (2) remains unchanged. In case one, define $S_{iv} \equiv s_{iv}L_{iv}$ where s is the share of household i's total labor L working at home. The collective household objective is then

$$\max_{s_{iv}}\ \mathbb{E}_t[p_{v,t+1}\sigma_{v,t+1}]K_v^\theta(s_{iv}L_{iv})^\phi R_{iv}^\beta + L_{iv}(1-s_{iv})(W_{vj,t+1} - C_{vj,t+1}),$$

subject to $Y_{ivt} \geq \tau_{vj}C_{vj,t+1}$, with the solution s_{iv}^* implying that household i finds migration profitable if $s_{iv}^*L_{iv}(W_{vj,t+1} - C_{vj,t+1}) > \mathbb{E}_t[p_{v,t+1}\sigma_{v,t+1}]\phi K_v^\theta S_{iv}^{\phi-1}R_{iv}^\beta$, which holds under equation (1). In case two, we appeal to the fact that $Y_{ivt} = \sum_\ell^{\mathcal{L}} y_{\ell ivt}$, where $y_{\ell ivt}$ is output per capita. Hence, household i has at least one migrant abroad in $t+1$ whenever $\mathbb{E}_t[p_{v,t+1}\sigma_{v,t+1}]y_{\ell iv,t+1} \leq (W_{vj,t+1} - C_{vj,t+1})$. Because labor is perfectly substitutable within the household and the technology is constant returns, this condition also holds under equation (1).

Extensive Margin. As discussed in Section 5, λ_v has an ambiguous effect on the extensive margin regardless of the formulation of the extreme landholding statistics. Under the finite sample formulation, the proof follows immediately from the derivative of the first equation in the footnote on page 15 with respect to λ_v, $N_v R_U^{-\lambda_v N_v}\ln R_U - N_v(1-R_L^{\lambda_v})^{N_v}R_L^{-\lambda_v}\ln R_L$, the sign of which cannot be determined without imposing *ad hoc* bounds on parameter values. The ambiguity similarly holds for the population-based order statistic approach. Meanwhile, the positive effect of population size N_v on the extensive margin follows from straightforward differentiation.

B Econometric Procedures

This section details the two-step estimating framework introduced in equations (8) in Section 5.

B.1 Parametric

The parametric approach due to Poirier (1980) presumes that $(u_{vt}, u_{v,t+1}, \Delta\varepsilon_{v,t+1})$ in equation (8) follow a trivariate normal distribution with mean zero, variances $(1, 1, \text{var}(\Delta\varepsilon))$, and pairwise correlation terms $(\rho_{u_t u_{t+1}},\ \rho_{u_t\Delta\varepsilon},\ \rho_{u_{t+1}\Delta\varepsilon})$. These assumptions imply that

$$\mathbb{E}\left[\Delta\varepsilon_{v,t+1}\middle|\mathbf{Z}'_{v,t-1}\boldsymbol{\phi}_{t-1} > -u_{vt}, \mathbf{Z}'_{vt}\boldsymbol{\phi}_t > -u_{v,t+1}\right] = \rho_{u_t\Delta\varepsilon}\kappa_{vt} + \rho_{u_{t+1}\Delta\varepsilon}\kappa_{v,t+1},$$

where κ_{vt} and $\kappa_{v,+1}$ are bivariate Mills ratio terms. Implementation proceeds in two steps. First, I estimate a bivariate probit model for the extensive margins in t and $t+1$. Since prices, rainfall and population size vary over time, the bivariate first stage has several sequential exclusion restrictions. Second, I augment an empirical specification for the change in the log migration rate with the estimated correction terms $\widehat{\kappa}_{vt}$ and $\widehat{\kappa}_{v,t+1}$, which enter with population coefficients equal to $\rho_{u_t\Delta\varepsilon}$ and $\rho_{u_{t+1}\Delta\varepsilon}$ respectively. Straightforward OLS then delivers a consistent estimate of second stage parameters. See Rochina-Barrachina (1999) for further theroetical background on the relationship between Poirier's original cross-sectional bivariate probit and the two-period panel implementation as described here.

B.2 Semiparametric

This section sketches a practical semiparametric procedure based on Das et al. (2003) for estimating the system of equations in (8) that is arguably more robust to distributional misspecification than the parametric Poirier approach. Rather than closed-form correction terms, the semiparametric approach relies on a double-index in the propensity scores $g(\mathbf{Z}'_{v,t-1}\boldsymbol{\phi}_{t-1}, \mathbf{Z}'_{vt}\boldsymbol{\phi}_t)$, where g is an unknown function of the latent variable indices.

Implementation proceeds as follows. First, rather than assuming bivariate normality of $(u_{vt}, u_{v,t+1})$, I use a seemingly unrelated linear probability models (SU-LPM) making no assumptions on the joint distribution of u_{vt} and $u_{v,t+1}$ (Zellner and Lee, 1965).[1]

Second, I use the estimates of $\boldsymbol{\phi}_t$ and $\boldsymbol{\phi}_{t+1}$ to approximate $g(\cdot)$. In practice, I employ an Lth-degree power series expansion in the propensity scores $\widehat{P}_s = \mathbf{Z}'_s\widehat{\boldsymbol{\phi}}_s$—linear predictions recovered from the bivariate SU-LPM estimator—for village v to have at least one migrant in period s.[2] Lastly, consistent second-stage estimates of Θ can be obtained from an OLS regression conditioning on the power series $g(\cdot)$ function so long as at least two variables in $\mathbf{Z}_{t-1} \cup \mathbf{Z}_t$ do not also appear in $\mathbf{X}_t$.

B.3 Inference

In both the parametric and semiparametric framework outlined above, the correction terms introduce added sampling variation into the second-stage.[3] Taking a conservative and unbiased approach to inference, I implement

[1] Results are similar albeit computationally costly using a semi-nonparametric pseudo-maximum likelihood (SNP-ML) procedure based on an approximation to the unknown latent error densities (Gallant and Nychka, 1987).

[2] This is essentially the approach suggested by Das et al. (2003) who recommend using a fully nonparametric estimator to estimate the propensity score. Newey (1988) argues that a first stage linear probability model provides consistent estimates in two-step selection models, though a semiparametric first stage estimator provides more efficient (second-stage) estimates (Newey, 2009). An important difference with Das et al., however, is that they assume $u_{vt} \perp u_{v,t+1}$ whereas the estimates of $\boldsymbol{\phi}$ obtained using bivariate SU-LPM explicitly allow for $\text{corr}(u_{vt}, u_{v,t+1}) \neq 0$. Results are robust to estimating two distinct LPMs with $\text{corr}(u_{vt}, u_{v,t+1}) = 0$.

[3] Although the Pareto parameters $\widehat{\lambda}_v$ are generated regressors, these fitted distributional terms are obtained from more than 55,000 regressions comprising the universe of agricultural households in Indonesia. Similar to other studies employing population measures of inequality, I treat the added sampling variation from these terms as negligible. This is reasonable here given that the Agricultural Census of 2003 purports to capture the full agricultural population of every village. Moreover, even if these distributional parameters are estimated with error, these errors

a bootstrap$-t$ procedure (also known as percentile$-t$) with clustering at the district level. All tables report the uncorrected standard errors, but the significance levels are computed based on the cluster bootstrap$-t$ procedure described in detail in Cameron et al. (2008).[4] Each second-step significance level is based on 999 bootstrap iterations, where I cluster the standard errors at each iteration and construct the iteration-specific Wald test statistic (t-stat) re-centered on the original point estimate. Using these 999 Wald statistics, I then compute the (possibly asymmetric) 90th, 95th, and 99th% confidence intervals in reporting the significance level $\alpha \in \{0.1, 0.05, 0.01\}$ associated with each point estimate.

The simulation results in Cameron et al. (2008) suggest that the empirical setup in this paper is well suited to the cluster bootstrap$-t$ procedure. In particular, the data comprise a large number of districts (> 200 in all specifications) with an unbalanced number of villages, several observable variables are relatively constant within district, and several binary regressors. Moreover, Yamagata (2006) finds that the bootstrap$-t$ procedure outperforms the conventional bootstrap$-se$ procedure in the context of estimating Heckman (1976)-type selection models similar to those in this paper.[5]

should not affect the standard errors on the first- and second-step coefficients in the model in equation (8) because (i) the moment conditions in the two village-level equations in (8) are orthogonal to the moment condition in the auxiliary OLS household-level regressions used to estimate λ_v for each village, and (ii) the λ_v terms enter linearly, and hence the added sampling variation can be ignored (see Newey and McFadden, 1994, pp. 2182-2183).

[4]In applications of the bootstrap$-t$ procedure, authors sometimes report p-values. While retaining the original biased standard errors, I report the unbiased significance levels when those p-values fall below 0.1. The underlying p-values are available upon request.

[5]The cluster bootstrap$-t$ procedure that I employ yields confidence intervals with correct coverage in addition to asymptotic refinement. In unreported results similar to Yamagata (2006), I also find that the 95% confidence intervals generated by a conventional cluster bootstrap$-se$ procedure fail to cover the original point estimate in more than 5% of iterations, suggesting important finite-sample shortcomings of the conventional bootstrap.

C Data Description

Variable	Source	Definition
population	*Podes* 2005/8	all people registered as residents for at least six months or less than six months with the intention of staying
migrants	*Podes* 2005/8	all people working abroad on a fixed wage for a fixed time period
$\widehat{\lambda}_v$	Agricultural Census 2003	estimate of the Pareto exponent λ_v for village v based on OLS estimation (Gabaix and Ibragimov, 2011); see Appendix D for details
share households above $\underline{R}$	Agricultural Census 2003	share of all households in village v reporting landholdings less than $\underline{R}$ where $\underline{R} = 0.1$ hectares in the baseline case
$\widehat{\lambda}_v^{63}$	Agricultural Census 1963	estimate of the Pareto exponent λ_v for district d based on maximum likelihood estimation given the reporting of landholdings frequencies in 8 bins
rice prices	Wimanda (2009) via BPS	see Appendix E for details
rainfall in year t	NOAA/GPCP	total amount of rainfall during the given growing/harvest season where (1) seasons are 12 month intervals beginning with the first month of the province-specific wet season in a given year (Maccini and Yang, 2009), and (2) rainfall at the village level is based on rainfall levels recorded interpolated down to 0.5 degree (latitude/longitude) pixels between rainfall stations
plurality destination fixed effect	*Podes* 2005	indicators for whether a plurality of migrants from village v were working in Malaysia, Hong Kong, Singapore, Taiwan, Japan, South Korea, UAE, Saudi Arabia, Jordan, Kuwait, USA and Other
no motorized land travel to district capital	*Podes* 2002	equals one if there is no direct travel to the district capital using motorized land-based vehicles
reporting frequency	*Podes* 2005	one of five ordered levels: no formal population register, non-routine reporting, annual reporting, quarterly reporting, monthly reporting
distance to nearest district capital	*Podes* 2005/8	the minimum of the travel distance in kilometers to the given district capital or the nearest capital in a neighboring district
distance to subdistrict capital	*Podes* 2005/8	travel distance in kilometers to the capital of the village's subdistrict
distance to nearest emigration center		great circle distance from the centroid of the district in which village is located to the centroid of the nearest of 17 cities capable of processing legal international contract migration; cities include Aceh, Medan, Pekanbaru, Palembang, Jakarta, Bandung, Semarang, Yogyakarta, Surabaya, Pontianak, Banjarbaru, Nunukan, Makassar, Mataram, Kupang, Tanjung Pinang, and Bali
urban	*Podes* 2005/8	a government-constructed indicator which equals one if the village has a population density greater than 5000 per square kilometer, a majority of the population recorded as non-farming households, and any number of public institutions which I do not observe directly in *Podes*
distance to Ho Chi Minh City/Bangkok (port)		great circle distance from the centroid of the village is located to the nearest Indonesian port plus the shipping distance abroad; geocoordinates of Indonesian port cities obtained from AtoBviaC and shipping distances from e-ships
Arab (Chinese) population share	Population Census 2000	the number of individuals claiming Arab (Chinese) descent as a share of village population
Muslim population share	*Podes* 2005/8	the number of individuals claiming adherence to Islamic faith as share of village population
post-primary education share	Population Census 2000	share of the population aged 5 and above that has completed junior secondary (*SLTP/setara*), senior secondary (*SLTA/setara*), *or* post-secondary (*Diploma/DIII/Akedemi/DII/DIV*)
share population aged 15-29	Population Census 2000	age range is chosen to correspond to the majority migration age of 18-34 in later years as reported in the Bank Indonesia (2009) survey
estimated mean household expenditure/capita	Suryahadi et al. (2005)	estimate of the average household expenditures per month, obtained from the poverty mapping exercise based on the 2000 Census
total rice output in tons per Ha	*Podes* 2002	total rice output recorded in village in 2001 divided by total area harvested
bank presence	*Podes* 2002/5	all formal banking institutions including rural people's banks (BPR) and commercial microfinance (BRI)
village land area	*Podes* 2005/8	total land area in hectares
wetland/total farmland	*Podes* 2005/8	the ratio of *sawah* or wetland to the total agricultural land available in the village; wetland is most suitable for rice production though it can be used to grow other crops such as tobacco and sugar as well
agricultural GDP/capita	Central Statistics Bureau	district-level nominal GDP

D Theory and Estimation with the Pareto Distribution

In this section, I provide additional background on the assumed Pareto distribution for land-holdings as well as details on the empirical content of the estimated Pareto shape parameters $\widehat{\lambda}_v$. Figure D.1 shows the familiar power law linearity in plots of the log complementary CDF against log wetland holding size for 16 randomly chosen districts. A more systematic analysis of Paretian properties at the village level requires estimating distributional parameters using the universal microdata from the the Agricultural Census.

I obtain estimates of λ_v for every village in Indonesia using the Gabaix and Ibragimov (2011) estimator. That is, for each village I regress the log rank minus 1/2 on the log of the given land-holding size. Given that some households within each village report the same land-holding size, ties are broken by taking the average rank.[1] Identical results obtain when using the log minimum, log maximum rank, or the log complementary CDF as the dependent variable (the measures have mutual correlations above 0.95). In terms of differences in $\widehat{\lambda}_v$ across the three different measures of land-holdings, Figure D.2 demonstrates that total agricultural land-holdings tend to yield the lowest estimates of λ_v (greater dispersion) whereas wetland holdings tend to yield the largest estimates of λ_v (less dispersion).[2] This is consistent with the existence of relatively smallholder rice agriculture and much larger plots used to grow other crops besides rice throughout the country.

Applying a test for departures from Paretian linearity suggested by Gabaix (2009), I find that the Pareto assumptions do not hold in around 25 percent of villages. Nevertheless, for reasons discussed in the paper, I maintain the view that the Pareto provides a reasonable approximation to the land-holdings distribution for the specific analytic purposes in this study. The goal is not to establish that land-holdings undeniably follow a power law, but rather that the formulation here provides a good fit to the data. And in Appendix F.5, I demonstrate that the key parameter estimates in the two-step model for flow migration rates are unaffected by imposing alternative choices of $\underline{R}$ in the estimation of λ_v.

The variation in $\widehat{\lambda}_v$ across villages contains information on the distribution of wealth and agricultural activities. Figure D.3 establishes further that λ_v is informative about the share of households engaged in the sale of agricultural products. Villages with lower $\widehat{\lambda}_v$ (i.e., higher mean and greater dispersion in R_{iv}) are more likely to have a majority of households selling agricultural output.

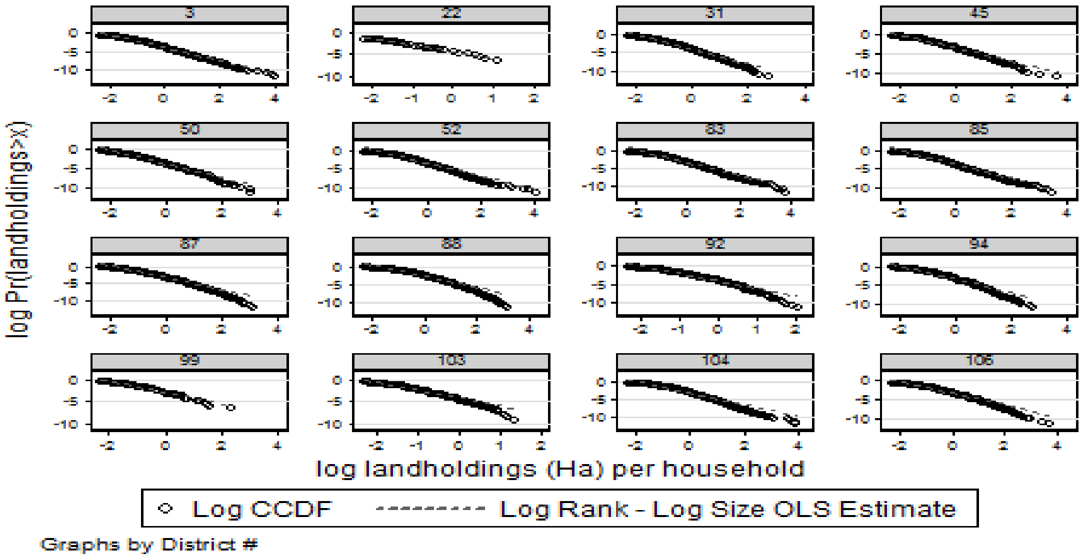
In closing, I mention several facts supporting the important assumption in the paper that the empirical land-holdings distribution in village v is predetermined with respect to migration in v. First, note that the Agricultural Census was enumerated in late 2003 (i) two years before we first observe migrant stocks in *Podes* 2005, and (ii) several months prior to the initial discussion and eventual implementation of the import ban. Thus, the observed heterogeneity in land-holdings could not be due to land transactions in expectation of or response to the price shock. Moreover, Benjamin (1995) demonstrates that farm sizes in Javanese villages are relatively fixed in the short-run due to imperfect land markets and long rental contracts. More recent *Susenas* data from 2005 covering the entire country confirm that less than one percent households engage in land transactions over a one year horizon. The same transaction rate holds in the data from one year prior, suggesting that households had not purchased land in expectation of rising prices.

[1] The discrete clumping at certain round land-holding sizes apparent in Figure D.1 in the paper is due in part to imperfect knowledge about plot sizes or boundaries. I therefore view the continuity of the Pareto distribution as a reasonable and innocuous approximation to the discrete land-holdings distribution—an assumption common in empirical work using the Pareto distribution (see Gabaix, 2009).

[2] In each case, there are a number of villages with $\widehat{\lambda}_v < 1$, which implies infinite mean land-holdings under the strict Pareto assumptions. When estimating λ_v using total agricultural land-holdings, for example, nearly 7 percent of villages have estimates of $\lambda_v < 1$. In all but 428 of these villages, however, the 95% upper confidence interval exceeds unity according to the unbiased standard error formula $\widehat{\lambda}_v \sqrt{2/N_v}$ given in Gabaix and Ibragimov (2011).

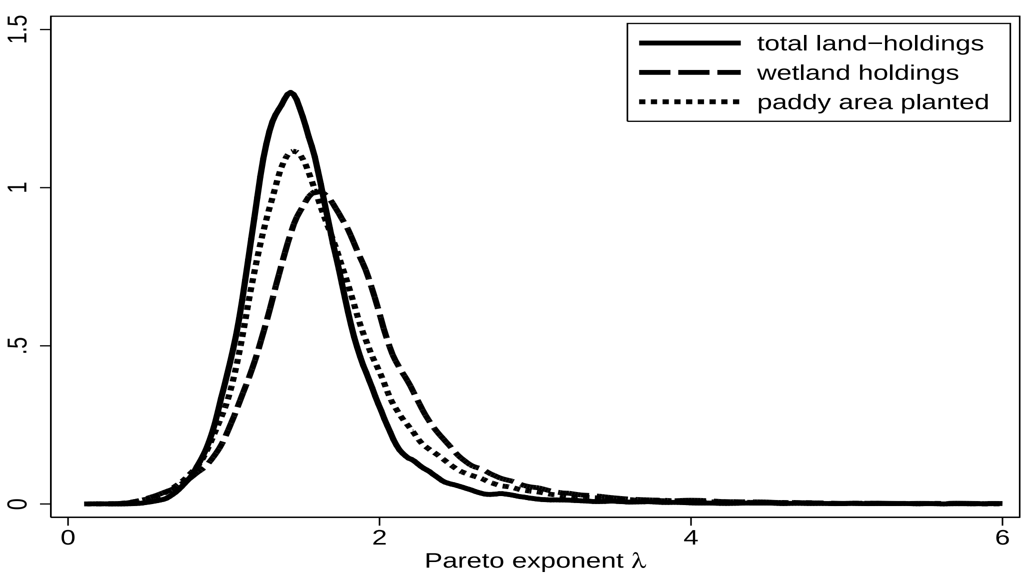
Figures

Figure D.1: Pareto Linearity in Log–Log Plots



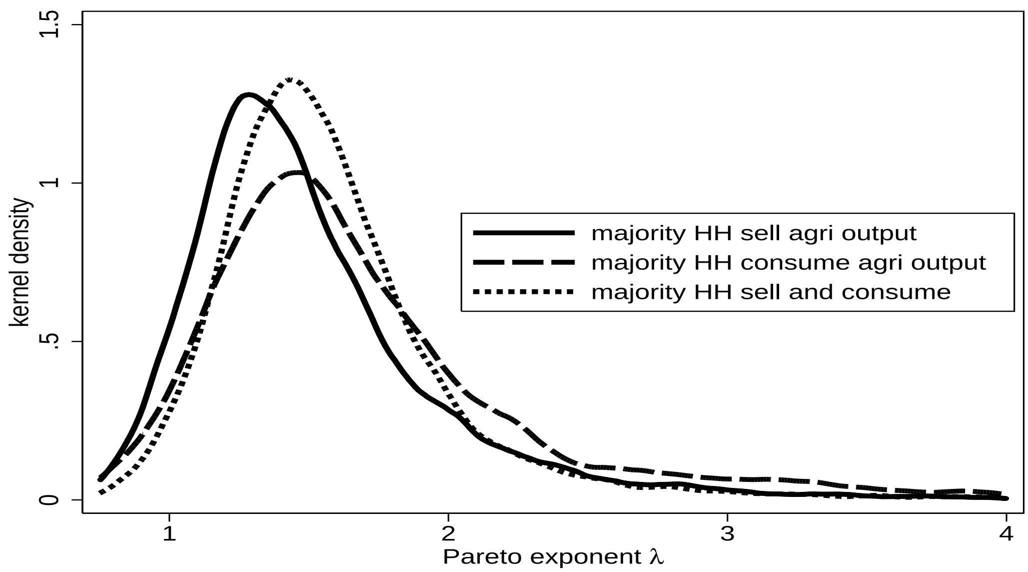
Notes: The figures report the log CCDF – log size observations for wetland holdings for Indonesian households recorded in 16 districts chosen at random from the Agricultural Census of 2003. The graphs impose lower thresholds of $\underline{R} = 0.1$ in estimating the CCDF. The line constitutes the best linear fit from the log rank — log size regression.

Figure D.2: The Distribution of $\widehat{\lambda}_v$ for Different Land-holdings Measures



Notes: The figure plots kernel densities for $\widehat{\lambda}_v$ estimated with $\underline{R} = 0.1$ Ha. The densities are based on an Epanechnikov kernel and a rule-of-thumb bandwidth.

Figure D.3: Probability Majority of Households in Village v Are Net Producers



Notes: The curves are kernel densities of $\widehat{\lambda}_v$ broken down by village depending depending on whether the village head reports in *Podes* 2005 that a majority of (agricultural) households in the village sell, subsist, or both conditional on that village reporting agriculture being the most prominent source of employment. The densities employ an Epanechnikov kernel, a rule-of-thumb bandwidth, and trimming of the top and bottom 1 percent of $\widehat{\lambda}_v$.

E Further Details on Agricultural Income Shocks

In this section, I provide further background on the rice price shock subsequent to the ban on imports in 2004, the time series properties of rainfall and rice prices, and the effect of these shocks on expenditures and wages.

E.1 Spatial Variation in the Rice Price Shock: Theory and a Simple Test

To understand how the import ban exerted differential pressure on local prices across regions, I first consider a simple model which micro-founds local rice prices based on the domestic market structure, imports, and the world price. The primary contribution of the model is to rationalize the lack of spatial arbitrage evident in Figure 2 in the paper.[1] I adapt the formulation for changes in national rice prices given in Warr (2008) to a model in which key parameters are allowed to vary across regions of the country. I assume that there are no strategic interactions among producers or consumers across villages, but local market power (among farmers) is possible in the sense of monopolistic competition.

The key prediction of the model is that changes in rice prices vary across villages according to a simple expression relating proportional changes in farmgate rice prices in village v in year t, p_v^d (d for domestic), to changes in world prices, p^m, (m for imported)

$$\widehat{p}_{vt}^d = \epsilon_v^m \widehat{p}_t^m + \epsilon_v^o \widehat{p}_t^o \tag{E.1}$$

where ϵ_v^m is the passthrough elasticity from world prices, and ϵ_v^o is the elasticity of domestic rice prices with respect to changes in prices p^o of an index of other goods consumed within Indonesia.[2] The partial equilibrium form of the village v passthrough elasticity ϵ_v^m is given by

$$\epsilon_v^m = S_v^m(\rho_v + \eta_v)/(\chi_v^d + \rho_v S_v^m - \eta_v S_v^d), \tag{E.2}$$

where $\eta_v \leq 0$ is the overall price elasticity of demand for rice (composite of domestic and imported) in the geographically delineated markets relevant to village v; S_v^m is the share of imported rice in total rice expenditures and $S_v^d = 1 - S_v^m$ is the expenditure share on domestically-produced rice; ρ_v is the Armington elasticity of substitution between domestic and imported rice; and χ_v^d is the elasticity of domestic supply with respect to prices p^d.

As world prices declined from 2005 to late 2007 (see Figure E.1), the model above suggests that, net of the effect of the change in other prices, domestic prices should also have fallen. Instead, the import ban effectively imposed $\epsilon_v^m = 0$ *for all* villages (see Figure E.2).[3] Conditional on other determinants of rice prices, the relevant counterfactual setting would be one in which villages with $\epsilon_v^m > 0$ before the ban experience a decline in real rice prices while villages with no import penetration in local markets ($\epsilon_v^m \approx 0$) experience no change at all. In other words, given the price stabilizing role of imports in villages with $\epsilon_v^m > 0$ before the ban, the model implies that the import ban should cause larger price increases in villages with a higher passthrough elasticity,

$$\widehat{p}_{vt}^d \Big|_{\epsilon_v^m > 0} > \widehat{p}_{v't}^d \Big|_{\epsilon_{v'}^m \approx 0} \quad , \tag{E.3}$$

[1] The delayed effect of the import ban evident in that figure has a straightforward explanation. Imported rice was especially important in the months around harvests at the end of growing seasons with particularly low rainfall. Because the spring 2004 harvest occurred after a season of high rainfall, the lack of imported rice in early 2004 had little effect on prices. In fact, it was not until just prior to the primary harvest in spring 2005 after a season of low rainfall in certain regions that the lack of imports proved important as domestic rice prices began to escalate across Indonesia.

[2] One concern with this approach is that Indonesia's import level directly affects world prices. Although there is some time series evidence that world prices are increasing in Indonesian imports, it is unclear whether the relationship is causal or due to the effect of climate shocks throughout Southeast Asia which reduce output in major rice-exporting countries and also increase demand for imports in Indonesia. By all accounts, Indonesia remains a price-taker in the world rice market. Dawe (2008), for example, identifies an optimal ad valorem tariff of around 4 percent, which is essentially indistinguishable from free trade.

[3] Small import shipments in late 2007 were undertaken as part of a limited government-licensed procurement from Thailand and Vietnam to be distributed largely through the Raskin program which provides heavily subsidized rice to households below and just above the poverty line.

while the counterfactual implies the opposite

$$\widehat{p}^{d}_{vt}\Big|_{\epsilon^m_v>0} \leq \widehat{p}^{d}_{v't}\Big|_{\epsilon^m_{v'}\approx 0} \quad . \tag{E.4}$$

The relevant empirical question, then, is what determines variation in ϵ^m_v across villages.

According to equation (E.2), the local intensity of world price passthrough is governed by four parameters: the share of imports in local rice consumption, the price elasticities of supply and demand, and the Armington elasticity of substitution between domestic and imported rice. The key implications are that ϵ^m_v should be decreasing in the local price elasticity of supply and increasing in the share of imported rice in the markets which purchase village v output.[4] The limited available estimates suggest that supply elasticities vary considerably across regions and land types—0.15 on Java, 0.4 in Sumatra, 1.25 in Sulawesi for wetland paddy, and dryland supply elasticities are approximately twice as large (Warr, 2005). Moreover, given prevailing transportation and trade costs, the local pre-ban import penetration ratio should be decreasing in (i) the distance to the nearest international port and major wholesale markets, and (ii) the shipping distance from the nearest port to Bangkok and Ho Chi Minh City, the two primary markets from which the majority of Indonesia's rice imports originate. Indonesia's unique geography generates substantial variation in these distance-driven components of ϵ^m_v.

Using the monthly consumer rice price index described in the paper, Table E.1 demonstrates that the empirical changes in rice prices from 2002–8 are consistent with the model sketched above. I control for lagged rainfall levels to account for local supply shocks, and the main proxy for ϵ^m_v is the log average shipping distance to Thailand and Vietnam via the nearest port city in Indonesia. Regardless of the growth horizon on the left hand side (monthly, semi-annual, or annual), the primary takeaway is that after the import ban in January 2004, prices grew slower in Indonesian cities farther removed from the main rice exporter shipping routes in Southeast Asia. Before the ban, the opposite was true. Figure E.3 graphically depicts this main finding, which is consistent with equations (E.3) and (E.4). As elaborated in the paper, the distinct lack of spatial arbitrage evident in these results can be explained in part by the disruption of path-dependent, international buyer-seller networks after the import ban.

E.2 On Measuring Rice Prices

A few issues concerning the price indices deserve mention. First, while the price index is only available in 44 cities across Indonesia, these data points are arguably representative of the average regional prices faced by rice producers in nearby rural villages. Relative to prices in local rural markets, these measures should be (i) less affected by supply shocks in small groups of villages, and, (ii) more likely to capture the general equilibrium impact of the import ban. Second, farmgate prices are not available at the regional level. Nevertheless, results would likely be unchanged if farmgate prices were used instead, given the high correlation between farmgate, wholesale, and consumer prices over the period under study (see Figure E.4). Third, in some regions of Indonesia, up to 15% the price index is actually comprised of cassava and other tubers. This does not pose a problem here since prices of cassava and other tubers were stagnant over the period under study and hence should have little effect on the overall index.

[4] There are two other predictions less relevant to the first order discussion here. First, ϵ^m_v is decreasing in the Armington elasticity of substitution between domestic and imported rice. This elasticity should be quite homogenous across the country and relatively high (Warr (2008) estimates around 5) given that nearly all Indonesian rice production is of the Indica type which is the predominant variety produced in Southeast Asia and traded on world markets (Dawe, 2008). Second, ϵ^m_v is increasing in the consumer price elasticity of demand for (all) rice in the regions relevant to village v. Estimates from the mid-1990s suggest that the price elasticity of demand is approximately -0.45 on average across all regions of Indonesia (Friedman and Levinsohn, 2001). Most of the variation in this estimate occurs within rather than across regions as the wealthy can more readily substitute away from rice staples when prices rise. The slight exception is that in some of the Outer Islands, availability of cassava and other tubers allow greater substitution away from rice and hence higher demand elasticities.

E.3 Spatial and Time Series Properties of Rainfall and Rice Prices

An important feature of rice prices is their approximate unit root properties. This is demonstrated in Figure E.5 which plots the p-values from augmented Dickey and Fuller (1979) tests of the null hypothesis that the *domestic* rice price index in region c has a unit root (the different color dots correspond to alternative lag structures). Since rice prices across Indonesian cities are not independent, I also apply the heterogeneous panel unit root tests of Im et al. (2003) and Fisher's meta-analytic test, and in both cases, I fail to reject reject the null hypothesis that rice prices follow a unit root in all cities. Recognizing further the possibility that the structural breaks in prices around late 2005 evident in Figure 2 in the paper might be mistaken for unit roots, I apply city-specific Zivot and Andrews (2002) unit root tests which allow for an endogenous break in both trends and intercepts. Doing so, I fail to reject the null of a unit root in 41 out of 44 cities and identify structural breaks between 2004m11 and 2006m4 for all but five cities.

Whereas rice prices tend to follow a unit root, rainfall levels are serially uncorrelated across seasons. Considering seasonal rainfall levels at the district level (adjusted for province-specific growing seasons) going back to 1953, I cannot reject the null hypothesis of covariance stationarity for any Indonesian district. Figure E.6 documents the spatial variation in the measures of cumulative rainfall shocks used in the paper.

E.4 Exogeneity of Price Shocks with Respect to Landholdings Distribution

In Table E.2, I rule out the concern that rice price shocks were more intense in regions with a greater mass of large landholders selling to the market. Columns 1 and 2 reveal a small, statistically and economically insignificant effect of the dispersion parameter, $\widehat{\lambda}_v$, on the growth in rice prices between 2002m1 and 2005m3. Columns 3 and 4 reveal similar results for the post period, 2005m4 through 2008m3. Finally, columns 5 and 6 consider the interperiod difference and recover similarly small and insignificant coefficients. In even-numbered columns, I control for rainfall shocks, which has little effect on the coefficient for $\widehat{\lambda}_v$. Note that in column 6, we find that positive rainfall shocks exert downward pressure on prices as expected. This comes entirely from the post-ban period when imports play little role in stabilizing prices, which are now more tightly connected to domestic weather shocks.

E.5 Effect of Rainfall and Rice Price Shocks on GDP, Wages, and Profits

Rainfall has a strong positive relationship with time-varying agricultural productivity. Using a panel of district-level agricultural GDP from 2000–10, I estimate an elasticity of agricultural GDP with respect to rainfall (in periods t and $t-1$) of around 0.15. This robust positive estimate is in line with results specific to rice output in Levine and Yang (2014) and Naylor et al. (2001).

As discussed in Section 3.5 of the paper, Table E.3 shows that household expenditures (as a proxy for permanent income) exhibit an elasticity around 0.25 with respect to rainfall shocks and an elasticity around 1 with respect to rice price shocks when instrumenting using the policy variation in Table E.1.[5] Interestingly, the elasticity for price shocks is much smaller when not exploiting the persistent policy shock as an instrument. Table E.4 meanwhile shows that agricultural wages (although noisy) are increasing after positive rainfall and rice price shocks.[6]

In Section 6, I note back of the envelope calculations for the increase in gross profits per harvest caused by the rice price shock. Here, I provide background on those calculations, which were based on village-specific profit margins. First, using data from Timmer (2008), I take the rice price in Jakarta to be 3,000 Rupiah per kilogram in March 2005. Second, using the local rice price indices from March 2005 and March 2008, I back out the price in Rupiah terms after reindexing to the value of the index in Jakarta in each of those months. Third, I apply a

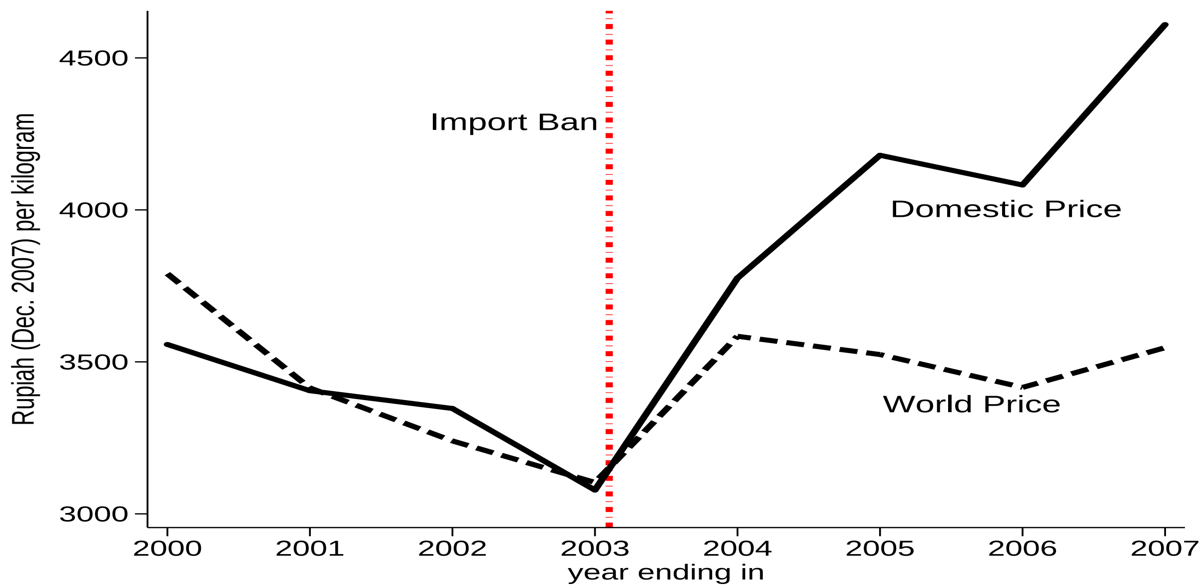
[5]These estimates are based on a district-level panel of average household expenditures per capita constructed from *Susenas* household survey conducted in July of every year and representative at that district level after applying probability weights.

[6]These estimates are also based on the *Susenas* survey data. However, the reporting structure of wages changed over the sample period and hence it is not appropriate to construct a district-level panel as was done with expenditures.

measure of village-specific total paddy output (in kilograms) per hectare in 2001 (see Appendix C) to all potential farmers in the village. Although unit-level productivity varies varies across households, the bulk of this variation is across rather than within villages and hence little information is lost in focusing on productivity differences arising purely from land area planted (see Bazzi, 2012b). Fourth, I convert wet paddy output to marketable rice output using a standard conversion factor of 0.55. Fifth, I convert Rupiah to USD at an exchange rate prevailing in late 2005 of 10,000 Rupiah to 1 USD. Finally, when accounting for own consumption, I assume that the household has two harvests per year and subtract 520 kilograms of rice (the recommended intake for a family of four) valued at the market price. Although I only reported the income boost for farmers with 0.25 and 0.75 Ha of landholdings, estimates for other landholding sizes are available upon request.

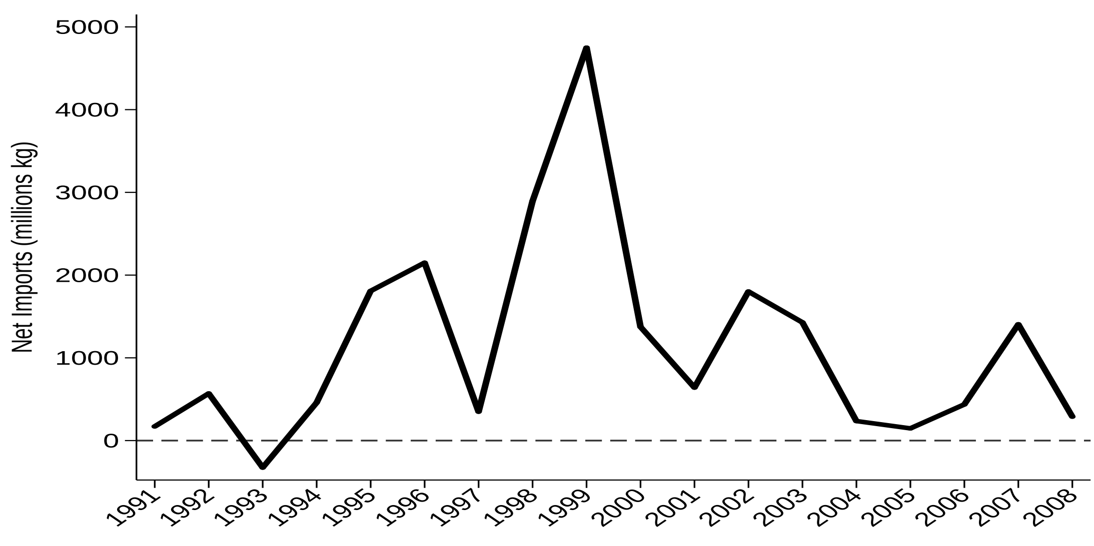
Figures

Figure E.1: World vs. Domestic Rice Prices (year-end)



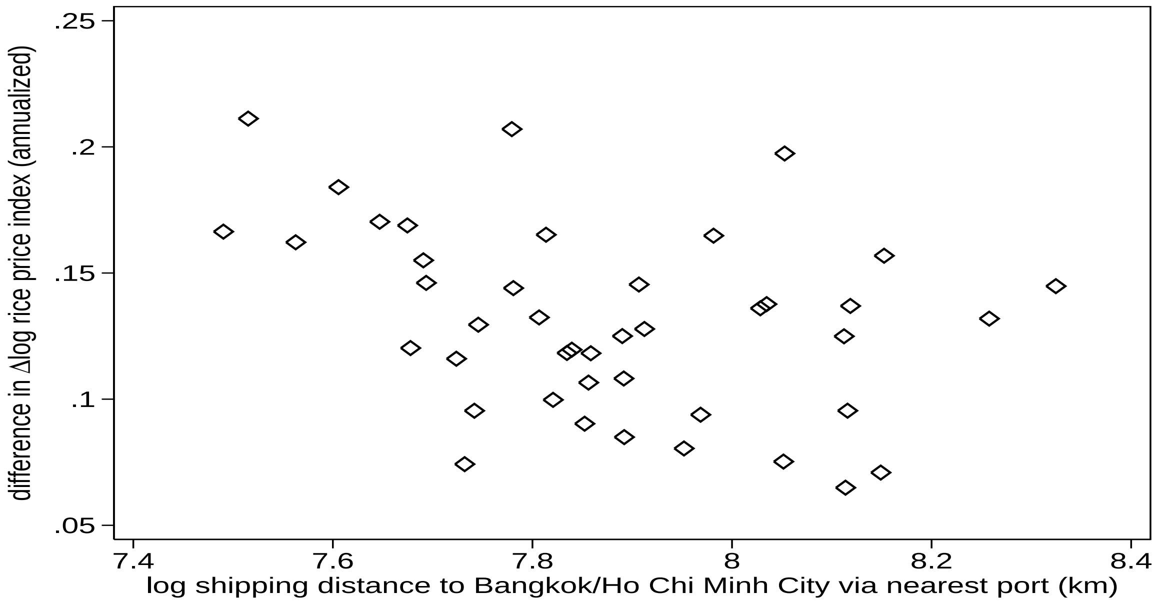
Notes: Year-end average farmgate/producer prices from 2000 to 2007 across Indonesia reported by the Food and Agriculture Organization (FAO). Nominal prices are deflated by the national CPI reported by Bank of Indonesia. Exchange rate and world price data are obtained from the IMF. Further adjustments are made as suggested in Dawe (2008): Thai 100B f.o.b. adjusted to retail level by USD 20 per ton and 10% markup from wholesale to retail, adjusted downward for quality by 20% from 1991-2000 and by 10% from 2001-2007 based on trends in quality preferences in the world market.

Figure E.2: Net Rice Imports in Indonesia, 1991-2008



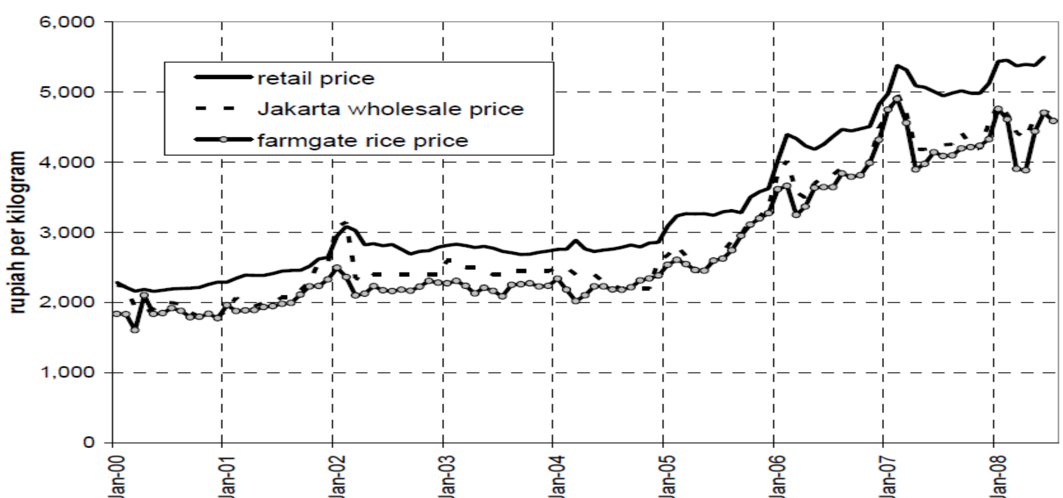
Notes: Data obtained from SITC Rev. 2 in the Comtrade-UN database on 5 December 2010. All categories of rice products are included in the figure. The uptick in 2007 is almost entirely due to emergency imports by the government logistical agency, Bulog.

Figure E.3: Spatial Variation in Rice Price Changes Before and After the Import Ban



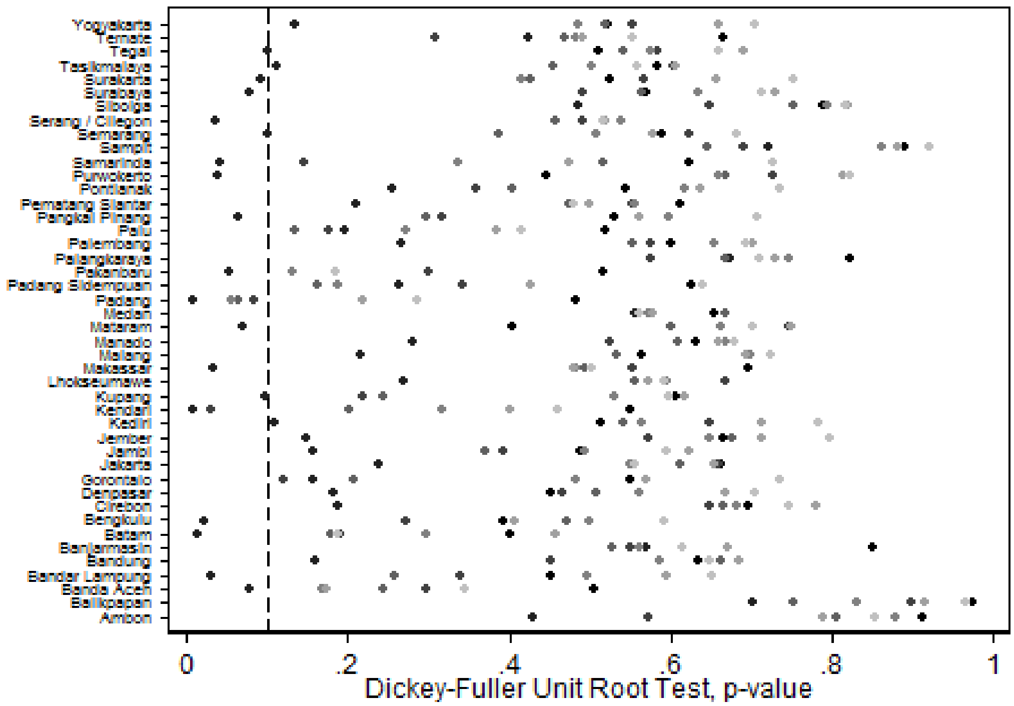
Notes: This figure demonstrates that rice prices grew faster in port cities closer to Bangkok and Ho Chi Minh City after the ban on rice imports. Monthly rice prices obtained from Wimanda (2009). Distances calculated as the sum of (i) the travel distance from the village to the district capital reported in *Podes* 2005, (ii) the great circle distance from the given Indonesian city to the nearest port, and (iii) the average shipping distance from the given Indonesian port to the port in Bangkok and Ho Chi Minh City. The port cities and shipping distances are obtained from `http://e-ships.net`. For cities with ports, I take the distance from the centroid of the city to the exact latitude/longitude of the port.

Figure E.4: Nominal Wholesale/Retail/Farmgate Prices (monthly 2000-2008)



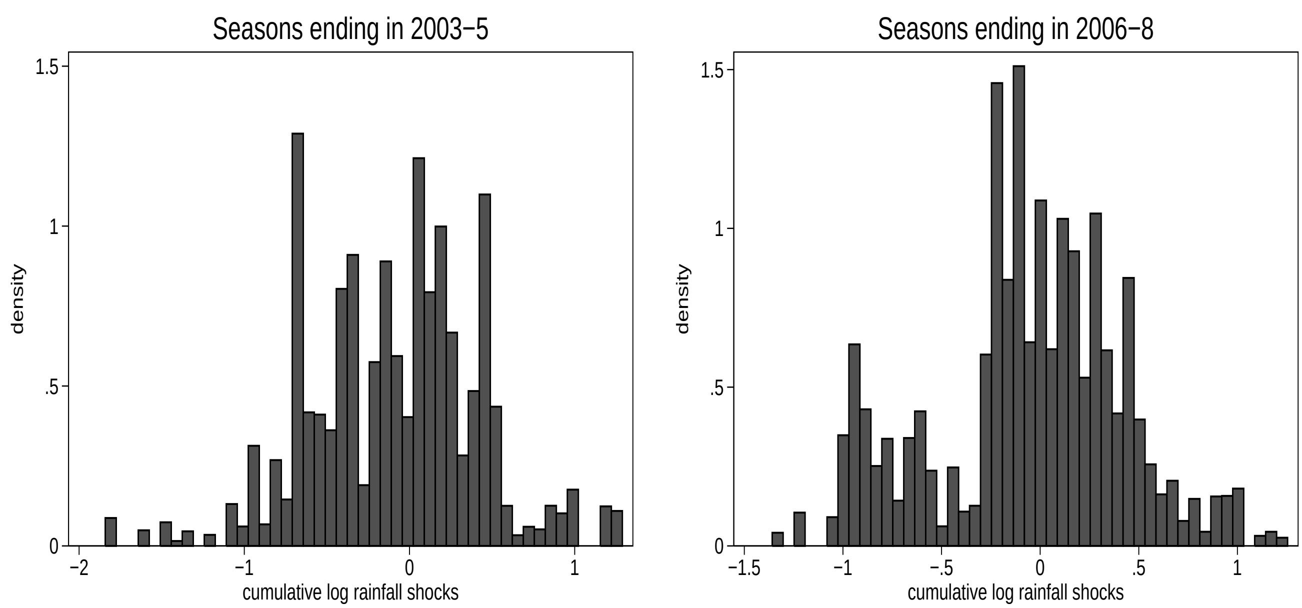
Notes: Prices from January 2000 to January 2008. Farmgate price quoted in terms of wet paddy. After drying and milling, 100 kg of wet paddy produces 55 kg of rice. The figure is reproduced from Timmer (2008), but the original source is Peter B. Rosner from Ministry of Trade and Central Bureau of Statistics (BPS).

Figure E.5: Domestic Rice Prices Follow a Unit Root Process



Notes: The monthly rice price index is from Wimanda (2009). The circles indicate p-values from augmented Dickey-Fuller unit roots for each of the 44 cities where the colors are lightest for those tests based on a larger number of lags. Using the more robust panel unit root test of Im et al. (2003), I additionally fail to reject the null hypothesis that all panels contain a unit root.

Figure E.6: Rainfall Shocks Across the Indonesian Archipelago, 2002-2008



Notes: The histograms show the spatial incidence of cumulative rainfall shocks over the growing seasons ending in 2003-5 and 2006-8. The shocks corresponding to each year are defined as the log difference between the given village's rainfall (measured at the district level) in the province-specific rice growing season and the long-run district-level mean rainfall excluding the given season from 1953-2008. Further details on the time series properties of rainfall can be found in Online Appendix E.

Tables

Table E.1: Predetermined Import Exposure and Spatial Variation in Rice Prices

	Dependent Variable: Δlog price, t					
	(1)	(2)	(3)	(4)	(5)	(6)
log shipping distance to THA/VNM	0.0086		0.0092		0.0023	
	(0.0027)***		(0.0054)*		(0.0139)	
log shipping distance to THA/VNM	-0.0073	-0.0102	-0.0214	-0.0239	-0.0320	-0.0360
$\times$ $1(year \geq 2004)$	(0.0039)*	(0.0060)*	(0.0101)**	(0.0116)**	(0.0121)**	(0.0126)***
log distance to emigration center	0.0003		0.0006		-0.0012	
	(0.0002)		(0.0005)		(0.0015)	
log distance to emigration center	-0.0006	-0.0007	-0.0002	-0.0003	-0.0013	-0.0012
$\times$ $1(year \geq 2004)$	(0.0003)*	(0.0006)	(0.0009)	(0.0010)	(0.0011)	(0.0012)
log price, $t-1$	-0.0287	-0.0780	-0.1035	-0.1352	-0.1592	-0.2014
	(0.0088)***	(0.0117)***	(0.0097)***	(0.0121)***	(0.0138)***	(0.0195)***
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
log monthly rainfall, $t, \ldots, t-12$	Yes	Yes	Yes	Yes	Yes	Yes
City Fixed Effects	No	Yes	No	Yes	No	Yes
City-Specific Linear Time Trend	No	No	Yes	Yes	Yes	Yes
City-Specific Quadratic Time Trend	No	No	No	No	Yes	Yes
Observations	2,606	2,606	2,606	2,606	2,606	2,606
R^2	0.5051	0.5195	0.5286	0.5390	0.5489	0.5596

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; Regressions based on monthly price series for major cities across Indonesia. Shipping distance is the sum of overland and sea-based travel to the major port cities in Thailand and Vietnam (taking the simple average of the two major rice exporters). Rainfall is calculated as the nearest observed monthly precipitation level. Standard errors are clustered by city. Time FE are month$\times$year.

Table E.2: Rice Price Shocks and the Landholdings Distribution

	Dependent Variable:					
	price shock (pre) 2002m1–2005m3		price shock (post) 2005m4–2008m3		Δ price shock post-pre	
	(1)	(2)	(3)	(4)	(5)	(6)
Pareto exponent $\widehat{\lambda}$	0.0052	0.0052	-0.0001	-0.0005	-0.0060	-0.0053
	(0.0046)	(0.0045)	(0.0050)	(0.0050)	(0.0086)	(0.0090)
(Δ) rainfall shock		0.0016		-0.0042		-0.0392
		(0.0035)		(0.0037)		(0.0077)***
Observations	382	382	382	382	382	382

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; District-level regressions with rice price shocks as the dependent variable where the shocks are annualized log growth rates between (columns 1/2) 2002m1–2005m3, (3/4) 2005m4–2008m3, and (5/6) the interperiod difference. *rainfall shock* is the cumulative annual log deviation from the district's long-run mean rainfall in the growing seasons ending in (columns 1/2) 2002–2005, (3/4) 2006–2008, and (5/6) the interperiod difference. *rice price shock* is the respective The estimated Pareto exponent $\widehat{\lambda}_v$ is obtained for wetland area in 2002 and averaged across all villages with population weights corresponding to total landholders in the village; higher values indicate less dispersion in landholding sizes. Results are unchanged for other types of landholdings and for λ estimated at the district-level. Robust standard errors in parentheses.

Table E.3: Rice Price Shocks, Rainfall Shocks, and Expenditures

	Dependent Variable: Δ log district avg. HH expenditures/capita			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Sample Restriction	Main Source of Income			
	Any Sector		Agriculture	
Δ price shock	0.1232	0.4235	0.1000	1.2920
	(0.0727)*	(0.3476)	(0.0917)	(0.5722)**
Δ rainfall shock	0.2363	0.2506	0.2547	0.3113
	(0.0432)***	(0.0470)***	(0.0645)***	(0.0719)***
Year FE	Yes	Yes	Yes	Yes
Observations	1,489	1,489	1,489	1,489
R^2	0.47	0.47	0.39	0.35
First Stage F Stat		11.00		11.00

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; First difference estimation on a district-level panel spanning 2002-7. The dependent variable is the average household expenditures per capita for all households (cols. 1-2) and households reporting agricultural activities (cols. 3-4) as their main source of income. The averages are representative at the district level based on sampling weights in the *Susenas* survey data. Columns 2 and 4 are estimated by IV using log shipping distance to THA/VNM $\times$ $1(\text{year} \geq 2004)$ as the instrument. Standard errors are clustered by district. Standard errors clustered at the district level.

Table E.4: Rice Price Shocks, Rainfall Shocks, and Wages

	Dependent Variable: log individual wages				
location sample restriction:	none	urban	rural	rural	rural
sectoral sample restriction:	none	none	none	non-agric.	agric.
	(1)	(2)	(3)	(4)	(5)
$\boldsymbol{\gamma}_1$: log annual rainfall	-0.1072	-0.1974	-0.0434	-0.0559	0.4677
	(0.0455)**	(0.0512)***	(0.0528)	(0.0517)	(0.2446)*
γ_2: log annual rainfall $\times$ $1(year \geq 2004)$	0.0935	0.1701	0.0266	0.0376	-0.1826
	(0.0423)**	(0.0448)***	(0.0460)	(0.0463)	(0.2253)
$\boldsymbol{\beta}_1$: Δ log rice price	-0.6161	-0.5680	-0.4966	-0.4060	-0.6623
	(0.1393)***	(0.1635)***	(0.1556)***	(0.1529)***	(0.5469)
β_2: Δ log rice price $\times$ $1(year \geq 2004)$	0.6338	0.6277	0.4482	0.3303	1.4103
	(0.1515)***	(0.1630)***	(0.1856)**	(0.1873)*	(0.5892)**
$\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$: rain shock, post	-0.014	-0.027	-0.048	-0.018	0.285
	(0.033)	(0.045)	(0.076)	(0.033)	(0.146)*
$\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$: price shock, post	0.018	0.060	-0.017	-0.075	0.748
	(0.059)	(0.080)	(0.035)	(0.078)	(0.250)***
Mincer Control Variables	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Number of Individuals	759,784	327,475	432,309	414,663	17,646
Number of Individuals (weighted)	286mn	108mn	177mn	170mn	7.8mn
R^2	0.28	0.24	0.27	0.27	0.47

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; Regressions estimated by weighted least squares with the dependent variable being log individual wages in Rupiah. The data span the years 2002-7 and the weights achieve representativeness at the district level. Wage reporting changed over the period and hence it is preferable to retain the individual-level observations rather than take a weighted average at the district level. Standard errors clustered at the district level.

F Further Empirical Results

This appendix presents several additional results beyond the main empirical analysis in the paper.

F.1 Heterogeneity in Land Quality

The main results in the paper are based on a measure of the landholdings distribution capturing rice growing in a planting season just prior to the ban on imports. In Table F.2, I show that the income elasticity of migration varies depending on the exposure of the given type of landholdings to rainfall and rice price variability. Column 1 reproduces the key estimates from column 4 in Table 5 of the paper. Column 2 shows that the heterogeneous effect of rainfall shocks ($\theta_{a\lambda}$) is muted and statistically insignificant for $\widehat{\lambda}_v$ based on wetland holdings. Compared to columns 1 and 3, this specification focuses on households (and villages) with landholdings that are relatively less reliant on rainfall.[1] The second takeaway is that the heterogeneous effect of price shocks ($\theta_{p\lambda}$) is considerably muted for $\widehat{\lambda}_v$ in column 3 based on landholdings used to grow crops besides rice. Compared to columns 1 and 2, this specification of λ_v includes non-rice producing households (and villages) for whom rice price increases have a negative effect on real income. Overall, the patterns in Table F.2 are consistent with the sort of passthrough from individual income shocks to aggregate migration suggested by the model.

As a model validation exercise, I use the quasi-structural interpretation of $\theta_{p\lambda} = 1/\beta$ in equation (4) in the paper to estimate a Cobb-Douglas coefficient on paddy landholdings that is consistent with results in the literature. The estimate in column 2 implies $\widehat{\beta} = 0.52$. In Bazzi (2012b), I estimate $\widetilde{\beta} = 0.55$ using auxiliary household survey data on wetland holdings and rice output from 2004. Using the delta method, I reject at the 95% level that $\widehat{\beta} = 0.34$, the lowest estimate of β in the literature (Fuglie, 2010), but I cannot reject that $\widehat{\beta} = 0.69$, the largest available estimate (Mundlak et al., 2004).[2]

F.2 Controlling for Agricultural Wage Shocks

Adding agricultural wage shocks to the baseline estimating equation should dampen the elasticities with respect to rainfall and rice price shocks if some of the effects of those shocks are operating through changes in the returns to (rice) farming. Table F.3 shows that this is indeed the case. I construct agricultural wage shocks as the difference in the growth of log average agricultural wages between 2002–5 and 2006–8 at the district level where the averages are representative from the annual *Susenas* household survey data. The elasticities on rainfall and rice price shocks fall slightly relative to the baseline in Table 5 in the paper. However, the key qualitative and quantitative results remain largely unchanged.

F.3 Assessing Exclusion Restrictions

Table F.5 varies the exclusion restrictions employed in estimating the key parameters $\widetilde{\Theta} \equiv (\theta_a, \theta_p, \theta_{a\lambda}, \theta_{p\lambda})$ using the Das et al. (2003) procedure and $\widehat{\lambda}_v$ based on wetland holdings. With four instruments and two first stage equations, I can assess the effect of treating at most two instruments as non-excludable. In the table, I compare baseline estimates of $\widetilde{\Theta}$ in columns 1-3 to estimates when including (i) the log number of villages in v's district (columns 4-6), (ii) the log number of villages in v's district and the log area of v's district less v (columns 7-9), (iii)

[1] Although rice is most productively grown on wetland, it is also grown across Indonesia on rainfed dry land. The measure of paddy area planted used in the main estimate of λ_v in column 1 captures rice grown on both land types. The measure used in column 3 comprises all land types used to grow a range of crops, many of which are rainfed.

[2] Of course, this exercise delivers less favorable results when using the baseline measure of $\widehat{\lambda}_v$ based on paddy area planted in column 1. However, two factors argue for taking the estimate based on wetland holdings: (i) planting in the pre-ban growing periods may not capture rice-growing potential given the possibility of crop switching, and (ii) attenuation bias due to classical measurement (or data entry) error in the Agricultural Census.

the log population of v (columns 10-12), (iv) the log population of v and area in v's district less v (columns 13-15), and (v) the log population of v and v's district less v (columns 16-18). Except for a few insignificant differences, I find no systematic departures from the baseline results in Table 5.

F.4 Alternative Specifications of Rainfall and Rice Price Shocks

Table F.6 shows that the primary conclusions regarding the effects of rainfall shocks are robust to the inclusion of period-specific shocks rather than the difference in shocks between t and $t-1$. Furthermore, I fail to reject that the coefficient on the rainfall shock in t equals the absolute value of the coefficient on the rainfall shock in $t-1$.

Table F.7 considers an alternative specification for rainfall shocks in which the annual shocks are fully elaborated from 2002–8 (i.e., the rainfall shock in each season s is assigned its own elasticity parameter θ_{as} for $200s = 3, \ldots, 8$ and $\theta_{as\lambda}$ for the interactions with $\widehat{\lambda}_v$). At the bottom of the table, I report the sum of coefficients for period t ($s = 3, 4, 5$), period $t-1$ ($s = 6, 7, 8$), and both t and $t-1$ ($s = 3, \ldots, 8$) as well as the associated p-value for the null hypothesis that the given sum equals zero. In columns 1-3, we draw the same conclusions as in Table F.6: (i) the sum of period t ($t-1$) rainfall shocks is positive (negative) and statistically significant, and (ii) the null hypothesis that $\theta_{a3} + \theta_{a4} + \theta_{a5} = -(\theta_{a6} + \theta_{a7} + \theta_{a8})$ cannot be rejected. Furthermore, in columns 4-6, we similarly rule out the possibility that the baseline specification of rainfall shocks leads to spurious conclusions regarding the key elasticity parameter $\theta_{a\lambda}$. That is, the sum of period t ($t-1$) coefficients on the interaction of rainfall shocks and $\widehat{\lambda}_v$ are positive (negative) and statistically significant.[3] In unreported results, I also show that the main results are robust to allowing negative rainfall shocks to have a different effect than positive rainfall shocks (i.e., rather than using a single continuous measure crossing zero).

Table F.8 presents alternative approaches to measuring the rice price shock. Columns 1-4 report estimates of θ_p and $\theta_{p\lambda}$ using $\widehat{\lambda}_v$ for wetland holdings. In columns 5-8, I specify the price "shock" as a difference in log *average* prices over 2005m4-2008m3 and 2002m2-2005m3 rather than a difference in annualized log growth rates between those two periods. This specification yields similar results. In columns 9-12, I adopt insights from the model for rice prices developed in Appendix E.1. Because the model predicts that the price shock should be *decreasing* in distance from port cities in Indonesia and the shipping routes to Thailand and Vietnam, a negative coefficient on the two distance terms would be consistent with a positive elasticity of migration flows with respect to rice price shocks. Columns 9-10 are consistent with this hypothesized relationship as are the negative coefficients on the interaction terms with $\widehat{\lambda}_v$ in columns 11-12. These results are effectively the reduced form of the IV results in Table F.4.

F.5 Alternative Choices of the Pareto Lower Bound

Although the λ_v parameters should be unaffected by the location of $\underline{R}$, in practice, the Pareto distribution is only an approximation, which works better in some villages than others (see Appendix D). In Table F.9, I show that the estimates of the key elasticity parameters generally do not change when imposing alternative $\underline{R} \in \{0.15, 0.2, 0.25\}$ in the estimation of λ_v (and the share of households above $\underline{R}$) for wetland holdings. The results are similar for λ_v estimated using total agricultural landholdings or paddy area planted in 2002.[4]

[3] An interesting feature of the fully elaborated specification is that the s and $s-1$ coefficients alternative in sign, with the period s contemporaneous with the *Podes* enumeration dates in 2005 and 2008 being positive. Two factors might explain this pattern: (i) the mean reverting properties of rainfall (see Appendix E.5), and/or (ii) a particular spatial distribution of two-year migration contract cycles. Nevertheless, the cumulative migration flows are what we observe in the data and hence the sum is what matters, not the individual years per se.

[4] Note that the sample sizes differ across columns because consistent (i.e., usable) estimates of λ_v require at least 3 distinct size measures above $\underline{R}^*$. Some villages do not satisfy this criteria for a given minimum threshold value and landholding type.

F.6 Accounting for Village Demographic Structure and Past Internal Migration

Table F.10 demonstrates that the main results are robust to controlling for (i) the share of the population that lived outside the village in 1995, (ii) the share of population aged 15-29, and (iii) the average household size in the village—each drawn from the 2000 Population Census. Variable (i) proxies for potential prior experience in and network connections to domestic labor markets outside the village. Variable (ii) captures to some extent labor market pressures induced by Indonesia's relatively recent demographic transition. Moreover, individuals within that given cohort are the most likely to have been potential migrants beginning 3-7 years later and hence recorded in 2005 and 2008 migrant stocks.[5] Although highly correlated with mean village income, mean household size also picks up variation in household labor supply, which may in turn affect the robustness of agricultural labor markets (i.e., off own-farm) and the capacity of households to diversify labor allocation across borders—both of which could have direct effects on flow migration rates.

F.7 On the (Non-)Effect of Measurement and Reporting Outliers

Tables F.11 and F.12 demonstrate that the key estimates of $\widetilde{\Theta} \equiv (\theta_a, \theta_p, \theta_{a\lambda}, \theta_{p\lambda})$ in the paper are robust to and arguably strengthened by accounting for outliers in the data along a few important dimensions. Column 2 controls for the frequency with which the village updates its population register (see Appendix C). This helps account for some of the measurement error in migration rates as well as potential misclassification bias arising from villages reporting no migrants when in fact there is at least one migrant from the village. Column 3 trims the bottom and top 1 percent of $\widehat{\lambda}_v$. Column 4 removes villages subject to censoring in reported migrant stocks in 2005 and/or 2008.[6] In column 5, I retain only those villages for which I did not have to rely on any fuzzy matching algorithms for merging villages across the 2005 and 2008 waves of *Podes* (see Section H). Although I have confidence in the matching algorithms, they may contribute to measurement error on both sides of the estimating equation. Last, column 6 simultaneously implements the prior four restrictions. In all cases, the main qualitative and quantitative interpretation of Θ remains unchanged.

In column 7 of Tables F.11 and F.12, I drop provinces identified in Bank Indonesia (2009) as having a large number of undocumented international migrants (primarily going to Malaysia). The Village Potential data, recall from Section 3.2, define international labor migrants as those working abroad for a fixed wage and time period. It is possible therefore that this count includes some undocumented migrants for which the determinants of migration choice and the nature of liquidity constraints may be somewhat different than for legal migrants. When dropping these provinces—which, keep in mind, still have a large number of legal international migrants—a few differences emerge with respect to the full sample results. First, in Table F.11, the elasticity parameters for rainfall and price shocks slightly increase. However, the estimates of $\theta_{a\lambda}$ and $\theta_{p\lambda}$ in column 7 fall in magnitude. The large, precisely estimated $\theta_{p\lambda}$ for $\widehat{\lambda}_v$ based on wetland holdings disappears entirely. It seems, then, that undocumented migrants may explain some of the stronger response of migration flows to price shocks in villages with a greater mass of small landholders.

F.8 Rainfall Shocks and Internal Migration

Here, I briefly discuss the effect of rainfall shocks on internal migration flows. Using weighted samples from Indonesian Population Censuses in 2000 and 2010 as well as Intercensal Population Surveys in 1985, 1995, and

[5] However, inclusion of this variable might introduce a source of bias in that villages with a large share of aged 15-29 in 2000 may be precisely those villages for which (i) the Asian financial crisis of 1997-8 led to a large return migration from urban areas, and/or (ii) the local economy was (expected to be) thriving as global agricultural commodity prices remained high through the early 2000s.

[6] The 2005 survey records separately the total number of male migrants and the total number of female migrants whereas the 2008 survey simply records the total number of migrants. Whether the different format of the question across years biases reporting is an open question. However, top coding poses a challenge in the following sense. In 2005, the separate reporting for male and female migrants allowed total migrant stocks to exceed 998 persons for 40 villages while villages could only record a maximum of 998 persons abroad in the 2008 survey.

2005, I am able to construct a bilateral district-level migration matrix in which each observation comprises the stock of individuals hailing from origin district o in year $t-5$ and currently residing in destination district d in year t.[7] I estimate the following quasi-gravity model for internal (h for home) migration flows as a function of origin and destination rainfall shocks:

$$\ln migrants^{h}_{odt} = \alpha rainfall\ shock_{ot} + \beta rainfall\ shock_{dt} + \upsilon_o + \upsilon_d + \upsilon_t + \epsilon_{odt}. \tag{F.1}$$

where, for $j = o, d$, $rainfall\ shock_{jt}$ captures (in logarithmic form) the cumulative annual rainfall shocks over the four years prior to t,[8] υ_j are geographic fixed effects, υ_t is a year fixed effect, ϵ_{odt} is an idiosyncratic error term.[9]

Estimating equation (F.1) by OLS for the entire period 1985-2010, I find $\widehat{\alpha} \approx -0.056$ (std. error of 0.022), which suggests that origin rainfall shocks reduce internal out-migration. Restricting to the period 2005-2010—roughly, the period over which I observe international migrants in the Village Potential data used in the paper—I obtain $\widehat{\alpha} \approx -0.452$ (std. error of 0.071).[10] (In both cases, I also find that $\beta > 0$ and statistically significant, which is consistent with migration being responsive to destination wage shocks.) Taken together, the negative estimates of α support the claim that positive rainfall shocks increase district population size and hence are likely also to increase village population size, presuming (i) inter-district migration is a lower bound for overall internal out-migration observed at the village level, and (ii) intra-district migration outside the home village follows similar processes. Such upward pressure on village population size in the denominator of the dependent variable in the paper (Δlog migrants/population) implies that the positive relationship between changes in *international* migration rates and rainfall cannot be explained by the unobservable internal migration flows at the village level.

F.9 Further Background on the Validation Exercise Using Micro Data

In the paper, I discuss results from estimating a migration choice model and using the implied marginal effects to recover an alternative measure of the village-level elasticity of flow migration rates with respect to income shocks. In this brief subsection, I provide a few additional details on the analysis therein.

First, note that in columns 3-4 of Table 2 in the paper, I report coefficient estimates from the following equation

$$\begin{aligned} migrate_{iv,t+1} \quad = \quad & \alpha + \beta\, rainfall\ shock_{vt} + \gamma\, price\ shock_{vt} + rainfall\ shock_{vt} \times (land_i \zeta^a_1 + land^2_i \zeta^a_2) \\ & + price\ shock_{vt} \times (land_i \zeta^p_1 + land^2_i \zeta^p_2) + \psi_i + \ \psi_t + \ e_{iv,t+1}, \end{aligned}$$

which, recall, I estimate using a conditional fixed effects (CFE) logit estimator, and where (i) $land_i$ comprises all landholdings owned, under rental, or rented out and used to grow rice, and (ii) column 3 imposes $\zeta^a_2 = 0$ and $\zeta^p_2 = 0$. Using these estimates, I then recover average marginal effects (AMEs) at each value of $land_i \in \{0.1, 0.2, \ldots, 2.5\}$ Ha, where (i) 2.5 Ha is the maximum in the sample, and (ii) and the calculation of AMEs requires imposing $\eta_i = 0\ \forall i$. Thus, we obtain AMEs for both rainfall and rice price shocks at each landholding size (at 0.1 Ha increments).[11]

I use these individual-level AMEs to construct implied aggregate village-level elasticities of migration rates

[7] The data were downloaded from the Integrated Public Use Microdata Series, International in August 2012. The district-level total migrant and population counts are based on summing the person-specific population weights provided by IPUMS-International and representative at the district-level. Details on the (Inter-)Census specific samples can be obtained on the IPUMS website for Indonesia. Further details on the panel construction are available upon request.

[8] For example, the shock in for origin district k in 2005 is simply the sum of the annual log deviations in 2001-2004 from the long-run district-level mean calculated over all years from 1948-2010 excluding 2001-2004.

[9] I use the log number of migrants rather than the migrant share or the odds of migration quite simply because the goal is to characterize changes in district population levels arising from internal migration (i.e., the denominator in the dependent variable in the paper).

[10] I cluster standard errors by origin×destination district pair. Standard errors increase slightly when using two-way clustering (Cameron et al., 2011) on both origin and destination district.

[11] Recall that the estimates are quantitatively similar when using the less-biased LPM approach to estimating AMEs with household fixed effects.

with respect to income shocks. I do so by applying the population shares to each landholding size-specific AME as implied by the village-level Pareto distribution. Consider, for example, the AMEs for rainfall shocks at landholding sizes 0.3 and 0.4 Ha. For each village v, I reweight the average of these two AMEs by the share of the population with landholding sizes$\in[0.3, 0.4]$ Ha as implied by the Pareto exponent $\widehat{\lambda}_v$.[12] I repeat this over all increments of landholding sizes in the village, apply the AME at 2.5 Ha to all households above 2.5 Ha (as implied by $\widehat{\lambda}_v$), and then sum the reweighted AMEs to recover an aggregate village-level elasticity. In Table 8, I then compared these implied elasticities to those from the actual village-level regressions in Table 5, which allowed the effect of income shocks on flow migration rates to vary with $\widehat{\lambda}_v$.

In recovering the elasticities of flow migration rates with respect to price and rainfall shocks, I take the baseline coefficient estimates of $\widetilde{\Theta} \equiv (\theta_a, \theta_p, \theta_{a\lambda}, \theta_{p\lambda})$ in column 2 of Table F.2 for $\widehat{\lambda}_v$ based on wetland holdings. I then assign to village v the average marginal effects of the price shock for all villages with $\widehat{\lambda}_v$ in the same percentile. That is, I calculate the average marginal effects of income shocks at each percentile of the distribution of $\widehat{\lambda}_v$ in the second-step sample of villages. Following this procedure makes it possible to compare the village-level elasticities with analogous elasticities recovered from an underlying migration choice model.

F.10 Further Background on the Estimation of Village-Specific Migration Costs

Having found strong empirical evidence of financial constraints to migration in testing the theory, I used the following structural equation (4) for the log flow migration rate to back out estimates of the migration costs:

$$\Delta \ln\left(\frac{M_{v,t+1}}{N_{v,t+1}}\right) = \frac{\lambda_v}{\beta}\Delta \ln p_{vt} + \Delta \ln\left[\left(\frac{\overline{\sigma}_v + a_{vt}}{\tau_{vj} C_{vj,t+1}}\right)^{\frac{\lambda_v}{\beta}} - \left(\frac{\overline{\sigma}_v \alpha_v \chi}{W_{vj,t+1} - C_{vj,t+1}}\right)^{\frac{\lambda_v}{\beta}}\right].$$

Here, I provide a few additional details on the calculation of these village-specific migration costs not mentioned in the paper.

First, I plug in the empirical analogues for rice prices and rainfall. I specify $\Delta \ln p_{vt}$ in the above equation as the log difference in the local rice price index over the entire period, 2002m1-2008m3. I set the rainfall level, $\sigma_{vt} \equiv \overline{\sigma}_v + a_{vt}$, equal to the average of the annual seasonal rainfall levels (in centimeters) over the three seasons prior to mid-2008 (mid-2005 for $\sigma_{v,t-1}$). I set $\overline{\sigma}_v$ equal to the average annual seasonal rainfall levels (in centimeters) over the 55 year period 1953-2008. The rainfall shocks a_{vt} capture the empirical difference between σ_{vt} and $\overline{\sigma}_v$.

Second, the prevailing wage offers W_{vjt} are calculated as follows. For villages with any migrants in 2005, W_{vjt} equals the two-year wage offered to Indonesians around 2005 in the plurality destination of migrants from that village. The wage in $t+1$ equals the two-year gross wage in 2008 in that same destination. For villages with no migrants in 2005, W_{vjt} equals the average among villages with any migrants in their district. Bank Indonesia (2009) and other available sources report the monthly wages for low-skill Indonesian workers in each of the destination countries as stipulated in bilateral Memoranda of Understanding and reported by recruiters. These typical wages fall between the very narrow range of actual wages received as reported by migrants in the Bank Indonesia (2009) survey. Wages increased in early 2007 for most of the plurality destinations in the Village Potential 2005 data, and for those that do not, I nevertheless increase the wages by 10 percent. The results are robust to other choices.

Plugging in the relevant empirical data into the above equation, I then solve for the fixed migration costs C_{vjt}. Obtaining an analytic solution, however, requires a few additional simplifications. First, I assume that migration costs are constant across periods. This assumption is conservative insomuch as migration costs likely

[12] One could also imagine reweighting nonparametrically by applying the observed shares in the Agricultural Census. The approach based on $\widehat{\lambda}_v$ is more consistent with the testable implications of the theoretical model and is moreover necessary for the purposes of comparison with the village-level elasticities of income shocks that vary with $\widehat{\lambda}_v$.

fell in response to (i) competitive pressures in the recruitment industry and (ii) improvements in transportation infrastructure including the addition of new legal emigrant processing centers in a few provinces. Second, I impose $W_{vjt} = W_{vj,t+1} = \widehat{W}_{vj}$, and I set $\widehat{W}_{vj}$ to be the average of the empirical wages across both periods.

F.11 What Role for Policy, Recruiters, and Networks?

Reductions in (upfront) costs can make it easier for poorer households and regions to access international labor markets even in the absence of large increases in own ability to finance. I use the estimated migration costs to provide two suggestive pieces of evidence on how intermediaries can reduce costs and dampen the income elasticity of migration.

Ethnic Networks. Several studies have shown how diaspora networks facilitate migration in low-income settings. In Indonesia, there are several ethnic groups that have longstanding religious ties to the Middle East or a migratory legacy across the Southeast Asia region. For example, the religiously conservative Sundanese have strong ties to Saudi Arabia (de Jonge and Kaptein, 2002). Although their homeland lies in West Java, Sundanese can be found in villages across the Indonesian archipelago as a result of several decades of inter-island population resettlement beginning in the 1950s (see Bazzi et al., forthcoming). The Buginese meanwhile hail from South Sulawesi but have a long seafaring history that led to a number of Buginese communities across Indonesia but also in Malaysia and Singapore (Lineton, 1975).

Here, I show that these longstanding Sundanese and Buginese networks may have reduced the fixed costs of *international* migration from rural areas of Indonesia. Although these ethnic communities may be isolated from native ethnic populations when residing outside their historical homelands, they still have strong ties to the broader Sundanese and Buginese networks with the potential for connections to international labor markets. I exploit this identifying variation by regressing the estimated migration cost (in USD) for village v on (i) indicators for whether the village has an ethnic Bugis or ethnic Sunda majority, and (ii) origin region and destination country fixed effects. The results succinctly summarized in the following equation,

$$\widehat{\text{cost}} = \underset{(28)^{***}}{2390}\,\text{other ethnic majority} - \underset{(133)^{***}}{419}\,\text{Bugis majority} - \underset{(49)^{***}}{167}\,\text{Sunda majority},$$

suggest that informal network intermediaries may have reduced the costs of migration over time.

Recruiters. The second way in which the poor are able to access costly migration opportunities is through recruiters, and in Indonesia these recruiters typically target female migrants. Hence, costs should be lower in villages where female migrants are the dominant participants in international labor markets. A simple regression analogous to the previous one and summarized as follows

$$\widehat{\text{cost}} = \underset{(31)^{***}}{2331}\,\text{only female migrants} + \underset{(88)^{***}}{269}\,\text{only male migrants} - \underset{(28)}{29}\,\text{female and male migrants},$$

suggests that migration costs are around 10 percent higher in villages with only male migrants compared to villages with only female migrants at baseline.

Although merely suggestive correlations, these results show that similar to other important migration channels today such as contemporary Mexico to the U.S. (McKenzie and Rapoport, 2007; Munshi, 2003) and historical Norway to the U.S. (Abramitzky et al., 2013), migrant networks and intermediaries may play a crucial role in shaping the income elasticity of migration.

Tables

Table F.1: Non-Landholdings Wealth Heterogeneity and the Income Elasticity of Migration

	(1)	(2)	(3)
Δ rainfall shock	0.035	-0.235	-0.450
	(0.239)	(0.343)	(0.397)
Δ rainfall shock $\times$ $\widehat{\lambda}$(land)	0.149		0.120
	(0.073)**		(0.070)*
Δ rainfall shock $\times$ $\widehat{\lambda}$(expenditures)		0.191	0.165
		(0.097)	(0.096)*
Δ price shock	-1.945	0.013	-2.496
	(1.162)	(1.367)	(1.546)
Δ price shock $\times$ $\widehat{\lambda}$(land)	1.088		1.054
	(0.426)***		(0.431)**
Δ price shock $\times$ $\widehat{\lambda}$(expenditures)		0.229	0.156
		(0.359)	(0.374)
Number of Villages	24,486	24,486	24,486

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; Column 1 re-estimates the main heterogeneous effects estimate from column 4 in Table 5 in the paper on the slightly smaller sample of villages used in the latter columns, which introduce the non-land-based measure of wealth dispersion recovered from the Poverty Mapping exercise of predicting household expenditures per capita using household demographics, education levels, and assets from 2000 (Suryahadi et al., 2005). The two measures of λ are directly comprable, and higher values indicate less dispersion. See the notes to Table 5 for additional details on the specification. Standard errors are clustered at the district level, and the significance levels are based on the block bootstrap-t procedure described in Appendix B.

Table F.2: Heterogeneity in Landholdings Quality and the Income Elasticity

Landholdings Measure for $\widehat{\lambda}$:	Paddy Planted (1)	Wetland (sawah) (2)	Total Agricultural (3)
Δ rainfall shock	0.167	0.262	0.225
	(0.176)	(0.168)	(0.169)
Δ rainfall shock $\times\ \widehat{\lambda}$	0.140	0.028	0.119
	(0.065)**	(0.052)	(0.074)**
Δ price shock	-1.031	-2.234	-0.016
	(0.822)	(0.709)*	(0.688)
Δ price shock $\times\ \widehat{\lambda}$	1.116	1.913	0.267
	(0.423)***	(0.335)***	(0.329)
Number of Villages	24,855	24,537	26,527
Mean $\widehat{\lambda}$	1.60	1.73	1.53
Std. Dev. $\widehat{\lambda}$	0.49	0.54	0.44

Notes: Significance levels: $*: 10\%$ $\quad **: 5\%$ $\quad ***: 1\%$; The dependent variable in all specifications is Δ log (emigrants/total population) between 2005 and 2008 and has mean 0.11. Column 1 reproduces the estimate in column 4 of Table 5 in the paper. Subsequent columns retain the same specification and change the measures of $\widehat{\lambda}$ (and the share of households above 0.1 ha) to that based on the given type of landholdings noted at the top of the column; higher values indicate less dispersion in landholding sizes. All estimates based on the Das et al. (2003) semiparametric correction procedure. Standard errors are clustered at the district level, and the significance levels are based on the block bootstrap-t procedure described in Appendix B.

Table F.3: Incorporating Agricultural Wage Shocks

	(1)	(2)	(3)
Δ agricultural wage shock	0.073	0.072	0.068
	(0.048)	(0.048)	(0.048)
Δ rainfall shock	0.242	-0.019	0.021
	(0.145)	(0.178)	(0.182)
Δ price shock	0.725	0.719	-0.636
	(0.481)	(0.484)	(0.855)
Δ rainfall shock $\times\ \widehat{\lambda}$		0.157	0.139
		(0.057)***	(0.062)**
Δ price shock $\times\ \widehat{\lambda}$			0.791
			(0.431)***
Joint Significance of Correction Terms	$[< 0.001]$	$[< 0.001]$	$[< 0.001]$
Number of Villages	23,970	23,970	23,970

Notes: Significance levels: $*: 10\%$ $\quad **: 5\%$ $\quad ***: 1\%$; The table augments the baseline specification in columns 2-4 Table 5 in the paper with Δ agricultural wage shock, which is the difference in log average district-level agricultural wages from *Susenas* household surveys over the same time horizon as the rainfall shocks. The sample is slightly smaller than the baseline due to missing observations in *Susenas* for a few districts in the early 2000s; however, this sample change does not affect the coefficients. Δ rainfall shock is the difference in cumulative log deviations from long-run mean rainfall between the growing seasons ending in 2006-2008 and 2002-2005. Δ rice price shock is the difference in annualized log growth rates between 2005m4-2008m3 and 2002m1-2005m3. The estimate of λ_v is based on paddy area planted; higher values indicate less dispersion in landholding sizes. All columns are based on the Das et al. (2003) procedure and include a 3rd degree polynomial in the propensity scores for the extensive margin in 2005 and 2008. Standard errors are clustered at the district level, and the significance levels are based on the block bootstrap-t procedure described in Appendix B.

Table F.4: Two-Step Estimates Instrumenting for the Price Shock

Instrumenting the Price Shock	No	Yes
	(1)	(2)
Δ rainfall shock	0.262	-0.046
	(0.168)	(0.205)
Δ price shock	-2.234	-11.405
	(0.709)*	(5.378)**
Δ rainfall shock $\times\ \widehat{\lambda}$	0.028	0.119
	(0.052)	(0.103)**
Δ price shock $\times\ \widehat{\lambda}$	1.913	2.986
	(0.335)***	(2.767)*
Import Exposure Instruments	No	Yes
Kleibergen-Paap Underidentification Test [p-value]	—	0.037
Kleibergen-Paap Wald Stat	—	2.9
Angrist-Pischke F Stat, Δ price shock	—	4.0
Angrist-Pischke F Stat, Δ price shock $\times\ \widehat{\lambda}$	—	5.1
Hansen J Stat [p-value]	—	0.614
Number of Villages	24,537	24,537

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; The dependent variable in all specifications is Δ log (emigrants/total population) between 2005 and 2008 and has mean 0.11. Column 1 reproduces the estimate from column 2 of Table F.2. Instruments in column 2 include (i) shipping distance from the nearest port to the average of Bangkok and Ho Chi Minh City and its interaction with $\widehat{\lambda}$, and (ii) distance to the nearest international port and its interaction with $\widehat{\lambda}$. All estimates are based on the Das et al. (2003) semiparametric correction procedure. Standard errors are clustered at the district level, and the significance levels are based on the block bootstrap-t procedure described in Appendix B.

Table F.5: Relaxing Sets of Exclusion Restrictions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pareto exponent $\hat{\lambda}_v$	0.072	0.073	-0.113	0.080	0.082	-0.111	0.090	0.091	-0.109
	(0.018)***	(0.018)***	(0.030)***	(0.018)***	(0.018)***	(0.030)**	(0.019)***	(0.019)***	(0.031)**
Δ rainfall shock	1.287	1.304	-2.234	1.299	1.317	-2.343	1.376	1.397	-2.381
	(0.502)*	(0.503)*	(0.709)*	(0.513)	(0.515)	(0.724)	(0.507)	(0.509)	(0.734)
Δ price shock	0.305	0.162	0.262	0.328	0.150	0.260	0.346	0.165	0.277
	(0.134)**	(0.167)	(0.168)	(0.129)**	(0.165)	(0.166)	(0.128)**	(0.166)	(0.167)
$\hat{\lambda}_v \times \Delta$ rainfall shock		0.082	0.028		0.103	0.040		0.105	0.040
		(0.049)	(0.052)		(0.048)*	(0.052)		(0.048)*	(0.053)
$\hat{\lambda}_v \times \Delta$ price shock			1.913			1.973			2.031
			(0.335)***			(0.334)***			(0.344)***
log # villages in district				-0.073	-0.076	-0.065	-0.127	-0.131	-0.114
				(0.025)*	(0.025)**	(0.025)*	(0.034)	(0.034)*	(0.033)
log district area less v							0.057	0.058	0.052
							(0.025)	(0.025)	(0.024)

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Pareto exponent $\hat{\lambda}_v$	0.064	0.065	-0.108	0.064	0.065	-0.108	0.067	0.069	-0.108
	(0.018)**	(0.019)**	(0.030)***	(0.018)**	(0.019)**	(0.030)**	(0.019)**	(0.019)**	(0.030)**
Δ rainfall shock	1.163	1.183	-2.131	1.155	1.173	-2.134	1.160	1.178	-2.168
	(0.518)	(0.519)	(0.726)	(0.521)	(0.522)	(0.726)	(0.521)	(0.523)	(0.724)*
Δ price shock	0.271	0.123	0.217	0.271	0.120	0.215	0.276	0.116	0.216
	(0.132)*	(0.168)	(0.169)	(0.132)	(0.168)	(0.169)	(0.131)	(0.168)	(0.169)
$\hat{\lambda}_v \times \Delta$ rainfall shock		0.086	0.034		0.087	0.035		0.093	0.036
		(0.050)*	(0.052)		(0.049)	(0.052)		(0.049)	(0.052)
$\hat{\lambda}_v \times \Delta$ price shock			1.787			1.785			1.805
			(0.337)***			(0.338)***			(0.337)***
log village pop., t	0.089	0.089	0.086	0.090	0.090	0.087	0.076	0.073	0.078
	(0.028)*	(0.028)*	(0.027)	(0.027)	(0.027)	(0.027)	(0.035)	(0.036)	(0.034)
log district pop. less v, t							-0.025	-0.030	-0.016
							(0.038)	(0.038)	(0.038)
log district area less v				-0.006	-0.006	-0.005			
				(0.019)	(0.019)	(0.018)			
Number of villages	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; The table reports estimates of the key elasticity parameters sequentially relaxing one or two of the four baseline exclusion restrictions. The estimates in columns 1-3 correspond to baseline results for $\widehat{\lambda}_v$ obtained for wetland holdings. All estimates are based on the Das et al. (2003) semiparametric correction procedure and the measure of $\widehat{\lambda}_v$ for wetland holdings. The results are similar for parametric Poirier (1980) correction procedure and others types of landholdings. The dependent variable in all specifications is $\Delta \ln(M_{v,t+1}/N_{v,t+1})$ and has mean 0.11. Additional covariates in all specifications but not reported here include all those in Table 5. The correction terms are jointly statistically significant in all specifications. Standard errors are clustered at the district level, and significance levels are based on a block bootstrap-t procedure. Sample sizes are identical across sub-columns within the super-column, as reported at the bottom of the table.

Table F.6: Breaking Out Rainfall Shocks in Periods t and $t-1$

Correction Procedure	None	Semiparam.	Parametric
1st Stage Estimator	None	SU-LPM	BiProbit
	(1)	(2)	(3)
Δ rainfall shock	0.098	0.415	0.296
	(0.127)	(0.133)***	(0.128)**
	(4)	(5)	(6)
rainfall shock, t	0.159	0.407	0.309
	(0.127)	(0.132)***	(0.126)**
rainfall shock, $t-1$	-0.111	-0.415	-0.297
	(0.124)	(0.133)***	(0.127)**
Number of Villages	26,529	26,527	26,527

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; The top panel is the baseline approach and takes the difference in cumulative log rainfall deviations between periods t (2006-8) and $t-1$ (2003-5). The bottom panel allows cumulative log rainfall deviations in periods t and $t-1$ to enter separately. The dependent variable in all specifications is $\Delta \ln M_{v,t+1}/N_{v,t+1}$ and has mean 0.11. Standard errors are clustered at the district level and significance levels are based on a block bootstrap$-t$ procedure. Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper.

Table F.7: Full Elaboration of Annual Rainfall Shocks

Correction Procedure	None	Semipar.	Param.	None	Semipar.	Param.
1st Stage Estimator	None	SU-LPM	BiProbit	None	SU-LPM	BiProbit
	(1)	(2)	(3)	(4)	(5)	(6)
Pareto exponent $\widehat{\lambda}_v$	-0.004	0.053	0.040	-0.008	-0.056	-0.046
	(0.017)	(0.017)**	(0.018)**	(0.034)	(0.034)	(0.033)
θ_{a3}: log rainfall deviation, 2003	0.049	-0.662	-0.315	-0.011	0.230	0.063
	(0.213)	(0.228)	(0.220)	(0.322)	(0.331)	(0.322)
θ_{a4}: log rainfall deviation, 2004	0.424	2.475	1.379	0.649	2.319	1.568
	(0.441)	(0.466)**	(0.467)**	(0.680)	(0.660)**	(0.641)**
θ_{a5}: log rainfall deviation, 2005	-0.709	-3.063	-1.839	-0.701	-2.803	-1.729
	(0.201)***	(0.298)***	(0.238)***	(0.288)**	(0.362)***	(0.303)***
θ_{a6}: log rainfall deviation, 2006	0.098	0.674	0.135	-0.211	-0.830	-1.150
	(0.394)	(0.403)	(0.401)	(0.537)	(0.541)	(0.531)*
θ_{a7}: log rainfall deviation, 2007	0.636	-1.680	-0.662	0.613	-0.775	0.135
	(0.454)	(0.547)	(0.510)	(0.610)	(0.665)	(0.640)
θ_{a8}: log rainfall deviation, 2008	-0.452	2.101	1.267	-0.579	0.888	0.501
	(0.380)	(0.461)**	(0.449)**	(0.623)	(0.674)	(0.659)
$\theta_{a3\lambda}$: log rainfall deviation, 2003 $\times\ \widehat{\lambda}_v$				0.043	-0.580	-0.243
				(0.169)	(0.174)**	(0.169)
$\theta_{a4\lambda}$: log rainfall deviation, 2004 $\times\ \widehat{\lambda}_v$				-0.149	0.100	-0.108
				(0.342)	(0.337)	(0.323)
$\theta_{a5\lambda}$: log rainfall deviation, 2005 $\times\ \widehat{\lambda}_v$				0.001	-0.167	-0.073
				(0.121)	(0.145)	(0.119)
$\theta_{a6\lambda}$: log rainfall deviation, 2006 $\times\ \widehat{\lambda}_v$				0.217	1.037	0.865
				(0.271)	(0.277)***	(0.268)***
$\theta_{a7\lambda}$: log rainfall deviation, 2007 $\times\ \widehat{\lambda}_v$				-0.014	-0.654	-0.596
				(0.265)	(0.279)*	(0.269)**
$\theta_{a8\lambda}$: log rainfall deviation, 2008 $\times\ \widehat{\lambda}_v$				0.091	0.794	0.534
				(0.348)	(0.370)**	(0.360)*
$\sum_{s=3}^{5}\theta_{as}$	-0.236	-1.250	-0.775	-0.063	-0.253	-0.099
H_0: $\sum_{s=3}^{5}\theta_{as}=0$ [p-value]	[0.571]	[0.003]	[0.080]	[0.914]	[0.655]	[0.859]
$\sum_{s=6}^{8}\theta_{as}$	0.283	1.094	0.740	-0.177	-0.717	-0.514
H_0: $\sum_{s=6}^{8}\theta_{as}=0$ [p-value]	[0.521]	[0.011]	[0.103]	[0.793]	[0.271]	[0.425]
$\sum_{s=3}^{8}\theta_{as}$	0.046	-0.155	-0.035	-0.240	-0.971	-0.612
H_0: $\sum_{s=3}^{8}\theta_{as}=0$ [p-value]	[0.632]	[0.141]	[0.733]	[0.268]	[0.0001]	[0.003]
$\sum_{s=3}^{5}\theta_{as\lambda}$				-0.105	-0.647	-0.424
H_0: $\sum_{s=3}^{5}\theta_{as\lambda}=0$ [p-value]				[0.699]	[0.024]	[0.103]
$\sum_{s=6}^{8}\theta_{as\lambda}$				0.293	1.178	0.804
H_0: $\sum_{s=6}^{8}\theta_{as\lambda}=0$ [p-value]				[0.369]	[0.001]	[0.011]
$\sum_{s=3}^{8}\theta_{as\lambda}$				0.189	0.531	0.379
H_0: $\sum_{s=3}^{8}\theta_{as\lambda}=0$ [p-value]				[0.119]	$[<0.001]$	[0.003]
Number of Villages	26,529	26,527	26,527	26,529	26,527	26,527

Notes: The table reports estimates of equation (13) (in columns 1-3) and (14) (in columns 4-6) in the text with a fully elaborated set of annual rainfall shocks instead of cumulating those shocks over three seasons into a single rainfall shock term. Standard errors are clustered at the district level and significance levels are based on a block bootstrap$-t$ procedure. The dependent variable in all specifications is $\Delta \ln M_{v,t+1}/N_{v,t+1}$ and has mean 0.11. Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper. The p-values in the bottom panel are based on F tests.

Table F.8: Alternative Rice Price Shock Measures

Correction Procedure	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.
1st Stage Estimator	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit
Price Shock Proxy	*Δannualized log rice price growth*				*Δlog average rice price*				*log shipping distance pass-through model*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Pareto exponent $\widehat{\lambda}_v$	0.072	0.050	-0.113	-0.070	0.058	0.043	-0.338	-0.118	0.070	0.049	0.933	1.640
	(0.018)***	(0.017)***	(0.030)***	(0.030)**	(0.019)**	(0.017)**	(0.128)**	(0.111)	(0.019)***	(0.018)***	(0.842)	(0.801)*
Δ price shock	1.287	0.625	-2.234	-1.503								
	(0.502)*	(0.451)	(0.709)*	(0.688)**								
Δ price shock × $\widehat{\lambda}_v$			1.913	1.155								
			(0.335)***	(0.327)***								
Δ avg. price					1.630	0.396	-0.311	-0.269				
					(0.613)**	(0.524)	(0.688)	(0.647)				
Δ avg. price × $\widehat{\lambda}_v$							1.158	0.442				
							(0.377)***	(0.320)**				
log shipping distance to THA/VNM									-0.852	-0.866	-0.718	-0.529
									(0.319)	(0.295)**	(0.383)	(0.357)
× $\widehat{\lambda}_v$											-0.041	-0.172
											(0.114)	(0.106)
log distance to nearest port									-0.116	-0.001	0.041	0.064
									(0.053)	(0.050)	(0.065)	(0.062)
× $\widehat{\lambda}_v$											-0.098	-0.042
											(0.023)**	(0.022)
Number of villages	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537	24,537

Notes: Columns 1-4 are the baseline. Columns 5-8 uses the log difference in average rice prices between periods as the measure of the "shock". Columns 9-12 apply the insights from the trade model in Appendix E.1 to use a distance-based proxy for the local intensity of the price shock. The dependent variable in all specifications is $\Delta \ln M_{v,t+1}/N_{v,t+1}$ and has mean 0.11. Standard errors are clustered at the district level and significance levels are based on a block bootstrap$-t$ procedure. Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper.

Table F.9: Alternative Choices of $\underline{R}$ in Estimating λ_v

Correction Procedure	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.
1st Stage Estimator	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	$\underline{R} = 0.1$ Hectares						$\underline{R} = 0.15$ Hectares					
Pareto exponent $\widehat{\lambda}_v$	0.072	0.050	0.073	0.052	-0.113	-0.070	0.057	0.041	0.053	0.040	-0.090	-0.057
	(0.018)***	(0.017)***	(0.018)***	(0.017)***	(0.030)***	(0.030)**	(0.015)***	(0.014)***	(0.015)***	(0.014)***	(0.023)***	(0.026)**
Δ rainfall shock	0.305	0.212	0.162	0.069	0.262	0.135	0.320	0.227	0.270	0.137	0.357	0.200
	(0.134)*	(0.130)	(0.167)	(0.166)	(0.168)	(0.161)	(0.135)**	(0.130)*	(0.167)	(0.161)	(0.168)*	(0.159)
Δ rainfall shock $\times\ \widehat{\lambda}_v$			0.082	0.085	0.028	0.048			0.025	0.047	-0.014	0.017
			(0.049)	(0.058)*	(0.052)	(0.051)			(0.042)	(0.046)	(0.045)	(0.043)
Δ price shock	1.287	0.625	1.304	0.627	-2.234	-1.503	1.319	0.615	1.326	0.609	-1.654	-1.265
	(0.502)*	(0.451)	(0.503)*	(0.452)	(0.709)*	(0.688)*	(0.503)*	(0.453)	(0.503)*	(0.452)	(0.657)	(0.682)
Δ price shock $\times\ \widehat{\lambda}_v$					1.913	1.155					1.451	0.908
					(0.335)***	(0.327)***					(0.266)***	(0.285)***
Number of Villages	26,527	26,527	26,527	26,529	26,527	26,527	26,435	26,435	26,435	26,435	26,435	26,435
	$\underline{R} = 0.2$ Hectares						$\underline{R} = 0.25$ Hectares					
Pareto exponent $\widehat{\lambda}_v$	0.034	0.036	0.030	0.034	-0.093	-0.057	0.026	0.025	0.024	0.027	-0.072	-0.065
	(0.013)*	(0.014)***	(0.012)**	(0.014)***	(0.023)***	(0.030)**	(0.012)*	(0.012)**	(0.012)**	(0.011)**	(0.022)***	(0.029)***
Δ rainfall shock	0.291	0.204	0.219	0.056	0.299	0.145	0.326	0.225	0.161	-0.000	0.288	0.116
	(0.136)*	(0.131)	(0.162)	(0.153)	(0.164)	(0.153)	(0.136)*	(0.131)*	(0.165)	(0.154)	(0.167)*	(0.158)
Δ rainfall shock $\times\ \widehat{\lambda}_v$			0.034	0.072	-0.000	0.033			0.074	0.106	0.023	0.055
			(0.036)	(0.039)**	(0.037)	(0.037)			(0.037)*	(0.037)***	(0.039)	(0.038)
Δ price shock	1.338	0.627	1.331	0.622	-1.361	-1.188	1.369	0.638	1.358	0.637	-0.998	-1.331
	(0.498)*	(0.447)	(0.497)*	(0.447)	(0.661)	(0.731)	(0.499)*	(0.444)	(0.498)**	(0.444)	(0.651)	(0.735)*
Δ price shock $\times\ \widehat{\lambda}_v$					1.235	0.819					0.985	0.825
					(0.248)***	(0.289)***					(0.229)***	(0.274)***
Number of Villages	26,346	26,346	26,346	26,346	26,346	26,346	26,242	26,242	26,242	26,242	26,242	26,242

Notes: The table reports estimates allowing for alternative $\underline{R}$ thresholds in the estimation of λ_v (and the share of households above $\underline{R}$). Baseline estimates using $\underline{R} = 0.1$ Ha are reported in columns 1-6 of the top panel. All estimates in the table are for wetland holdings. The dependent variable in all specifications is $\Delta \ln M_{v,t+1}/N_{v,t+1}$ and has mean 0.11. Standard errors are clustered at the district level, and significance levels are based on a block bootstrap$-t$ procedure. Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper.

Table F.10: Controlling for Omitted Demographic Variables

Correction Procedure	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.	Semipar.	Param.
1st Stage Estimator	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit	SU-LPM	BiProbit
Landholdings measure:	Total Agricultural Landholdings				Wetland Holdings				Paddy Planted, 2002			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Pareto exponent $\widehat{\lambda}_v$	0.039	0.037	0.005	0.004	0.073	0.048	-0.084	-0.044	0.044	0.048	-0.057	-0.053
	(0.017)*	(0.018)*	(0.034)	(0.034)	(0.017)***	(0.017)***	(0.028)**	(0.027)	(0.016)**	(0.016)**	(0.043)	(0.039)*
Δ rainfall shock	0.392	0.299	0.199	0.188	0.266	0.201	0.195	0.120	0.352	0.248	-0.522	-1.238
	(0.134)**	(0.127)**	(0.169)	(0.162)	(0.134)*	(0.129)	(0.167)	(0.160)	(0.141)**	(0.135)*	(0.797)	(0.745)*
Δ price shock	0.493	0.333	0.318	-0.168	1.321	0.609	-1.628	-1.005	0.937	0.450	0.137	0.106
	(0.462)	(0.437)	(0.690)	(0.679)	(0.510)*	(0.464)	(0.694)	(0.651)	(0.497)	(0.455)	(0.174)	(0.173)
Δ rainfall shock $\times\ \widehat{\lambda}_v$			0.120	0.074			0.042	0.050			0.132	0.086
			(0.072)**	(0.071)			(0.051)	(0.049)			(0.060)**	(0.060)
Δ price shock $\times\ \widehat{\lambda}_v$			0.106	0.327			1.585	0.872			0.823	1.001
			(0.326)	(0.331)			(0.312)***	(0.294)***			(0.392)**	(0.370)***
average household size, 2000	-0.003	-0.014	-0.003	-0.014	-0.006	-0.023	-0.007	-0.023	-0.009	-0.022	-0.010	-0.020
	(0.023)	(0.022)	(0.023)	(0.022)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.022)
15-29 year old population share, 2000	0.220	0.239	0.263	0.261	0.528	0.698	0.618	0.743	0.385	0.520	0.434	0.583
	(0.301)	(0.303)	(0.301)	(0.303)	(0.300)	(0.298)	(0.297)	(0.297)*	(0.304)	(0.299)	(0.301)	(0.297)
internal migrant share, 2000	0.655	0.594	0.662	0.597	1.068	0.821	0.978	0.785	0.864	0.705	0.841	0.665
	(0.159)***	(0.164)***	(0.159)***	(0.163)***	(0.157)***	(0.165)***	(0.156)***	(0.162)***	(0.164)***	(0.171)***	(0.164)***	(0.171)***
Number of Villages	26,527	26,527	26,527	26,527	24,537	24,537	24,537	24,537	24,476	24,476	24,476	24,476

Notes: The table augments the baseline estimates with the mean of average household size in the village, the share of the population aged 15-29 in 2000, and the share of individuals that resided in a different district in 1995—all of which are obtained from the Population Census of 2000. The dependent variable in all specifications is $\Delta \ln M_{v,t+1}/N_{v,t+1}$ and has mean 0.11. Standard errors are clustered at the district level, and significance levels are based on a block bootstrap$-t$ procedure. Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper.

Table F.11: Accounting for Measurement and Reporting Outliers in Estimating θ_a and θ_p

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Landholdings measure	Total Agricultural Landholdings						
Pareto exponent $\widehat{\lambda}_v$	0.039	0.036	0.058	0.037	0.076	0.091	-0.029
	(0.017)*	(0.017)	(0.024)	(0.017)	(0.019)**	(0.025)**	(0.021)
Δ price shock	0.409	0.430	0.394	0.378	0.425	0.386	0.781
	(0.448)	(0.447)	(0.449)	(0.446)	(0.491)	(0.490)	(0.482)
Δ rainfall shock	0.415	0.403	0.415	0.420	0.368	0.369	0.575
	(0.133)**	(0.133)**	(0.134)***	(0.133)***	(0.138)*	(0.139)**	(0.170)***
Number of Villages	26,527	26,527	26,294	26,482	23,539	23,296	19,031
Landholdings measure	Wetland Holdings						
Pareto exponent $\widehat{\lambda}_v$	0.072	0.069	0.086	0.071	0.109	0.104	0.017
	(0.018)***	(0.018)***	(0.022)***	(0.018)***	(0.021)***	(0.025)***	(0.020)
Δ price shock	1.287	1.321	1.256	1.256	1.299	1.309	1.378
	(0.502)*	(0.500)*	(0.496)	(0.499)*	(0.532)*	(0.535)**	(0.548)*
Δ rainfall shock	0.305	0.293	0.328	0.311	0.307	0.315	0.445
	(0.134)**	(0.134)*	(0.135)*	(0.134)**	(0.141)*	(0.143)*	(0.172)**
Number of Villages	24,537	24,537	24,304	24,493	21,929	21,705	17,286
	Paddy Planted, 2002						
Pareto exponent $\widehat{\lambda}_v$	0.043	0.041	0.065	0.044	0.074	0.088	0.038
	(0.016)**	(0.016)*	(0.021)**	(0.016)**	(0.019)***	(0.024)**	(0.018)
Δ price shock	0.919	0.955	0.989	0.914	0.820	0.945	1.294
	(0.487)	(0.485)	(0.489)	(0.485)	(0.523)	(0.525)	(0.527)*
Δ rainfall shock	0.390	0.376	0.398	0.393	0.377	0.379	0.541
	(0.139)**	(0.140)**	(0.139)**	(0.140)**	(0.148)**	(0.149)**	(0.179)***
Number of Villages	24,855	24,855	24,650	24,812	22,136	21,924	17,615
Reporting Frequency Indicators		Yes				Yes	
λ Trimmed			Yes			Yes	
Migration Reporting Outliers Removed				Yes		Yes	
Perfect Match Stage					Yes	Yes	
High Illegal Migration Provinces Removed							Yes

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; Standard errors are clustered at the district level in all specifications, and significance levels are based on a block bootstrap$-t$ procedure. Column 1 reproduces the baseline estimates from Table 5; column 2 includes five indicators for the frequency of population register updating in the village (see Appendix C in the paper); column 3 trims the bottom 1 and top 99 percentiles of the distribution of $\widehat{\lambda}_v$; column 4 removes those villages for which the reporting format of *Podes* 2005 and/or 2008 results in top-censoring of migrant stocks in certain villages; column 5 retains only those villages for which *Podes* 2005 and/or 2008 could be matched exactly on administrative codes and village name (see Appendix H); column 6 combines the previous four restrictions; column 7 drops villages in East Java, West Nusa Tenggara and provinces in Kalimantan, all of which are conjectured to have high illegal emigration outflows according to Bank Indonesia (2009). Additional covariates in all specifications but not reported here include all those reported under Table 5 in the paper or mentioned in the notes therein.

Table F.12: Accounting for Outliers in Estimating $\theta_{a\lambda}$ and $\theta_{p\lambda}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Land-holdings measure	Total Agricultural Land-holdings						
Pareto exponent $\widehat{\lambda}_v$	-0.007	-0.011	-0.038	-0.012	0.063	0.016	-0.032
	(0.035)	(0.035)	(0.055)	(0.035)	(0.044)	(0.060)	(0.041)
Δ price shock	-0.016	-0.016	-0.779	-0.101	0.537	-0.270	0.741
	(0.688)	(0.687)	(0.909)	(0.689)	(0.808)	(0.932)	(0.769)
Δ rainfall shock	0.225	0.204	0.268	0.227	0.261	0.219	0.425
	(0.169)	(0.169)	(0.203)	(0.169)	(0.182)	(0.214)	(0.210)**
Δ price shock $\times$ $\widehat{\lambda}_v$	0.267	0.284	0.719	0.302	-0.089	0.347	0.041
	(0.329)	(0.329)	(0.520)	(0.332)	(0.409)	(0.540)	(0.380)
Δ rainfall shock $\times$ $\widehat{\lambda}_v$	0.119	0.126	0.062	0.122	0.061	0.055	0.102
	(0.074)**	(0.073)*	(0.107)	(0.074)**	(0.077)	(0.109)	(0.087)
Number of Villages	26,527	26,527	26,294	26,482	23,539	23,296	19,031
Land-holdings measure	Wetland Holdings						
Pareto exponent $\widehat{\lambda}_v$	-0.113	-0.113	-0.098	-0.121	-0.081	-0.072	-0.028
	(0.030)***	(0.030)**	(0.041)*	(0.029)***	(0.032)**	(0.045)	(0.035)
Δ price shock	-2.234	-2.132	-1.998	-2.360	-2.079	-1.672	0.145
	(0.709)*	(0.701)*	(0.833)	(0.705)*	(0.752)*	(0.899)	(0.780)
Δ rainfall shock	0.262	0.237	0.129	0.281	0.187	0.192	0.477
	(0.168)	(0.168)	(0.207)	(0.168)	(0.181)	(0.221)	(0.209)**
Δ price shock $\times$ $\widehat{\lambda}_v$	1.913	1.876	1.738	1.965	1.801	1.522	0.603
	(0.335)***	(0.333)***	(0.431)**	(0.335)***	(0.354)***	(0.463)***	(0.366)
Δ rainfall shock $\times$ $\widehat{\lambda}_v$	0.028	0.037	0.111	0.021	0.066	0.062	-0.043
	(0.052)	(0.052)	(0.084)	(0.051)	(0.056)	(0.091)	(0.055)
Number of Villages	24,537	24,537	24,304	24,493	21,929	21,705	17,286
	Paddy Planted, 2002						
Pareto exponent $\widehat{\lambda}_v$	-0.083	-0.087	-0.101	-0.082	-0.020	-0.056	-0.030
	(0.046)**	(0.044)**	(0.055)**	(0.046)*	(0.051)	(0.064)	(0.051)
Δ price shock	-1.031	-1.042	-1.937	-1.035	-0.465	-1.447	-0.306
	(0.822)	(0.804)	(0.899)	(0.824)	(0.915)	(1.050)	(0.889)
Δ rainfall shock	0.167	0.131	0.252	0.173	0.177	0.201	0.561
	(0.176)	(0.177)	(0.198)	(0.176)	(0.202)	(0.223)	(0.220)**
Δ price shock $\times$ $\widehat{\lambda}_v$	1.116	1.144	1.696	1.115	0.711	1.336	0.841
	(0.423)***	(0.411)***	(0.506)***	(0.426)**	(0.482)*	(0.605)**	(0.456)**
Δ rainfall shock $\times$ $\widehat{\lambda}_v$	0.140	0.155	0.086	0.137	0.121	0.098	-0.055
	(0.065)**	(0.065)**	(0.087)	(0.064)**	(0.081)*	(0.102)	(0.075)
Number of Villages	24,855	24,855	24,650	24,812	22,136	21,924	17,615
Reporting Frequency Indicators		Yes				Yes	
λ Trimmed			Yes			Yes	
Migration Reporting Outliers Removed				Yes		Yes	
Perfect Match Stage					Yes	Yes	
High Illegal Migration Provinces Removed							Yes

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; Standard errors are clustered at the district level in all specifications, and significance levels are based on a block bootstrap$-t$ procedure. See the Notes to Table F.11

G Zeros, Balls, Bins, and Traveling Salesman

This section provides further background on the extensive margin, elaborating on the discussion of the balls-and-bins test and the stylized model of recruiters in the paper.

G.1 Ruling out a Balls-and-Bins Interpretation of the Extensive Margin

I adapt a simple probabilistic balls-and-bins test developed in Armenter and Koren (2014) to show that the extensive margin cannot be explained as a purely random phenomenon arising from the existing distribution of village sizes. The basic idea is to compare the empirical incidence of zeros with that arising from a model in which villages receive migrants (balls) randomly but with the probability proportional to village population size. Suppose that each migrant m is a ball. There are $M \in \mathbb{N}$ total migrants comprised of the sum across all villages, $M = \sum_{v=1}^{V} m_v$. Also, suppose that each village is a bin, the width of which is given by the share of that village's population in the total population of Indonesia. Formally, the size of bin v is given by $s_v = N_v/N$, where N_v is village v population and $N = \sum_{v=1}^{V} N_v$. The joint probability of migrants across villages follows a multinomial distribution

$$\mathbb{P}(m_1, \ldots, m_V) = \frac{M!}{m_1! \cdots m_V!} s_1^{m_1} \cdots s_V^{m_V},$$

in which the expected number of *nonzero* migration villages $\mathcal{V}^*$ (or non-empty bins) is given by

$$\mathbb{E}[\mathcal{V}^*|M] = \sum_{v=1}^{V} \left[1 - (1 - s_v)^M\right]. \tag{G.1}$$

Calculating the sample analogue of equation (G.1) for 2008, I find that the balls-and-bins model predicts nonzero migration in 64,457 villages out of the total 65,966 villages. (If population sizes were uniform across villages, the balls-and-bins model predicts that every village would almost surely have at least one migrant.) In other words, only 5.5 percent of the 27,297 zero migration village in the empirical data can be explained away as an atheoretical statistical regularity in sparse data. Figure G.1 compares the predicted probability of having any migrants under the balls-and-bins model (dashed curve) with the actual share of villages with any migrants (solid curve). Both are plotted against log village population size. The incidence of zeros in the data is much higher than would be predicted on the basis of a random balls-and-bins allocation of migrants across villages. The vertical distance between the two curves constitutes the scope for the theory and empirics in the paper to address the substantive economic forces behind the extensive margin including, among others, the role of recruiters.

G.2 A Heuristic Framework for Recruiter Location Choice

If the market of potential migrants is too small, recruiters will not serve v, and upfront cost will be accordingly high. Given the difficulty of initial (first-mover) migration from villages without outside intermediaries, recruiter location choice should be highly correlated with the extensive margin. To add structure, one can think of recruiters as "traveling salesmen" tasked with identifying the least cost method of visiting a set number of locations within a defined area. Consistent with evidence in Bachtiar (2011), suppose that these agencies must obtain operating licenses in district capitals and face a fixed cost of entering villages (e.g., making royalty payments to village officials). In order to maximize potential migrants reached and minimize fixed entry and variable travel costs, recruiters must first select districts within which to operate and then the order in which villages are visited.

To illustrate the logic behind these implications, first consider two districts k and k' with equal populations and inter-village travel distances. District k has two equally populated villages, while district k' has three villages: village $1_{k'}$ has equal population with the two villages in k, while villages $2_{k'}$ and $3_{k'}$ are equally populated with the total equal to the population of $1_{k'}$. Assuming (i) constant fixed costs of establishing agency presence in equally sized districts and (ii) constant fixed costs of entering villages, a given recruiter would be more likely to enter district k than k'. If, however, recruiters choose to visit district k' for other (unobserved) reasons, then village $1_{k'}$ would be more likely visited than $2_{k'}$ or $3_{k'}$. Now, add one identical village to each district with population greater than all existing villages in each district. Assuming recruitment agencies are subject to budget constraints preventing visits to all villages within k, it is straightforward to show within this framework that recruiters would only visit the newly added village in k.

I now provide a sketch of the general traveling salesman model underlying the proposed instruments for the extensive margin. Begin by considering the problem of recruiters selecting a district within which to operate, retaining the assumption that licensing and other fees are paid to district government officials.[1] Let the cost of traveling between villages v and v' within district k be denoted by $d^k_{vv'} > 0$. Suppose further that there are $\mathcal{V}^k$ villages in district k and that the population of the district less village v is given by N^k_{-v}. For empirical tractability, additionally assume that the least-cost path of visiting every village within a district is approximately proportional to the area of the district less village v, A^k_{-v}.[2] Now suppose that the fixed cost of entering a village, f, is identical across all villages and that the ex ante probability of successfully recruiting any individual potentials migrant is identical across villages. The optimization problem for the recruitment agency is to maximize the number of potential migrants, $\widetilde{M}_v$, reached (with advertisements and contract offers) and to minimize the costs (and hence maximize expected revenue). The objective is then

$$\max_k \sum_{v=1}^{\mathcal{V}^k} \widetilde{M}_v \quad \text{s.t. } f\mathcal{V}^k + A^k_{-v} \leq \mathcal{B}, \;\; \widetilde{M}_v \leq N_v \; \forall v$$

where $\mathcal{B}$ is the exogenously given budget of the recruitment agency. The (heuristic) solution function (i.e., the optimal district) should be increasing in N^k_{-v} and decreasing in A^k_{-v} and $\mathcal{V}^k$. Once inside a given district k, all else equal, budget-constrained recruiters are relatively more likely to visit villages with larger populations since the unconditional probability of successfully recruiting a single migrant is higher.

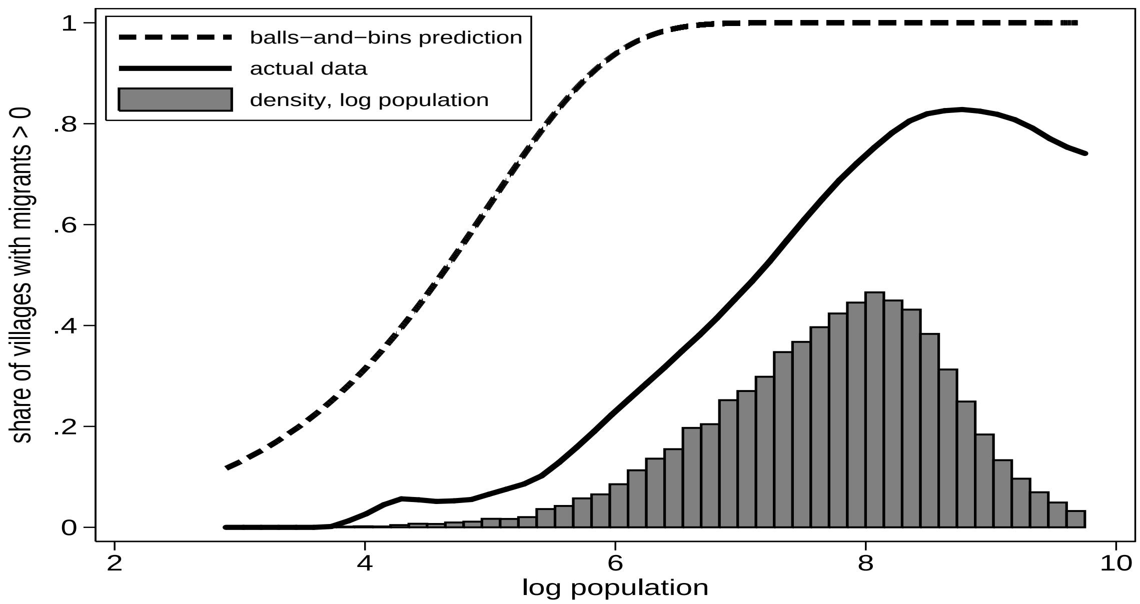
In Table G.1, I test the above hypotheses using the only available proxy for recruiter visits at the village level—a measure in *Podes* indicating whether any recruiters targeting female migrants are based in the village. Conditional on the presence of any migrants in the prior period (i.e., as recorded mid-2005), the likelihood that village v has a recruiter in period $t+1$ (i.e., prior to mid-2008) is (i) increasing in the population of village v and the population of v's district less v, and (ii) decreasing in the number of village in v's district. These findings are consistent with the predictions of the traveling salesman framework sketched above. The positive albeit statistically null correlation with the area of v's district is not. However, that the probability is decreasing in the distance from the subdistrict capital (albeit not the district capital) suggests that some of the distance components of the model hold. A more rigorous test would require computing the actual distances between villages and using some of the available methods for solving the traveling salesman problem. This is beyond the scope of the present study as the patterns observed in Table G.1 bear out indications that recruiter decisions follow some approximation to the model described above. Other important results in Table G.1 include (i) the positive correlation with the share of ethnic Arabs in the total population of v in 2000 and (ii) the statistically null correlation with agricultural income shocks. Finding (i) is consistent with the role of local ethnic intermediaries in facilitating migration to Middle

[1]This assumption follows from evidence on the procedures through which recruitment agencies engage with government institutions in the process of authorizing legal contract emigrants (Bank Indonesia, 2009). The engagement with local government officials has been increasing in recent years as decentralization has resulted in a devolution of authorities and regulatory power to the regions (see Bachtiar, 2011). Under Law 39/2004, recruitment agencies are only permitted to recruit and place prospective labor migrants who are registered at the local Ministry of Manpower and Transmigration. Of course agency field workers often bypass local governments and enter villages directly, but the agencies must still liaise with officials in the district capital for the purposes of document preparation and other predeparture certification processes.

[2]That is, $\sum_{v'\neq v}^{\mathcal{V}^k} \sum_{v\neq v'}^{\mathcal{V}^k} d^k_{vv'} \propto A^k_{-v}$.

Eastern destinations. Finding (ii) suggests that recruiters do not respond to income shocks in deciding where to target their contract offers.

Figure G.1: Comparing the Actual Incidence of Zeros with the Balls-and-Bins Prediction



Notes: The "actual data" estimate is based on a local linear probability regression of an indicator for any migrants in village v on log population, using an Epanechnikov kernel and trimming the top 1 percent of villages for presentational purposes. The "balls-and-bins" prediction is based on the model described in Section G.1.

Tables

Table G.1: On the Determinants of Recruiter Visits

	$\mathbb{P}$(female migrant recruiter present)	
	(1)	(2)
1(migrants>0)	0.116	0.109
	(0.011)***	(0.009)***
log village population	0.048	0.046
	(0.004)***	(0.004)***
log district population less v	0.041	0.047
	(0.013)***	(0.015)***
log number of villages in district	-0.018	-0.026
	(0.014)	(0.015)*
log district area less v	0.011	0.011
	(0.008)	(0.007)
Pareto exponent $\hat{\lambda}_v$		-0.001
		(0.005)
share households above $\underline{R}$		-0.022
		(0.018)
log distance to subdistrict capital		-0.012
		(0.003)***
log distance to district capital		-0.004
		(0.006)
Arab population share		1.023
		(0.464)**
Chinese population share		0.697
		(2.221)
Muslim population share		-0.041
		(0.030)
price shock		-0.212
		(0.174)
rainfall shock		-0.018
		(0.014)
Number of Villages	51,593	51,593

Notes: Significance levels: $*$: 10% $**$: 5% $***$: 1%; The table reports marginal effects at the mean from a probit regression of all covariates shown in the respective columns as well as the following additional covariates in column 2: wetland area as a share of total farmland, log distance to nearest emigration center, and an indicator for government-prescribed urban status. Standard errors are clustered at the district level.

H Panel Data Construction

In this subsection, I describe the process of constructing a panel dataset of Indonesian villages comprised of data collected in 2000, 2002, 2003, 2005, and 2008. Starting from the baseline 65,966 villages in Table 1 in the paper, the final sample of villages is reduced further by two factors. First, because this paper focuses on heterogeneous income shocks in agricultural areas, I exclude urban villages without land-holdings entries in the Agricultural Census. There are other practical reasons for doing so as well. In Indonesia, agricultural commodity price increases generally have homogeneous, negative effects on real income in urban areas, and rainfall shocks tend to have no effect on rice production in nominally urban areas (Levine and Yang, 2014). Second, changes in administrative boundaries over the period 2000-2008 required dropping a small number of villages with missing data from one or more of the additional sources, including the Population Census of 2000. I ultimately treat these villages as missing at random. In the late 1990s and early 2000s, responding to a range of political and economic incentives in the wake of decentralization, government officials set about proliferating administrative units across the country and at varying levels of government (see Fitrani et al., 2005). The proliferation was relatively more common in the Outer Islands than on Java. This process has created difficulties for researchers attempting to link administrative units over time in the *Podes* and other surveys. Most researchers work with district-level aggregates and take districts in some base year and aggregate backwards to achieve minimum comparative areas (MCA) (e.g., Vothknecht and Sumarto, 2009). For studies such as the present one, however, it is crucial to retain the village as the unit of analysis.

The remainder of this appendix details the matching of villages across multiple waves of *Podes* (2002, 2005 and 2008), the 2003 Agricultural Census, and the 2000 Population Census. Prior to beginning, I exclude villages from the islands of Papua in Eastern Indonesia and Nias off the West coast of Sumatra. I exclude Papua because the data quality is questionable and moreover the social and economic conditions do not lend themselves to the issues addressed in this study. I exclude Nias since the special post-Indian Ocean tsunami *Podes* survey administered in this region in mid-2005 did not include questions on migration.

Panel construction proceeds with the merging of villages recorded in *Podes* 2005 and 2008. The Central Statistics Bureua (BPS) does not provide exact concordances between villages across these two survey rounds. As such, I manually construct a mapping between these two data sources, which contain the main dependent variables of interest, using a combination of exact and fuzzy merge-matching algorithms combining information on province ID, district ID, subdistrict ID, village ID, village name, and village land area. In the initial step, I combine 2008 villages with identical 2007 village IDs as made available in *Podes* 2008. I also remove all non-diacritic characters from village names prior to implementing the algorithm. The resulting panel is comprised of 65,966 MCAs. A detailed breakdown by province of the number of villages matched at each stage of the algorithm can be seen in Table H.1. Around 700 villages could not be merged into a reliable MCA across years.[1] Nevertheless, I view these villages as missing at random insomuch as the timing of elections resulting in the splitting of districts and subsequently villages has been shown elsewhere to be orthogonal to baseline observables of interest (Skoufias et al., 2010).

At the next stage of matching, I incorporate data from the 2000 Population Census using the unique administrative IDs available in *Podes* 2005. The merge-matching algorithm proceeds analogously to that described above. Given the relatively longer period of possible administrative proliferation between 2000 and 2005, the resulting success rate in matching villages was lower than that obatined for *Podes* 2005 and 2008. I then repeat the matching procedure for villages recorded in *Podes* 2002, which contains the requisite information to construct the commodity price index (sans rice) and the measure of overall village-level rice productivity. The resulting match rate is again less favorable than that obtained for *Podes* 2005 and 2008. Table H.2 shows the match rates by source and

[1] These villages account for 4,570 migrants in 2005 and 12,746 migrants in 2008.

province.

Lastly, I perform a similar matching procedure for the Agricultural Census conducted in August 2003. In rural areas, all agricultural households were enumerated in every village. In urban areas as defined by the government, households in a sample of villages among those with any agricultural activities received enumerators. Additionally, due to security concerns at the time, only a subsample of all households in a few villages were enumerated in Aceh Province in May 2004. Due to a lack of village names for certain areas in a few provinces in the raw data provided by BPS, I was unable to merge a number of villages in the Agricultural Census using the *Podes* 2005 IDs (or in unreported results, the *Podes* 2002 IDs). Ultimately, however, the resulting panel consisting of data from 2000, 2002, 2003, 2005, 2008 comprises the overwhelming majority of Indonesian villages and particularly those in rural areas where international migration constitutes an important labor market opportunity.

Table H.1: Merge-Matching Procedure for Linking *Podes* 2005 and 2008

	Matching Stage								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
Matching Variable	*Type of Matching*								
Province ID	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	
Kabupaten ID	Exact	Exact	Exact	Exact	Exact		Exact		
Kecamatan ID	Exact	Exact	Exact	Exact					
Village ID	Exact								
Village Name	Exact	Fuzzy	Exact	Fuzzy	Exact	Exact	Fuzzy	Fuzzy	
Land Area		Exact							
Aceh	5047	153	64	99	161	174	128	109	5935
North Sumatra	3216	223	65	162	161	588	250	179	4844
West Sumatra	522	165	79	79	2	0	30	3	880
Riau	1183	20	116	34	29	0	81	4	1467
Jambi	871	25	57	39	66	0	151	5	1214
South Sumatra	1553	21	181	46	202	34	521	104	2662
Bengkulu	813	9	18	16	61	1	281	6	1205
Lampung	1661	37	33	58	68	84	188	45	2174
Kepulauan Bangka Belitung	305	6	1	8	0	0	0	0	320
Kepulauan Riau	170	2	5	8	11	0	49	0	245
DKI Jakarta	264	0	0	3	0	0	0	0	267
West Java	5036	22	392	53	43	52	91	103	5792
Central Java	8442	7	18	49	16	0	28	1	8561
Yogyakarta	438	0	0	0	0	0	0	0	438
East Java	8302	35	13	75	11	0	38	1	8475
Banten	1154	2	77	10	38	39	129	28	1477
Bali	681	2	2	5	8	0	3	0	701
West Nusa Tenggara	688	11	26	39	8	0	45	0	817
East Nusa Tenggara	1828	69	123	47	176	181	129	160	2713
West Kalimantan	1258	24	13	22	21	57	26	72	1493
Central Kalimantan	1087	16	22	23	63	0	106	8	1325
South Kalimantan	1634	27	122	11	74	0	75	3	1946
East Kalimantan	978	23	146	46	28	1	74	16	1312
North Sulawesi	942	2	8	11	22	97	52	121	1255
Central Sulawesi	1308	19	0	24	102	0	67	3	1523
South Sulawesi	2257	196	34	113	80	156	79	282	3197
Sulawesi Tenggara	1089	13	43	27	70	65	248	101	1656
Gorontalo	351	0	0	10	5	8	23	49	446
Maluku	621	42	80	54	21	3	17	27	865
North Maluku	533	2	16	7	48	0	144	11	761
Total	54232	1173	1754	1178	1595	1540	3053	1441	65966

Notes: This table reports the number of villages matched at each stage of the algorithm I devised in order to link *Podes* 2005 and 2008. Fuzzy matching was done using the `reclink` program with a minimum match score of 0.6 followed by visual inspection and manual matching at each stage of the process.

Table H.2: Percentage of Villages in 2005 with Missing Data by Source and Province

Province (2005)	Population, 2005	# Villages, 2005	Agri. Census, 2003	Podes, 2002	Census, 2000	Pov. Map, 2000
Aceh	4,115,642	5,935	0.986	0.062	0.645	0.681
North Sumatra	11,258,698	4,844	0.104	0.151	0.057	0.069
West Sumatra	4,464,086	880	0.374	0.094	0.256	0.275
Riau	4,647,206	1,467	0.076	0.087	0.166	0.174
Jambi	2,675,052	1,214	0.056	0.050	0.081	0.081
South Sumatra	6,684,582	2,662	0.099	0.231	0.092	0.102
Bengkulu	1,579,034	1,205	0.285	0.469	0.075	0.083
Lampung	7,184,793	2,174	0.024	0.032	0.068	0.070
Kepulauan Bangka Belitung	1,023,689	320	0.459	0.547	0.116	0.131
Kepulauan Riau	1,082,151	245	1.000	1.000	0.335	0.331
DKI Jakarta	7,484,573	267	0.004	0.000	0.007	0.007
West Java	37,355,255	5,792	0.010	0.013	0.014	0.018
Central Java	32,771,370	8,561	0.012	0.002	0.004	0.007
Yogyakarta	3,407,430	438	0.014	0.000	0.000	0.000
East Java	35,907,891	8,475	0.015	0.002	0.004	0.013
Banten	8,809,337	1,477	0.006	0.003	0.005	0.005
Bali	3,271,583	701	0.021	0.021	0.034	0.036
West Nusa Tenggara	4,227,864	817	0.716	0.140	0.147	0.214
East Nusa Tenggara	4,265,106	2,713	0.095	0.118	0.091	0.164
West Kalimantan	3,973,249	1,493	0.132	0.169	0.073	0.128
Central Kalimantan	1,897,154	1,325	0.026	0.063	0.081	0.279
South Kalimantan	3,203,372	1,946	0.032	0.145	0.011	0.079
East Kalimantan	2,870,860	1,312	0.072	0.079	0.212	0.252
North Sulawesi	2,168,461	1,255	0.255	0.325	0.116	0.122
Central Sulawesi	2,362,476	1,523	0.099	0.135	0.099	0.101
South Sulawesi	8,338,093	3,197	0.121	0.116	0.101	0.118
Sulawesi Tenggara	1,975,490	1,656	0.193	0.365	0.091	0.124
Gorontalo	907,503	446	0.368	0.413	0.184	0.188
Maluku	1,343,401	865	0.101	0.351	0.086	0.135
North Maluku	875,599	761	0.419	0.717	0.252	0.355

Notes: This table reports the rate of failed matches between villages reported in *Podes* 2005 and the given source of data listed in the top row of the table.

References

Abramitzky, R., L.P. Boustan, and K. Eriksson, "Have the poor always been less likely to migrate? Evidence from inheritance practices during the Age of Mass Migration," *Journal of Development Economics*, 2013, *102*, 2–14.

Armenter, R. and M. Koren, "A Balls-and-Bins Model of Trade," *American Economic Review*, 2014, *104* (7), 2127–2151.

Bachtiar, P. B., "The Governance of Indonesian Overseas Employment in the Context of Decentralization," *SMERU Research Institute Working Paper*, 2011.

Bank Indonesia, "Survei Nasional Pola Remitansi TKI," *Special Report*, 2009.

Bazzi, S., "A Note on Heterogeneity and Aggregation Using Data on Agricultural Productivity," *Unpublished note*, 2012b.

_ , A. Gaduh, A. Rothenberg, and M. Wong, "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, forthcoming.

Benjamin, D., "Can unobserved land quality explain the inverse productivity relationship?," *Journal of Development Economics*, 1995, *46*, 51–84.

Cameron, A. C., J. B. Gelbach, and D. L. Miller, "Bootstrap-based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

_ , _ , and _ , "Robust inference with multi-way clustering," *Journal of Business and Economic Statistics*, 2011, *28*, 238–249.

Chambers, M.J. and R.E. Bailey, "A theory of commodity price fluctuations," *Journal of Political Economy*, 1996, pp. 924–957.

Das, M., W.K. Newey, and F. Vella, "Nonparametric estimation of sample selection models," *Review of Economic Studies*, 2003, *70* (1), 33–58.

Dawe, D., "Can Indonesia Trust the World Rice Market?," *Bulletin of Indonesian Economic Studies*, 2008, *44* (1), 115–132.

de Jonge, H.M.C. and N.J.G. Kaptein, *Transcending borders: Arabs, politics, trade and Islam in Southeast Asia*, Leiden: KITLV Press, 2002.

Dickey, D. A. and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 1979, *74*, 427–431.

Fitrani, F., B. Hofman, and K. Kaiser, "Unity in Diversity? The Creation of New Local Governments in a Decentralizing Indonesia," *Bulletin of Indonesian Economic Studies*, 2005, *41*, 57–79.

Friedman, J. and J. A. Levinsohn, "The Distributional Impacts of Indonesia's Financial Crisis on Household Welfare: A "Rapid Response" Methodology," *NBER Working Paper No. 8564*, 2001.

Fuglie, K. O., "Productivity growth in Indonesian agriculture, 1961-2000," *Bulletin of Indonesian Economic Studies*, 2010, *40*, 209–225.

Gabaix, X., "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1*, 255–293.

_ and R. Ibragimov, "Rank-1/2: A Simple Way to Improve the Estimation of Tail Exponents," *Journal of Business and Economic Statistics*, 2011, *29* (1), 24–39.

Gallant, A. R. and D. W. Nychka, "Semi-nonparametric maximum likelihood estimation," *Econometrica*, 1987, *55* (2), 363–390.

Heckman, J. J., "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *NBER Chapters*, 1976, pp. 120–137.

Im, K.S., M.H. Pesaran, and Y. Shin, "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 2003, *115*, 53–74.

Jayachandran, S., "Selling labor low: Wage responses to productivity shocks in developing countries," *Journal of Political Economy*, 2006, *114* (3), 538–575.

Levine, D. and D. Yang, "The Impact of Rainfall on Rice Output in Indonesia," *NBER Working Paper No. 20302*, 2014.

Lineton, Jacqueline, "Pasompe'Ugi': Bugis migrants and wanderers," *Archipel*, 1975, *10* (1), 173–201.

Maccini, S. and D. Yang, "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review*, 2009, *99* (3), 1006–1026.

McKenzie, D. and H. Rapoport, "Network effects and dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, 2007, *84* (1), 1–24.

Mundlak, Y., D. Larson, and R. Butzer, "Agricultural dynamics in Thailand, Indonesia and the Philippines," *Aus-*

tralian Journal of Agricultural and Resource Economics, 2004, *48* (1), 95–126.

Munshi, K., "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

Naylor, R. L., W. P. Falcon, D. Rochberg, and N. Wada, "Using El Nino/Southern Oscillation climate data to predict rice production in Indonesia," *Climatic Change*, 2001, *50*, 255–265.

Newey, W. K., "Two-step series estimation of sample selection models," *Unpublished manuscript*, 1988.

—, "Two-step series estimation of sample selection models," *The Econometrics Journal*, 2009, *12*, S217–S229.

— **and D. McFadden**, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 1994, *4*, 2111–2245.

Poirier, D., "Partial Observability in Bivariate Probit Models," *Journal of Econometrics*, 1980, *12* (2), 209–217.

Rochina-Barrachina, M. E., "A new estimator for panel data sample selection models," *Annales d'Économie et de Statistique*, 1999, *55-56*, 153–181.

Skoufias, E., A. Narayan, K. Kaiser, and B. Dasgupta, "Electoral Accountability, Fiscal Decentralization and Service Delivery in Indonesia," *World Bank Policy Research Working Paper*, 2010.

Suryahadi, A., W. Widyanti, R. Prama Artha, D. Perwira, and S. Sumarto, "Developing a Poverty Poverty Map for Indonesia: A Tool for Better Targeting in Poverty Reduction and Social Protection Programs, Book 1: Technical Report," *SMERU Research Report*, 2005.

Timmer, C. P., "Causes of High Food Prices," *Asian Development Bank Working Paper*, 2008, *128.*

Vothknecht, M. and S. Sumarto, "Violent Conflict and Economic Growth: Evidence from Indonesian Villages," *Unpublished manuscript*, 2009.

Warr, P., "Food policy and poverty in Indonesia: a general equilibrium analysis," *Australian Journal of Agricultural and Resource Economics*, 2005, *49* (4), 429–451.

—, "The transmission of import prices to domestic prices: an application to Indonesia," *Applied Economics Letters*, 2008, *15*, 499–503.

Wimanda, R. E., "Price variability and price convergence: Evidence from Indonesia," *Journal of Asian Economics*, 2009, *20* (4), 427–442.

Yamagata, T., "The small sample performance of the Wald test in the sample selection model under the multicollinearity problem," *Economics Letters*, 2006, *93* (1), 75–81.

Zellner, A. and T. H. Lee, "Joint estimation of relationships involving discrete random variables," *Econometrica*, 1965, *33* (2), 382–394.

Zivot, E. and D. W. K. Andrews, "Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis," *Journal of Business and Economic Statistics*, 2002, *20*, 25–44.