NBER WORKING PAPER SERIES

VOLATILITY AND THE GAINS FROM TRADE

Treb Allen
David Atkin

Working Paper 22276
http://www.nber.org/papers/w22276

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2016, Revised December 2021

We thank Costas Arkolakis, Kyle Bagwell, Dave Donaldson, Jonathan Eaton, Marcel Fafchamps, Pablo Fajgelbaum, Chang-Tai Hsieh, Sam Kortum, Yu-Jhih Luo, Rocco Machiavello, Kiminori Matsuyama, John McLaren, Nina Pavcnik, Steve Redding, Andres Rodriguez-Clare, Esteban Rossi-Hansberg, Andy Skrzypacz, Bob Staiger, Jon Vogel and seminar participants at Columbia University, Dartmouth College, George Washington University, Harvard University, University of Maryland, the NBER ITI Winter Meetings, Pennsylvania State University, Princeton University, Princeton IES Summer Workshop, Purdue University, Stanford University, University of British Columbia, University of California -Berkeley, University of California - Davis, University of North Carolina, University of Toronto, and University of Virginia. We thank Scott Fulford for kindly providing the rural bank data we use. Rodrigo Adao, Fatima Aqeel, Masao Fukui, Annekatrin Lüdecke, Daniel O'Connor, Saptarshi Majumdar, and Yuta Takahashi provided exceptional research assistance. Part of this paper was completed while Allen was a visitor at the Stanford Institute for Economic Policy Research (SIEPR), whose hospitality he gratefully acknowledges. This material is based upon work supported by the National Science Foundation under grant SES-1658838. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Volatility and the Gains from Trade
Treb Allen and David Atkin
NBER Working Paper No. 22276
May 2016, Revised December 2021
JEL No. F1,G11,O13,O18

ABSTRACT

Trade liberalization changes the volatility of returns by reducing the negative correlation between local prices and productivity shocks. In this paper, we explore these second moment effects of trade. Using forty years of agricultural micro-data from India, we show that falling trade costs due to expansions of the Indian highway network reduced the responsiveness of local prices to local rainfall but increased the responsiveness of local prices to prices elsewhere. In response, farmers shifted their production toward crops with less volatile yields, especially so for those with poor access to risk mitigating technologies such as banks. We then characterize how volatility affects farmer's crop allocation using a portfolio choice framework where returns are determined in general equilibrium by a many-location, many-good Ricardian trade model with flexible trade costs. Finally, we structurally estimate the model—recovering farmers' risk-return preferences from the gradient of the mean-variance frontier at their observed crop choices—to quantify the second moment effects of trade. We find that first moment gains from specialization dominate second moment effects on average and that improvements in risk-mitigating technologies would encourage farmers to take advantage of higher-risk higher-return allocations, with the strength of these effects hinging on whether the riskiest crops are also the comparative advantage ones.

Treb Allen
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
treb@dartmouth.edu

David Atkin
Department of Economics, E52-550
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
atkin@mit.edu

1 Introduction

While trade liberalization increases average returns through specialization, it also affects the volatility of returns by reducing the negative correlation between local prices and productivity shocks. When production is risky, producers are risk averse, and insurance markets are incomplete—as is the case for farmers in developing countries—the interaction between trade and volatility may have important welfare implications. Yet we have a limited empirical understanding of the relationship between trade and volatility. In particular, does volatility magnify or attenuate the gains from trade; how do agents respond to changes in the risk they face arising from falling trade costs; and can complementary policies ensure that the gains from trade are maximized?

In this paper, we empirically, analytically, and quantitatively explore the second moment effects of trade. Using forty years of agricultural micro-data from India, we show empirically that expansions of the Indian highway network reduced the responsiveness of local prices to local rainfall but increased the responsiveness of local prices to prices elsewhere. In response, farmers not only moved toward crops in which they had a comparative advantage, they also shifted their production toward crops with less volatile yields, an effect that was especially strong for farmers with poor access to the formal banking sector. We then incorporate a portfolio allocation framework—where producers optimally allocate resources (land) across risky production technologies (crops)—into a many location, many good, general equilibrium Ricardian trade model. The model yields analytical expressions for the equilibrium prices and crop allocations and generates straightforward relationships between observed equilibrium outcomes and underlying structural parameters, allowing us to quantify the second moment welfare effects of trade. Structural estimates suggest that first moment gains from specialization outweigh any second moment losses and that improvements in risk mitigating technologies would encourage farmers to choose higher-risk higher-return crop allocations that they would otherwise have been unwilling to pursue.

Rural India is our empirical setting, home to roughly one-third of the world's poor and an environment where agricultural producers face substantial risk. Even today, less than half of agricultural land is irrigated, with realized yields driven by the timing and intensity of the monsoon and other more-localized rainfall variation. Access to agricultural insurance is limited, forcing farmers—who comprise more than three quarters of the economically active population—to face the brunt of the volatility. Furthermore, many are concerned that the substantial fall in trade costs over the past forty years (due, in part, to expansions of the Indian highway network as well as reductions in tariffs) has amplified the risk faced by farmers. As the The New York Times writes:

> "When market reforms were introduced in 1991, the state scaled down subsidies and import barriers fell, thrusting small farmers into an unforgiving global market. Farmers took on new risks, switching to commercial crops and expensive, genetically modified seeds... They found themselves locked in a whiteknuckle gamble, juggling ever larger loans at exorbitant interest rates, always hoping a bumper harvest would allow them to clear their debts, so they could take out new ones. This pattern has left a trail of human wreckage." ("After Farmers Commit Suicide, Debts Fall on Families in India", 2/22/2014).

These concerns, and the importance of better understanding the link between trade and volatility, are encapsulated by the fact that the Doha round of global trade negotiations collapsed in 2008 (and remains stalled today) precisely because of India and China's insistence on special safeguard mechanisms to protect their farmers from excessive price volatility. More recently, in September 2020 the Indian government attempted to improve the efficacy of its agricultural markets by liberalizing a 60 year old system restricting intranational trade which resulted in year-long country-wide protests of hundreds of thousands of farmers, hundreds of deaths, and ultimately government capitulation, all over the concern that such a reform would erode existing farmer protections.[1]

Using a dataset containing the annual price, yield, and area planted for 15 major crops across 311 districts and 40 years matched to bilateral travel times along the evolving national highway network, we document three sets of stylized facts. First, reductions in trade costs due to the expansion of the highway network reduced the elasticity of local prices to local supply shocks and increased the elasticity of local prices to prices elsewhere. Second, this fall in trade costs not only caused farmers to reallocate toward crops for which they had a comparative advantage—as traditional trade models would predict—it also caused farmers to reallocate away from risky crops that had more volatile yields and/or yields that had higher covariances with other crops, an effect that was particularly pronounced in districts with poor bank access. Third, the combination of the previous two effects increased the volatility of farmers' nominal incomes, an effect only partially offset by a decline in price index volatility.

We next develop a general equilibrium Ricardian model of trade and volatility that both captures many of the key features of agricultural trade in India and explains the three sets of stylized facts. In the model, heterogeneous traders engage in the buying and selling of homogeneous agricultural goods to take advantage of price differences between local villages and a central market. To circumvent the familiar difficulties arising from corner solutions for prices and patterns of specialization, we assume that the distribution of trade costs these traders face takes a convenient Pareto form. Consistent with the first stylized fact, this assumption allows equilibrium prices to be written as a log-linear function of the local yield and the market price, with the relative magnitude of these elasticities governed by the shape parameter of the Pareto distribution of trade costs. This model-implied relationship between prices and yields more closely matches the patterns in the data compared to the "kinked" relationship between prices and yields implied by traditional price arbitrage models with homogeneous trade costs. Furthermore, in the absence of volatility, this model generates a simple expression for the equilibrium pattern of specialization—highlighting that, as trade costs fall, farmers will reallocate their crops away from those they wish to consume and toward those in which they have a comparative advantage in production.

Incorporating volatility into the model poses additional challenges. To derive the equilibrium pattern of specialization in the presence of volatility we embed a portfolio choice problem from the finance literature (see e.g. Campbell and Viceira (2002)) into our Ricardian trade framework. In contrast to finance applications, the general equilibrium nature of our trade model means that each farmer's decision depends on the distribution of yields of all crops in all locations and the crop

[1] "Why has Narendra Modi abandoned cherished plans to overhaul Indian farming?" *The Economist*, 11/19/2021.

choices of all other farmers. Despite this complication, our expression for the pattern of specialization remains tractable and is a straightforward generalization of the no volatility case. Consistent with the second stylized fact, as trade costs fall, farmers re-allocate their land toward crops for which they have a risk-adjusted comparative advantage. In doing so, they balance traditional "first moment" gains from trade against "second moment" changes in volatility, with the trade-off governed by their level of risk aversion. The model also allows us to sign the effect of a fall in trade costs on the variance of farmers' nominal incomes and the variance of their price index, with the former rising and the latter falling, consistent with the third stylized fact.

Finally, we extend the framework to create a "quantitative" version of the model that adds realism by incorporating a number of additional features of the empirical setting (e.g. a hierarchical trading network featuring many different regional markets, arbitrary correlations in yields across crops and districts, and a manufacturing sector). We then estimate this extended model and use it to quantify the welfare effects of the expansion of the Indian highway network. Despite the added complexity coming from these extensions, the tractability of the model allows us to recover the key model parameters from the data in a transparent manner. First, as the model implies that the relative magnitude of the elasticities of local prices to local yields shocks and prices elsewhere are governed by the distribution of traders' costs, we can recover unobserved trade costs via a linear regression. These trade costs fall with the increases in market access resulting from highway expansions. Second, as farmers' unobserved risk-return preferences shape the gradient of the mean-variance frontier at the observed crop choices, we can estimate farmers' risk aversion from a linear regression derived from farmers' first order conditions. We find that these risk aversion estimates fall as rural bank access improves, consistent with banks providing a risk mitigating technology that allows farmers to behave in a less risk averse manner.

We use these parameter estimates to quantify the welfare effects of the expansion of the Indian highway network. Between the 1970s and 2000s, we estimate the expansion of the Indian highways resulted in the mean real income of farmers increasing by 2.2%. This would have been accompanied by a small decline in the volatility of farmers' real incomes driven by improved market integration elsewhere stabilizing market prices, with expected welfare rising by 2.3% on net. Finally we ask how concurrent improvements in risk mitigating technologies through the expansion of rural bank access affect farmers' gains. We find that real income would rise by an additional 27% (2.8% vs. 2.2%) on average—driven by farmers pursuing riskier crop allocations that, in the absence of such improvements in risk mitigating technologies, they would have been unwilling to undertake—and welfare almost doubles, with the strength of the complementarities hinging on whether the riskiest crops are also the comparative advantage ones.

This paper relates to a number of strands of literature in both international trade and economic development. There is a longstanding theoretical literature on trade and volatility; see Helpman and Razin (1978) and references cited therein. In a seminal paper, Newbery and Stiglitz (1984) develop a stylized model where trade can reduce welfare in the absence of insurance (although to obtain this stark result they assume farmers and consumers differ in their preferences and do not consume what they produce).[2] Our baseline model abstracts from the risk associated with different types of agent producing different types of goods in a location and shows using a revealed preference argument a la Dixit and Norman (1980) that when farmers are able to produce all goods they consume, trade always increases their welfare even in the presence of volatility. That said, the lack of risk sharing between agents producing different types of goods—either within the agricultural sector because farmers are endowed with different types of land or across sectors—is an important mechanism through which trade may have deleterious second-moment effects; see e.g. Rodrik (1997). We partially address this concern in our quantification by extending the model to include an urban manufacturing sector that allows the possibility of welfare losses for farmers who demand manufactures that they cannot obtain in autarky. More generally, our paper incorporates the intuition developed in these seminal works into a quantitative trade model that is sufficiently flexible (e.g. by incorporating many goods with arbitrary variances and covariances of returns and flexible trade costs) to be taken to the data.

Recently, several papers have explored the links between macro-economic volatility and trade, see e.g. Easterly et al. (2001); di Giovanni and Levchenko (2009); Karabay and McLaren (2010); Lee (2018). Our paper, in contrast, focuses on the link between micro-economic volatility—i.e. good-location specific productivity shocks—and trade. Similar in this regard, and most closely related to our paper, are the works of Burgess and Donaldson (2010, 2012) and Caselli et al. (2019). Burgess and Donaldson (2010, 2012) use an Eaton and Kortum (2002) framework to motivate an empirical strategy that studies the relationship between famines and railroads in colonial India finding, like us, declines in the responsiveness of local prices and increases in the responsiveness of real income to rainfall shocks.[3] Caselli et al. (2019) also use an Eaton and Kortum (2002) framework to quantify the relative importance of sectoral and cross-country specialization in a world of globally sourced intermediate goods. We see our paper as having three distinct contributions relative to these papers. First, we depart from the Eaton and Kortum (2002) framework and develop an alternative quantitative general equilibrium framework that allows us to analyze the pattern of trade while more closely matching several important characteristics of the empirical setting we consider (e.g. homogenous goods, a hierarchical trading network, and heterogeneous traders). Second, by embedding a portfolio allocation decision where real returns are determined in a general equilibrium trade setting, we characterize the endogenous response of agents to trade-induced changes in their risk profile. Third, we empirically validate that farmers are responding as the model predicts.

The paper is also related to a growing literature applying quantitative trade models to the study of agriculture in the absence of volatility. Sotelo (2020), Costinot and Donaldson (2016),

[2]Eaton and Grossman (1985) and Dixit (1987, 1989a,b) extend the theoretical analysis of Newbery and Stiglitz (1984) to incorporate imperfect insurance and incomplete markets.

[3]Despite focusing on intra-national trade in the same country, India, there are also important differences between modern India and the colonial setting studied by Burgess and Donaldson (2010, 2012), most notably that trade costs seem if anything to have risen between the tail end of the Colonial period and the start of our sample, 1970. As evidence for this claim, we find that local rainfall shocks affect local prices at the start of our sample period (consistent with substantial barriers to trade across locations), while Donaldson (2018) finds they did not post railway construction in his Colonial India sample (consistent with low barriers to trade across locations).

Costinot et al. (2016), and Bergquist et al. (2019) examine how trade affects crop choice. In these models, locations grow multiple crops due to heterogeneity in the productivity of different plots (in contrast to wanting to diversify against risk, as in our model). As in Allen (2014), we relax the no-arbitrage condition, although here we do so by allowing for heterogeneous traders with varying trade costs rather than information frictions. As in Chatterjee (2020), traders play an important role in determining equilibrium prices, although here we abstract from farmer-trader bargaining and instead focus on the role of volatility and its affect crop choices.

Finally, the paper relates to three strands of the economic development literature. First, we follow a long tradition of modeling agricultural decisions as portfolio allocation problems (see e.g. Fafchamps (1992); Rosenzweig and Binswanger (1993); Kurosaki and Fafchamps (2002)). Second, we build on a substantial development literature examining the effect that access to formal credit has on farmers (see e.g. Burgess and Pande (2005) and Jayachandran (2006)). Third, we add to a primarily reduced form literature analyzing the impacts of infrastructure investment (e.g. Duflo and Pande (2007) for dams and Asher and Novosad (2020) for village roads, both in India). We contribute to these literatures in three ways: first, our rich data allows us to characterize the optimal crop choice using the observed mean, variance, and covariance of yield shocks across crops; second, we demonstrate that rural bank access leads farmers to choose riskier crop portfolios; and third, we examine the interaction between rural bank access and domestic infrastructure policy.

The remainder of the paper is organized as follows. In Section 2, we describe the empirical context and the data we have assembled. Section 3 presents three new stylized facts relating trade to volatility and the resulting responses by farmers. In Section 4, we present the baseline model, show that it is consistent with the reduced form results, and analytically characterize the second moment welfare effects of trade. In Section 5, we structurally estimate the extended "quantitative" version of the model and quantify these welfare effects. Section 6 concludes.

2 Empirical context and data

2.1 Rural India over the past forty years

This paper focuses on rural India over a forty year period spanning 1970 to 2009. The majority of rural households derive income from agriculture; 85% of the rural workforce was in agriculture in the 1971 Census and 72% in the 2011 Census. Over this period, there were three major developments that had substantial impacts on the welfare of rural Indians. The first set of changes were to the technology of agricultural production. Increased use of irrigation and high-yield varieties (HYV) raised mean yields and altered the variance of yields.[4] The second major change was the policy-driven expansion of formal banking into often unprofitable rural areas (see Burgess and Pande

[4]Irrigation coverage rose from 23 to 49% of arable land and HYV use rose from 9 to 32% in the VDSA data we introduce shortly. Some HYV crops had lower variance due to greater resistance to pests and drought, others higher due to greater susceptibility to weather deviations, see Munshi (2004). While we find no evidence that the adoption of HYV is correlated with improvements market access (see footnote 22), our analysis will flexibly incorporate observed changes in the means, correlations, and covariances of crop yields over time.

(2005) and Fulford (2013)).[5] The availability of credit helped farmers smooth income shocks and so provided a form of insurance.[6]

The third set of changes relate to reductions in inter- and particularly intra-national trade costs. The reductions were driven by two types of national policy changes. The first—which we will exploit extensively in the empirical analysis—were major expansions of the Indian interstate highway system including the construction of the 'Golden Quadrilateral' between Mumbai, Chennai, Kolkata and Delhi and the 'North South and East West Corridors'.[7] The result was that over the sample period, India moved from a country where most freight was shipped by rail to one dominated by roads—in 1970 less than a third of total freight was trucked on roads, four decades later road transport accounted for 64% of total freight.[8] The second policy change was the broad economic liberalization program started in 1991 that gradually reduced agricultural tariffs both across-states within India (see discussion in Atkin (2013)) and with the outside world. This paper focuses on domestic trade, that is the inter-state and intra-state trade that constituted the overwhelming majority of India's agricultural trade over our sample period.[9]

2.2 Agricultural trade in rural India

Agricultural trade in rural India has remained relatively unchanged since the 1960s, when the Agricultural Produce Marketing Committee (APMC) Acts were passed by Indian states. The APMC Acts established state-level marketing boards to regulate the trade of agricultural commodities, which in turn created state-regulated markets for agricultural trade called *mandis*—located in large towns near production centers—where farmers were legally required to sell their goods.[10]

The basic structure of the trading process is as follows. Upon harvest, farmers either consume their produce directly or sell it to local traders in their village who transport it to the district mandi.[11] At the mandi, the local traders sell the produce to (larger) regional traders who transport it to terminal markets in the state (or in some cases outside the state), which are typically located

[5] Basu (2006) and Shah et al. (2007) document that this expansion increased both the number of loans taken out and the deposits made in rural areas, and the share of rural household debt from banks rose from 2.4% to 29% between 1971 and 1991. By 2003, 44% of large farmers (55% of India's agricultural land), 31% of small farmers (40% of land) and 13% of marginal farmers (15% of land) had an outstanding loan from a formal bank.

[6] India also has a subsidized crop insurance scheme although, even today, only 6% of farmers voluntarily purchase coverage (a further 11% have agricultural loans with mandatory insurance requirements, see Mahul et al. (2012)).

[7] See Datta (2012); Ghani et al. (2016); Asturias et al. (2018) for estimates of the effect of the "Golden Quadrilateral" on firm inventories, manufacturing activity, and firm competition, respectively.

[8] These figures are Indian government estimates from the 10th, 11th and 12th five-year-plans.

[9] External agricultural trade remained subject to a restrictive license system until April 2001. Focusing on the three most traded products—rice, sugar and wheat—external trade (international exports plus imports) equaled 0.5, 0.3 and 11% of production by weight in the 1970s, and 2.8, 0.7 and 3% in the 2000s, respectively. Unfortunately, India only records internal trade by rail, river and air (recall road accounted for between one and two thirds of freight); and then only for trade between 40 or so large trading blocks in India. Using rail, river and air data that likely severely underestimate inter-district trade, internal trade equaled 3.8, 1.3 and 21.4% of production by weight in the 1970s, and 10.2, 0.9 and 16.3% in the 2000s.

[10] The typical Indian district has one mandi; e.g. in 2006 there were 610 mandis selling rice (paddy) across 600 districts. In 2003, the Indian government proposed that states allow producers to sell outside the mandi system, but most states opted not to implement major changes to their APMC Acts (see e.g. Gautam (2015)). As mentioned in the introduction, subsequent attempts at reform by the national government have led to ongoing protests and ultimately capitulation by the government.

[11] Chatterjee et al. (2020) find that only the largest farmers transport their produce to the mandi themselves.

in large cities where the produce is processed for retail consumption.[12] The result is a *hierarchical* trading network illustrated in panel (a) of Figure 1. Many farmers trade at the village level, many villages trade at the mandi level, many mandis trade at the state level, and many states trade at the national level, intermediated by traders at all but the bottom level. Unlike models where all locations trade directly with each other, our model below incorporates this more realistic hierarchical structure.

In addition to the hierarchical trading structure, there are several other important characteristics of agricultural trade in rural India that should be emphasized. First, these agricultural goods are best viewed as homogeneous. In each mandi, there is a market price for each type of good and that price exhibits very little variation across transactions on a given day.[13] Second, traders not only engage in arbitrage when purchasing farmers' production, they also engage in arbitrage when selling (processed) agricultural goods for consumption.[14] Third, farmers take market prices as given, with traders earning any profits resulting from arbitrage (see Goyal (2010), Chand (2012), Mitra et al. (2018), and Chatterjee (2020)). Fourth, traders exhibit a large degree of heterogeneity in their scale, varying from small traders who have no capital and incur large trade costs to transport goods (e.g. renting a tractor to transport produce to the mandi) to large multinational corporations (see, e.g. Upton and Fuller (2004)).[15] In our model below, we will incorporate each of these aspects of India's trading system: modeling goods as homogeneous (instead of as aggregates of a product with infinite varieties, as commonly assumed), having farmers take prices as given and traders earning profits (instead of perfect competition in the transport sector, as commonly assumed), and allowing traders to have heterogeneous trade costs (instead of homogeneous trade costs, as commonly assumed). Finally, we note that while the model developed below has been tailored to our empirical context, the characteristics above are common in agricultural settings throughout the developing world, suggesting its broader applicability.[16]

2.3 Data

We assemble the following dataset on agricultural production and trade costs covering the entirety of the forty year period 1970-2009:

[12]For staple crops, local traders also have the option of selling to the public distribution systems (PDS), which guarantees a minimum support price (MSP). In practice, however, there are costs associated with selling to the PDS, including the uncertainty of when payment arrives. As a result, the MSP is typically not binding. Appendix Figures A.4–A.7 plot the distribution of log prices alongside the MSPs for applicable crops. There is little evidence of price heaping just at or above the MSPs, and substantial mass below the MSPs, suggesting any attenuation from abstracting from MSPs in our model is limited.

[13]For example, for mustard, paddy and wheat where we observe daily mandi-level prices from 2006 onward, the median difference between the max and min price divided by the mode on a given day was 0.04, 0.06, and 0.07.

[14]The traders engaged in arbitrage when selling goods to farmers for consumption need not be the same individuals as those engaged in arbitrage when buying from farmers, but sometimes they are. For example, Chatterjee et al. (2020) find that 39% of local traders in their sample also own shops in the village.

[15]While trader size increases with the level of the hierarchy, there also exists substantial heterogeneity within the same level; see e.g. Chand (2012), and Kapur and Krishnamurthy (2014).

[16]For example, Bergquist et al. (2021) document hierarchical trading networks for maize in Uganda, Grant and Startz (2021) highlight the prevalence of chains of intermediation in Nigeria, Bergquist and Dinerstein (2020) and Dhingra and Tenreyro (2020) document imperfect competition amongst agricultural traders in Kenya, and Allen (2014) documents the presence of heterogeneous agricultural traders in the Philippines.

Agricultural Data: Data on district-level cropping patterns (i.e. the area allocated), crop prices[17] and crop yields come from the ICRISAT Village Dynamics in South Asia Macro-Meso Database (henceforth VDSA) which is a compilation of various official government datasources. Cropping patterns, prices, and yields are all observed at the district $\times$ crop $\times$ year level for 311 districts (using time-invariant 1966 district and state boundaries) in 19 states that contain 95% of India's population. The database spans the 1966-67 crop year through 2009-10 and covers the 15 major crops for which farm harvest prices are available.[18] For comparability, all Rupee values are deflated by the all-India CPI.

Trade Costs: We obtained all editions of the government-produced *Road Map of India*, published in the years 1962, 1969, 1977, 1988, 1996, 2004 and 2011. We digitized and geo-coded these maps and identified the highways using an algorithm based on the color of digitized pixels. Figure 2 depicts the substantial expansion of the the Indian highway network over the forty year period. Using these maps, we construct a "speed image" of India, assigning a speed of 60 miles per hour on highways and 20 miles per hour elsewhere. This image allows us to calculate travel times between any two districts in each year using the Fast Marching Method (see Sethian (1999)).[19]

Rural Bank Data: Data on rural bank access, an important source of insurance in India, come from RBI bank openings by district assembled by Fulford (2013).

Consumer Preferences: Consumption data come from the National Sample Survey (NSS) Schedule 1.0 Surveys produced by the Central Statistical Organization.

Rainfall Data: Gridded weather data come from Willmott and Matsuura (2012) and were matched to each district by taking the inverse distance weighted average of all the grid points within the Indian subcontinent.

3 Trade and volatility: Stylized facts

In this section, we present three sets of stylized facts.

3.1 Prices and trade

We first demonstrate the key mechanism linking trade and volatility.

[17]These are the farm harvest prices—i.e. the farm gate price a farmer receives.

[18]The 15 crops are barley, chickpea, cotton, finger millet, groundnut, linseed, maize, pearl millet, pigeon pea, rice, rape and mustard seed, sesame, sorghum, sugarcane, and wheat. Where paddy rather than rice prices are available we convert to rice prices using the average ratio of the paddy to rice price in that state and decade. These 15 crops accounted for an average of 73% of total cropped area across districts and years. The data are at the annual level and combine both the rabi and kharif cropping seasons. Data coverage across crops with districts is good: in the median district-decade pair, we observe at least one year of production data for 13 of the 15 crops and at least one year of price data for 11 of the 15 crops. In the 5% of cases where we observe prices but yields are missing, we interpolate yields by taking weighted averages of non-missing yields in all districts for the same crop and year using inverse squared-log-distance weights. To deal with extreme yield values that are likely erroneous, we winsorize the resulting yields at the 1st and 99th percentiles. We do not interpolate missing prices, as they are equilibrium outcomes, an issue we return to when calculating total revenues in Section 3.3.

[19]We linearly interpolate travel times in years between editions of the *Road Map of India*. See Allen and Arkolakis (2014) for a previous application of the Fast Marching Method to estimate trade costs.

Stylized Fact 1(a): As trade costs with other locations in the state fall, prices respond less to local yields...

We first show that district-level prices are inversely related to district-level yield shocks, as a supply and demand model would predict, and that this responsiveness is attenuated as trade costs fall.

To do so, we regress district-level log prices on log yields and explore how the yield coefficient—the elasticity of price to yield—changes with reductions in the costs of trading with other locations:

$$\ln p_{igtd} = \beta_1 \ln A_{igtd} + \beta_2 \ln A_{igtd} \times MA_{id}^{instate} + \gamma_{gtd} + \gamma_{igd} + \gamma_{it} + \nu_{igdt}, \tag{1}$$

where $\ln p_{igtd}$ is the price in district i of good g in year t decade d, and $\ln A_{igtd}$ is the local yield. The variable $MA_{id}^{instate}$ captures district i's trade openness in decade d—as measured by market access to other districts in the state with the precise definition provided below. To control for confounds, we include three sets of fixed effects: a crop-year fixed effect γ_{gtd} that controls for changes in national or world prices of the good; a district-crop-decade fixed effect γ_{igd} that controls for slow-moving changes in crop-specific costs, in the area allocated to the crop, in preferences, or in technologies; and a district-year fixed effect γ_{it} that controls for local cost or demand shocks common to all crops (and sweeps out the level effect of market access). Finally, here and in the later facts we make our results representative of rural India by weighting observations by the total area planted with our 15 crops in each district. We note that our specification, including the choice of fixed effects, will match the expression we derive for equilibrium prices in Section 4 below.

Our district-decade measure of openness derives from the digitized road maps described in Section 2.3. Recall that the digitized maps allow us estimate the travel time between any two points in India in any year. Motivated by the hierarchical structure of India's trading network described in Section 2.2, we consider within-state market access. (We explore the relevance of national market access below.) Following Donaldson and Hornbeck (2016), we construct an annual within-state market access measure for district i in year t by taking a weighted sum of the (inverse) bilateral travel times to other districts in i's state, the set S_i, as follows:

$$MA_{it}^{instate} = \sum_{j \in S_i} \left(\frac{1}{travel\ time_{ijt}^{\phi}} \right) Y_{jt}, \tag{2}$$

where Y_{jt} is the income of district j in period t (proxied by the total agricultural revenues from our dataset) and $\phi > 0$ determines how quickly market access declines with travel time. Higher values of market access correspond to greater trade openness as districts are able to trade more cheaply with districts where demand is high. Averaging district-year values within each decade provides us with our $MA_{id}^{instate}$ interaction term.[20]

To parameterize ϕ we draw on the gravity literature that measures how rapidly log trade flows decline with log distance. Following the meta-analyses of Disdier and Head (2008) and Head and Mayer (2014), we set $\phi = 1.5$—the average gravity coefficient for developing country samples—in our

[20] We take decadal averages to align with the later stylized facts and our quantitative analysis. Results are robust to using annual market access variation.

preferred market access specification.[21] We also consider $\phi = 1$, a natural benchmark and close to the average of 1.1 found for the all country sample, as well as alternate estimates of the off-highway speed of travel (1/4 of highway speed rather than 1/3) for robustness.

Since farmers may invest more care harvesting crops that have high prices, yields are likely to respond positively to price shocks, exerting an upward bias on the yield elasticity. To deal with this endogeneity concern, we instrument local yields with rainfall-predicted yields. Specifically, we regress log yields on local rainfall shocks in each month of that year interacted with state-crop fixed effects and include the same fixed effects as in the specification above. This generates a predicted yield measure that, after conditioning on the fixed effects, depends only on rainfall realizations and time-invariant parameters (and hence is unaffected by changes in the production technology over time). Predicted yields interacted with market access serve as our instrument for the interaction term. The instruments are very strong with a Kleibergen-Paap first stage F-stat above 2000.

In order for the coefficients on the interaction between yield and market access to be interpreted causally, we further require that road building does not respond to changes in the elasticity of yields to prices after controlling for the rich set of fixed effects. Such endogeneity concerns are mitigated by the fact that, as detailed in Section 2.1, much of the highway construction was part of centrally-planned national programs designed to connect larger regions rather than improve within-state market access. Reassuringly, changes in our market access measure are not associated with changes in relevant district characteristics; see Appendix Table A.1.[22] Yet concerns may still remain, so to address potential confounders we interact yields with various sets of fixed effects below.

Columns 1 and 2 of Table 1 present the OLS and IV estimates of the regression in equation (1). As we would expect, a positive shock to supply lowers prices ($\beta_1 < 0$), with the coefficient becoming more negative after instrumenting yields (consistent with the upward bias in the OLS discussed above). More central to our analysis, the elasticity of local prices to local yields increases significantly—from negative values towards zero—with improvements in market access. That is, as trade costs fall, the role that local prices play in insuring against yield shocks (i.e. prices rising when yields are low) is weakened. In terms of magnitudes, using our IV specification, a rise in market access equal to the median 1970-2009 change in within-state market access raises the elasticity by 0.017 (from a mean in the 1970s of -0.047).

These findings are robust to alternative market access measures, either setting $\phi = 1$ (column 3) or lowering the ratio we assume between on- and off-highway speeds (column 4). They are also robust to controlling for crop-specific technological changes or differences in crop suitability across districts that are correlated with market access and potentially affect the yield elasticity: columns

[21] Head and Mayer (2014) report an average coefficient on log distance of -1.1 across 159 papers and 2,508 regressions while Disdier and Head (2008) reports that estimates from developing country samples are lower by an average of 0.44—consistent with distance being more costly in developing countries as found in Atkin and Donaldson (2015).

[22] Specifically, Appendix Table A.1 uses the same specification we use to analyze district-decade level outcomes below and finds no significant associations between within-state market access and district-decade level characteristics including banks per capita, the (crop-share weighted) mean and variance of log yields, or the proportion of land planted with high yield varieties or under irrigation. Even if there were, the main effect of market access is swept out by the district-decade fixed effects and so such associations would not necessarily lead to bias.

5 and 6 include interactions with log yield and either the full set of crop-decade fixed effects or the full set of crop-district fixed effects.[23]

Finally, we explore whether the reduced responsiveness of prices to local yields also depends on changes in trade costs with locations outside the state. Column 7 supplements the specification in (1) with an interaction between local log yields and outside-state market access, $MA_{id}^{outstate}$, calculated identically to within-state market access but now summing the inverse bilateral distances over all locations outside the state rather than inside. The coefficient on the interaction is small and insignificant, consistent with restrictions on interstate commerce that motivate the hierarchical structure of India's trading network described in Section 2.2 and suggest that district prices should respond foremost to market conditions elsewhere within the same state.

Stylized Fact 1(b): ... and prices respond more to yields elsewhere, particularly elsewhere within the state.

Reductions in trade costs also raise the responsiveness to yields in other districts, particularly those within the same state. To demonstrate this, in column 8 of Table 1 we amend the specification in equation (1) to further include the log of the area-weighted average yields in the other districts within the same state, $\ln A_{-i,sgtd}$, as well as its interaction with within-state market access. Local prices decline with high yields elsewhere, with prices becoming significantly more responsive to yields elsewhere (i.e. decline more) with increases in market access.

Column 9 further includes the log of the area-weighted average of yields in districts in other states and its interaction with outside-state market access (alongside the $\ln A_{igtd} \times MA_{id}^{outstate}$ interaction introduced above). We find that when access to markets in other states improves, local prices respond less to local yields and respond more to other state's yields. But the former effect is again not significant, and the increased responsiveness to national yields is smaller in magnitude than that to state yields, consistent with the hierarchical trading network.

3.2 Crop choices and trade

Our second set of stylized facts provides evidence that farmers respond to declines in trade costs by trading off traditional first-moment gains from specialization with second moment risk-reduction strategies, consistent with a portfolio choice model.

[23] As running a specification with 100s of endogenous variables and 100s of instruments is both infeasible and inadvisable, here we present the reduced form that replaces yields with predicted yields. That said, we note that the coefficient on the interaction between predicted yield and market access potentially combines the effects of market access on the elasticity of price to yield and any effect of market access on yield-increasing technologies. This is not an issue for the coefficients on the interaction between yield and market access in the OLS and IV regressions since in those cases such yield-improving technologies are already captured through measured yields (OLS) or the first stage regressions (IV). As we find a positive coefficient on the interaction between predicted yield and market access in the first stage for log yields corresponding to column 5 and a negative one for column 6, the reduced form will tend to be too small and too large, respectively.

Stylized Fact 2(a): As trade costs fall, farmers reallocate their land toward crops for which they have a comparative advantage and away from crops that are more risky.......

We first regress the share of land allocated to each crop on the mean and variance of yields of that crop, both interacted with our within-state market access measure $MA_{id}^{instate}$ introduced above:

$$\text{arcsinh}\theta_{igd} = \beta_1 \mu_{igd}^{A} + \beta_2 \sigma_{igd}^{2,A} + \beta_3 \mu_{igd}^{A} \times MA_{id}^{instate} + \beta_4 \sigma_{igd}^{2,A} \times MA_{id}^{instate} + \gamma_{gd} + \gamma_{id} + \gamma_{ig} + \varepsilon_{igd}, \quad (3)$$

where $\text{arcsinh}\theta_{igd}$ is the inverse hyperbolic sine of the decade-d average share of cropped land planted with crop g in district i, μ_{igd}^{A} is the mean of log yields in that district-crop-decade, and $\sigma_{igd}^{2,A}$ is the variance of log yields in that district-crop-decade.[24] We saturate the model by including crop-decade, district-decade, and district-crop fixed effects. These control for both national crop-specific trends, district-decade level shocks, and persistent differences in local agroclimatic conditions that could potentially be related to local agricultural technologies and hence bias the β coefficients. As crop choices are not independent, standard errors are clustered at the district-decade level. Finally, to ensure our results are representative at the district-decade level, we again weight observations by the total area planted in the district.

As in Fact 1 above, our choice of specification—including the mean and variance of log yields as independent variables and the choice of fixed effects—arises directly from the expression we will derive for the equilibrium crop choice in Section 4. The one departure is the use of the inverse hyperbolic sine transformation in lieu of logging crop shares given that 19% of crop share observations in our regression sample are equal to zero.[25]

To allay worries about endogenous movements in yields in response to cropping decisions—for example cropping more marginal lands which alters the mean and variance of yields on a representative plot—we instrument for the mean and variance of log yields with the mean and variance of log yields as predicted by rainfall variation.[26] These instruments interacted with market access serve as instruments for the two interaction terms in equation (3). Reassuringly, the instruments are strong with a Kleibergen-Paap first stage F-stat of 117.

The OLS and IV regression coefficients are shown in Columns 1 and 2 of Table 2. The significant positive β_3 coefficient for both the OLS and IV implies that as trade costs fall—and hence market access improves—farmers respond by reallocating land toward crops in which they are relatively more productive. The significant negative β_4 coefficient indicates that a fall in trade costs also leads

[24]Note that, unlike the variance of yields, the variance of log yields is mean independent. As the variance of log yields terms display extreme values, we winsorize this variable at the 1st and 99th percentiles. Appendix Table A.2 shows that while our instruments have a weaker first stage F-stat without this adjustment, the coefficients are essentially unaffected and remain statistically significant.

[25]Logging crop choice—thereby dropping the zero-share observations—will generate a selected sample (unlike in our model where no crop will be allocated exactly zero land). Appendix Table A.2 reports very similar results using either the actual crop share as the dependent variable or the log of the crop share with zero-crop-share values replaced by the first percentile value of non-missing log crop shares (along with non-missing values below the first percentile).

[26]We use a method analogous to the predicted yields in Fact 1: we regress log yields on local rainfall shocks for each month interacted with state-crop fixed effects, controlling for the fixed effects in specification (3), and use the mean and variance of these predicted values. We then winsorize the variance terms at the 1st and 99th percentiles as we do for the endogenous variances.

farmers to reallocate towards crops that have lower variances of yields. In terms of magnitudes, using the IV coefficients, a fall in trade costs equal to the median 1970-2009 change in within-state market access increases the responsiveness to mean log yields by one fifth and changes the responsiveness of crop choice to the variance of log yields from a slightly positive one (coefficient 0.010) to a slightly negative one (coefficient -0.004).[27] Our baseline focuses on the mean and volatility of yields, which are the exogenous variables in our theory, but similar results obtain when replacing yields with the value of production in column 3,[28] and for the two alternative market access measures (see Appendix Table A.3).

Together, these results suggest that farmers are not only responding to trade cost declines by specializing in high yield crops in which they have a comparative advantage—the traditional "first moment" effects—but also by reallocating land toward crops that are less risky—a "second moment" effect of trade on risk mitigation that our portfolio allocation model below will emphasize.

Farmers may also engage in hedging and allocate more land to crops whose yields are less correlated with other crops in order to mitigate the increase in risk they face due to reductions in trade costs. To test this additional "second moment" effect, we supplement equation (3) with $\sum_{g'\neq g}\sigma^A_{igg'd}$—the sum of the covariance of log yields of crop g with the log yields of each of the other 14 crops—and its interaction with market access.[29] As a portfolio choice model would predict, column 4 of Table 2 shows a negative and significant coefficient on the interaction between the covariance term and market access, with a magnitude similar to that on the variance interaction.

Finally, column 5 extends equation (3) by including interactions with outside-state market access. Echoing earlier findings, increases in market access outside of the statemove farmers further toward low variance crops but to a lesser degree than increases in within-state market access (although movements to higher-mean crops appear more pronounced).

Stylized Fact 2(b): ... with the riskiness of yields mattering more in locations with worse access to risk mitigation technologies.

The second fact in this section shows that the degree to which farmers trade off the "first moment" and "second moment" forces when choosing their crop allocation depends on their access to risk mitigation technologies. As discussed in Section 2.1, the presence of local rural bank branches provides an important form of insurance as farmers can take out loans in bad times and repay them in good times. Fact 2(b) explores how this insurance technology affects crop choices by allowing the responsiveness to the variance (and covariance) of log yields to depend on bank access.

Specifically, we extend specification (3) to include the triple interaction of the variance of log yields, market access and bank access—measured as the decadal average of rural banks per capita

[27] As we show in Section 4.3, that farmers allocate more land to high volatility crops in the 1970s when trade costs are highest is consistent with high volatility crops also being the crops that farmers particularly liked to consume.

[28] We calculate crop value by multiplying yields by (crop-area-weighted) state-level average prices, both because of missing price data issues (see footnote 32) and because Section 4.3 will show that this is the theoretically-consistent price to use. As above, we instrument means and variances of log values with means and variances of predicted yields multiplied by state average prices that leave out own-district prices to avoid reverse causality.

[29] As for the variance terms, we instrument with the sum of the covariances of rainfall-predicted log yields and this sum interacted with market access having first winsorized the sum of covariance terms at the 1st and 99th percentiles.

in a district—alongside the double interaction of the variance and banks term (with the bank and market access interaction swept out by the district-decade fixed effects). Once again, we instrument the interaction terms with similar terms that replace the variance of log yields with the variance of log predicted yields.[30] The first stage F-stat remains strong with a value of 78.6.

The estimates are shown in Column 6 of Table 2 with the triple interaction positive and significantly different from zero at the 1% level. Similar results obtain for the two alternative market access measures in Appendix Table A.3.[31] Consistent with farmers being willing to bear more risk if insured, the presence of more insurance options attenuated the movement into less risky crops that resulted from reductions in trade costs. That is, the better the bank access, the more important the "first moment" effects of trade on specialization and the less important the "second moment" effects of trade on risk mitigation. In terms of magnitudes, the increase in the responsiveness to the variance of yields that result from better market access shrinks by 40% when going from the 25th percentile of banks per capita to the 75th.

Column 7 repeats this exercise with the covariance of log yields terms introduced in Fact 2(a) with the triple interaction of the covariance with banks and market access also positive and significant at the 10% level. Finally, column 8 includes both within-state and outside-state market access measures. These insurance interactions are absent for market access outside of the state, providing further evidence for the primacy of within-state market access.

3.3 Volatility and trade

Our third set of stylized facts captures the net impact of the mechanisms highlighted in Facts 1 and 2 by exploring the offsetting effects of reductions in trade costs on income and price index volatility.

Stylized Fact 3(a): As trade costs fall, farmers' revenue volatility increases...

First, we calculate nominal (gross) income—i.e. the total revenue from the production of all 15 crops—using annual data on agricultural revenues per hectare.[32] Of course, these are gross of crop costs which may change over time—an issue we confront head on in the structural estimation below.

[30] A remaining concern is that bank branch placement is endogenous. One potential mechanism is that rural banks are attracted to locations with improved market access (although to bias our estimate of the triple interaction, given our main effects and fixed effects, this entry would need to be focused on places where district-crop-decade preference shocks increased relative demand for more volatile crops). Empirically, however, it turns out the two are uncorrelated in our context: across the 311 districts, the correlation between the change in market access and rural banks per capita between the 1970s and 2000s is -0.0832 and the association is negative and insignificant when using district-decade variation as shown in Appendix Table A.1 (in part because of government mandated bank expansions into less profitable locations documented in Burgess and Pande (2005)).

[31] Appendix Table A.3 also shows that these results are robust to including interactions between banks, market access and the mean of log yields. Appendix Table A.2 reports similar results not winsorizing the variance terms and using alternative transformations of crop shares.

[32] Unlike in Facts 1 and 2 where missing price data generates missing district-crop level observations, here missing prices are more pernicious as they result in missing components of revenue. As Fact 1 shows, prices are endogenous to local yields and market access, making imputation using prices elsewhere infeasible. To mitigate the most troubling cases, we restrict attention to the 96.4% of district-decade pairs where at least 25% of area cropped has non-missing prices. Instead of restricting attention to a subset of district-decades with better data, Appendix Table A.5 follows our Fact 2 strategy of winsorizing the variances of log nominal income at the 1st and 99th percentile (along with the variance of the price index and real income we explore below). Results are qualitatively similar although the coefficient shrinks for the variance of nominal income and is magnified for the variance of the price index.

To explore how the volatility of nominal income—i.e. revenue volatility—responded to reductions in trade costs, for each district and decade we calculate $var(\ln nominal\ income)_{id}$. We then project this object onto within-state market access:[33]

$$var(\ln nominal\ income)_{id} = \beta_1 MA_{id}^{instate} + \gamma_i + \gamma_{sd} + \varepsilon_{id}. \quad (4)$$

District fixed effects γ_i control for persistent differences in volatility while state-decade fixed effects γ_{sd} control for temporal changes common to markets within a state. Note that here we are unable to exploit variation across crops within a district and time period as we did in Facts 1 and 2, making endogeneity concerns more substantial even with the inclusion of time trends at the lowest possible level, i.e. state-decade. Thus, we should be more cautious in interpreting the following results as causal (and these concerns motivate the need for the quantitative results in Section 5 that isolate the effects of trade cost reductions alone). That said, it is reassuring that $MA_{id}^{instate}$ is uncorrelated with banks, yields, or yield-improving technologies; see Appendix Table A.1.

Column 1 of Table 3 reports the estimated β_1 coefficient.[34] Consistent with planting reallocations (Fact 2) only partially mitigating the reduced responsiveness of prices to yield shocks (Fact 1), the variance of log nominal income rises significantly with increases in market access. In terms of magnitudes, a rise in within-state market access equal to the median change in market access between the 1970s and 2000s increases the variance of log revenue by an amount equal to 78% of the mean 1970s variance. (The average variance rose 3.2 fold over this period with our reductions in trade costs accounting for 45% this rise.) Column 4 repeats the exercise but further including outside-state market access. Echoing earlier findings, improvements in within-state rather than outside-state market access drive these results.

Stylized Fact 3(b): ... and the volatility of their price index declines...

To explore the impact of reductions in trade costs on the volatility of farmers' price indices, we construct for each district a Cobb-Douglas price index over the 15 crops in our sample. We obtain district-level expenditure shares for each of these crops from the 1987 household-level NSS surveys and these (constant) shares serve as the weights in the Cobb Douglas price index.[35]

Column 2 of Table 3 replaces the dependent variable in (4) with $var(\ln CD\ Price\ Index)_{id}$, the district-decade variance of the log price index. Consistent with the reduced responsiveness of prices to yields documented in Fact 1—and in contrast to the rising volatility of nominal income—the coefficient on market access is negative, i.e. the price index becomes less volatile with reductions in trade costs. While only about a third the size of the effect on revenue volatility, the coefficient is significantly different from zero with a p-value of 0.110.[36] Column 5 shows that, once again, it is

[33]As above, we make our results representative by weighting observations by the total area planted with our 15 crops in each district-decade. Recall that we only have revenue data for the 73% of arable land devoted to these 15 crops. Thus, this analysis provides a somewhat incomplete picture of the revenue volatility faced by Indian farmers.

[34]Appendix Table A.5 presents qualitatively similar results using our two alternative measures of market access.

[35]Specifically, $\ln CD\ Price\ Index_{it} = \sum_g bshare_{ig} \ln p_{igt}$ where $CD\ Price\ Index_{it}$ is the Cobb Douglas price index.

[36]As foreshadowed in footnote 32, the p-value falls to 0.087 when we winsorize the variance of the price index rather than dropping district-decades with poor data in Appendix Table A.5.

within-state rather than outside-state market access driving this drop.

Stylized Fact 3(c): ... with the volatility of real income rising on net.

Finally, we turn to impacts of reductions in trade costs on the volatility of real income, the ratio of nominal income and the Cobb-Douglas price index introduced above. Consistent with the observed rise in the volatility of nominal income coupled with a smaller decline in the volatility of the price index, column 3 of Table 3 shows that real income volatility increases with within-state market access. The coefficient on market access falls by 38% compared to the nominal income specification but the estimate is still positive and statistically significant. Column 6 shows that this rise is once again driven by within-state rather than outside-state market access.

To sum up, we have shown that falling trade costs reduce the responsiveness of prices to local yields but increased the responsiveness of local prices to prices elsewhere (Fact 1). Farmers respond by changing their crop allocations—trading off "first moment" gains from specialization against "second moment" strategies to mitigate risk (Fact 2), which led overall to an increase in the volatility of farmers' real income (Fact 3). We now turn to presenting a model that is both sufficiently tractable to generate these comparative statics and sufficiently flexible to quantify the welfare impact of the Indian highway expansion.

4 Modeling trade and volatility

In this section, we develop a general equilibrium model of trade and volatility. Matching our empirical context described in Section 2, farmers in many villages produce and consume a finite number of homogeneous agricultural goods while traders engage in price arbitrage between villages and markets. We circumvent difficulties due to corner solutions by assuming these traders face heterogeneous transportation costs. This assumption yields tractable expressions for prices and patterns of specialization and allows us to incorporate volatility by applying tools from the portfolio allocation literature. The resulting model generates qualitative predictions consistent with the three stylized facts and delivers intuitive expressions that uncover our key structural parameters. Section 5 extends this model to add greater realism before bringing it to the data in order to quantify how volatility affects the gains from trade.

4.1 Model setup

Geography: There are a large number of locations ("villages") indexed by $i \in \mathcal{N}$ and a central market. Each village i is inhabited by a measure L_i of identical farmers who produce and consume goods in their village i. The central market is inhabited by a set of heterogeneous traders who engage in an arbitrage process (described below) and drivers who are hired by the traders to ship goods between the central market and each of the villages.

Production: There are a finite number of homogenous goods ("crops") indexed by $g \in \{1,...,G\} \equiv \mathcal{G}$ that can be produced in each location i. Land is the only factor of production. Each farmer in each location is endowed with a unit of land and chooses how to allocate that land across the production of each of the G crops. Let θ_{ig}^f denote the fraction of land farmer f living in village i allocates to good g, where $\sum_{g\in\mathcal{G}}\theta_{ig}^f=1$; we refer to $\left\{\theta_{ig}^f\right\}_{g\in\mathcal{G}}$ as farmer f's *crop choice.*

Production is risky. Let the (exogenous) yield of a unit of land in location i for good g be $A_{ig}(s)$, where $s\in S$ is the state of the world. Given her crop choice, the nominal income farmer f receives in state $s\in S$ is:

$$Y_i^f(s)=\sum_{g\in\mathcal{G}}\theta_{ig}^f A_{ig}(s)p_{ig}(s), \tag{5}$$

where $p_{ig}(s)$ is the price of good g in location i in state of the world s.

We note that these assumptions abstract from idiosyncratic risk as all farmers within a given location in a particular state of the world face the same yield realization for each good. However, an alternative (mathematically-equivalent) interpretation is that farmers face idiosyncratic risk but engage in perfect risk sharing arrangements with other farmers in the same location. Consistent with this interpretation being a reasonable approximation of reality in rural India, Appendix Table A.6 analyzes consumption responses to rainfall-induced shocks to farm income, echoing the seminal work of Townsend (1994). Combining four NSS survey rounds spanning 1987–2005 with our VDSA data, we find that household consumption is more responsive to district-level income shocks than to household-level income shocks.

Preferences: Farmers have constant relative risk aversion preferences with *effective risk aversion* parameter $\rho_i>0$:

$$U_i^f(s)\equiv\frac{1}{1-\rho_i}\left(\left(Z_i^f(s)\right)^{1-\rho_i}-1\right), \tag{6}$$

where $Z_i^f(s)\equiv\prod_{g\in\mathcal{G}}c_{ig}^f(s)^{\alpha_{ig}}$ is a Cobb-Douglas aggregate of goods, $c_{ig}^f(s)$ denotes the quantity consumed of good g in state s, and $\alpha_{ig}>0$ is the expenditure share spent on good g with $\sum_{g\in\mathcal{G}}\alpha_{ig}=1$. As $Z_i^f(s)$ can be written in its indirect utility form as nominal income divided by a price index, in what follows we refer to $Z_i^f(s)$ as a farmer's *real income.* Traders and drivers are assumed to have the same Cobb-Douglas preferences over goods.

Following Eswaran and Kotwal (1990), we refer to ρ_i as the *effective risk aversion* and interpret it as combining both the innate risk preferences of the farmer and any access the farmer has to ex-post risk mitigating technologies (savings, borrowing, insurance, etc.). In Appendix A.3.2, we micro-found this interpretation by allowing farmers to purchase insurance from perfectly competitive local money-lenders ("banks"). In the spirit of this interpretation, Appendix Table A.6 further shows that as local bank access improves, the responsiveness of household consumption to household shocks shrinks and the response to district shocks rises.

Trade: A large number of traders arbitrage prices across locations subject to (ad valorem) trade costs. Rather than assuming that all traders face the same costs as is implicit in standard trade models, we assume that traders are heterogeneous in their trading technology and capacity constrained.[37] As a result, the standard no-arbitrage equation—that the trade costs bound the price

[37] The assumption that traders are capacity constrained is made only for convenience: Appendix A.3.1 shows that

ratio—is replaced by an alternative condition (equation 11 below) that has the intuitive property that more goods flow toward a destination when its relative price is higher.

We now describe the trading process that delivers this key arbitrage equation, although our results hold for any process that micro-founds this equation. For example, Appendix A.3.3 shows it can arise from iceberg trade costs that are increasing and convex in the quantity shipped.

Every farmer wishing to buy or sell a good is randomly matched to a trader. If a farmer wishes to sell a unit of good g, the trader she is matched to pays her the local market price $p_{ig}(s)$ and then decides whether to sell the good locally or export it to the central market. If the trader decides to sell the good locally, he sells it for $p_{ig}(s)$, making zero profit. If the trader exports the good, he sells it for the central market price $\bar{p}_g(s)$, incurs an (iceberg) trade cost τ_{ig}, and earns a profit of $\bar{p}_g(s) - \tau_{ig} p_{ig}(s)$.[38]

The process works in reverse for a farmer wishing to buy some quantity of good g. She is randomly matched to a trader and buys for the local price $p_{ig}(s)$. The trader previously decided whether to import the good from the central market (paying $\bar{p}_g(s)$ but incurring iceberg trade cost τ, for a profit of $p_{ig}(s) - \tau_{ig}\bar{p}_g(s)$) or to source it locally (paying $p_{ig}(s)$, earning zero profit).

Trade costs τ_{ig} to ship good g between village i and the central market (in either direction) are heterogenous across traders and drawn from a Pareto distribution with shape parameter $\varepsilon_i \in (0,\infty)$:

$$\Pr\{\tau_{ig} \le \bar{\tau}\} = 1 - \bar{\tau}^{-\varepsilon_i}. \tag{7}$$

The greater the value of ε_i, the lower the average trade costs between the village and the central market (in particular, as $\varepsilon_i \to 0$ trade becomes infinitely costly for all traders and as $\varepsilon_i \to \infty$ trade becomes costless for all traders).

Discussion: We draw three distinctions between this setup and the workhorse trade model based on the seminal work Eaton and Kortum (2002) that has been applied to agricultural trade by Donaldson (2018), Costinot et al. (2016), Sotelo (2020), (Bergquist et al., 2019) and many others.

First, in Eaton and Kortum (2002), the "smoothness" necessary to avoid corner solutions arises from each location drawing different productivities for each of a continuum of varieties of a crop (with Frechet distributed draws); here, smoothness arises from trade costs being heterogeneous (with Pareto distributed draws). As discussed in Section 2.2, there is substantial evidence for trader heterogeneity in India whereas the agricultural goods we consider are relatively homogeneous.

Second, in Eaton and Kortum (2002), every location trades directly with every other location (see panel (b) of Figure 1 for an illustration); here, trade between villages occurs only indirectly through the traders in the central market. Panel (c) of Figure 1 illustrates such a trading network (incorporating the full hierarchical network structure we introduce in Section 5 below), which closely matches the realities of Indian trade presented in panel (a).

a model where better traders can offer greater capacity is isomorphic to the model presented here.

[38] We assume the trade cost is paid to agents ("drivers") that, along with traders, inhabit the central market and for whom moving goods provides all their income. Note that, by assuming drivers have no other income source, we abstract from the gains from trade arising from the direct reduction of resources necessary to move goods, instead focusing on gains arising from comparative advantage and specialization. In what follows, we will present combined welfare results for all residents of the central market.

Third, in Eaton and Kortum (2002), buyers alone engage in price arbitrage by choosing the seller offering the lowest price (inclusive of trade costs). Here, traders engage in price arbitrage both when buying from farmers and when selling to them which again closely matches the reality of agricultural trade in India described in Section 2.2.

4.2 Trade and prices

We begin by characterizing equilibrium trade and prices.

Villages: Consider first a trader selling produce to a farmer and deciding from where to source the good. If the village price is lower than the central market price, i.e. $p_{ig}(s) \leq \bar{p}_g(s)$, then no arbitrage opportunity exists and all traders will source the good locally. But if the central market price is lower than the local price, i.e. $p_{ig}(s) > \bar{p}_g(s)$, some traders will engage in price arbitrage, buying in the central market and selling (for a profit) in the village.

Now consider a trader buying produce from a farmer and deciding where to sell it. If the village price is greater than the central market price, i.e. $p_{ig}(s) \geq \bar{p}_g(s)$, then no arbitrage opportunity exists and all traders will sell the good locally. But if the central market price is greater than the village price, i.e. $\bar{p}_g(s) > p_{ig}(s)$, then some traders will engage in price arbitrage, buying in the village and selling (for a profit) in the central market.

Thus, for the market of good g in village i to clear when $p_{ig}(s) > \bar{p}_g(s)$, i.e. when good g is flowing into the village, it must be the case that the quantity produced by the village is equal to the total quantity consumed locally multiplied by the probability that traders source from the village:

$$C_{ig} \times \Pr\{p_{ig}(s) \leq \tau_{ig}\bar{p}_g(s)\} = Q_{ig}. \tag{8}$$

For the market of good g in village i to clear when $p_{ig}(s) < \bar{p}_g(s)$, i.e. when good g is flowing out of the village, it must be the case that the quantity consumed locally is equal to the total quantity produced locally multiplied by the probability that the traders sell to the village:

$$Q_{ig} \times \Pr\{\tau_{ig} p_{ig}(s) \geq \bar{p}_g(s)\} = C_{ig}. \tag{9}$$

Combining equations (8) and (9) with the assumed Pareto distribution of trade costs from equation (7), we immediately see that—regardless of the relative prices in the village and the central market (and hence regardless the direction of trade)—the relationship between relative prices, and quantities consumed and produced can be written as:

$$C_{ig}(s) = \left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i} Q_{ig}(s). \tag{10}$$

Intuitively, equation (10) states that trader arbitrage results in the good flowing toward locations with higher relative prices with an elasticity governed by the distribution of trade costs ε_i.

An alternative way of interpreting this no-arbitrage equation is to consider how local prices respond to the local quantity produced. Combining equation (10) with the farmers' Cobb-Douglas demand we obtain:

$$\ln p_{ig}(s) = -\left(\frac{1}{1+\varepsilon_i}\right)\ln Q_{ig}(s) + \frac{\varepsilon_i}{1+\varepsilon_i}\ln \bar{p}_g(s) + \frac{1}{1+\varepsilon_i}\ln(\alpha_{ig}Y_i(s)). \tag{11}$$

Equation (11) shows how a village's openness (summarized by its Pareto shape parameter ε_i) determines how its own production affects its equilibrium prices. In autarky ($\varepsilon_i = 0$), the price elasticity is one, consistent with the Cobb-Douglas demand. But as trade costs fall (ε_i increases), the elasticity of prices to own quantity produced falls, with the elasticity tending to zero as trade becomes costless ($\varepsilon_i \to \infty$). This fall in the elasticity of local prices to local production occurs simultaneously with an increase in the elasticity of local prices to the prices in the central market.

It is useful to contrast this relationship between prices and quantity with the relationship in a model in which trade costs are assumed to be homogeneous: In such a model, local prices equal autarky village prices as long as the absolute price gap between the autarky village price and the central market price is less than or equal to the costs of trading. But whenever the price gap exceeds this value, traders engage in arbitrage and the local price is pinned down by the central market price net of trade costs. This results in a "kinked" demand curve—illustrated in panel (a) of Figure 3. In our model—illustrated in panel (b) of Figure 3—there are no such kinks in the demand curve: instead, trader heterogeneity ensures a smooth relationship (log linear given the Pareto assumption) between prices and quantities. Of course, whether the "kinked" or "smooth" model better reflects the actual nature of price arbitrage is an empirical question. In panel (c) of Figure 3, we compare the two arbitrage models' abilities to explain the observed relationship between prices and (rainfall-predicted) quantities (see Appendix A.6 for details). The "smooth" model substantially outperforms the more-standard "kinked" model, explaining a larger fraction of the observed variation with an average R^2 of 0.15 (versus 0.11 in a "kinked" model) across district-decades.

Central Market: The quantity consumed in the central market is equal to the total net inflows of goods from each village:

$$\bar{C}_g(s) = \sum_{i\in\mathcal{N}}\left(1 - \left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i}\right)Q_{ig}(s), \tag{12}$$

and the income of central market residents (i.e. truckers and drivers) is equal to the total arbitrage revenues earned across all crops and villages:

$$\bar{Y}(s) = \sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{N}}(\bar{p}_g(s) - p_{ig}(s))\left(1 - \left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i}\right)Q_{ig}(s). \tag{13}$$

Combining the arbitrage equation (11) with equations (12) and (13)—and imposing the Cobb-Douglas demands of central market residents—one can calculate the equilibrium prices in the central market, and hence each village via equation (11). Formally:

Definition 1. Given any set of preferences $\{\alpha_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, trade costs $\{\varepsilon_i\}_{i\in\mathcal{N}}$, the population distribution $\{L_i\}_{i\in\mathcal{N}}$, and any state of the world $s \in S$ such that quantity produced is $\{Q_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, a *state* *equilibrium* is a set of village prices $\{p_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, village consumption $\{C_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, central market prices $\{\bar{p}_g(s)\}_{g\in\mathcal{G}}$, and central market consumption $\left\{\bar{C}_g(s)\right\}_{g\in\mathcal{G}}$ such that:

1. Markets clear within each village, i.e. (a) farmers' income equals the value of their produce; (b) farmers maximize their utility given their income; and (c) traders optimally engage in arbitrage.

2. Markets clear within the central market, i.e. (a) traders' and drivers' combined income is equal to arbitrage revenue; and (b) central market prices equate demand and supply.

The following proposition shows that the equilibrium is well defined.

Proposition 1. *Given any set of preferences $\{\alpha_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, trade costs $\{\varepsilon_i\}_{i\in\mathcal{N}}$, and any state of the world $s\in S$ such that quantity produced is $\{Q_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$:*

1. *There exists a state equilibrium.*

2. *If the trade costs $\{\varepsilon_i\}_{i\in\mathcal{N}}$ are sufficiently close to 1, then that equilibrium is unique.*

Proof. See Appendix A.2.1. □

Part 1 of Proposition 1 shows that the equilibrium is well defined for any geography of trade costs and realized quantities produced. Part 2 of Proposition 1 provides sufficient conditions for uniqueness by establishing conditions under which the excess demand function in the central market satisfies the gross substitutes property. As gross substitutes is itself a sufficient but not necessary condition for uniqueness, we expect uniqueness to be a more general phenomenon; consistent with this conjecture, an iterative algorithm based on equation (12) converges rapidly to an equilibrium for a wide variety of $\{\varepsilon_i\}$; see Appendix A.5 for details.

4.3 Optimal crop choice

We now derive a convenient and intuitive expression showing how farmers' optimal crop choice is affected by trade in the presence of volatility. To provide intuition for this expression, we begin by considering the special case where productivity is constant.

No volatility In the absence of volatility, and taking prices as given, a farmer will equalize her income per unit of land (i.e. her factor price) across all goods she produces:[39]

$$p_{ig}A_{ig}=\lambda_i \ \forall g\in\{1,...,G\}, \tag{14}$$

where $\lambda_i>0$ is the shadow value of land. Substituting in equation (11) for the equilibrium price, imposing symmetry across farmers so that $\theta_{ig}^f=\theta_{ig}$ for all farmers f in village i, and applying the constraint that land shares sum to one yields:

$$\theta_{ig}=\frac{\alpha_{ig}(A_{ig}\bar{p}_g)^{\varepsilon_i}}{\sum_{h\in\mathcal{G}}\alpha_{ih}(A_{ih}\bar{p}_h)^{\varepsilon_i}}. \tag{15}$$

[39] It is straightforward to show that all goods will be produced in all locations in equilibrium as equation (11) implies that the price of a good will become infinite as the land allocated to that good tends to zero.

Farmers specialize more in the production of good g the greater their own demand for that good (α_{ig}) and the greater the market returns from producing that good ($A_{ig}\bar{p}_g$), with the relative weights of the two considerations depending on the degree of openness to trade (ε_i). As a village becomes more open (i.e. ε_i increases), farmers allocate a greater fraction of their land toward goods that have high market returns rather than goods they wish to consume.

What about the gains from trade? Farmers' utility is:

$$U_i = \frac{1}{1-\rho_i}\left(\prod_g (\alpha_{ig}A_{ig})^{\alpha_{ig}} - 1\right)^{1-\rho_i}, \tag{16}$$

which, in the absence of volatility, does not depend on the degree of openness. As in a standard Ricardian model, opening up to trade increases the returns to goods that a location has a comparative advantage in, causing farmers to grow more of those crops. Unlike a standard Ricardian trade model, however, local prices fall as more comparative advantage crops are grown since not all of the excess production is exported by the heterogenous traders. Farmers continue to reallocate toward their comparative advantage crops up to the point that their returns per unit land are equalized across crops, resulting in the same autarkic relative prices and leaving their welfare unchanged.[40] This is not to say that there are no gains from specialization: there are. But these gains are captured entirely by the traders engaging in price arbitrage—a feature of the model that we believe is not unrealistic given the large literature documenting the substantial market power traders have over small farmers referenced in Section 2.2 above. This result also helps contrast how volatility affects the gains from trade for farmers, a point we turn to next.

Volatility We now turn to the more general case where productivity is volatile, due for example to variation in rainfall. Rather than farmers equalizing their marginal nominal income across crops as before, farmers now equalize their marginal expected utility, necessitating a characterization of the distribution of farmers' real income that we integrate over all states of the world. To do so, we combine techniques from the portfolio choice literature in finance with the general equilibrium trade framework developed above. This general equilibrium structure adds substantial complication to the problem since the distribution of farmers' real incomes depends on the geography of trade costs, the distribution of yields, and the crop choices of all other farmers. Despite this complexity, however, we are still able to derive an explicit expression for farmers' equilibrium crop choice that is a straightforward generalization of equation (15).

We begin by positing the following distribution of crop yields across states of the world:

Assumption 1 (Log normal distribution of yields)**.** *Assume that the joint distributions of yields across goods are log normal within village i and are independently distributed across villages. In particular, define $\mathbf{A}_i(s)$ as the $G \times 1$ vector of $A_{ig}(s)$. Then $\ln \mathbf{A}_i \sim N\left(\boldsymbol{\mu}^{A,i}, \boldsymbol{\Sigma}^{A,i}\right)$ for all $i \in \{1,...,N\}$,*

[40] A crucial assumption underlying this result is that each farmer takes the market prices as given. If, instead, farmers internalized the effect of their crop allocation choice on equilibrium prices—say through the formation of an agricultural collective—they would choose to restrict the degree to which they specialize, increasing the price of their comparative advantage goods and improving their terms of trade. See Appendix A.3.4 for further details.

where $\boldsymbol{\mu}^{A,i} \equiv \left[\mu_g^{A,i}\right]$ *is a* $G \times 1$ *vector and* $\boldsymbol{\Sigma}^{A,i} \equiv \left[\Sigma_{gh}^{A,i}\right]$ *is a* $G \times G$ *variance-covariance matrix.*

That yield realizations are independently distributed across many locations implies that the (endogenous) central market price is state invariant, i.e. shocks to yields in individual villages "average out" in the aggregate. While helpful for simplifying the exposition in this section, we relax this independence assumption in the quantification in Section 5 by allowing yield shocks to be correlated across villages and central market prices to be state dependent.

We next follow the finance literature (see, e.g. Campbell and Viceira (2002)) and approximate the real income of farmer f by taking a second-order approximation around its (log) mean (see Appendix A.1 for derivations of this and later expressions in this section):[41]

$$\ln Z_i^f(s) \approx \mu_i^Z + \sum_{g \in \mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,g}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ig} \right) \left(\ln A_{ig}(s) - \mu_g^{A,i} \right), \tag{17}$$

where μ_i^Z—defined in Appendix equation (32)—is a scalar that depends on crop choice and equilibrium market prices, but does not depend on the particular state of the world s.

Equation (17) states that a positive yield deviation $(\ln A_{ig}(s) - \mu_g^{A,i})$ benefits a farmer proportionally to a weighted average of their share of land allocated to producing the good $(\theta_{i,g}^f)$ and their consumption share of the good (α_{ig}), with the weight determined by the degree of openness of the village (ε_i). Intuitively, the more open a village (i.e. the higher the ε_i), the more a farmer is engaged with buying and selling goods in the market, and the more a farmers' real income depends on the realized local yields of the crops she grows than the yields of the crops she consumes.

Next, we calculate the expected utility of farmers as a function of their crop choice (and the crop choices of all other farmers). From equation (17), it immediately follows that farmer real income is (approximately) log normally distributed across states of the world:

$$\ln Z_i^f \sim N\left(\mu_i^Z, \sigma_i^{2,Z}\right), \tag{18}$$

where the variance of her log real income $\sigma_i^{2,Z}$—defined in Appendix equation (33)—depends on her equilibrium crop choice. This in turn implies the expected utility takes the convenient form:

$$\mathbb{E}\left[U_i^f\right] = \frac{1}{1-\rho_i} \left(\exp\left((1-\rho_i) \left(\ln \mathbb{E}\left(Z_i^f \right) - \rho_i \sigma_i^{2,Z} \right) \right) - 1 \right), \tag{19}$$

where $\mathbb{E}\left(Z_i^f\right) = \exp\left(\mu_i^Z + \frac{1}{2}\sigma_i^{2,Z}\right)$ since Z_i^f is log-normally distributed. Thus, farmer f trades off the (log of the) mean of her real income with the variance of her (log) real income, with the exact trade-off governed by the degree of effective risk aversion ρ_i.

Having characterized the expected utility of farmers, we can then derive the following first order conditions from the farmer's crop choice problem that state that the marginal contribution of each

[41] The second order approximation implies that the sum of log normal variables is itself approximately log normal. Campbell and Viceira (2002) approximate around zero returns, which is valid for assets over a short period of time. Because our time period is a year, we instead approximate around the mean log yields. This comes at a slight cost to tractability, but substantially improves the approximation—in the quantitative results in Section 5.4 below, we find that the approximated expected utility is highly correlated (exceeding 0.999) with the actual expected utility.

crop to expected utility should be equalized:

$$\mu_{ig}^{Z} - \rho_i \left(\frac{\varepsilon_i}{1+\varepsilon_i}\right) \sum_{h\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right) \theta_{i,h}^{f} + \left(\frac{1}{1+\varepsilon_i}\right) \alpha_{ih} \right) \Sigma_{gh}^{A,i} = \lambda_i, \tag{20}$$

where $\mu_{ig}^{Z} \equiv \frac{\partial \ln \mathbb{E}\left(Z_i^f\right)}{\partial \theta_{ig}^{f}}$ is the marginal contribution of crop g to the log of the mean real income and λ_i is the shadow value of land. Equation (20)—which generalizes the indifference condition (14) to accommodate volatility—is again intuitive: a good with a higher marginal contribution to the variance of real returns must have higher marginal contribution to the mean real returns (i.e. a high μ_{ig}^{Z}) to compensate for the additional risk. This expression will prove essential when estimating farmers' effective risk aversion from their observed crop choices in Section 5.3 below.

Finally, by combining farmers' first order conditions (20), imposing symmetry across farmers within village, and imposing that crop shares sum to one, we can derive an expression for the equilibrium crop choice that generalizes equation (15) to incorporate volatility:

$$\theta_{ig} = \frac{\alpha_{ig} \left(B_{ig} \bar{p}_g\right)^{\varepsilon_i}}{\sum_{h\in\mathcal{G}} \alpha_{ih} \left(B_{ih} \bar{p}_h\right)^{\varepsilon_i}}, \tag{21}$$

where B_{ig} is the *risk adjusted productivity* of farmers in location i producing crop g.[42] In the absence of volatility, the risk adjusted productivity is simply the actual productivity, i.e. $B_{ig} = A_{ig}$, and equation (21) collapses to (15). But in the presence of volatility (and for sufficiently high risk aversion ρ_i), B_{ig} is smaller the greater g's marginal contribution to the aggregate volatility of real returns, i.e. $\sum_{h\in\mathcal{G}} \left(\frac{\varepsilon_i}{1+\varepsilon_i} \theta_{i,h} + \frac{1}{1+\varepsilon_i} \alpha_{ih}\right) \Sigma_{gh}^{A,i}$, so that farmers trade off traditional "first moment" benefits from specializing in crops with higher mean yields against "second moment" benefits of specializing in less risky crops.

4.4 Equilibrium

We can now characterize the full equilibrium of the model.

Definition 2. Given any set of preferences $\{\alpha_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, trade costs $\{\varepsilon_i\}_{i\in\mathcal{N}}$, and distributions of yields across states of the world $\left\{\boldsymbol{\mu}^{A,i}, \boldsymbol{\Sigma}^{A,i}\right\}_{i\in\mathcal{N}}$, an *equilibrium* is a set of crop allocations $\{\theta_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ and, for each state of the world $s \in S$, a set of village prices $\{p_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, village consumption $\{C_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, central market prices $\{\bar{p}_g(s)\}_{g\in\mathcal{G}}$, and central market consumption $\left\{\bar{C}_g(s)\right\}_{g\in\mathcal{G}}$ such that:

1. Each state of the world $s \in S$ is in a state equilibrium.

2. Farmers optimally choose their crop allocation to maximize their expected utility across all states, i.e. crop choice satisfies equation (21).

Because Proposition 1 holds for any realized quantities produced $\{Q_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$—including those that would arise from the optimal crop allocation—it immediately implies the following corollary:

[42]Specifically, $B_{ig} \equiv \exp(\mu_g^{A,i}) / \left(\lambda_i - \left(\frac{1}{2}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2 \Sigma_{gg}^{A,i} + \frac{\varepsilon_i}{(1+\varepsilon_i)^2} \sum_{h\in\mathcal{G}} \alpha_{ih} \Sigma_{gh}^{A,i} - \rho_i \left(\frac{\varepsilon_i}{1+\varepsilon_i}\right) \sum_{h\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right) \theta_{i,h} + \left(\frac{1}{1+\varepsilon_i}\right) \alpha_{ih}\right) \Sigma_{gh}^{A,i}\right)\right)$.

Corollary 1. *For any set of preferences* $\{\alpha_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$*, trade costs* $\{\varepsilon_i\}_{i\in\mathcal{N}}$*, and distributions of yields across states of the world* $\left\{\boldsymbol{\mu}^{A,i},\boldsymbol{\Sigma}^{A,i}\right\}_{i\in\mathcal{N}}$*, there exists an equilibrium and it is unique if each* $\{\varepsilon_i\}_{i\in\mathcal{N}}$ *is sufficiently close to one.*

Having characterized the equilibrium of the model, we now turn to its qualitative implications.

4.5 Qualitative implications

Explaining the stylized facts

We now show that the model is consistent with the three stylized facts presented in Section 3:

Proposition 2. *Consider a small increase in village* i'*s openness to trade* ε_i*:*

(a) [Stylized Fact 1] Any increase in openness: (1a) decreases the responsiveness of local prices to local yield shocks; and (1b) increases the responsiveness of local prices to the central market price:

$$\frac{d}{d\varepsilon_i}\left(-\frac{\partial \ln p_{ig}(s)}{\partial \ln A_{ig}(s)}\right)<0 \text{ and } \frac{d}{d\varepsilon_i}\left(\frac{\partial \ln p_{ig}(s)}{\partial \ln \bar{p}_g}\right)>0.$$

(b) [Stylized Fact 2] Starting from autarky, an increase in openness: (2a) causes farmers to reallocate production toward crops with higher mean and less volatile yields (as long as $\rho_i>1$*, i.e. farmers are sufficiently risk averse); and (2b) the reallocation toward less volatile crops is attenuated the greater the access to insurance (i.e. the lower* ρ_i*). Formally, for any two crops* $g\neq h$*:*

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial(\ln\theta_{ig}-\ln\theta_{ih})}{\partial\left(\mu_g^{A,i}-\mu_h^{A,i}\right)}\right)|_{\varepsilon_i=0}>0,\ \frac{d}{d\varepsilon_i}\left(\frac{\partial\ln\theta_{ig}-\partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i}-\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0}<0,$$

$$\text{and } -\frac{d^2}{d\varepsilon_i d\rho_i}\left(\frac{\partial\ln\theta_{ig}-\partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i}-\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0}>0.$$

(c) [Stylized Fact 3] Consider a decomposition of the variance of real income as follows:

$$\sigma_i^{2,Z}=\sigma_i^{2,Y}+\sigma_i^{2,P}-2cov_i^{Y,P},$$

where

$$\sigma_i^{2,Y}\equiv var(\ln Y_i(s)-c_i(s)),\ \sigma_i^{2,P}\equiv var\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s)+c_i(s)\right), \text{ and}$$

$$cov_i^{Y,P}\equiv cov\left(\ln Y_i^f(s)-c_i(s),\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s)+c_i(s)\right)$$

are the variance of farmers' nominal income volatility, the variance of farmers' nominal price volatility, and the co-variance between the two (and $c_i(s)$ *is a nuisance term capturing the aggregate scale of both nominal prices and incomes which affects neither aggregate real income nor its volatility). Any increase in openness: (3a) increases the volatility of farmers' nominal income volatility; (3b) decreases the volatility of farmers' nominal price volatility; and (3c) has an ambiguous effect* *on farmers' real income volatility. Formally, we have:*

$$\frac{\partial \sigma_i^{2,Y}}{\partial \varepsilon_i} > 0, \frac{\partial \sigma_i^{2,P}}{\partial \varepsilon_i} < 0, \text{ and } \frac{\partial \sigma_i^{2,Z}}{\partial \varepsilon_i} \lesseqgtr 0,$$

where a sufficient condition for farmers' real income volatility to increase with openness, i.e. $\frac{\partial \sigma_i^{2,Z}}{\partial \varepsilon_i} \geq 0$, *is* $\sum_{g \in \mathcal{G}} \theta_{i,g} \left(\sum_{h \in \mathcal{G}} \Sigma_{gh}^{A,i} \alpha_{ih} \right) \geq \sum_{g \in \mathcal{G}} \alpha_{ig} \left(\sum_{h \in \mathcal{G}} \Sigma_{gh}^{A,i} \alpha_{ih} \right)$.

Proof. See Appendix A.2.2. □

As trade costs fall and a village becomes more open, more traders engage in price arbitrage which reduces the responsiveness of prices to local yields and increases the responsiveness of local prices to the central market price—consistent with Stylized Fact 1. Farmers react to the increase in openness by changing their crop allocation, placing less weight on crops they consume and more weight on those in which they have a comparative advantage. But at the same time, to mitigate the increased risk farmers now face due to local prices being less responsive to local yields, farmers respond by moving into crops with less volatile yields. The trade-off between these traditional "first moment" gains from specialization and "second moment" efforts to reduce risk is governed by their level of risk aversion—consistent with Stylized Fact 2. Because prices become less responsive to local yields, farmers face more volatile nominal incomes at the same time as more stable consumption prices—consistent with Stylized Fact 3, with the net effect on the volatility of real income depending on the extent to which a farmer's crop allocation is more risky than her expenditure allocation.

Volatility and the gains from trade

We now turn to the welfare implications of the model. We summarize the relationship between welfare, trade costs and volatility in the following proposition:

Proposition 3. *1) In the presence of volatility, moving from autarky to costly trade improves farmer welfare, i.e. the gains from trade are positive; 2) moving from a world with no volatility to one with volatility amplifies farmers' gains from trade; but 3) increasing the volatility in an already volatile world may attenuate farmers' gains from trade.*

Proof. See Appendix A.2.3. □

Part (1) of Proposition 3 arises from a standard revealed preference argument (see, e.g. Dixit and Norman (1980)). Because all farmers in a location are identical, in autarky each consumes what she produces in all states of the world. With trade, a farmer always has the option to make the same planting decisions; moreover, because the farmer both buys and sells to traders at the local price, she always has the option to consume what she produces. Hence, in all states of the world, a farmer can always achieve the same level of utility as in autarky, so her expected utility must be at least as great. Furthermore, given the model structure, the expected utility gains are strictly positive, as prices with trade will differ from autarkic prices with probability one.

Combining Part (1) with the result above that farmers gains from trade are zero in the absence of volatility (see equation (16)), Part (2) follows immediately. Intuitively, volatility amplifies the gains from trade via two mechanisms: first, on the consumption side, farmers are now able to maintain a more balanced consumption basket by trading crops with relatively good yield realizations to purchase crops with relatively bad yield realizations; second, on the production side, by decoupling production and consumption decisions, trade allows farmers to alter their planting decisions in order to reduce their risk exposure. However, part (3) shows that additional volatility—for example making "safe" crops more volatile—can attenuate the gains from trade by reducing farmers' ability to use their crop allocation to reduce their risk; Appendix Table A.7 provides such an example.

It is important to emphasize that Proposition 3 hinges on the assumption that farmers are able to produce all that they wish to consume; if, for example, farmers also consume manufacturing goods that they cannot produce, as in Newbery and Stiglitz (1984) gains from trade in the presence of volatility need not be positive—a possibility we explicitly introduce in the quantitative version of our model below.[43]

5 Quantifying the welfare effects of trade and volatility

We now bring the framework developed above to the rural Indian data to quantify the welfare effects of trade in the presence of volatility.

5.1 Extending the baseline model

We first extend the basic framework above to create a "quantitative" model that adds realism by incorporating a number of additional features.

Constant elasticity of substitution preferences

In the baseline model above, we assumed that agents consumed a Cobb-Douglas aggregate of goods; we now generalize to constant elasticity of substitution (CES) preferences with elasticity of substitution σ ($\bar{\sigma}$) for village (market) residents.

A Manufacturing Good

In the baseline model, we assumed that farmers are able to produce all goods in the economy; while convenient, certain goods (such as services or manufacturing) are less commonly produced in rural India. As noted above, the presence of such goods has potentially important implications for the gains from trade. We extend the model to incorporate a numeraire manufacturing/services good $g = 0$ that is produced in markets but not in the villages. This numeraire good is costlessly traded and agents have Cobb-Douglas preferences across the good and the (CES) consumption bundle of agricultural goods (with β_i equal to the agricultural expenditure share).

Finite number of villages with correlated productivity shocks

We amend Assumption 1 and now allow for arbitrary correlations of realized yields across crops and a (now finite) number N of villages:

[43] Newbery and Stiglitz (1984) provide an extreme example of Pareto inferior trade where farmers only wish to consume non-farm goods whose productivity is not volatile while non-farm producers only wish to consume volatile farm goods.

Assumption 2 (Log normal distribution of yields (generalized))**.** *Assume that the joint distributions of yields across all goods and villages are log normally distributed across states of the world. In particular, define* $\mathbf{A}(s)$ *as the* $(G\times N)\times 1$ *vector of* $\{A_{ig}(s)\}_{g\in\mathcal{G},i\in\mathcal{N}}$*. Then* $\ln\mathbf{A}\sim N\left(\mu^A,\mathbf{\Sigma}^A\right)$*, where* $\boldsymbol{\mu}^A\equiv\left[\mu_{ig}^A\right]$ *is a* $GN\times 1$ *vector and* $\mathbf{\Sigma}^A\equiv\left[\Sigma_{ig,jh}^A\right]$ *is a* $GN\times GN$ *variance-covariance matrix.*

With a finite number of villages and correlated yield shocks, equilibrium market prices are now state dependent. As a result, the volatility of famers' real income will be affected not only by changes in a village's own trade costs but also changes in trade costs elsewhere in the network. As we will see below, this new second-moment effect will have important quantitative implications.

Multiple markets

In the baseline model, we assume that all villages trade with the same central market. To better capture India's hierarchical trading network (described in Section 2.2 and panel (a) of Figure 1), we now incorporate multiple layers of markets. While in principal the model can be extended to include an arbitrary number of layers, given data limitations we consider a three-layer hierarchy where each village $i\in\mathcal{N}$ (an Indian district in our empirics) engages in trade with a regional market $m\in\mathcal{M}\equiv\{1,...,M\}$ (the largest city within each Indian state in our empirics), which in turn engages in trade with a central market (Delhi in our empirics). Panel (c) of Figure 1 depicts the resulting trading network.[44]

5.2 The quantitative model: A summary

We briefly summarize how the results presented in Section 4 change with these model extensions; Appendix A.4 provides further details.

Equilibrium prices

Conditional on the equilibrium regional market prices, the arbitrage process between villages and their regional markets remains unchanged allowing us to generalize the equilibrium price equation (11) to:

$$\ln p_{ig}(s)=-\frac{1}{\sigma+\varepsilon_i}\ln A_{ig}(s)+\frac{\varepsilon_i}{\sigma+\varepsilon_i}\ln\bar{p}_{m(i)g}(s)+\delta_{ig}+\delta_i(s), \tag{22}$$

where δ_{ig} is a location-good term that is constant across all states of the world and $\delta_i(s)$ is a location-state of world term that is the same across all goods.[45] Echoing equation (11), the equilibrium price in a location responds less to local yield shocks and more to prices in its regional market as trade costs fall (i.e. ε_i increases).

[44]While there is also a trade across villages within a district (see panel (a) of Figure 1), comprehensive agricultural data at the sub-district level do not exist. The limited sub-district data that do exist suggest the price variability within district is small relative to across-district price variation. For example, in the 1971 REDS data, restricting attention to the 51 districts where at least two households sold rice, the (log) paddy price has an overall standard deviation of 0.50 and a (median) within-district standard deviation of 0.15. We also omit an easy-to-accommodate international trade layer linking the central market with a world market given India's highly restrictive agricultural trade regime during the majority of our sample period (see footnote 9).

[45]In particular, $\delta_{ig}\equiv\frac{1}{\sigma+\varepsilon_i}\ln(\beta_i\alpha_{ig}/L_i\theta_{ig})$ and $\delta_i(s)\equiv\frac{1}{\sigma+\varepsilon_i}\ln(Y_i(s)/\sum_{h=1}^G\alpha_{ih}(p_{ih}(s))^{1-\sigma})$.

The fractal nature of the hierarchical trading network means that a very similar expression governs the equilibrium regional market prices:

$$\ln\bar{p}_{mg}(s) = -\frac{1}{\bar{\sigma}+\varepsilon_m}\ln\bar{Q}_{mg}(s) + \frac{\varepsilon_m}{\bar{\sigma}+\varepsilon_m}\ln p_g^*(s) + \delta_{mg} + \delta_m(s), \tag{23}$$

where $\bar{\sigma}$ is the regional market resident's elasticity of substitution across agricultural goods, ε_m is the Pareto shape parameter governing the distribution of trade costs across traders engaging in arbitrage between regional market m and the central market, $\bar{Q}_{mg}(s)$ is the net quantity of a good that arrives to the market, δ_{mg} is a market-good term that is constant across all states of the world, and $\delta_m(s)$ is a market-state of world term that is the same across all goods.[46] Similar to village level prices, regional market prices depend both on the quantity of goods that arrive at the market and the central market prices, and, as above, lower trade costs (i.e. higher ε_m) increase the responsiveness to the latter relative to the former.

Equilibrium crop choice

Finally, we examine how a farmer's crop choice affects her distribution of real income. As in the baseline model, we apply the same second order approximation from the finance literature. To incorporate the fact that equilibrium regional market prices are state dependent, we also apply a first order log-linear approximation of the equilibrium regional market prices around their (log) mean yield. Together, these two approximations imply that the real income of farmer f in location i is approximately log normally distributed with mean μ_i^Z and variance $\sigma_i^{2,Z}$ defined by equations (70) and (72) in Appendix A.4. As a result, the first order condition from the farmer's crop choice again follows equation (20), where the marginal effect of increasing the share of land allocated to crop g on the log of mean real income and variance of log real income, μ_{ig}^Z and σ_{ig}^Z, are defined in equations (73) and (74) in Appendix A.4. And just as in the baseline model, the intuition is that farmers allocate land to a crop up to the point that any increase in total mean income is fully offset by the increase in volatility, where the mean-variance trade-off is determined by the farmer's effective level of risk aversion ρ_i. Applying the farmer's first order conditions generalizes the equilibrium crop choice (i.e. equation 21) to

$$\theta_{ig} = \frac{\alpha_{ig} B_{ig}^{\varepsilon_i+\sigma-1} \bar{p}_{m(i)g}^{\varepsilon_i}}{\sum_{h=1}^{G} \alpha_h B_{ih}^{\varepsilon_i+\sigma-1} \bar{p}_{m(i)h}^{\varepsilon_i}}, \tag{24}$$

where B_{ig} is again the risk-adjusted productivity (defined in equation (76) in Appendix A.4).

To summarize, the quantitative model remains tractable while being a more realistic description of India. And as in the baseline model, two sets of structural parameters play key roles in determining the strength of the central economic forces. First, the distribution of trade costs ($\{\varepsilon_i\}$ and $\{\varepsilon_m\}$) determines the relative responsiveness of local prices to local shocks and prices elsewhere—and hence how trade affects volatility. Second, the effective risk aversion parameters ($\{\rho_i\}$) determine how farmers trade off risk versus return—and hence how they respond to changes

[46] In particular, $\delta_{mg} \equiv \frac{1}{\bar{\sigma}+\varepsilon_m}\ln\alpha_{mg}$ and $\delta_m(s) \equiv \frac{1}{\bar{\sigma}+\varepsilon_m}\ln\left(\beta_m \bar{Y}_m(s) / \sum_{h=1}^{G} \alpha_{mh}(\bar{p}_{mh}(s))^{1-\bar{\sigma}}\right)$.

in volatility. We turn now to the estimation of these key parameters.

5.3 Estimation of structural parameters

We now estimate the structural parameters—with particular attention paid to the key trade cost distributions and effective risk aversion parameters highlighted above.

Observed parameters: Budget shares, market sizes, and the distribution of yields

We choose district-specific agricultural expenditure shares β_i and district-crop-specific CES demand shifters α_{ig} to match observed district-average expenditure shares from the 1987-88 NSS described in Section 2.3; see Appendix Table A.8 for summary statistics. Regional and central market preferences are set equal to the average preferences of their constituent districts.

We set the size of each district in each decade d to its average total cropped area that decade. We set the size of each regional market—which determines the quantity of the numeraire good it produces—so that its size relative the total size of all its constituent districts matches the observed urban-rural population ratio in the state, thereby ensuring that a person in India either grows crops on one unit of land or produces one unit of the numeraire good. We set the size of the central market to match the relative size of Delhi compared to the total urban population of India.

We determine the distribution of (log) yields in each decade by treating each year within the decade as an independent draw from a common underlying distribution.[47] Maintaining Assumption 2 that the yields across all 15 crops and 311 districts are multivariate log-normally distributed, we equate the mean log yield to the average log yield observed across years within each decade.[48] Appendix Figure A.2 depicts the distribution of mean (log) yields across districts and crops for the 1970s. There is substantial variation across districts. For example, southern India is relatively more productive in sugarcane while northern India is relatively more productive in wheat.

Similarly, we calculate the full variance covariance matrix $\mathbf{\Sigma}_d^A \equiv [\sigma_{igd,jhd}]$ from annual yield variation across crops and districts within decades, where $\sigma_{igd,jhd}$ is the decade-d covariance between the log yields of crop g in district i and the log yields of crop h in district j. Appendix Figure A.3 provides a graphical depiction of the matrix. There is substantial correlation of yields across crops within districts, across districts within crops, and even across different crops in different districts, highlighting the importance of incorporating such flexibility in the quantitative model.

A small digression on missing yield data. Unlike in Section 3.2, we now require estimates of the spatial covariances across locations. Thus, we cannot rely on spatial interpolation of missing yields which mechanically imposes a specific spatial correlation structure and so we estimate the distribution of yields from the 70.6% of yield observations that are non-missing.[49] Any bias in our

[47]Consistent with this assumption, we find no serial correlation in (log) yields conditional on crop-district-decade fixed effects.

[48]The division of the forty years of data into four different decades is not only convenient, it also offers the maximum likelihood of observing the realized yields across the 1,771 possible partitions of the data into four distinct time periods of at least five years in length.

[49]To reduce the risk of outliers driving subsequent results, we only calculate district i-crop g $\times$ district j-crop h covariances for which we observe both sets of yields for all years within a decade, otherwise setting their covariance equal to zero.

structural estimates is minimized by the fact that, to make results representative, we weight each observation by its cropped area and missing yields are typically associated with zero or negligible cropped areas. Finally, we note that conditional on our estimates of the structural parameters, our choice of crop cultivation costs will rationalize observed crop choices—mitigating the concern that our structural results will be sensitive to mis-measurement of the mean and variance-covariance of log yields of crops with small cropped areas.

Estimating the district-level trade openness ε_{id} and elasticity of substitution σ

Treating each year as a different realized state of the world, the empirical analog of equation (22) provides a simple and intuitive equation for estimating district i trade openness each decade d, ε_{id}:

$$\ln p_{igtd} = -\frac{1}{\sigma+\varepsilon_{id}} \ln A_{igtd} + \frac{\varepsilon_{id}}{\sigma+\varepsilon_{id}} \ln \bar{p}_{m(i)gtd} + \delta_{igd} + \delta_{itd} + \nu_{igtd}, \tag{25}$$

where δ_{igd} and δ_{itd} are district-crop-decade and district-year-decade fixed effects, respectively, and the residual ν_{igtd} captures measurement error in district prices p_{igtd}, district yields A_{igtd}, and regional market prices $\bar{p}_{m(i)gtd}$. Consistent with the empirical context described in Section 2.2, we treat each Indian state as its own regional market. Because we do not directly observe regional market prices, we set $\bar{p}_{m(i)gtd}$ equal to quantity-weighted average state prices.

Intuitively, districts are more open to trade the less responsive their local prices are to local yield shocks and the more responsive their local prices are to regional market prices (conditional on the appropriate set of fixed effects). Similarly to Section 3, we instrument with the rainfall-predicted yields and state-level average prices leaving out own-district prices to: (a) correct for potential endogeneity in yields (e.g. farmers putting more care into harvesting high price crops); (b) avoid the mechanical reflection problem in the the market level price; and (c) correct for any (classical) measurement error in yields and market prices.

As a first pass, we recover a common trade openness parameter, i.e. $\varepsilon_{id} = \varepsilon$, along with the elasticity of substitution σ directly from the estimated regression coefficients. The IV specification is reported in column 2 of panel (a) of Table 4 and implies $\varepsilon = 2.1$ and $\sigma = 6.2$. However, these averages belie substantial variation across space and time. Echoing Stylized Fact 1, columns 3 and 4 interact yields and prices with within-state market access $MA_{id}^{instate}$ and find that prices are both less responsive to local yield shocks and more responsive to state-market prices when the highway system expands. Thus, to estimate district-decade openness ε_{id}, we impose the parameterization $\varepsilon_{id} = \beta_0 + \beta_1 MA_{id}^{instate}$ and estimate β_0 and β_1 using GMM and the same moment conditions as our IV specification.[50] Column 6 of panel (a) of Table 4 presents these results. Consistent with districts becoming more open with highway improvements, we find average values of ε_{id} growing from 1.9 in the 1970s (with an interquartile range of 0.2) to 2.2 in the 2000s (with an interquartile range of 0.3). We also estimate an elasticity of substitution of $\sigma = 6.0$ across crops.[51]

[50]While in principle we could estimate the district level trade openness ε_{id} non-parametrically, the small number of time periods and goods relative to the number of districts means such estimates are extremely noisy.

[51]This estimate is similar to the large literature that estimates trade elasticities that imply an elasticity of substitution of around five; see e.g. Simonovska and Waugh (2014). In the Indian context, Van Leemput (2021) and Tomar

Estimating the market-level trade openness $\bar{\varepsilon}_{md}$ and elasticity of substitution $\bar{\sigma}$

As discussed above, equilibrium regional market level prices are characterized much like district level prices given the fractal nature of the hierarchical trading structure. Accordingly, our estimation of the market-level trade openness $\bar{\varepsilon}_{md}$ and elasticity of substitution $\bar{\sigma}$ proceed similarly to their district-level analogs, with equation (23) yielding the following empirical specification:

$$\ln\bar{p}_{mgtd} = -\frac{1}{\bar{\sigma}+\bar{\varepsilon}_{md}}\ln\bar{Q}_{mgtd} + \frac{\bar{\varepsilon}_{md}}{\bar{\sigma}+\bar{\varepsilon}_{md}}\ln p^*_{gtd} + \delta_{mgd} + \delta_{mtd} + \bar{\nu}_{mgtd}, \tag{26}$$

where δ_{mgd} and δ_{mtd} are market-crop-decade and market-year-decade fixed effects, respectively, and $\bar{\nu}_{mgtd}$ captures measurement error in market-level prices $\bar{p}_{mgtd}$, market-level quantities $\bar{Q}_{mgtd}$, and central market prices p^*_{gtd}. We measure the central market price as the quantity-weighted average price across all of India.

Intuitively, the more open a region is, i.e. $\bar{\varepsilon}_m$ is higher, the less responsive the regional market price is to the quantity produced within the region and the more responsive the regional price is to prices in the rest of India. Once again, we pursue an instrumental variables strategy, instrumenting with the predicted quantities from rainfall-predicted yields throughout the state and with the average price in all districts outside of the state.

Mimicking the district-level analysis, the first two columns of Table 4 panel (b) assume a common level of openness in all markets and periods, i.e. $\bar{\varepsilon}_{md} = \bar{\varepsilon}$. The IV implies $\bar{\varepsilon} = 2.0$ and $\bar{\sigma} = 4.9$. Parameterizing market openness as a function of travel time to Delhi, and using IV GMM , column 6 estimates that regional markets with greater travel times to Delhi are less open—although the estimated coefficient is small and statistically insignificant. Thus, the average values of $\bar{\varepsilon}_{im}$ increase only slightly from 1.91 in the 1970s to 1.94 in the 2000s as a result of the expansion of the Indian highway network. We estimate an elasticity of substitution of $\bar{\sigma} = 4.8$ across crops.

Estimating the effective risk aversion ρ_{id} and costs of cultivation

Recall that farmers choose a land allocation along the frontier of the (log) mean real returns and the variance of (log) real returns, with the gradient of the frontier at the chosen allocation equal to their effective risk-aversion parameter ρ_{id}. This relationship is summarized by the farmer's first order conditions (equation 20), which we re-write as:

$$\mu^Z_{igd} = \rho_{id}\sigma^Z_{ig} + \delta_{id} + \delta_{ig} + \delta_{gd} + \zeta_{igd}, \tag{27}$$

where the marginal contribution to the mean and variance of real returns, μ^Z_{ig} and σ^Z_{ig}, are calculated from the mean and variance-covariance of the (observed) nominal gross yields $\boldsymbol{\mu}^A$ and $\boldsymbol{\Sigma}^A$ (see equations (73) and (74) in the Appendix). The δ_{id} fixed effect is the district-decade level Lagrange multiplier λ_{id}; while a district-good fixed effect δ_{ig}, a crop-decade fixed effect δ_{gd}, and an idiosyncratic district-good-decade error term ζ_{igd} together capture any unobserved differences in the cost of cultivation across crops. Note that, given these recovered crop costs and the other estimated structural parameters, the farmers' first order conditions will hold with equality at their

(2016) find an elasticity of substitution in agriculture of 2.3 and 3.3, respectively.

observed land allocation. In other words, we calibrate the unobserved crop costs so that farmers in all districts and all decades are producing at the optimal point along their mean-variance frontier.[52]

Under the assumptions regarding crop costs above, ρ_{id} can be estimated using equation (27) via ordinary least squares. However, given that our variance-covariance matrix is itself an estimate, to correct for (classical) measurement error, we instrument for the marginal contribution to the variance term σ_{ig}^{Z} with an instrument constructed using the rainfall predicted variance-covariance matrix of log-yields (also appropriately transformed using equation (74) in the Appendix).[53]

Table 5 begins by presenting results assuming a common effective risk aversion parameter $\rho_{id}=\rho$. Mean real returns are increasing in the variance of real returns with the IV estimates implying an effective risk aversion parameter slightly greater than one ($\rho = 1.3$), consistent with previous estimates of risk aversion of Indian farmers (e.g. Rosenzweig and Wolpin (1993)).

In Stylized Fact 2, however, we saw that farmers with greater access to rural banks placed less emphasis on second moment concerns when making their crop choice. And as the effective risk aversion parameter captures both the inherent risk aversion of farmers and their access to risk-mitigating technologies, we incorporate such heterogeneity by assuming that ρ_{id} is a function of rural bank access, i.e. $\rho_{id} = \rho^{A} bank_{id} + \rho^{B}$, where we expect $\rho^{A} < 0$.[54] Columns 3 and 4 of Table 5 provide support for this hypothesis with farmers now accepting lower mean real returns to compensate for the same amount of volatility when they have bank access, i.e. they choose less conservative crop allocations. Our preferred IV specification implies an average effective risk aversion of 2.2 (with an interquartile range between 1.9 and 2.7) in the 1970s, which falls to to 1.2 in the 2000s (with an interquartile range between 0.7 and 1.6).

Reassuringly, the combination of the fixed effects and residuals from regression (27)—which we interpret as the unobserved crop costs that ensure the crop choices we observe are optimal from the farmer's perspective—positively correlate with the actual crop costs we observe at the state level for a subset of our sample period; see Appendix Table A.9 for further details.[55]

[52] Because of the presence of the Lagrange multiplier, crop costs are only identified up to scale. In the results that follow, we normalize the cost of the first crop (barley) to zero in each district-decade; this normalization does not affect the estimated change in welfare.

[53] As in Section 3.2, to mitigate outlier concerns, we winsorize μ_{ig}^{Z} and σ_{ig}^{Z} at the 1% and 99% level.

[54] While the assumed linear relationship with $bank_{id}$ is consistent with the empirical evidence above, it is agnostic about the particular mechanism through which access to rural banks serves as a risk mitigating technology. The modeling of banks' decisions regarding lending, expansion, entry, pricing, etc.—see e.g. Salim (2013)—is beyond the scope of this paper, although we explore in Section 5.5 how the welfare impacts of India's highway expansions vary under alternative assumptions regarding the evolution of bank access. One limitation of assuming linearity is that for 11% of district-decades, the linear relationship implies $\rho_{id} < 0$. In our counterfactuals, we truncate these district-decades to have $\rho_{id}=0$, i.e. we impose they are risk-neutral. Non-parametric estimation of the risk aversion parameter by quartile of bank access confirms that $\rho_{id} \geq 0$ for each quartile.

[55] This is particularly reassuring given that our static model abstracts from dynamic crop choice considerations, including crop-rotation, switching costs, etc; see Scott (2013). Note, however, that such dynamic concerns are more relevant when considering year-to-year changes in crop choices than when examining decade-to-decade changes as we do here. That said, our crop choice results from Stylized Fact 2a are essentially unchanged when we allow the adjustment to take place annually within each decade (see columns 5–6 of Appendix Table (A.4)). Because we calibrate the unobserved crop cultivation costs so that the observed crop choice is the equilibrium crop choice, abstracting from dynamic considerations does not affect the model fit; however, there remains the possibility that the model incorrectly predicts the extent to which farmers' crop allocations respond to changes in trade costs. Reassuringly, we find a strong positive correlation between the observed change in crop choices and the model-predicted change resulting from the

5.4 The welfare impacts of the expansion of India's highway network

We now use our structural estimates to quantify the welfare effects of the expansion of the Indian highway network. To do so, we hold all structural parameters, including the distribution of yields, constant at their 1970s levels except for the district- and state-level trade costs ε_{id} and $\bar{\varepsilon}_{md}$. We allow these trade costs to evolve to match observed changes in within-state market access and travel time to Delhi in each decade as described in Section 5.3. We then calculate the equilibrium distributions of real incomes and equilibrium crop choices in all districts; see Appendix A.5 for details. Finally, we calculate the equilibrium realized real income in all locations using the observed realized yields in each year in the 1970s (which ensures that effects depend on the log normal approximation above only through farmers' optimal crop choice).

Panel (a) of Table 6 presents the results, reported as the mean percentage change relative to the 1970s across districts.[56] The expansion of the Indian highway network between the 1970s and 2000s increased mean real incomes for farmers by 2.2%, whereas the variance (of the log) of real income decreased with an average decline across districts of 0.05. Combining these, expected welfare rose by 2.3%. Lower trade costs, and the associated decline in arbitrage revenue going to traders, resulted in declines in the average mean real income of market residents—including traders, drivers, and producers of the homogeneous good—of 0.9% and small declines in the variability of their real income of 0.008.

These average effects belie substantial spatial heterogeneity. Panel (a) of Figure 4 plots changes in state-level market access between the 1970s and 2000s. Panel (b) shows that districts whose state-level market access grew the most experienced greater increases in the mean of their real income—these locations enjoyed first moment gains from specializing in their comparative advantage crops and from being able to consume the comparative advantage crops of other regions more cheaply. To see this claim more precisely, column 1 of Table 7 projects the district-level gains for each decade on within-state market access and the crop-area weighted average of within-state market access for all other districts in the same state. Consistent with these gains coming primarily through improvements in a district's own market access, the coefficient on own market access is more than five times larger than that on market access improvements elsewhere in the state.

A different pattern emerges for the impact of the highway expansion on volatility. Even though real income volatility declines on average, an analogous analysis—see panel (c) of Figure 4 and column 2 of Table 7—shows that districts with the greatest improvements in their own within-state market access actually saw their real incomes become *more* volatile. As Sections 3 and 4 highlight, declines in one's own trade costs reduce the insurance that the response of local prices to local yield shocks naturally provides. But greater integration elsewhere has an opposite effect, reducing volatility by making market prices less susceptible to idiosyncratic shocks. As a result of these opposing forces, real income volatility increased in 112 of 311 districts. Finally, as panel (d) of

expansion of Indian's highways; see Appendix Table A.10.

[56]We report welfare as the percentage increase in nominal income that an agent receives with certainty, holding all parameters at their 1970s values, that would yield the equivalent change in expected utility as from the counterfactual, i.e. the certainty equivalent variation (CEV). See equation (50) and the surrounding discussion in Appendix A.3.2.

Figure 4 and column 3 of Table 7 illustrate, combining the first and second moment effects into welfare, the gains derive from both improvements in own market access and those elsewhere.

5.5 Improvements in risk-mitigating technologies and the gains from trade

We now turn to examining how the growth in rural bank access—a risk mitigation technology—altered the impacts of the highway expansion. As we saw both in Stylized Fact 2 and Section 5.3, farmers were willing to incur greater risk in their crop allocations as their access to rural banks improved. How did this fall in farmer's effective risk aversion affect the gains from trade? To answer this question, panel (b) of Table 6 examines the combined impact of highways and bank access by allowing both trade costs, ε_{id} and $\bar{\varepsilon}_{md}$, and the effective risk aversion parameters, ρ_{id}, to evolve together based on the observed expansions of the highway network and rural banks. Increases in the number of rural banks per capita encouraged farmers to pursue more risky crop allocations than they would have with the highway expansion alone. Relative to our previous counterfactual that held banks at their 1970s levels, mean real incomes rise by an additional 0.6 percentage points—a 27% increase. To achieve these greater mean incomes, farmers incurred greater risk, substantially increasing the volatility of real income with the variance of log real income now rising by 0.7. The welfare gains rise by a substantial 2.1 percentage points to 4.4%.

Panel (c) of Table 6 further asks how much greater would the gains from trade have been if rural India had uniformly-good bank access as the highways were constructed. To do so, we bring any district below the 75th percentile of bank access in a particular decade up to the 75th. Both the mean and variance of incomes rise further as farmers pursue even higher-risk higher-return cropping strategies and thewelfare reach 5.9%.

From where did these additional welfare gains in panels (b) and (c) arise? Improved bank access both makes volatility less costly from a welfare perspective and allows farmers to pursue riskier crop allocations in order to achieve greater mean returns. But improvements in bank access and infrastructure may also be substitutes, as in isolation they may encourage farmers to reallocate crops in incompatible ways. To assess the relative magnitudes of these effects, we first calculate the direct impact of bank access on welfare by changing the effective risk aversion parameter to incorporate the improved bank access, denoted by $\rho = \rho_B$, but holding crop allocations and trade costs at their 1970s levels, with the 1970s crop allocations denoted by $\theta = \theta_{\not R,\not B}$; column 4 presents these results.[57] We then assess the total impact of bank access on welfare by also allowing crop allocations to respond to the change in ρ still holding trade costs at their 1970's levels (i.e. $\theta = \theta_{\not R,B},\ \rho = \rho_B$); column 5 presents these results. Focusing on panel (b) and comparing these numbers to the 2.1 percentage point increase in welfare relative to panel (a), we find that most of the additional welfare gains arise from the direct impact of banks on the welfare cost of volatility. And, if anything, there is a small amount of substitution on average between improvements in bank access and infrastructure.

[57] As in Section 5.4, we report the CEV. In these bank access counterfactuals, we hold fixed farmers' innate risk aversion (i.e. their preferences) but allow their effective risk aversion to change with technological improvements (i.e. bank access). Appendix A.3.2 microfounds this approach including the expression for the CEV.

As above, these mean effects belie substantial heterogeneity. Column 1 of Table 8 projects the additional welfare gain in each district from the combination of bank access and highways relative tohighways alone on within-state market access, the change in effective risk aversion from the improved bank access (i.e. $\rho_{i,d} - \rho_{i,70s}$) and the interaction of the two. Improvements in market access and declines in risk aversion both increase welfare, but act as substitutes on average. However, whether highways and banks are substitutes or complements hinges on whether the riskiest crops are also the comparative advantage ones. In column 2, we show that districts where riskier crops have higher yields (measured by a positive correlationbetween the mean and variance of log yields), banks and highways are complements. Intuitively, in these locations farmers need to act more risk loving to take full advantage of traditional first moment gains from trade. Conversely, in column 3 we find that highways and banks are substitutes when riskier crops have lower yields as here reallocating toward comparative advantage crops also reduces volatility. Column 4 of Table 8 confirms this heterogeneity via a triple interaction between market access, the change in risk aversion, and the mean variance correlation.

If allocations that increased mean real income were available, why did farmers not pursue them without the bank expansion? To answer this question, we evaluate the change in welfare from the chosen crop allocations in each panel assuming that effective risk aversion parameters were fixed at the level consistent with 1970s bank access (i.e. $\theta = \theta_{R,B}$, $\rho = \rho_{B}$). Column 6 presents these results. In both panels (b) and (c), the welfare gains from the more aggressive crop allocations would have been smaller than for the crop allocation chosen in panel (a). That is, farmers were only willing to pursue the riskier crop allocations necessary to achieve greater first-moment returns if they also had access to better risk-mitigation technologies.

6 Conclusion

This paper examines the relationship between trade and volatility in the context of Indian agriculture. We first document that reductions in trade costs due to the expansion of the Indian highway network reduced the elasticity of local prices to local yields, leading farmers to reallocate their land toward crops with lower yield volatility, especially those with worse access to banks. We then embed a portfolio allocation decision into a novel many-location Ricardian trade model. Risk averse producers choose their optimal allocation of resources across goods. This allocation, along with the distributions of trade costs and yields, determines the general equilibrium distribution of real incomes. The model yields tractable equations governing prices and farmers' resource allocations and matches well 40 years of district-level data on yields, prices and cropping patterns.

The model provides intuitive and transparent estimating equations that identify both trade costs—using the relationship between local prices, yield shocks, and prices elsewhere—and farmers' risk preferences—using the slope of the mean-variance frontier at the observed crop choices. Using these estimates, we show that first moment gains from specialization dominate second moment effects and that improvements in risk mitigating technologies allow farmers to achieve greater first moment gains by pursuing riskier crop reallocations.

References

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

— **and Costas Arkolakis**, "Trade and the topography of the spatial economy," *The Quarterly Journal of Economics*, 2014.

Arrow, Kenneth J and Gerard Debreu, "Existence of an equilibrium for a competitive economy," *Econometrica: Journal of the Econometric Society*, 1954, pp. 265–290.

Asher, Sam and Paul Novosad, "Rural roads and local economic development," *American Economic Review*, 2020, *110* (3), 797–823.

Asturias, Jose, Manuel Garcia-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 11 2018, *17* (6), 1881–1940.

Atkin, David, "Trade, Tastes, and Nutrition in India," *American Economic Review*, 2013, *103* (5), 1629–63.

— **and Dave Donaldson**, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Working Paper 21439, National Bureau of Economic Research July 2015.

Basu, Priya, *Improving Access to Finance for India's Rural Poor*, World Bank Publications, 2006.

Bergquist, Lauren, Craig McIntosh, and Meredith Startz, "Search Cost, Intermediation, and Trade: Experimental Evidence from Ugandan Agricultural Markets," *Working paper*, 2021.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and entry in agricultural markets: Experimental evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–47.

— **, Benjamin Faber, Thibault Fally, Matthias Hoelzlein, Edward Miguel, and Andres Rodriguez-Clare**, "Scaling Agricultural Policy Interventions: Theory and Evidence from Uganda," *Unpublished manuscript, University of California at berkeley*, 2019.

Burgess, Robin and Dave Donaldson, "Can openness mitigate the effects of weather shocks? Evidence from India's famine era," *The American Economic Review*, 2010, *100* (2), 449–453.

— **and** — , "Railroads and the Demise of Famine in Colonial India," Technical Report, Working Paper 2012.

— **and Rohini Pande**, "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment," *American Economic Review*, 2005, *95* (3), 780–795.

Campbell, John Y and Luis M Viceira, *Strategic asset allocation: portfolio choice for long-term investors*, Oxford University Press, 2002.

Caselli, Francesco, Miklos Koren, Milan Lisicky, and Silvana Tenreyro, "Diversification Through Trade*," *The Quarterly Journal of Economics*, 09 2019, *135* (1), 449–502.

Chand, Ramesh, "Development Policies and Agricultural Markets," *Economic and Political Weekly*, 2012, *47* (52), 53–63. Publisher: Economic and Political Weekly.

Chatterjee, Shoumitro, "Market Power and Spatial Competition in Rural India," 2020.

— **, M Krishnamurthy, D Kapur, and M Bouton**, "A study of the agricultural markets of Bihar, Odisha and Punjab: Final report," *University of Pennsylvania*, 2020.

Costinot, Arnaud and Dave Donaldson, "How Large Are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997," WP 22946, NBER December 2016.

— **,** — **, and Cory B Smith**, "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world," *Journal of Political Economy*, 2016, *124*, 205–248.

Datta, Saugato, "The impact of improved highways on Indian firms," *Journal of Development Economics*, 2012, *99* (1), 46–57.

Dhingra, Swati and Silvana Tenreyro, "The rise of agribusiness and the distributional consequences of policies on intermediated trade," 2020.

di Giovanni, Julian and Andrei A Levchenko, "Trade openness and volatility," *The Review of Economics and Statistics*, 2009, *91* (3), 558–585.

Disdier, Anne-Célia and Keith Head, "The puzzling persistence of the distance effect on bilateral trade," *The Review of Economics and statistics*, 2008, *90* (1), 37–48.

Dixit, Avinash, "Trade and insurance with moral hazard," *Journal of International Economics*, 1987, *23* (3), 201–220.

_ , "Trade and insurance with adverse selection," *The Review of Economic Studies*, 1989, *56* (2), 235–247.

_ , "Trade and insurance with imperfectly observed outcomes," *The Quarterly Journal of Economics*, 1989, *104* (1), 195–203.

_ **and Victor Norman**, *Theory of international trade: A dual, general equilibrium approach*, Cambridge University Press, 1980.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Duflo, Esther and Rohini Pande, "Dams," *The Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Easterly, William, Roumeen Islam, and Joseph E Stiglitz, "Shaken and stirred: explaining growth volatility," in "Annual World Bank conference on development economics," Vol. 191 2001, p. 211.

Eaton, J. and S. Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eaton, Jonathan and Gene M. Grossman, "Tariffs as Insurance: Optimal Commercial Policy When Domestic Markets Are Incomplete," *The Canadian Journal of Economics*, 1985, *18* (2), pp. 258–272.

Eswaran, Mukesh and Ashok Kotwal, "Implications of Credit Constraints for Risk Behaviour in Less Developed Economies," *Oxford Economic Papers*, 1990, *42* (2), 473–482.

Fafchamps, Marcel, "Cash crop production, food price volatility, and rural market integration in the third world," *American Journal of Agricultural Economics*, 1992, *74* (1), 90–99.

Fulford, Scott L., "The effects of financial development in the short and long run: Theory and evidence from India," *Journal of Development Economics*, 2013, *104* (0), 56 – 72.

Gautam, Abhijeet Kumar, *A National Market for Agricultural Commodities—Some Issues and the Way Forward*, Ministry of Finance, Government of India, 2015.

Ghani, E., A. G. Goswami, and W. R. Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 2016, *126* (591), 317–357.

Goyal, A., "Information, direct access to farmers, and rural market performance in central India," *American Economic Journal: Applied Economics*, 2010, *2* (3), 22–45.

Grant, Matthew and Meredith Startz, "Cutting out the middleman: The structure of chains of intermediation," *Working paper*, 2021.

Hanson, Gordon H, "Market potential, increasing returns and geographic concentration," *Journal of international economics*, 2005, *67* (1), 1–24.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse, Toolkit, and Cookbook," in "Handbook of International Economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Helpman, Elhanan and Assaf Razin, *A Theory of International Trade Under Uncertainty*, Academic Press, 1978.

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

Kapur, Devesh and Mekhala Krishnamurthy, "Understanding mandis: market towns and the dynamics of India's rural and urban transformations," *Univ. of Pennsylvania*, 2014.

Karabay, Bilgehan and John McLaren, "Trade, offshoring, and the invisible handshake," *Journal of international Economics*, 2010, *82* (1), 26–34.

Kurosaki, Takashi and Marcel Fafchamps, "Insurance market efficiency and crop choices in Pakistan," *Journal of Development Economics*, 2002, *67* (2), 419–453.

Lee, Iona Hyojung, "Industrial output fluctuations in developing countries: General equilibrium consequences of agricultural productivity shocks," *European Economic Review*, 2018, *102*, 240–279.

Leemput, Eva Van, "A passage to India: Quantifying internal and external barriers to trade," *Journal of International Economics*, 2021, *131*, 103473.

Mahul, Olivier, Niraj Verma, and Daniel J. Clarke, *Improving Farmers' Access to Agricultural Insurance in India*, The World Bank, 2012.

Mas-Colell, A., M.D. Whinston, and J.R. Green, *Microeconomic theory*, New York: Oxford University Press, 1995.

Mitra, Sandip, Dilip Mookherjee, Maximo Torero, and Sujata Visaria, "Asymmetric Information and Middleman Margins: An Experiment with Indian Potato Farmers," *The Review of Economics and Statistics*, March 2018, *100* (1), 1–13.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of development Economics*, 2004, *73* (1), 185–213.

Newbery, D.M.G. and J.E. Stiglitz, "Pareto inferior trade," *The Review of Economic Studies*, 1984, *51* (1), 1–12.

Redding, Stephen and Anthony J Venables, "Economic geography and international inequality," *Journal of international Economics*, 2004, *62* (1), 53–82.

Redding, Stephen J and Daniel M Sturm, "The costs of remoteness: Evidence from German division and reunification," *American Economic Review*, 2008, *98* (5), 1766–97.

Rodrik, Dani, *Has Globalization Gone Too Far?*, Washington, DC: Institute for International Economics, 1997.

Rosenzweig, Mark R and Hans P Binswanger, "Wealth, weather risk and the composition and profitability of agricultural investments," *The Economic Journal*, 1993, *103* (416), 56–78.

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

Salim, Mir M, "Revealed objective functions of microfinance institutions: evidence from Bangladesh," *Journal of Development Economics*, 2013, *104*, 34–55.

Scott, Paul T, "Dynamic discrete choice estimation of agricultural land use," *Tolouse School of Economics*, 2013.

Sethian, J.A., *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Vol. 3, Cambridge University Press, 1999.

Shah, Mihir, Rangu Rao, and PS Vijay Shankar, "Rural credit in 20th century India: overview of history and perspectives," *Economic and Political Weekly*, 2007, pp. 1351–1364.

Simonovska, Ina and Michael E Waugh, "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 2014, *92* (1), 34–50.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, 2020, *128* (7), 2690–2738.

Tomar, Shekhar, "Gains from Agricultural Market Reform: Role and Size of Intermediaries," *Indian School of Business*, 2016.

Townsend, Robert M, "Risk and insurance in village India," *Econometrica*, 1994, *62* (3), 539–591.

Upton, David and Virginia Fuller, "The ITC eChoupal initiative," *Harvard Business School Case*, 2004, *9*, 1–20.

Willmott, Cort and Kenji Matsuura, "Terrestrial Precipitation: 1900-2010 Gridded Monthly Time Series, Version 3.02," Technical Report, University of Delaware 2012.

Table 1: Price-Yield Elasticities and Openness

Dependent variable:	Log Price								
	(1) OLS	(2) IV	(3) IV	(4) IV	(5) RF	(6) RF	(7) IV	(8) IV	(9) IV
Log(Yield)	-0.033***	-0.069***	-0.071***	-0.068***			-0.082***	-0.064***	-0.080***
	(0.004)	(0.010)	(0.012)	(0.010)			(0.015)	(0.016)	(0.020)
Log(Yield) × State MA	0.036*	0.154***					0.137***	0.227***	0.213***
	(0.019)	(0.038)					(0.040)	(0.061)	(0.060)
Log(Yield) × State MA (phi=1)			0.067***						
			(0.023)						
Log(Yield) × State MA (alt. speed)				0.116***					
				(0.033)					
$\text{Log}(\widehat{\text{Yield}})$ × State MA					0.070**	0.246***			
					(0.032)	(0.060)			
Log(Yield) × National MA							0.035		0.048
							(0.033)		(0.034)
Log(StateYield)								-0.009	-0.001
								(0.013)	(0.012)
Log(StateYield) × State MA								-0.124**	-0.131**
								(0.052)	(0.052)
Log(NationalYield)									0.136***
									(0.039)
Log(NationalYield) × National MA									-0.103*
									(0.054)
District-Crop-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop-Decade Yield Interactions	No	No	No	No	Yes	No	No	No	No
Crop-District Yield Interactions	No	No	No	No	No	Yes	No	No	No
R-squared	0.946	0.001	0.001	0.001	0.946	0.950	0.000	0.001	0.001
Observations	86,811	86,811	86,811	86,811	86,811	86,811	86,811	86,172	86,172
First-Stage F Stat	.	2449.4	2559.0	2442.9	.	.	1631.3	664.6	430.8

Notes: Regressions of local log prices on local log yields and log yields interacted with within-state market access (i.e. access to districts in the same state). Each observation is a district-crop-year triplet. Column (1) reports OLS regression while columns (2)–(4) and (7)–(9) report IV regressions instrumenting local yield terms with equivalents calculated with predicted yields and predicted yield equivalents interacted with market access. Predicted yield obtained from a regression of log yield on local rainfall shocks for each month interacted with state-crop fixed effects and controlling for district-crop-decade, crop-year, and district-year fixed effects. Columns (3) and (4) replace within-state market access with alternate within-state market access measures. Columns (5) and (6) report reduced form results that include crop-decade and crop-district interactions with predicted yields, respectively. Column (7) includes interactions with local log yields and outside-state market access (i.e. access to districts in other states). Columns (8) and (9) include additional interactions of within-state market access with state-level yields (i.e. cropped-area weighted averages of yields in other districts in the same state), and national-level market access with national yields (i.e. other state's cropped-area average yields). Observations are weighted by district-year total cropped area divided by the number of observations in a district year. Market access multiplied by 100,000. Robust standard errors reported in parentheses. Stars indicate statistical significance: * $p<.10$ ** $p<.05$ *** $p<.01$.

Table 2: Crop Choice and Openness

Dependent variable:	IHS fraction of land planted by crop							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	IV	IV	IV	IV	IV	IV
Mean(log Yield)	0.001	0.004	-0.002	0.005*	-0.006**	0.002	0.002	-0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Var(log Yield)	0.008*	0.028**	0.021***	0.006	0.038**	0.080***	0.004	0.066**
	(0.004)	(0.012)	(0.007)	(0.012)	(0.016)	(0.023)	(0.026)	(0.027)
Mean × State MA	0.012***	0.010***	0.015***	0.010**	0.001	0.012***	0.012***	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Var × State MA	-0.034**	-0.125***	-0.056***	-0.074**	-0.080***	-0.224***	-0.083	-0.174***
	(0.016)	(0.031)	(0.014)	(0.034)	(0.028)	(0.062)	(0.075)	(0.058)
Covar(log Yield)				0.028***			0.066***	
				(0.009)			(0.020)	
Covar × State MA				-0.076**			-0.133**	
				(0.030)			(0.060)	
Mean × National MA					0.021***			0.022***
					(0.005)			(0.005)
Var × National MA					-0.044*			0.002
					(0.026)			(0.044)
Var × Bank						-13.319***	-3.019	-10.835**
						(3.665)	(4.025)	(5.053)
Var × State MA × Bank						22.719***	7.370	16.277**
						(8.327)	(9.956)	(7.709)
Covar × Bank							-8.646***	
							(3.013)	
Covar × State MA × Bank							13.646*	
							(8.045)	
Var × National MA × Bank								-1.066
								(4.820)
Crop-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Crop FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.972	-0.001	-0.000	-0.015	0.005	-0.006	-0.034	0.001
Observations	18,639	18,626	15,503	18,626	18,626	18,626	18,626	18,626
First-Stage F Stat	.	117.1	216.0	37.7	84.3	78.6	14.8	22.9

Notes: The dependent variable is the inverse hyperbolic sine of the fraction of land planted with a particular crop. Each observation is a district-crop-decade triplet. Columns (2)-(8) instrument for mean log yields and the variance of log yields with the mean and variance of log predicted yields from a regression of log yield on local rainfall shocks for each month interacted with state-crop fixed effects and controlling for crop-decade, district-decade, and district-crop fixed effects. Interactions with market access are instrumented with the predicted yield instruments interacted with market access. Columns (6)-(8) include additional interactions with district banks per capita. Column (3) replaces functions of yields with functions of the value of production, priced at state-average prices (and instrumented using functions of predicted yields multiplied by district-leave-out state prices). Columns (4) and (7) includes the sum of the covariance of yields with the other 14 crops plus interactions with within-state market access (instrumented with the covariance of predicted yields and interactions with within-state market access). Columns (5) and (8) repeat the interaction analysis with outside-state market access (i.e. access to districts in other states). Market access variables multiplied by 100,000 and banks per capita multiplied by 1000. Observations are weighted by the district-decade total cropped area divided by the number of observations in a district decade. Standard errors clustered at the district-decade level reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table 3: Real Income and Openness

Dependent variable:			Components of Real Income			
	(1)	(2)	(3)	(4)	(5)	(6)
	Var Log Nominal Y	Var Log P Index	Var Log Real Y	Var Log Nominal Y	Var Log P Index	Var Log Real Y
State Market Access	1.719**	-0.475	1.062***	1.858**	-0.510	1.123***
	(0.709)	(0.297)	(0.377)	(0.748)	(0.313)	(0.398)
National Market Access				-0.473	0.117	-0.206
				(0.805)	(0.337)	(0.428)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.542	0.755	0.489	0.542	0.755	0.489
Observations	1166	1166	1166	1166	1166	1166

Notes: Regressions of the variance of the log of real income and its components on within-state market access (i.e. access to districts in the same state) multiplied by 100,000. Columns (4)-(6) additionally include outside-state market access (i.e. access to districts in other states) multiplied by 100,000. Each observation is a district-decade pair and includes all observations with at least 25% of cropped area with observed prices. Observations are weighted by district-decade total cropped area divided by the number of observations in a district decade. Robust standard errors reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table 4: Estimated openness to trade

Panel (a): District Level Openness (ϵ_i)

Dependent variable:	District price ($\ln p_{igt}$)					
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) GMM	(6) GMM
Log yield	-0.034***	-0.120***	-0.040***	-0.151***		
	(0.002)	(0.006)	(0.004)	(0.010)		
State MA × Log yield			0.322*	1.576***		
			(0.178)	(0.420)		
Log state price	0.385***	0.256***	0.382***	0.227***		
	(0.009)	(0.014)	(0.013)	(0.021)		
State MA × Log state price			0.142	1.375**		
			(0.438)	(0.616)		
District trade openness (ϵ_i)	11.315***	2.134***			2.134***	1.705***
	(0.913)	(0.190)			(0.190)	(0.240)
District trade openness (ϵ_i) × State MA						16.860**
						(7.215)
District elasticity of substitution (σ)	18.084***	6.196***			6.196***	5.969***
	(1.309)	(0.307)			(0.307)	(0.284)
Observations	85918	85918	85918	85918	85918	85918
First Stage F-statistic		7293.04		3095.02	291.99	150.50
District-Crop-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
District-Year-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel (b): State Market Access (ϵ_m)

Dependent variable:	State price ($\ln \bar{p}_{mgt}$)					
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) GMM	(6) GMM
Log state quantity	-0.097***	-0.146***	-0.094***	-0.125		
	(0.021)	(0.048)	(0.036)	(0.082)		
Travel time to Delhi × Log state quantity			-0.000	-0.001		
			(0.001)	(0.004)		
Log India price	0.549***	0.287***	0.616***	0.299**		
	(0.064)	(0.058)	(0.134)	(0.127)		
Travel time to Delhi × Log India price			-0.003	-0.000		
			(0.004)	(0.004)		
State trade openness (ϵ_m)	5.643***	1.967**			1.967***	2.073
	(1.445)	(0.818)			(0.546)	(1.439)
State trade openness (ϵ_m) × Travel time to Delhi						-0.005
						(0.035)
State elasticity of substitution (σ)	4.645***	4.881***			4.881***	4.848***
	(1.147)	(1.602)			(1.181)	(1.607)
Observations	6870	6870	6870	6870	6870	6870
First Stage F-statistic		651.22		320.44	8.49	4.29
State-Crop-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Year-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each observation is a district-crop-year triplet (panel (a)) or a state-crop-year triplet (panel (b)). The dependent variable in columns (1)-(4) is the log price in the district (panel (a)) or state (panel (b)), where the state price is the total value produced in the state (at district level prices) divided by the total quantity produced in the state. In columns (2) and (4), yields and quantities are instrumented with rainfall predicted yields and quantities, respectively, and prices are instrumented with prices in the rest of the state (panel (a)) or the rest of the country (panel (b)). Columns (5) and (6) use a GMM specification, where column (5) replicates the results of column (2) and column (6) allows for the implied openness measures to vary with within-state market access (panel (a)) or distance to Delhi (panel (b)). Each observation is weighted by the total cropped area in the district (panel (a)) or state (panel (b)) within a decade. Market access variables are multiplied by 100,000. Robust standard errors are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table 5: Estimated effective risk aversion

Dependent variable:	Mean real returns (μ_{ig}^Z)			
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Variance of real returns (σ_{ig}^Z)	0.554**	1.324***	1.710***	3.265***
	(0.224)	(0.429)	(0.443)	(1.111)
Variance of real returns (σ_{ig}^Z) $\times$ Banks			-0.310***	-0.454**
			(0.098)	(0.217)
District-Decade FE	Yes	Yes	Yes	Yes
District-Crop FE	Yes	Yes	Yes	Yes
Crop-Decade FE	Yes	Yes	Yes	Yes
First stage F-stat		421.491		76.946
R-squared	0.969	-0.004	0.969	-0.005
Observations	14916	14916	14916	14916

Notes: Each observation is a district-crop-decade triplet. The dependent variable is the marginal contribution of a crop to log mean real returns (μ_{ig}^Z). The independent variable is the marginal contribution of a crop to the variance of log real returns (σ_{ig}^Z) and, in columns (3) and (4), its interaction with rural banks per capita. In IV columns, the variance of real returns is instrumented using the variance-covariance matrix of rainfall predicted yields instead of the actual variance-covariance matrix. Both the dependent and independent variables are winsorized at the 1% and 99% level. Each observation is weighted by the total area allocated to the crop within a district-decade. Robust standard errors are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table 6: Welfare impact of the expansion of the Indian highway network

Panel (a): Highway expansion only

	Districts						Markets	
			Welfare					
	(1) Mean	(2) Variance	(3) $\theta_{R,B}, \rho_B$	(4) $\theta_{\not{R},\not{B}}, \rho_B$	(5) $\theta_{\not{R},B}, \rho_B$	(6) $\theta_{R,B}, \rho_{\not{B}}$	(7) Mean	(8) Variance
	2.240***	-0.048***	2.300***	0.000	0.000	2.300***	-0.924	-0.008**
	(0.178)	(0.008)	(0.177)	(.)	(.)	(0.177)	(0.699)	(0.003)
N	311	311	311	311	311	311	17	17

Panel (b): Highway and bank expansion

	Districts						Markets	
			Welfare					
	(1) Mean	(2) Variance	(3) $\theta_{R,B}, \rho_B$	(4) $\theta_{\not{R},\not{B}}, \rho_B$	(5) $\theta_{\not{R},B}, \rho_B$	(6) $\theta_{R,B}, \rho_{\not{B}}$	(7) Mean	(8) Variance
	2.846***	0.692***	4.391***	1.874***	2.161***	2.085***	-0.902	0.004
	(0.201)	(0.123)	(0.241)	(0.176)	(0.192)	(0.204)	(0.721)	(0.002)
N	311	311	311	311	311	311	17	17

Panel (c): Highway and (counterfactual) improved bank expansion

	Districts						Markets	
			Welfare					
	(1) Mean	(2) Variance	(3) $\theta_{R,B}, \rho_B$	(4) $\theta_{\not{R},\not{B}}, \rho_B$	(5) $\theta_{\not{R},B}, \rho_B$	(6) $\theta_{R,B}, \rho_{\not{B}}$	(7) Mean	(8) Variance
	3.133***	1.311***	5.930***	3.041***	3.705***	1.637***	-0.951	0.019**
	(0.217)	(0.238)	(0.326)	(0.258)	(0.302)	(0.281)	(0.770)	(0.008)
N	311	311	311	311	311	311	17	17

Notes: This table reports the estimated effects of the Indian highway expansion. In panel (a), we hold the effective risk-aversion parameter in each district at its 1970s value. In panel (b), we allow each district's effective risk-aversion parameter to change based on the observed change in bank access. In panel (c), we consider a counterfactual where all districts are given effective risk-aversion parameters consistent with being in the upper quartile of the distribution of actual 2000s rural bank access. In columns 1 and 2, we report the average change across districts in the log of mean real income and the variance of the log of real income, respectively, multiplied by 100. Columns 3–6 report the change in welfare measured as the certainty equivalent variation (CEV), i.e. the percentage increase in income that an agent receives with certainty that would generate the equivalent change in expected utility as the counterfactual in question, with the θ and ρ denoting crop choice and effective risk aversion, respectively, the subscript R ($\not{R}$) indicating that the road expansion did (did not) occur, and B ($\not{B}$) indicating that the bank expansion considered in the panel did (did not occur). That is, column 3 reports the actual CEV from both the road and bank expansion considered in the panel header; column 4 reports the CEV using the 1970s crop allocation and trade costs but allowing bank expansion to change the effective risk aversion parameters; column 5 reports the CEV from further allowing bank expansion to change crop allocations; and column 6 reports the CEV using the same crop allocation as in column 3 but evaluating welfare using the 1970s effective risk-aversion parameters. Robust standard errors are reported in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Explaining the heterogeneity across districts in the gains from the expansion of the Indian highway network

Dependent variable:	Mean (1)	Variance (2)	Welfare (3)
State MA	138.234***	0.895**	137.061***
	(19.451)	(0.450)	(19.517)
State MA elsewhere in state	25.934*	-1.989***	28.620*
	(15.381)	(0.674)	(15.318)
District FE	Yes	Yes	Yes
R-squared (within)	0.853	0.016	0.858
Observations	1244	1244	1244

Notes: Each observation is a district-decade pair; there are 4 decades and 311 districts. The dependent variables are the effect of the Indian highway expansion on the log of the mean real returns (column (1)), the variance of the log of real returns (column (2)) and the expected welfare (column (3)), respectively, holding all other parameters constant at 1970s levels. State market access elsewhere in the state is the crop-area weighted average within-state market access in that decade for all other districts within the state. Market access variables are multiplied by 100,000. Units are in log basis points (i.e. approximately percentage points). Standard errors clustered at the district level are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

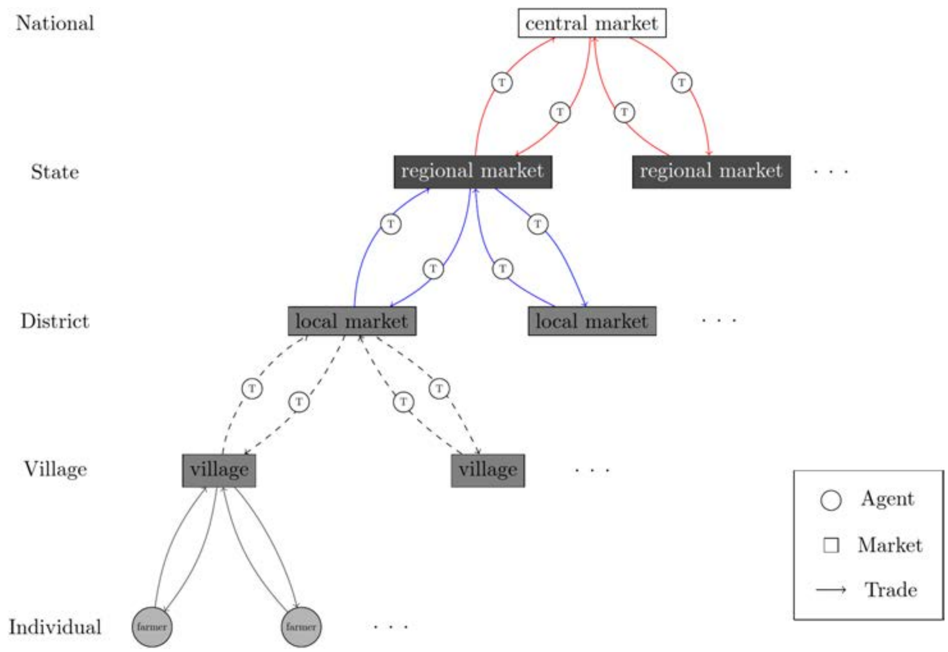
Table 8: The additional gains from the Indian highway network expansion with improved bank access

Dependent variable:	Additional Welfare			
	(1)	(2)	(3)	(4)
State MA	10.033*	-40.996*	14.163**	-23.806**
	(5.578)	(22.589)	(6.932)	(11.112)
Δ Effective risk aversion $(\rho_{i,d}-\rho_{i,70s})$	-2.549***	-0.863***	-3.049***	-1.849***
	(0.310)	(0.217)	(0.392)	(0.231)
State MA $\times$ $(\rho_{i,d}-\rho_{i,70s})$	27.046***	-31.811*	38.013***	-3.605
	(8.385)	(16.586)	(10.090)	(9.511)
State MA $\times$ $\text{Corr}\left(\mu_{i,g}^A,\sigma_{i,g}^A\right)$				-128.825***
				(43.171)
$(\rho_{i,d}-\rho_{i,70s})$ $\times$ $\text{Corr}\left(\mu_{i,g}^A,\sigma_{i,g}^A\right)$				3.151***
				(1.162)
State MA $\times$ $(\rho_{i,d}-\rho_{i,70s})$ $\times$ $\text{Corr}\left(\mu_{i,g}^A,\sigma_{i,g}^A\right)$				-122.274***
				(45.035)
District FE	Yes	Yes	Yes	Yes
R-squared (within)	0.354	0.518	0.374	0.387
Observations	1244	244	1000	1244
Sample	Full	$\text{Corr}\left(\mu_{i,g}^A,\sigma_{i,g}^A\right)>0$	$\text{Corr}\left(\mu_{i,g}^A,\sigma_{i,g}^A\right)<0$	Full

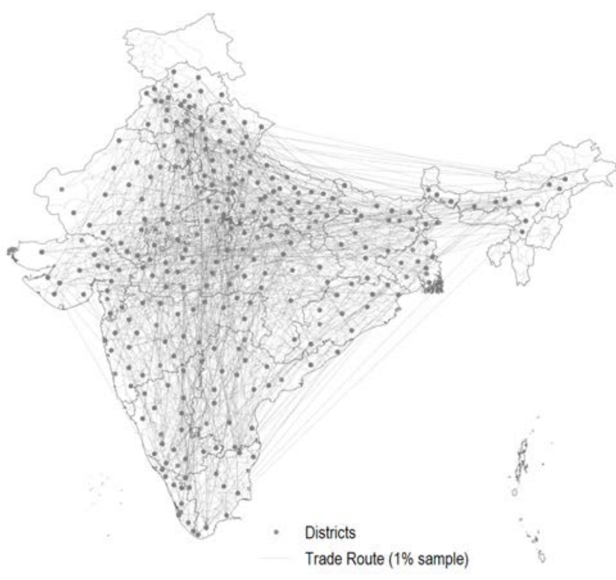
Notes: Each observation is a district-decade pair; there are 4 decades and 311 districts. The dependent variable is the additional impact of the combined Indian highway expansion and rural bank expansion on welfare relative to the Indian highway expansion alone holding all other parameters constant at 1970s levels. Column 2 (3) only includes districts where the correlation across crops within district of the log mean yield and the variance of log yields is positive (negative) in the 1970s, i.e. districts where the high (low) return crops are more riskier. Welfare is measured as the percentage increase in nominal income that an agent receives with certainty that would yield the equivalent change in expected utility as from the counterfactual, i.e. the certainty equivalent variation (CEV). Standard errors clustered at the district level are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Figure 1: A NEW (MORE REALISTIC) MODEL OF THE AGRICULTURAL TRADE NETWORK

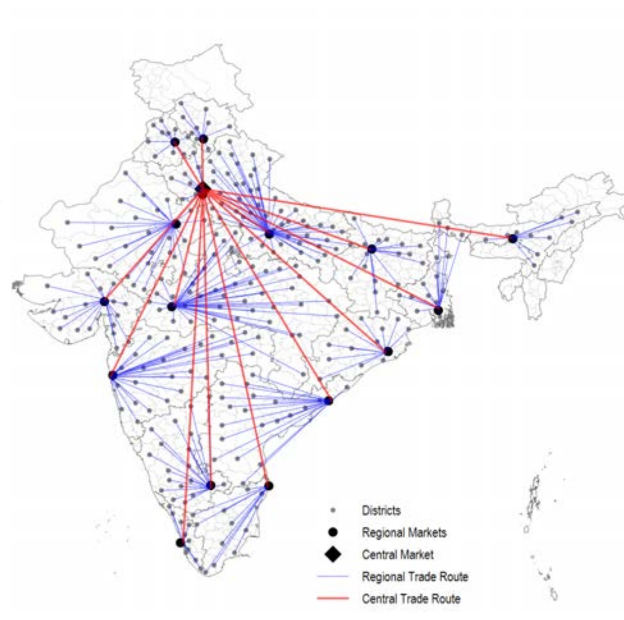
(a) The Indian agricultural trade network



(b) Standard "horizontal" trade network

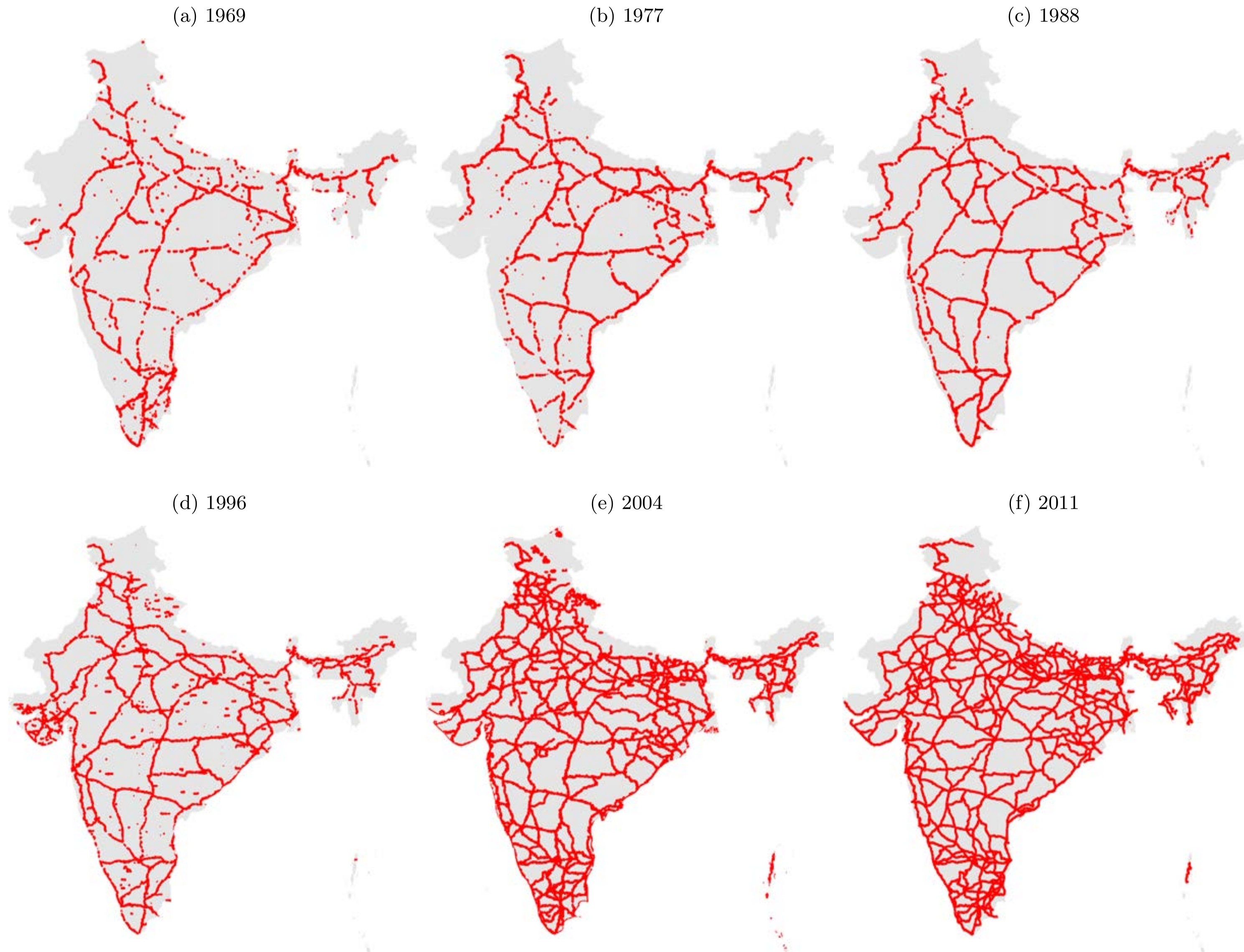


(c) New "hierarchical" trade network



Notes: This figure depicts the Indian agricultural trading network and compares it to the network assumed in a standard trade model and that assumed in our model. Panel (a) illustrates the actual structure of a typical Indian agricultural trading network. Panel (b) depicts the trading network of a standard trade model where each location can trade directly with all other locations (for readability, only a random 1% sample of links are shown). Panel (c) depicts the "hierarchical" trading network in our model, where each district only trades directly with a regional market, which in turn trades with a central market. Note panels (a) and (c) coincide except for the village-to-district trading links, which are excluded in the model due to the absence of village level data.

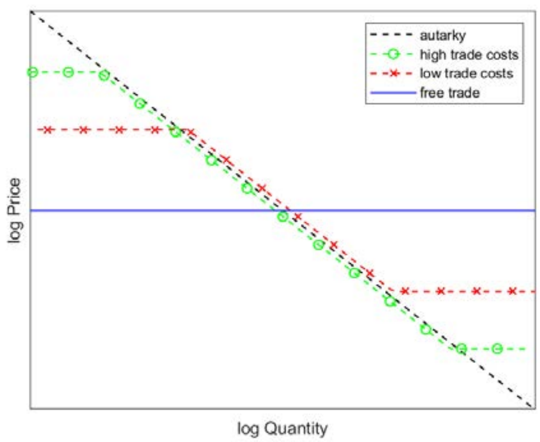
Figure 2: The Indian Highway Network over Time



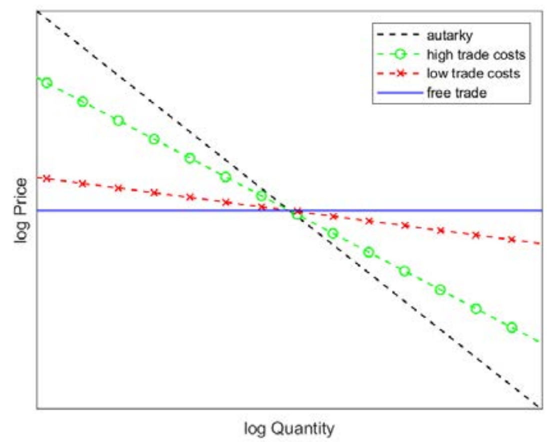
Notes: This figure shows the expansion of the Indian highway network over time. The networks are constructed by geocoding the scanned *Road Map of India* described in Section 2.3 for each of the above years and using image processing to identify the pixels associated with highways. Bilateral distances between all districts are then calculated by applying the Fast Marching Method algorithm (see Sethian (1999)) to the resulting speed image.

Figure 3: A new (more realistic) model of price arbitrage

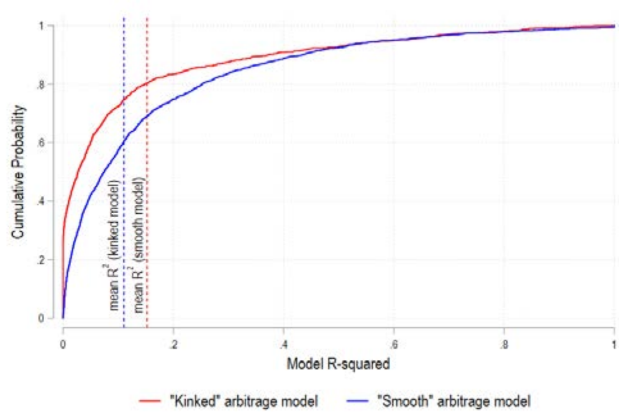
(a) Standard "kinked" arbitrage model



(b) New "smooth" arbitrage model



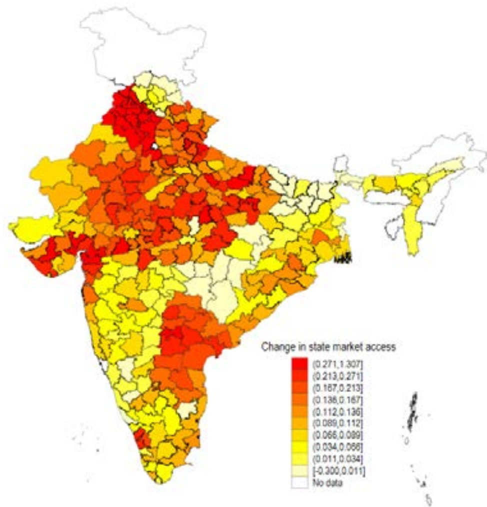
(c) Explaining the observed price-yield relationship



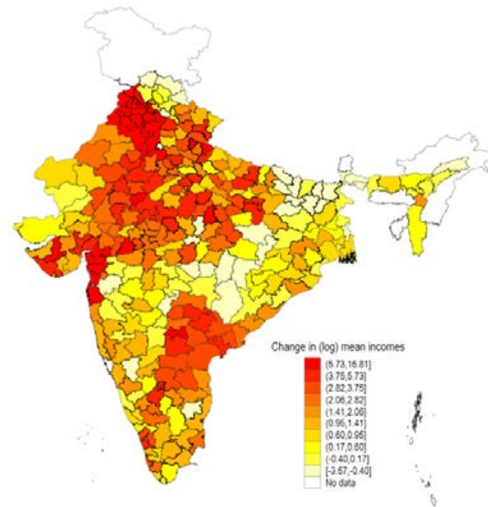
Notes: This figure compares our model to a standard model of price arbitrage. Panel (a) depicts the "kinked" relationship between local prices and local quantities produced in a standard trade model, where (log) local prices are equal to the (log) world price plus/minus an iceberg trade cost other than in a narrow range where relative prices are sufficiently similar that no trade occurs and prices are determined by autarkic demand. Panel (b) depicts the "smooth" relationship between local prices and local quantities in our model, where heterogeneous trade costs ensure that some trade occurs at all prices, and the distribution of trade costs across traders determines the elasticity of local prices to local quantities produced. Panel (c) compares the fit of the two models to Indian data on rainfall-predicted quantities and observed yields and reports the distribution of R^2 for each model across all district-decade pairs in our sample; see Section 4.2 for details.

Figure 4: THE SPATIAL DISTRIBUTION OF THE GAINS FROM TRADE

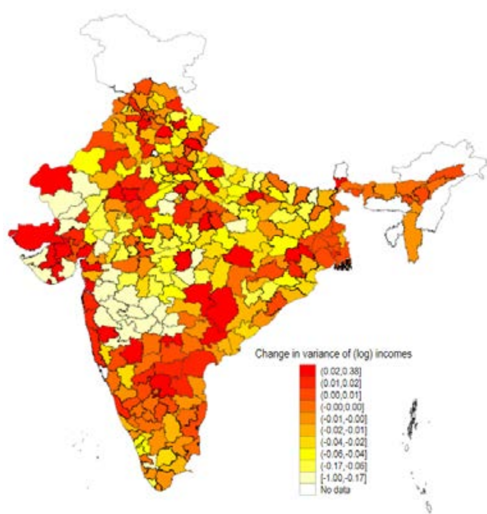
(a) Change in market access



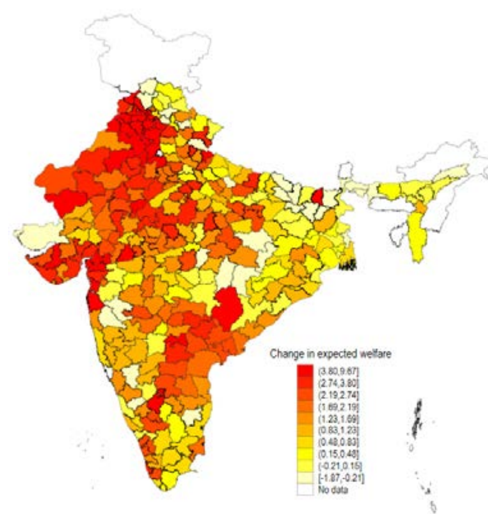
(b) Change in mean real income



(c) Change in variance of (log) real income



(d) Change in expected welfare



Notes: This figure presents the spatial distribution of the gains from trade resulting from the expansion of the Indian highway network from the 1970s to the 2000s. Panel (a) depicts the change in the observed (within-state) market access; panel (b) depicts the change in the (log of) mean real income; panel (c) depicts the change in the variance (of the log) of real income; and panel (c) depicts the change in expected welfare. The units of panels (b), (c), and (d) are log basis points (i.e. approximately percentage points). In all panels, reds (yellow) indicate higher (lower) deciles of changes.

Volatility and the Gains from Trade: Appendix

Treb Allen and David Atkin

A Appendix

A.1 Model derivations

In this subsection, we present the derivations of several results in the main paper.

Approximation of real returns (Equation 17) First, to calculate the income of farmers in a village, we combine equation (5) and (11) to yield:

$$Y_i(s) = \left(\sum_{g\in\mathcal{G}} \alpha_{ig}\left(\frac{\bar{p}_g(s)Q_{ig}(s)}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\frac{1+\varepsilon_i}{\varepsilon_i}}. \tag{28}$$

Similarly, combining equation (6) with (11) yields the following expression for the period welfare of a farmer in village i:

$$Z_i^f(s) = \frac{1}{L_i} \times \left(\sum_{g\in\mathcal{G}} \alpha_{ig}\left(\frac{\bar{p}_g(s)Q_{ig}(s)}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right) \times \prod_{g\in\mathcal{G}}\left(Q_{ig}(s)\left(\frac{\alpha_{ig}}{\bar{p}_g(s)Q_{ig}(s)}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\alpha_{ig}} \tag{29}$$

In the autarky (i.e. $\varepsilon_i = 0$), equation (29) simplifies to $Z_i^{f,aut}(s) \equiv \frac{1}{L_i}\prod_{g\in\mathcal{G}}(Q_{ig}(s))^{\alpha_{ig}}$, as farmers consume what they produce. In free trade (i.e. $\varepsilon_i \to \infty$), equation (29) simplifies to $Z_i^{f,free}(s) \equiv \frac{1}{L_i} \times \left(\sum_{g\in\mathcal{G}}\bar{p}_g(s)Q_{ig}(s)\right) \times \prod_{g\in\mathcal{G}}\left(\frac{\alpha_{ig}}{\bar{p}_g(s)}\right)^{\alpha_{ig}}$,as farmers sell what they produce and purchase what they consume at the central market prices.

We now note that with a large number of villages and idiosyncratic shocks that $\bar{p}_g(s) = \bar{p}_g$, i.e. the central market prices is state invariant. Taking logs of equation 29 then yields:

$$\begin{aligned}\ln Z_i^f(s) = &\ln\left(\sum_{g\in\mathcal{G}} \theta_{ig}^f \times \frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g(s)\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)\\ &+\left(\frac{1}{1+\varepsilon_i}\right)\sum_{g\in\mathcal{G}}\alpha_{ig}\ln A_{ig}(s) + \sum_{g\in\mathcal{G}}\alpha_{ig}\left(\ln\left(\alpha_{ig}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\right)^{\frac{1}{1+\varepsilon_i}}\right) - \ln\bar{p}_g\right)\end{aligned} \tag{30}$$

We then apply the following second-order approximation implying that the sum of log normal variables is itself approximately log normal (see, e.g. Campbell and Viceira (2002)). Suppose that $\ln\boldsymbol{x}_i(s) \sim N(\boldsymbol{\mu}_i^x, \boldsymbol{\Sigma}_i)$ and $X_i(s) \equiv \ln\left(\sum_{g\in\mathcal{G}} w_{i,g}x_{i,g}(s)\right)$ for some weights $\sum_{g\in\mathcal{G}} w_{i,g} = 1$. Then a second order approximation around the mean log returns is:

$$X_i(s) \approx \ln\left(\sum_{g\in\mathcal{G}} w_{i,g}\exp(\mu_{i,g}^x)\right) + \sum w_{i,g}(\ln x_{i,g}(s) - \mu_{i,g}^x) - \frac{1}{2}\sum_{h\in\mathcal{G}}\sum_{g\in\mathcal{G}} w_{i,g}w_{i,h}\sigma_{i,gh}^x + \frac{1}{2}\sum_{g\in\mathcal{G}} w_{i,g}\sigma_{i,gg}^x. \tag{31}$$

In our case, we have:

$$\ln x_{ig}(s) \equiv \ln\left(\frac{\alpha_{ig}}{\theta_{ig}}\right) + \frac{\varepsilon_i}{1+\varepsilon_i}\ln\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\right) + \frac{\varepsilon_i}{1+\varepsilon_i}\ln(A_{ig}(s))$$

and $w_{i,g} \equiv \theta_{i,g}^f$ which implies that $\mu_{i,g}^x = \ln\left(\frac{\alpha_{ig}}{\theta_{ig}}\right) + \frac{\varepsilon_i}{1+\varepsilon_i}\ln\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\right) + \frac{\varepsilon_i}{1+\varepsilon_i}\mu_g^{A,i}$ and $\sigma_{i,gh}^x = \left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2\sigma_{i,gh}^A$.

Applying the approximation (31) to the real returns (30) results in:

$$\ln Z_i^f(s) \approx \mu_i^Z + \sum_{g\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,g}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ig} \right) \left(\ln A_{ig}(s) - \mu_g^{A,i} \right),$$

where

$$\begin{aligned}
\mu_i^Z \equiv & \sum_{g\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,g}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ig} \right) \mu_g^{A,i} + \ln \left(\sum_{g\in\mathcal{G}} \theta_{ig}^f \times \frac{\alpha_{ig}}{\theta_{ig}} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \exp\left(\mu_g^{A,i}\right) \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) \\
& - \frac{\varepsilon_i}{1+\varepsilon_i} \sum_{g\in\mathcal{G}} \theta_{i,g}^f \mu_g^{A,i} + \sum_{g\in\mathcal{G}} \alpha_{ig} \left(\ln \left(\alpha_{ig} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \right)^{\frac{1}{1+\varepsilon_i}} \right) - \ln \bar{p}_g \right) \\
& + \frac{1}{2} \left(\frac{\varepsilon_i}{1+\varepsilon_i} \right)^2 \left(\sum_{g\in\mathcal{G}} \theta_{i,g}^f \Sigma_{gg}^{A,i} - \sum_{h\in\mathcal{G}} \sum_{g\in\mathcal{G}} \theta_{i,g}^f \theta_{i,h}^f \Sigma_{gh}^{A,i} \right),
\end{aligned} \tag{32}$$

as required.

It immediately follows that farmer utility is (approximately) log normally distributed across states of the world:

$$\ln Z_i^f \sim N\left(\mu_i^Z, \sigma_i^{2,Z}\right),$$

where

$$\sigma_i^{2,Z} \equiv \sum_{g\in\mathcal{G}} \sum_{h\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,g}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ig} \right) \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,h}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ih} \right) \Sigma_{gh}^{A,i}. \tag{33}$$

Optimal crop choice first order conditions (equation 20) Beginning with the maximization problem:

$$\max_{\{\theta_{ig}^f\}} \mu_i^Z + \frac{1}{2}(1-\rho_i)\sigma_i^{2,Z} \text{ s.t. } \Sigma_{g\in\mathcal{G}} \theta_{ig}^f = 1$$

and substituting in the expressions for μ_i^Z and $\sigma_i^{2,Z}$ from equation (18) results in:

$$\begin{aligned}
\max_{\{\theta_{ig}^f\}} & \ln \left(\sum_{g\in\mathcal{G}} \theta_{ig}^f \frac{\alpha_{ig}}{\theta_{ig}} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \exp\left(\mu_g^{A,i}\right) \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) + \left(\frac{1}{1+\varepsilon_i} \right) \sum_{g\in\mathcal{G}} \alpha_{ig} \mu_g^{A,i} + \sum_{g\in\mathcal{G}} \alpha_{ig} \left(\ln \left(\alpha_{ig} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \exp\left(\mu_g^{A,i}\right) \right)^{\frac{1}{1+\varepsilon_i}} \right) - \ln \bar{p}_g \right) \\
& + \frac{1}{2} \left(\frac{\varepsilon_i}{1+\varepsilon_i} \right)^2 \left(\sum_{g\in\mathcal{G}} \theta_{i,g}^f \Sigma_{gg}^{A,i} - \sum_{h\in\mathcal{G}} \sum_{g\in\mathcal{G}} \theta_{i,g}^f \theta_{i,h}^f \Sigma_{gh}^{A,i} \right) \\
& + \frac{1}{2}(1-\rho_i) \sum_{g\in\mathcal{G}} \sum_{h\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,g}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ig} \right) \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,h}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ih} \right) \Sigma_{gh}^{A,i}
\end{aligned}$$

subject to:

$$\Sigma_{g\in\mathcal{G}} \theta_{ig}^f = 1.$$

Taking the first order conditions with respect to θ_{ig}^f (note that each farmer makes her crop choice taking the crop choice of other farmers as given) results in the following first order conditions:

$$\begin{aligned}
& \frac{\frac{\alpha_{ig}}{\theta_{ig}} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \exp\left(\mu_g^{A,i}\right) \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}} \theta_{ig}^f \times \frac{\alpha_{ig}}{\theta_{ig}} \left(\frac{\bar{p}_g \theta_{ig}}{\alpha_{ig}} \exp\left(\mu_g^{A,i}\right) \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} + \frac{1}{2} \left(\frac{\varepsilon_i}{1+\varepsilon_i} \right)^2 \Sigma_{gg}^{A,i} + \frac{\varepsilon_i}{(1+\varepsilon_i)^2} \sum_{h\in\mathcal{G}} \alpha_{ih} \Sigma_{gh}^{A,i} \\
& \qquad - \rho_i \left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \sum_{h\in\mathcal{G}} \left(\left(\frac{\varepsilon_i}{1+\varepsilon_i} \right) \theta_{i,h}^f + \left(\frac{1}{1+\varepsilon_i} \right) \alpha_{ih} \right) \Sigma_{gh}^{A,i} = \lambda_i
\end{aligned}$$

or equivalently:

$$\mu_{ig}^{Z}-\rho_i\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\sum_{h\in\mathcal{G}}\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,h}^{f}+\left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ih}\right)\Sigma_{gh}^{A,i}=\lambda_i,$$

where $\mu_{ig}^{Z}\equiv\frac{1}{\theta_{ig}}\frac{\alpha_{ig}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\exp\left(\mu_g^{A,i}\right)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}}\alpha_{ig}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\exp\left(\mu_g^{A,i}\right)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}+\frac{1}{2}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2\Sigma_{gg}^{A,i}+\frac{\varepsilon_i}{(1+\varepsilon_i)^2}\sum_{h\in\mathcal{G}}\alpha_{ih}\Sigma_{gh}^{A,i}$, as required.

Equilibrium crop choice (equation 21) We re-write the first order conditions as:

$$\frac{\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\exp\left(\mu_g^{A,i}\right)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}}\theta_{ig}^{f}\times\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\exp\left(\mu_g^{A,i}\right)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}=\lambda_i-\Big(\frac{1}{2}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2\Sigma_{gg}^{A,i}+\frac{\varepsilon_i}{(1+\varepsilon_i)^2}\sum_{h\in\mathcal{G}}\alpha_{ih}\Sigma_{gh}^{A,i}$$

$$-\rho_i\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\sum_{h\in\mathcal{G}}\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,h}+\left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ih}\right)\Sigma_{gh}^{A,i}\Big)\iff$$

$$\theta_{ig}\propto\alpha_{ig}(\bar{p}_gB_{ig})^{\varepsilon_i}\implies$$

$$\theta_{ig}=\frac{\alpha_{ig}(\bar{p}_gB_{ig})^{\varepsilon_i}}{\sum_{h\in\mathcal{G}}\alpha_{ih}(\bar{p}_hB_{ih})^{\varepsilon_i}},$$

where $B_{ig}\equiv\frac{\exp\mu_g^{A,i}}{\left(\lambda_i-\left(\frac{1}{2}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2\Sigma_{gg}^{A,i}+\frac{\varepsilon_i}{(1+\varepsilon_i)^2}\sum_{h\in\mathcal{G}}\alpha_{ih}\Sigma_{gh}^{A,i}-\rho_i\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\sum_{h\in\mathcal{G}}\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,h}+\left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ih}\right)\Sigma_{gh}^{A,i}\right)\right)^{\frac{1+\varepsilon_i}{\varepsilon_i}}}$, as required.

A.2 Proofs

This subsection contains the proofs of Propositions 1 and 2.

A.2.1 Proof of Proposition 1

We first restate the proposition:

Proposition. *Given any set of preferences* $\{\alpha_{ig}\}_{g\in\mathcal{G}}$*, trade costs* $\{\varepsilon_i\}_{i\in\mathcal{N}}$*, and any state of the world* $s\in S$ *such that quantity produced is* $\{Q_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$*:*
(a) There exists a state equilibrium.
(b) If the trade costs $\{\varepsilon_i\}_{i\in\mathcal{N}}$ *are sufficiently close to* 1*, then that equilibrium is unique.*

Proof of part (a) (existence)

Proof. In what follows, we omit dependence of prices $p_{ig}(s)$ and quantities $Q_{ig}(s)$ on state s for clarity. To prove existence, we first show that it is sufficient to focus on the excess demand function of the central market. We then show that the central market excess demand function satisfies all conditions necessary to guarantee existence from Proposition 17.C.1 of Mas-Colell et al. (1995).

We first note that given quantities $\{Q_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ and the equilibrium central market prices $\{\bar{p}_g\}_{g\in\mathcal{G}}$, village level incomes $\{Y_i\}_{i\in\mathcal{N}}$ are given immediately from equation prices (28); in turn, given village incomes $\{Y_i\}_{i\in\mathcal{N}}$, village level prices $\{p_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ are then given immediately from equation (11); and finally, given village level prices $\{p_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$, village level consumption $\{C_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ are given immediately from equation (10). That is, given quantities $\{Q_{ig}\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ and the equilibrium central market prices $\{\bar{p}_g\}_{g\in\mathcal{G}}$, it is straightforward to find a set of village prices $\{p_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ and village consumption $\{C_{ig}(s)\}_{i\in\mathcal{N}}^{g\in\mathcal{G}}$ such that markets clear within each village (and condition 1 of the state equilibrium is satisfied). Hence, all that remains to determine the full state equilibrium is the set of equilibrium central market prices $\{\bar{p}_g\}_{g\in\mathcal{G}}$ such that the central market clears.

To find the equilibrium central market prices, we consider the following central market excess demand function $Z \equiv \{Z_g\}_{g\in\mathcal{G}} : \mathbb{R}^G \to \mathbb{R}^G$:

$$Z_g\left(\{\bar{p}_g\}_{g\in\mathcal{G}}\right) : \frac{\bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\frac{\bar{p}_h}{p_{ih}}\right)^{-1}\right)\left(1-\left(\frac{\bar{p}_h}{p_{ih}}\right)^{-\varepsilon_i}\right) Q_{ih}}{\bar{p}_g} - \sum_i \left(1-\left(\frac{\bar{p}_g}{p_{ig}}\right)^{-\varepsilon_i}\right) Q_{ig} \iff$$

$$Z_g\left(\{\bar{p}_g\}_{g\in\mathcal{G}}\right) : \frac{\bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1-\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}}{\bar{p}_g} \tag{34}$$

$$-\sum_i \left(1-\left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_g^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig}, \tag{35}$$

where the first term of Z_g is the quantity of good g demanded by the central market at price vector $\{\bar{p}_g\}_{g\in\mathcal{G}}$ (see equation (13)) and the second term is the quantity of good g supplied to the central market at price vector $\{\bar{p}_g\}_{g\in\mathcal{G}}$ (see equation (12)) and the second line uses equations (11) and (28) to substitute out for village level prices.

We now verify that the excess demand function defined by (35) satisfies conditions (i) to (v) of Proposition 17.B.2 of Mas-Colell et al. (1995), which from Proposition 17.C.1 of Mas-Colell et al. (1995) guarantees the existence of a set of central market prices $\{\bar{p}_g(s)\}_{g\in\mathcal{G}}$ and central market consumption $\left\{\bar{C}_g(s)\right\}_{g\in\mathcal{G}}$ that clear the central market (i.e. satisfy condition 2 of the state equilibrium).

Condition (i): Continuity. This is self evident from equation (35).

Condition (ii): Homogeneity of degree zero in prices. For any $C > 0$, we have:

$$Z_g(\{C\bar{p}_g\}) = \frac{\bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_h)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_l)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1-\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_h)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_l)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}}{\bar{p}_g}$$

$$-\sum_i \left(1-\left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (C\bar{p}_h)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig}$$

$$= \frac{\bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1-\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}}{\bar{p}_g}$$

$$-\sum_i \left(1-\left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_g^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig}$$

$$= Z_g(\{\bar{p}_g\}),$$

as required.

Condition (iii): Walras' law. We have:

$$
\begin{aligned}
\sum_g \bar{p}_g Z_g =& \sum_g \bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1-\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih} \\
&-\sum_g \bar{p}_g \sum_i \left(1-\left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_g^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig} \\
=& -\sum_h \sum_i \left(\frac{\left(\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\frac{1}{\varepsilon_i}}}{\alpha_{ih}^{-\frac{1}{1+\varepsilon_i}} Q_{ih}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{-\frac{\varepsilon_i}{1+\varepsilon_i}}}\right) + \sum_h \sum_i \left(\frac{\left(\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon}}\right)^{\frac{1+\varepsilon_i}{\varepsilon_i}}}{\alpha_{ih}^{-1}}\right) \\
=& \left(\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\frac{1}{\varepsilon_i}} \left[-\sum_h \sum_i \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} + \sum_h \sum_i \alpha_{ih} \sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right] \\
=& \left(\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\frac{1}{\varepsilon_i}} \left[-\sum_h \sum_i \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} + \sum_i \sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right] \\
=& 0,
\end{aligned}
$$

as required.

Condition (iv): Bounded below. In particular, we need that there is an $s>0$ such that $Z_g(p)>-s$ for all p and all goods g. This is straightforward as the first sum must be nonnegative. To see this, note that in each term we have something of the form$\bar{p}_h(1-x)\left(1-x^{\frac{1}{\varepsilon_i}}\right)$ with $x>0$. If $x>1$, both $1-x$ and $1-x^{\frac{1}{\varepsilon_i}}$ are negative and the term is positive. Similarly, if $x<1$, both terms are positive. If $x=0$, It is zero. For the second sum, we have something of the form $1-x$ for each term with $x>0$. Therefore,

$$
\begin{aligned}
Z_g(\{\bar{p}_g\}) =& \frac{\bar{\alpha}_g \sum_h \sum_i \bar{p}_h \left(1-\left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1-\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_l^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}}{\bar{p}_g} \\
&-\sum_i \left(1-\left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_g^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \bar{p}_h^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig} \\
\geq& -\sum_i Q_{ig}
\end{aligned}
$$

Then we can take $s = \max_g \sum_i Q_{ig}$, and $Z_g(\{\bar{p}_g\}) \geq -s$ for all g and $\{\bar{p}_g\}$.

Condition (v): Limiting behavior as prices go to zero. Condition (v) requires that if $p^n \to p$, where $p \neq 0$ and $p_g = 0$ for some g, then $\max_g \lim_{n\to\infty} Z_g(p^n) \to \infty$. To see this, choose g such that $\lim_n \frac{p_g^n}{p_h^n} < \infty$ for all h; intuitively, p_g^n goes to 0 as fast or faster than any other price p_h^n. Since $p \neq 0$, there must be an h' such that $\lim_n \frac{p_g^n}{p_{h'}^n} = 0$. We have that

$$Z_g(p^n) = \frac{\bar{\alpha}_g \sum_h \sum_i p_h^n \left(1 - \left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_h^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_l^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1 - \alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_h^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_l^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}}{p_g^n}$$

$$- \sum_i \left(1 - \left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_g^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} (p_h^n)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig}$$

$$= \bar{\alpha}_g \sum_h \sum_i \frac{p_h^n}{p_g^n} \left(1 - \left(\alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\frac{p_l^n}{p_h^n}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)^{\frac{1}{\varepsilon_i}}\right)\left(1 - \alpha_{ih}\left(\frac{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\frac{p_l^n}{p_h^n}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right) Q_{ih}$$

$$- \sum_i \left(1 - \left(\alpha_{ig}\left(\frac{\alpha_{ig}^{\frac{1}{1+\varepsilon_i}} Q_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\frac{p_h^n}{p_g^n}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}\right)^{-1}\right)\right) Q_{ig}$$

This goes to ∞ as $n \to \infty$. To see this, consider the h such that $\lim_n \frac{p_h^n}{p_g^n} = \infty$. Then to guarantee $Z_g(p^n) \to \infty$, we simply need that

$$\alpha_{ih} \frac{\sum_l \alpha_{il}^{\frac{1}{1+\varepsilon_i}} Q_{il}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\frac{p_l}{p_h}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\alpha_{ih}^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}}}$$

does not equal 1 for one of those h and i. If there is any l and h such that $\lim_n \frac{p_l^n}{p_h^n} = 0$ and $\lim_n \frac{p_h^n}{p_g^n} = \infty$, then clearly this must be the case as $\frac{p_l}{p_h} = \infty$. The alternative is that there are is some subset $(p_{h_1}, \dots, p_{h_n})$ such that $0 < \frac{p_{h_i}}{p_{h_j}} < \infty$ and $\frac{p_g}{p_{h_i}} = 0$ for all of the other goods. For Z_g to not explode, these must all equal 0. That gives n equations for a given i

$$\alpha_{ih_j}^{\frac{1}{1+\varepsilon_i}} Q_{ih_j}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h_j}^{\frac{\varepsilon_i}{1+\varepsilon_i}} = \alpha_{ih_j} \sum_k \alpha_{ih_k}^{\frac{1}{1+\varepsilon_i}} Q_{ih_k}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h_k}^{\frac{\varepsilon_i}{1+\varepsilon_i}}, \forall j$$

The only solution to this linear system is $\alpha_{ih_j}^{\frac{1}{1+\varepsilon_i}} Q_{ih_j}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h_j}^{\frac{\varepsilon_i}{1+\varepsilon_i}} = 0$. This contradicts the fact that $p \neq 0$. Therefore, we must have that one of these does not equal to 1, meaning that $Z_g(p^n) \to \infty$.

Since the excess demand function $Z_g\left(\{\bar{p}_g\}_{g\in\mathcal{G}}\right)$ satisfies conditions (i)-(v), recall from above that Proposition 17.C.1 of Mas-Colell et al. (1995) guarantees the existence of a set of central market prices $\{\bar{p}_g(s)\}_{g\in\mathcal{G}}$ and central market consumption $\left\{\bar{C}_g(s)\right\}_{g\in\mathcal{G}}$ that clear the central market (i.e. satisfy condition 2 of the state equilibrium). As condition 1 is then trivially satisfied (see above), this establishes the existence of a state equilibrium. □

Proof of part (b) (uniqueness)

Proof. To establish sufficient conditions for uniqueness, we show that the excess demand function $Z_g\left(\{\bar{p}_g\}_{g\in\mathcal{G}}\right)$ defined in equation (35) satisfies the gross substitutes property $\partial Z_g\left(\{\bar{p}_g\}_{g\in\mathcal{G}}\right)/\partial\bar{p}_h > 0$ for all $h' \neq g$ as long as $\{\varepsilon_i\}$ is sufficiently close to one for all $i \in \mathcal{N}$. Then from Proposition 17.F.3 of Mas-Colell et al. (1995), there exists at most one equilibrium, which, when combined with part (a) (existence) of this proposition, implies that the equilibrium is unique.

We have:

$$p_g \frac{\partial Z_g(p)}{\partial p_{h'}} = \sum_i \bar{\alpha}_g Q_{ih'} - \bar{\alpha}_g \frac{Q_{ih'}}{1+\varepsilon_i} \alpha_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} p_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right)$$

$$- \bar{\alpha}_g \frac{\varepsilon_i}{1+\varepsilon_i} \left(\sum_h \alpha_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_h^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{1}{1+\varepsilon_i}} \right) \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} + \frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}}$$

$$= \sum_i \bar{\alpha}_g Q_{ih'} - \bar{\alpha}_g \frac{Q_{ih'}}{1+\varepsilon_i} \alpha_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} p_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right)$$

$$- \bar{\alpha}_g \frac{\varepsilon_i}{1+\varepsilon_i} Q_{ih'} \left(\sum_h \alpha_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_h^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{1}{1+\varepsilon_i}} \right) \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} Q_{ih'}^{-\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} + \frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}}$$

$$\geq \sum_i \bar{\alpha}_g Q_{ih'} \frac{\varepsilon_i}{1+\varepsilon_i} \left[\alpha_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} p_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) - \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} Q_{ih'}^{-\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_h^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{1}{1+\varepsilon_i}} \right) \right]$$

$$+ \frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}}$$

When $\varepsilon_i = 1$ for all $i \in \mathcal{N}$ we then have:

$$p_g \frac{\partial Z_g(p)}{\partial p_{h'}} \geq \frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \iff$$
$$p_g \frac{\partial Z_g(p)}{\partial p_{h'}} > 0,$$

since $\frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} > 0$. Moreover, by continuity, the exists a $\delta > 0$ where, for all ε_i such that $|\varepsilon_i - 1| < \delta$: we have

$$\frac{\varepsilon_i}{1+\varepsilon_i} \alpha_{ig}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ig}^{\frac{1}{1+\varepsilon_i}} p_g^{\frac{1}{1+\varepsilon_i}} \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} Q_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \geq \left| \sum_i \bar{\alpha}_g Q_{ih'} \frac{\varepsilon_i}{1+\varepsilon_i} \left[\begin{array}{c} \alpha_{ih'}^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} p_{h'}^{-\frac{\varepsilon_i}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{1}{1+\varepsilon_i}} p_h^{\frac{\varepsilon_i}{1+\varepsilon_i}} Q_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) \\ - \alpha_{ih'}^{\frac{1}{1+\varepsilon_i}} Q_{ih'}^{-\frac{1}{1+\varepsilon_i}} p_{h'}^{-\frac{1}{1+\varepsilon_i}} \left(\sum_h \alpha_{ih}^{\frac{\varepsilon_i}{1+\varepsilon_i}} p_h^{\frac{1}{1+\varepsilon_i}} Q_{ih}^{\frac{1}{1+\varepsilon_i}} \right) \end{array} \right] \right|$$

so that $p_g \frac{\partial Z_g(p)}{\partial p_{h'}} > 0$ for all ε_i such that $|\varepsilon_i - 1| < \delta$, as claimed. □

A.2.2 Proof of Proposition 2

We first restate the proposition:

Proposition. *Consider a village i which increases its openness to trade, i.e. ε_i increases by a small amount. Then:*

(1) [Stylized Fact 1] Any increase in openness: (1a) decreases the responsiveness of local prices to local yield shocks; and (1b) increases the responsiveness of local prices to the central market price:

$$\frac{d}{d\varepsilon_i}\left(-\frac{\partial \ln p_{ig}(s)}{\partial \ln A_{ig}(s)}\right) < 0 \text{ and } \frac{d}{d\varepsilon_i}\left(\frac{\partial \ln p_{ig}(s)}{\partial \ln \bar{p}_g}\right) > 0.$$

(2) [Stylized Fact 2] Starting from autarky, an increase in openness: (2a) causes farmers to reallocate production toward crops with higher mean and less volatile yields (as long as $\rho_i > 1$, i.e. farmers are sufficiently risk averse); and (2b) the reallocation toward less volatile crops is attenuated the greater the

access to insurance (i.e. the lower ρ_i). Formally, for any two crops $g \neq h$:

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial(\ln\theta_{ig}-\ln\theta_{ih})}{\partial\left(\mu_g^{A,i}-\mu_h^{A,i}\right)}\right)|_{\varepsilon_i=0}>0,\ \frac{d}{d\varepsilon_i}\left(\frac{\partial\ln\theta_{ig}-\partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i}-\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0}<0,$$

$$\text{and } -\frac{d^2}{d\varepsilon_i d\rho_i}\left(\frac{\partial\ln\theta_{ig}-\partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i}-\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0}>0.$$

(3) [Stylized Fact 3] Consider a decomposition of the variance of real returns as follows:

$$\sigma_i^{2,Z}=\sigma_i^{2,Y}+\sigma_i^{2,P}-2cov_i^{Y,P},$$

where

$$\sigma_i^{2,Y}\equiv var(\ln Y_i(s)-c_i(s))$$

is the variance of farmers' nominal income volatility,

$$\sigma_i^{2,P}\equiv var\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s)+c_i(s)\right)$$

is the variance of farmers' nominal price volatility,

$$cov_i^{Y,P}\equiv cov\left(\ln Y_i^f(s)-c_i(s),\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s)+c_i(s)\right)$$

is the co-variance between the two and $c_i(s)$ is nuisance a term capturing the aggregate scale of both nominal prices and incomes, which does not affect the aggregate real returns nor the volatility of the real returns. Any increase in openness increases the volatility of farmers' nominal income volatility (3a); decreases the volatility of farmers' nominal price volatility (3b); and has an ambiguous effect on farmers' real income volatility (3c). Formally, we have:

$$\frac{\partial\sigma_i^{2,Y}}{\partial\varepsilon_i}>0,\frac{\partial\sigma_i^{2,P}}{\partial\varepsilon_i}<0,\ \text{and}\ \frac{\partial\sigma_i^{2,Z}}{\partial\varepsilon_i}\lesseqgtr 0.$$

As sufficient condition for farmers' real income volatility to increase with openness, i.e. $\frac{\partial\sigma_i^{2,Z}}{\partial\varepsilon_i}\geq 0$, is $\sum_{g\in\mathcal{G}}\theta_{i,g}\left(\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ih}\right)\geq\sum_{g\in\mathcal{G}}\alpha_{ig}\left(\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ih}\right)$, which (loosely speaking) occurs when a farmers' crop allocation is more risky than her expenditure allocation.

Proof. **Stylized Fact 1.** From equation (11) we have:

$$\ln p_{ig}(s)=-\left(\frac{1}{1+\varepsilon_i}\right)\ln Q_{ig}(s)+\frac{\varepsilon_i}{1+\varepsilon_i}\ln\bar{p}_g(s)+\frac{1}{1+\varepsilon_i}\ln(\alpha_{ig}Y_i(s))\iff$$
$$\ln p_{ig}(s)=-\left(\frac{1}{1+\varepsilon_i}\right)\ln A_{ig}(s)-\left(\frac{1}{1+\varepsilon_i}\right)\ln\theta_{ig}-\left(\frac{1}{1+\varepsilon_i}\right)\ln L_i+\frac{\varepsilon_i}{1+\varepsilon_i}\ln\bar{p}_g+\frac{1}{1+\varepsilon_i}\ln\alpha_{ig}+\frac{1}{1+\varepsilon_i}\ln Y_i(s)$$

so that:

$$\frac{\partial\ln p_{ig}(s)}{\partial\ln A_{ig}(s)}=-\frac{1}{1+\varepsilon_i}$$

and hence:

$$\frac{d}{d\varepsilon_i}\left(-\frac{\partial\ln p_{ig}(s)}{\partial\ln A_{ig}(s)}\right)=\frac{d}{d\varepsilon_i}\left(\frac{1}{1+\varepsilon_i}\right)=-\frac{1}{(1+\varepsilon_i)^2}<0.$$

Similarly:

$$\frac{\partial \ln p_{ig}(s)}{\partial \ln \bar{p}_g} = \frac{\varepsilon_i}{1+\varepsilon_i}$$

and hence:

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial \ln p_{ig}(s)}{\partial \ln \bar{p}_g}\right) = \frac{d}{d\varepsilon_i}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right) = \frac{1}{(1+\varepsilon_i)^2} > 0,$$

as claimed.

Stylized Fact 2a. From equation (21) we have:

$$\theta_{ig} = \frac{\alpha_{ig}(\bar{p}_g B_{ig})^{\varepsilon_i}}{\sum_{h\in\mathcal{G}} \alpha_{ih}(\bar{p}_h B_{ih})^{\varepsilon_i}},$$

where $B_{ig} \equiv \frac{\exp \mu_g^{A,i}}{\left(\lambda_i - \left(\frac{1}{2}\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2 \Sigma_{gg}^{A,i} + \frac{\varepsilon_i}{(1+\varepsilon_i)^2}\sum_{h\in\mathcal{G}}\alpha_{ih}\Sigma_{gh}^{A,i} - \rho_i\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\sum_{h\in\mathcal{G}}\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,h} + \left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ih}\right)\Sigma_{gh}^{A,i}\right)\right)^{\frac{1+\varepsilon_i}{\varepsilon_i}}}$ so that:

$$\ln\theta_{ig} - \ln\theta_{ih} = \ln(\alpha_{ig}) - \ln(\alpha_{ih}) + \varepsilon_i(\ln\bar{p}_g - \ln\bar{p}_h) + \varepsilon_i(\ln B_{ig} - \ln B_{ih})$$

Differentiating this expression with respect to ε_i and evaluating at $\varepsilon_i = 0$ yields:

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial(\ln\theta_{ig} - \ln\theta_{ih})}{\partial\left(\mu_g^{A,i} - \mu_h^{A,i}\right)}\right)|_{\varepsilon_i=0} = 1 > 0,$$

as claimed.

Stylized Fact 2b. We proceed similarly. Differentiating respect to ε_i and evaluating at $\varepsilon_i = 0$ yields:

$$\frac{d}{d\varepsilon_i}(\partial\ln\theta_{ig} - \partial\ln\theta_{ih})|_{\varepsilon_i=0} = \frac{1}{\lambda_i}(1-\rho_i)\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\left(\partial\Sigma_{gh'}^{A,i} - \partial\Sigma_{hh'}^{A,i}\right)\right)$$

so that:

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial\ln\theta_{ig} - \partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i} - \sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0} = \frac{1}{\lambda_i}(1-\rho_i),$$

$$\frac{d}{d\varepsilon_i}\left(\frac{\partial\ln\theta_{ig} - \partial\ln\theta_{ih}}{\partial\Sigma_{gg}^{A,i}}\right)|_{\varepsilon_i=0} = \frac{1}{\lambda_i}(1-\rho_i)\alpha_{ig},$$

which is negative as long as $\rho_i > 1$, as claimed.

Stylized Fact 2c. From the previous expression, we immediately have:

$$\frac{d^2}{d\varepsilon_i d\rho}\left(\frac{\partial\ln\theta_{ig} - \partial\ln\theta_{ih}}{\partial\left(\sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{g,h'}^{A,i} - \sum_{h'\in\mathcal{G}}\alpha_{h'}\Sigma_{h,h'}^{A,i}\right)}\right)|_{\varepsilon_i=0} = -\frac{\alpha_{h'}}{\lambda_i}.$$

Stylized Fact 3. Let us first decompose the distribution of real returns into a price term, and income term, and a covariance term. We have:

$$\begin{aligned}Z_i^f(s) &= \prod_{g\in\mathcal{G}} (c_{ig}(s))^{\alpha_{ig}}\\ &= \prod_{g\in\mathcal{G}} \left(\frac{\alpha_{ig}Y_i^f(s)}{p_{ig}(s)}\right)^{\alpha_{ig}}\\ &= Y_i^f(s)\times\prod_{g\in\mathcal{G}}(\alpha_{ig})^{\alpha_{ig}}\times\prod_{g\in\mathcal{G}}(p_{ig}(s))^{-\alpha_{ig}}\end{aligned}$$

so that:

$$\ln Z_i^f(s) = \ln Y_i^f(s) + \sum_{g\in\mathcal{G}}\alpha_{ig}\ln\alpha_{ig} - \sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s).$$

Hence, we can decompose the variance of the real returns as follows:

$$\sigma_i^{2,Z} = \sigma_i^{2,Y} + \sigma_i^{2,P} + 2cov_i^{Y,P}, \tag{36}$$

where:

$$\sigma_i^{2,Y} \equiv var\left(\ln Y_i^f(s) - c_i(s)\right)$$

$$\sigma_i^{2,P} \equiv var\left(-\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s) + c_i(s)\right)$$

$$cov_i^{Y,P} \equiv cov\left(\ln Y_i^f(s) - c_i(s), -\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s) + c_i(s)\right)$$

and $c_i(s) \equiv \ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}}$ term captures the aggregate scale of both prices and incomes, which because it affects both terms with opposite signs, does not affect the aggregate returns nor the volatility of the real returns. Let us examine each term in turn.

Focusing first on the income term we have:

$$\begin{aligned}\ln Y_i^f(s) - \ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}} &= \ln\left(\sum_{g\in\mathcal{G}}\theta_{ig}^f A_{ig}(s)p_{ig}(s)\right) - \ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}} \iff\\ \ln Y_i^f(s) - \ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}} &= \ln\left(\sum_{g\in\mathcal{G}}\left(\frac{\theta_{ig}^f}{\theta_{ig}}\right)\times\alpha_{ig}\left(\frac{\bar{p}_g A_{ig}(s)\theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)\end{aligned}$$

Applying the same second order approximation as in the main text we have:

$$\begin{aligned}\ln Y_i^f(s) - \ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}} \approx{}& \ln\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\left(\frac{\bar{p}_g\exp\left(\mu_g^{A,i}\right)\theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right) - \sum_{g\in\mathcal{G}}\theta_{ig}\ln\left(\theta_{ig}^{-1}\alpha_{ig}\left(\frac{\bar{p}_g\exp\left(\mu_g^{A,i}\right)\theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)\\ &+\sum_{g\in\mathcal{G}}\theta_{ig}\ln\left(\theta_{ig}^{-1}\alpha_{ig}\left(\frac{\bar{p}_g A_{ig}(s)\theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right) - -\frac{1}{2}\sum_{h\in\mathcal{G}}\sum_{g\in\mathcal{G}}\theta_{ig}\theta_{ih}\Sigma_{gh}^{A,i} + \frac{1}{2}\sum_{g\in\mathcal{G}}\theta_{ig}\Sigma_{gh}^{A,i}\end{aligned}$$

so that:

$$\sigma_i^{2,Y} = \left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)^2\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\theta_{ig}\theta_{ih}\Sigma_{gh}^{A,i} \tag{37}$$

Now focusing on the price term we have:

$$-\sum_{g\in\mathcal{G}}\alpha_{ig}\ln p_{ig}(s)+\ln\left(\frac{Y_i(s)}{L_i}\right)^{\frac{1}{1+\varepsilon_i}}=\sum_{g\in\mathcal{G}}\alpha_{ig}\ln\left(\left(\frac{\bar{p}_g A_{ig}(s)\theta_{ig}}{\alpha_{ig}}\right)^{\frac{1}{1+\varepsilon_i}}(\bar{p}_g)^{-1}\right)\iff$$
$$=\left(\frac{1}{1+\varepsilon_i}\right)\sum_{g\in\mathcal{G}}\alpha_{ig}\ln A_{ig}(s)+\sum_{g\in\mathcal{G}}\alpha_{ig}\ln\left(\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}\right)^{\frac{1}{1+\varepsilon_i}}(\bar{p}_g)^{-1}\right)$$

so that the variance of the prices can be written as:

$$\sigma_i^{2,P}=\left(\frac{1}{1+\varepsilon_i}\right)^2\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ig}\alpha_{ih} \tag{38}$$

$$p_{ig}=(A_{ig}\theta_{ig}L_i)^{-\frac{1}{1+\varepsilon_i}}(\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}(\alpha_{ig}Y_i)^{\frac{1}{1+\varepsilon_i}}$$

Finally, the covariance between prices and incomes can be written as:

$$cov_i^{Y,P}=\frac{\varepsilon_i}{(1+\varepsilon_i)^2}\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\theta_{ig}\alpha_{ih}\Sigma_{gh}^{A,i}. \tag{39}$$

It is straightforward to verify that applying the decomposition (36) to expressions (37), (38), and (39) immediately yields expression (33) for the variance of the total real returns.

Now consider a small increase in the openness of a location. How does it affect the variance of farmers' incomes, prices, and the co-variance between the two? We immediately have:

$$\frac{\partial\sigma_i^{2,Y}}{\partial\varepsilon_i}=2\frac{\varepsilon_i^2}{(1+\varepsilon_i)^3}\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\theta_{ig}\theta_{ih}\Sigma_{gh}^{A,i}>0$$

$$\frac{\partial\sigma_i^{2,P}}{\partial\varepsilon_i}=-2\frac{1}{(1+\varepsilon_i)^3}\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ig}\alpha_{ih}<0,$$

as required.

Let us turn now to the variance of the total real returns. Recall from equation (33) that the variance of real returns is:

$$\sigma_i^{2,Z}\equiv\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,g}^f+\left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ig}\right)\left(\left(\frac{\varepsilon_i}{1+\varepsilon_i}\right)\theta_{i,h}^f+\left(\frac{1}{1+\varepsilon_i}\right)\alpha_{ih}\right)\Sigma_{gh}^{A,i}\iff$$
$$=\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\omega_i\theta_{i,g}^f+(1-\omega_i)\alpha_{ig}\right)\left(\omega_i\theta_{i,h}^f+(1-\omega_i)\alpha_{ih}\right)\Sigma_{gh}^{A,i},$$

where $\omega_i\equiv\frac{\varepsilon_i}{1+\varepsilon_i}$. Note that $\frac{\partial\omega_i}{\partial\varepsilon_i}=\frac{1}{1+\varepsilon_i}-\frac{\varepsilon_i}{1+\varepsilon_i}\frac{1}{1+\varepsilon_i}=\frac{1}{1+\varepsilon_i}\left(1-\frac{\varepsilon_i}{1+\varepsilon_i}\right)=\frac{1}{(1+\varepsilon_i)^2}$, so that $\frac{\partial\sigma_i^{2,Z}}{\partial\varepsilon_i}=\frac{1}{(1+\varepsilon_i)^2}\frac{\partial\sigma_i^{2,Z}}{\partial\omega_i}$. We then have:

$$\frac{\partial\sigma_i^{2,Z}}{\partial\varepsilon_i}=\frac{1}{(1+\varepsilon_i)^2}\frac{\partial}{\partial\omega_i}\left(\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\omega_i\theta_{i,g}^f+(1-\omega_i)\alpha_{ig}\right)\left(\omega_i\theta_{i,h}^f+(1-\omega_i)\alpha_{ih}\right)\Sigma_{gh}^{A,i}\right)\iff$$
$$=\frac{2}{(1+\varepsilon_i)^2}\left(\omega_i\left(\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\theta_{i,g}^f-\alpha_{ih}\right)\left(\theta_{i,h}^f-\alpha_{ih}\right)\Sigma_{gh}^{A,i}\right)+\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\theta_{i,g}^f-\alpha_{ig}\right)\alpha_{ih}\Sigma_{gh}^{A,i}\right)$$

Because $\Sigma_{gh}^{A,i}$ is positive definite, we know that $\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\theta_{i,g}^f-\alpha_{ih}\right)\left(\theta_{i,h}^f-\alpha_{ih}\right)\Sigma_{gh}^{A,i}\geq 0$ for any crop allocation $\left\{\theta_{i,g}^{f}\right\}$ and expenditure shares $\{\alpha_{ig}\}$. Hence, $\frac{\partial\sigma_i^{2,Z}}{\partial\varepsilon_i}\geq 0$ if:

$$\sum_{g\in\mathcal{G}}\sum_{h\in\mathcal{G}}\left(\theta_{i,g}^{f}-\alpha_{ig}\right)\alpha_{ih}\Sigma_{gh}^{A,i}\geq 0\iff$$
$$\sum_{g\in\mathcal{G}}\theta_{i,g}^{f}\left(\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ih}\right)\geq\sum_{g\in\mathcal{G}}\alpha_{ig}\left(\sum_{h\in\mathcal{G}}\Sigma_{gh}^{A,i}\alpha_{ih}\right),$$

as required. □

A.2.3 Proof of Proposition #3

We first restate the proposition:

Proposition. *1) In the presence of volatility, moving from autarky to costly trade improves farmer welfare, i.e. the gains from trade are positive; 2) moving from a world with no volatility to one with volatility amplifies farmers' gains from trade; but 3) increasing the volatility in an already volatile world may attenuate farmers' gains from trade*

Proof. **Part 1**. From equation (29), the real income of farmer f in village $i\in\mathcal{N}$ in state $s\in S$ with crop allocation $\left\{\theta_{ig}^{f}\right\}_{g\in\mathcal{G}}$ can be written as:

$$Z_i^f\left(s;\left\{\theta_{ig}^{f}\right\}_{g\in\mathcal{G}}\right)=\frac{\left(\sum_{g\in\mathcal{G}}\theta_{ig}^{f}\times\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)\prod_{g\in\mathcal{G}}\left(\alpha_{ig}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{1}{1+\varepsilon_i}}\right)^{\alpha_{ig}}}{\prod_{g\in\mathcal{G}}(\bar{p}_g)^{\alpha_{ig}}}. \quad (40)$$

Consider first the case of autarky, where $\varepsilon_i=0$. From equation (21), a farmers' optimal autarkic crop allocation is simply equal to her expenditure share, i.e. $\theta_{ig}^{f}=\alpha_{ig}$,so that from equation (40) her autarkic welfare is:

$$Z_i^{f,aut}(s)=\prod_{g\in\mathcal{G}}(\alpha_{ig}\times A_{ig}(s))^{\alpha_{ig}}.$$

Now consider the case of (costly) trade, where $\varepsilon_i>0$ but farmer f chooses her autarkic crop allocation. Then from equation (21), her real income is:

$$Z_i^f\left(s;\{\alpha_{ig}\}_{g\in\mathcal{G}}\right)=\frac{\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\times\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)\prod_{g\in\mathcal{G}}\left(\alpha_{ig}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{1}{1+\varepsilon_i}}\right)^{\alpha_{ig}}}{\prod_{g\in\mathcal{G}}(\bar{p}_g)^{\alpha_{ig}}}. \quad (41)$$

Note that from the generalized mean inequality we have:

$$\sum_{g\in\mathcal{G}}\alpha_{ig}\times\frac{\alpha_g}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_g}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\geq\prod_{g\in\mathcal{G}}\left(\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\alpha_{ig}},$$

with equality only in the case where $\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}=c_i$ for all $g\in\mathcal{G}$. Substituting this inequality into equation (41) immediately implies

$$Z_i^f\left(s;\{\alpha_{ig}\}_{g\in\mathcal{G}}\right)\geq Z_i^{f,aut}(s),$$

again with equality only in the case where $\frac{\alpha_{ig}}{\theta_{ig}}\left(\frac{\bar{p}_g\theta_{ig}}{\alpha_{ig}}A_{ig}(s)\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}=c_i$ for all $g\in\mathcal{G}$. Intuitively, as long as the equilibrium price vector is not exactly equal to the slope of the production possibility frontier, farmers can gain by selling goods for which they are relatively more productive and buying goods for which they are relatively less productive. As the productivity realizations are log-normally distributed across states of the world, this equality only occurs with measure zero. Hence, for almost all $s \in S$, we have $Z_i^f\left(s;\{\alpha_{ig}\}_{g\in\mathcal{G}}\right) > Z_i^{f,aut}(s)$, which in turn implies that the expected utility of a farmer choosing her autarkic allocation with costly trade is strictly greater than the expected utility of a farmer in autarky choosing her autarkic allocation, i.e. $\mathbb{E}\left[U_i^f\left(\{\alpha_{ig}\}_{g\in\mathcal{G}}\right)\right] > \mathbb{E}\left[U_i^{f,aut}\right]$. Finally, as farmers make their crop choice to maximize their expected utility, their actual expected welfare with costly trade is at least as great as their expected utility holding their crop choice at the autarkic allocation, so that $\mathbb{E}\left[\max_{\{\theta_{ig}\}_{g\in\mathcal{G}}} U_i^f\left(\{\theta_{ig}\}_{g\in\mathcal{G}}\right)\right] > \mathbb{E}\left[U_i^{f,aut}\right]$, i.e. the gains from trade are strictly positive, as claimed.

Part 2. From equation (16), in the absence of volatility, farmers' utility is invariant to ε_i, i.e. there are zero gains from trade. From Part 1, in the presence of volatility, there are strictly positive gains from trade. Taken together, this implies that the presence of volatility amplifies the gains from trade, as claimed.

Part 3. We prove the statement by example, illustrated in Appendix Table A.7. Consider a world where there are two types of villages (1 and 2) and two crops (A and B). Suppose both villages have equal expenditure shares on both crops in equal proportions and the means of both crops in both villages is identical. Suppose first that crop A in village 1 and crop B in village 2 are "risky" (i.e. have equally volatile yields), whereas crop B in village 1 and crop A in village 2 are "safe" (i.e. have zero yield volatility). In autarky, both village types grow equal amounts of both crops, but with trade, the two types of villages can specialize in the "safe" crops, achieving positive gains from trade (Case 1 in Appendix Table A.7). Suppose now that we increase the volatility of the safe crop in both village types so that it receives the same yield shock as the risky crop (i.e. the two crops have perfectly correlated yields within each village, although independent yield realizations across villages). As the relative yields between the two crops are always equal in both types of villages, there are no gains from trade (Case 2 in Appendix Table A.7), illustrating that increasing the volatility in an already volatile world can reduce the gains from trade, as required. □

A.3 Model isomorphisms and extensions

In this subsection, we present alternative interpretations for the model presented in the main paper,

A.3.1 Endogenous capacity constraints

In this subsection, we show how the framework presented in the paper is isomorphic to one in which better traders exchange greater amounts of goods, i.e. have greater capacity for arbitrage. To do so, we suppose that traders with lower trade costs (i.e. lower τ's) are able to offer greater capacity, with the following constant elasticity function:

$$Q(\tau) = c_i \tau^{-\lambda}$$

When $\lambda = 0$, capacity is fixed, but for $\lambda > 0$ we have the intuitive result that better traders (with lower τ) are able to engage in greater amounts of trade. The constant elasticity form – while analytically convenient – can be viewed as a first-order log-linear approximation to any function where better traders have greater capacity. The scalar c_i is determined to ensure that a single trader handles each unit of production (if traders are buying goods in the village to sell to the market) or consumption (if traders are buying goods in the market to sell to the village). We consider each case in turn.

Suppose first that $\bar{p} \geq p_i$ so that traders buy goods produced in the village and sell them in the market. In this case, it must be that each unit produced in the village is handled by a trader, i.e.:

$$Q_i = \int Q(\tau) dF(\tau).$$

Maintaining the assumption in the main text that the distribution of traders is Pareto distributed with shape parameter ε_i, we have:

$$Q_i = c_i \varepsilon_i \int_1^\infty \tau^{-\lambda - \varepsilon_i - 1} d\tau \iff$$
$$Q_i \left(\frac{\lambda + \varepsilon_i}{\lambda}\right) = c_i$$

It is straightforward to calculate the quantity of units the traders purchase in the village and sell to the market:

$$Q_{im} = \int_1^{\frac{\bar{p}}{p_i}} Q(\tau) dF(\tau) \iff$$

$$Q_{im} = \left(1 - \left(\frac{\bar{p}}{p_i}\right)^{-(\lambda+\varepsilon_i)}\right) Q_i$$

And the remainder of the production is sold to consumers locally so that:

$$C_i = \left(\frac{\bar{p}}{p_i}\right)^{-(\lambda+\varepsilon_i)} Q_i. \tag{42}$$

Suppose now that $\bar{p} < p_i$ so that traders buy goods in the market and sell them to farmers in the village. In this case, it must be that each unit consumed in the village is handled by a trader, i.e.:

$$C_i = \int Q(\tau) dF(\tau),$$

which yields through an identical derivation as above:

$$C_i \left(\frac{\lambda+\varepsilon_i}{\lambda}\right) = c_i.$$

It is then straightforward to calculate the quantity of units the traders purchase in the market and sell to the village:

$$Q_{mi} = \int_1^{\frac{p_i}{\bar{p}}} Q(\tau) dF(\tau) \iff$$

$$Q_{im} = \left(1 - \left(\frac{p_i}{\bar{p}}\right)^{-(\lambda+\varepsilon_i)}\right) C_i.$$

The remainder of the consumption in the village comes from local production, i.e.:

$$Q_i = \left(\frac{p_i}{\bar{p}}\right)^{-(\lambda+\varepsilon_i)} C_i. \tag{43}$$

Equations (42) and (43) are identical and isomorphic to equation (10) in the main text. This demonstrates that the shape parameter of the Pareto distribution ε_i (where traders are assumed to be infinitely capacity constrained) can be equivalently thought of as a combination of the exogenous heterogeneity of the trade costs across traders and an endogenous component related to the fact that better traders are able to engage in greater amounts of arbitrage.

A.3.2 A microfoundation for insurance

In the baseline model presented in Section 4, the farmer's utility function is given by equation (6):

$$U_i^f(s) \equiv \frac{1}{1-\rho_i} \left(\left(Z_i^f(s) \right)^{1-\rho_i} - 1 \right)$$

where ρ_i is the "effective" risk aversion parameter and we show in equation (18) that $\ln Z_i^f(s) \sim N\left(\mu_i^Z, \sigma_i^{2,Z}\right)$, which then implies that farmers' expected utility can be written as in equation (19):

$$\mathbb{E}\left[U_i^f\right] = \left(\frac{1}{1-\rho_i}\right) \left(\exp\left((1-\rho_i) \left(\mu_i^Z + \frac{1}{2}(1-\rho_i)\sigma_i^{2,Z} \right) \right) - 1 \right). \tag{44}$$

In what follows, we will show that there exists a micro-foundation for the "effective" risk aversion parameter ρ_i whereby farmers purchase insurance from perfectly competitive lenders ("banks"). In this micro-foundation, the "effective" risk aversion parameter ρ_i can then be written as a function of the (fundamental) risk aversion of farmers and a (technological) parameter governing the efficiency of the insurance market. As a result, we can interpret changes to the "effective" risk aversion parameter as technological changes in the access to banks, allowing us to perform normative counterfactual analysis.

Suppose that all farmers have identical and fundamental risk aversion parameters ρ_0 and have access to banks that offer insurance at perfectly competitive rates. To save on notation, in what follows, we will omit the location of the farmer and denote states of the world with subscripts, the probability of state of the world s with π_s. Suppose that the insurance allows pays out one unit of the consumption bundle in state of the world s for price p_s.[58] Hence, consumption in state of the world s will be the sum of the realized consumption in that state and the insurance payout less the cost of insurance: $C_s = Z_s + q_s - \sum_t p_t q_t$, where q_s is the quantity of insurance for state s purchased by the farmer. A farmer's expected utility function ex-post insurance is then:

$$\mathbb{E}\left[U^{f,ins}\right] = \sum_s \pi_s \frac{1}{1-\rho_0}\left((C_s)^{1-\rho_0} - 1\right).$$

Farmers purchase their insurance from a large number of "money-lenders" (or, equivalently, banks). Money-lenders have the same income realizations and preference-structure as farmers and face the same prices, but are distinct from farmers in that they are less risk averse. Let money-lenders' risk aversion parameter be denoted by $\lambda \leq \rho_0$, where we view λ as a technological parameter governing the quality/access farmers have to credit: the better farmers' access to credit, the lower the risk aversion of money-lenders.

Because lenders are also risk averse, farmers will not be able to perfectly insure themselves. Money lenders compete with each other to lend money, and hence the price of purchasing insurance in a particular state of the world is determined by the marginal cost of lending money. We first calculate the price of a unit of insurance in state of the world s. Since the price of insurance is determined in perfect competition, it must be the case that each money lender is just indifferent between offering insurance and not:

$$\sum_{t \neq s} \pi_t \frac{1}{1-\lambda}(Z_t + \varepsilon p_s)^{1-\lambda} + \pi_s \frac{1}{1-\lambda}(Z_t + \varepsilon p_s - \varepsilon)^{1-\lambda} = \sum_t \pi_t \frac{1}{1-\lambda} Z_t^{1-\lambda},$$

where the left hand side is the expected utility of a money-lender offering an small amount ε of insurance (which pays εp_s with certainty but costs ε in state of the world s) and the left hand side is expected utility of not offering the insurance. Taking the limit as ε approaches zero yields that the price ensures that the marginal utility benefit of receiving $p_s \varepsilon$ in all other states of the world is equal to the marginal utility cost of paying $\varepsilon(1-p_s)$ in state of the world s.

$$p_s \varepsilon \sum_{t \neq s} \pi_t Z_t^{-\lambda} = \varepsilon(1-p_s)\pi_s Z_s^{-\lambda} \iff$$

$$p_s = \frac{\pi_s Z_s^{-\lambda}}{\sum_t \pi_t Z_t^{-\lambda}}. \tag{45}$$

Equation (45) is intuitive: it says that the price of insuring states of the world with low aggregate income is high.

Now consider the farmer's choice of the optimal level of insurance. Farmers will choose the quantity of insurance to purchase in each period in order to maximize their expected utility:

$$\max_{\{q_s\}} \sum_s \pi_s \frac{1}{1-\rho_0}\left(\left(Z_s + q_s - \sum_t p_t q_t\right)^{1-\rho_0} - 1\right)$$

[58] For simplicity – and without loss of generality as the state of the world defines the price index – we measure both the insurance payout and the prices in real (i.e. price index adjusted) units.

which yields the following FOC with respect to q_s:

$$\pi_s\left(Z_s+q_s-\sum_t p_t q_t\right)^{-\rho_0}=p_s\sum_t \pi_t\left(Z_t+q_t-\sum_t p_t q_t\right)^{-\rho_0} \iff$$
$$\frac{\pi_s C_s^{-\rho_0}}{\sum_t \pi_t C_t^{-\rho_0}}=p_s. \tag{46}$$

Substituting the equilibrium price from equation (45) into equation (46) and noting that $\mathbb{E}[C^{-\rho_0}]=\sum_t \pi_t C_t^{-\rho_0}$ and $\mathbb{E}\left[I^{-\lambda}\right]=\sum_t \pi_t I_t^{-\lambda}$ yields:

$$\frac{C_s^{-\rho_0}}{\mathbb{E}[C^{-\rho_0}]}=\frac{Z_s^{-\lambda}}{\mathbb{E}[Z^{-\lambda}]}. \tag{47}$$

Because the first order conditions (46) are homogeneous of degree zero in consumption, they do not pin down the scale of ex-post real income, so to ensure that access to insurance only affects welfare through the second moment of returns, we assume that access to insurance does not affect the log mean real returns of farmers, i.e. $\mathbb{E}[\ln C_s]=\mu^Z$. As a result, we can write:

$$C_s=Z_s^{\frac{\lambda}{\rho_0}}\left(\exp\left(\mu^Z\right)\right)^{1-\frac{\lambda}{\rho_0}}, \tag{48}$$

i.e. access to insurance means that the ex-post realized real returns after insurance payouts are a Cobb-Douglas combination of the ex-ante realized returns prior to insurance payouts and the (log) mean real returns. This is intuitive: when money lenders have the same level of risk aversion as the farmers (i.e. $\lambda=\rho_0$), farmers' ex-post returns are equal to their ex-ante returns, i.e. there is no scope for insurance. Conversely, when money lenders are risk-neutral (i.e. $\lambda=0$), farmers' ex-post returns are simply equal to their mean real returns, i.e. they are perfectly insured. When money-lenders are still risk averse but less so than farmers, there is scope for imperfect insurance, where the scope depends on the degree of risk aversion of the money-lenders. Indeed, equation (48) can be viewed as a first-order log-linear approximation of any insurance technology that reduces the variance of ex-post realized returns around its mean.

Given that the ex-ante realized returns are log-normally distributed $\ln Z_s \sim N\left(\mu^Z,\sigma^{2,Z}\right)$, the ex-post realized returns are also log-normally distributed with:

$$\ln C_s \sim N\left(\mu^Z,\left(\frac{\lambda}{\rho_0}\right)^2\sigma^{2,Z}\right)$$

so that farmers' expected utility ex post insurance can be written as:

$$\mathbb{E}\left[U^{f,ins}\right]=\frac{1}{1-\rho_0}\left(\exp\left((1-\rho_0)\left(\mu^Z+\frac{1}{2}(1-\tilde{\rho})\sigma^{2,Z}\right)\right)-1\right), \tag{49}$$

where

$$\tilde{\rho}=1+(\rho_0-1)\left(\frac{\lambda}{\rho_0}\right)^2$$

is the effective level of risk aversion. As a result, we have now shown that the effective level of risk aversion can be written as a function of the innate risk aversion of farmers (ρ_0) and the technological parameter governing their access to insurance markets (as captured by λ), as claimed.

Finally, consider the evaluation of the welfare impact of some counterfactual that changes potentially both the access to insurance markets and the distribution of real returns from $\left\{\lambda_A,\mu_A^Z,\sigma_A^{2,Z}\right\}$ to $\left\{\lambda_B,\mu_B^Z,\sigma_B^{2,Z}\right\}$.

The change in expected utility is:

$$(1-\rho_0)\left(\mathbb{E}\left[U_B^{f,ins}\right]-\mathbb{E}\left[U_A^{f,ins}\right]\right)=\exp\left((1-\rho_0)\left(\mu_B^Z+\frac{1}{2}(1-\rho_0)\left(\frac{\lambda_B}{\rho_0}\right)^2\sigma_B^{2,Z}\right)\right)$$
$$-\exp\left((1-\rho_0)\left(\mu_A^Z+\frac{1}{2}(1-\rho_0)\sigma_A^{2,Z}\left(\frac{\lambda_A}{\rho_0}\right)^2\right)\right).$$

We now define what we call the certainty equivalent variation (CEV), which is the hypothetical percentage increase in income an individual would need to receive with certainty that would yield an equivalent change in expected welfare as the counterfactual, holding constant all prices and parameters constant at the baseline. It is straightforward to show that the CEV can be written as:

$$CEV=\left(\mu_B^Z+\frac{1}{2}(1-\rho_0)\left(\frac{\lambda_B}{\rho_0}\right)^2\sigma_B^{2,Z}\right)-\left(\mu_A^Z+\frac{1}{2}(1-\rho_0)\left(\frac{\lambda_A}{\rho_0}\right)^2\sigma_A^{2,Z}\right), \tag{50}$$

or, equivalently, we can write the CEV in terms of the effective risk aversion:

$$CEV=\left(\mu_B^Z+\frac{1}{2}(1-\tilde{\rho}_B)\sigma_B^{2,Z}\right)-\left(\mu_A^Z+\frac{1}{2}(1-\tilde{\rho}_A)\sigma_A^{2,Z}\right),$$

where $\tilde{\rho}_A \equiv 1+(\rho_0-1)\left(\frac{\lambda_A}{\rho_0}\right)^2$ and $\tilde{\rho}_B \equiv 1+(\rho_0-1)\left(\frac{\lambda_B}{\rho_0}\right)^2$ are the effective risk aversion parameters we estimate in Section 5.3. This is the welfare metric we report in Section 5.

A.3.3 Convex transportation costs

In equation (10), we show that under the appropriate set of assumptions, heterogeneous traders and a market clearing condition imply the following no-arbitrage condition:

$$\frac{C_{ig}(s)}{Q_{ig}(s)}=\left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i}$$

i.e. goods flow toward locations with higher relative prices. In this subsection, we provide an alternative setup that generates the same no-arbitrage condition assuming that transportation costs are increasing and convex in the quantity traded.[59] For notational simplicity, we omit the good g and state s notation in what follows.

As in the model in the paper, suppose there is a (small) village i engaging in trade with a (large) market subject to trade costs. Unlike the model in the paper where the trade costs are heterogeneous across traders, suppose now that they increase convexly with the quantity shipped between the village and the market. In particular, let $\bar{M}_i$ denote the quantity of goods imported by village i from the market and $\bar{X}_i$ denote the quantity of goods exported by village i to the market. Suppose that the iceberg trade cost τ_i between the village i and its market can be written as:

$$\ln\tau_i=\frac{1}{\varepsilon_i}\ln\left(1+\frac{\bar{M}_i}{Q_i}+\frac{\bar{X}_i}{C_i}\right), \tag{51}$$

where Q_i and C_i are the quantity produced and consumed in village i, respectively. Intuitively, equation (51) says that the greater the flows of goods between the village and the market – relative to the quantity produced in i for imports and relative to the quantity consumed in i for exports – the greater the iceberg trade costs incurred.

Now consider what equation (51) implies when combined with a no-arbitrage condition. Suppose first that the market price exceeds the village price, i.e. $\bar{p}\geq p_i$. In this case, the village will only export the good

[59] We are grateful to Rodrigo Adao for pointing out this alternative setup.

to the market, i.e. $\bar{M}_i = 0$ and $\bar{X}_i \geq 0$ and the following no-arbitrage condition will hold:

$$
\begin{aligned}
\ln\bar{p} - \ln p_i &= \ln\tau_i \iff \\
\ln\bar{p} - \ln p_i &= \frac{1}{\varepsilon_i}\ln\left(1 + \frac{\bar{X}_i}{C_i}\right) \iff \\
1 + \frac{\bar{X}_i}{C_i} &= \left(\frac{\bar{p}}{p_i}\right)^{\varepsilon_i}
\end{aligned}
\tag{52}
$$

Now consider the case where the village price exceeds the market price, i.e. $p_i \geq \bar{p}$. In this case, the village will only import the good from the market, i.e. $\bar{M}_i \geq 0$ and $\bar{X}_i = 0$ and the following no-arbitrage condition will hold:

$$
\begin{aligned}
\ln p_i - \ln\bar{p} &= \ln\tau_i \iff \\
\ln p_i - \ln\bar{p} &= \frac{1}{\varepsilon_i}\ln\left(1 + \frac{\bar{M}_i}{Q_i}\right) \iff \\
1 + \frac{\bar{M}_i}{Q_i} &= \left(\frac{p_i}{\bar{p}}\right)^{\varepsilon_i}
\end{aligned}
\tag{53}
$$

Finally, we impose market clearing in village i, which requires that the total quantity consumed in village i is equal to the total quantity it produces less the net quantity it exports to the market:

$$C_i = Q_i + \bar{M}_i - \bar{X}_i.$$

Combined with either equation (52) or (53), the market clearing condition immediately yields the same equation:

$$\frac{C_i}{Q_i} = \left(\frac{p_i}{\bar{p}}\right)^{\varepsilon_i},$$

which is identical to equation (10) in the main text, as claimed.

A.3.4 Farmer cooperative

In the baseline model, we assume that each farmer makes her crop choice taking the prices as given. Here we explore what would occur if a farmer takes into account the effect of her crop choice on prices, e.g. if all the farmers worked together to form a cooperative. In this case, the cooperative will maximize:

$$\max_{\theta_g}(Y_i(\{\theta_{ig}\}))\prod_g\left(\frac{\alpha_{ig}}{p_{ig}(\{\theta_{ig}\})}\right)^{\alpha_{ig}}$$

subject to:

$$\sum_g \theta_{ig} = 1.$$

Recall:

$$p_{ig} = (A_{ig}\theta_{ig}L_i)^{-\frac{1}{1+\varepsilon_i}}(\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}(\alpha_{ig}Y_i)^{\frac{1}{1+\varepsilon_i}}$$

$$Y_i(s) = \left(\sum_{g\in\mathcal{G}}\alpha_{ig}\left(\frac{\bar{p}_g(s)Q_{ig}(s)}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}\right)^{\frac{1+\varepsilon_i}{\varepsilon_i}}$$

so that we have

$$Z_i = (Y_i)\prod_g \left(\frac{\alpha_{ig}}{(A_{ig}\theta_{ig}L_i)^{-\frac{1}{1+\varepsilon_i}} (\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}} (\alpha_{ig}Y_i)^{\frac{1}{1+\varepsilon_i}}} \right)^{\alpha_{ig}} \iff$$

$$Z_i = Y_i^{\frac{\varepsilon_i}{1+\varepsilon_i}} \prod_g \left(\frac{\alpha_{ig}(A_{ig}\theta_{ig}L_i)^{\frac{1}{1+\varepsilon_i}}}{(\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right)^{\alpha_{ig}} \iff$$

$$Z_i = \left(\sum_{g\in\mathcal{G}} \alpha_{ig} \left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}} \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) \prod_g \left(\frac{\alpha_{ig}(A_{ig}\theta_{ig}L_i)^{\frac{1}{1+\varepsilon_i}}}{(\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right)^{\alpha_{ig}} \iff$$

$$Z_i = \left(\sum_{g\in\mathcal{G}} \alpha_{ig} \left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}} \right)^{\frac{\varepsilon_i}{1+\varepsilon_i}} \right) \prod_g \left(\frac{\alpha_{ig}(A_{ig}\theta_{ig}L_i)^{\frac{1}{1+\varepsilon_i}}}{(\bar{p}_g)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right)^{\alpha_{ig}}$$

Relative to the case where prices are taken as given, the first order conditions of the farmer cooperative are a little more involved. We have:

$$\frac{\partial Z_i}{\partial \theta_{ig}} = r_i \iff$$

$$\theta_{ig} \propto \varepsilon_i \left(\frac{\alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}} \alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right) + \alpha_{ig} \implies$$

$$\theta_{ig} = \frac{\varepsilon_i \left(\frac{\alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}} \alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right) + \alpha_{ig}}{\sum_g \left(\varepsilon_i \left(\frac{\alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}}{\sum_{g\in\mathcal{G}} \alpha_{ig}\left(\frac{\bar{p}_g A_{ig} L_i \theta_{ig}}{\alpha_{ig}}\right)^{\frac{\varepsilon_i}{1+\varepsilon_i}}} \right) + \alpha_{ig} \right)}.$$

Recall that when farmers take prices as given, their equilibrium crop choice is given my equation (15):

$$\theta_{ig} = \frac{(A_{ig}\bar{p}_g)^{\varepsilon_i}\alpha_{ig}}{\sum_{h\in\mathcal{G}}(A_{ih}\bar{p}_h)^{\varepsilon_i}\alpha_{ih}},$$

so this demonstrates that the farmer cooperative chooses a different optimal crop allocation. In particular, the elasticity of the relative crop choice to the central market price $\bar{p}_g$ is smaller for the cooperative (where it is bounded above by $\frac{\varepsilon_i}{1+\varepsilon_i}$) than for the price taking farmers (where it is equal to ε_i). Intuitively, the cooperative purposefully restricts the quantity produced of its high value (high $\bar{p}_g$) crops to ensure greater local prices.

A.4 The quantitative model

In this section, we provide a complete description of the quantitative model used in Section 5 to quantify the welfare impacts of the expansion of the Indian highway network. Recall there are four innovations in the quantitative model relative to the baseline model presented in Section 4: 1) more general (CES) preferences; 2) the presence of an tradable manufacturing good not produced in villages; 3) a finite number of villages with arbitrarily correlated yield realizations; and 4) multiple markets arranged hierarchically in which arbitrage occurs.

As in the baseline model, suppose there are N villages, indexed by $i \in \{1,...,N\} \equiv \mathcal{N}$ and G agricultural goods. Suppose now though that there is a single costlessly traded (numeraire) good 0 and M regional markets indexed by $m \in \{1,...,M\} \equiv \mathcal{M}$, and a single central market. Let $m(i)$ denote the regional market with which village $i \in \mathcal{N}$ engages in trade. Each market $m \in \mathcal{M}$ is inhabited by traders, truckers, and L_m producers of the numeraire good. We assume each producer can produce one unit of the numeraire good in any state of the world, so that $Q_{0m} = L_m$.

A.4.1 Equilibrium village prices and incomes

We begin by noting that the arbitrage process between villages and their regional markets is unchanged by the various model extensions. So while the equilibrium regional market prices will be affected by the addition of the manufacturing good, the more general preferences, and the presence of other regional markets, conditional on the equilibrium regional market prices, the arbitrage process between village $i \in \mathcal{N}$ and its regional market $m(i) \in \mathcal{M}$ for any agricultural good $g \in \{1,...,G\}$ continues to satisfy equation (10):

$$C_{ig}(s) = \left(\frac{p_{ig}(s)}{\bar{p}_{m(i)g}(s)}\right)^{\varepsilon_i} Q_{ig}(s),$$

for any quantity $Q_{ig}(s)$ produced. We can then combine this arbitrage with the demand equation $p_{ig}(s)C_{ig}(s) = \beta_i \frac{\alpha_{ig}(p_{ig}(s))^{1-\sigma}}{\sum_{h\in\mathcal{G}}\alpha_{ih}(p_{ih}(s))^{1-\sigma}} Y_i(s)$ and supply equation $Q_{ig}(s) = L_i \theta_{ig} A_{ig}(s)$ and take logs to yield the following expression for equilibrium prices:

$$\ln p_{ig}(s) = -\frac{1}{\sigma+\varepsilon_i}\ln A_{ig}(s) + \frac{\varepsilon_i}{\sigma+\varepsilon_i}\ln \bar{p}_{m(i)g}(s) + \frac{1}{\sigma+\varepsilon_i}\ln\left(\frac{\beta_i \alpha_{ig} Y_i(s)}{L_i \theta_{ig}\sum_{h=1}^{G}\alpha_{ih}(p_{ih}(s))^{1-\sigma}}\right), \tag{54}$$

which generalizes equation (11) to incorporate CES preferences over agricultural goods and the presence of a manufacturing good. Substituting equation (54) into the income expression $Y_i(s) = \sum_{g=1}^{G} p_{ig}(s) A_{ig}(s)\theta_{ig} L_i$ and solving simultaneously with the price index component $\sum_{h=1}^{G}\alpha_{ih}(p_{ih}(s))^{1-\sigma}$ allows us to express all endogenous variables in the village as functions of the realized yields, crop choice, and the market prices as follows:

$$\begin{aligned}\left(\frac{Y_i(s)}{L_i}\right) = &\left(\sum_{g=1}^{G}(\theta_{ig}A_{ig}(s))^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}} A_i(s)\theta_{ig}\right)^{\frac{\varepsilon_i+1}{\varepsilon_i}}\\ &\times\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\left((\theta_{ig}A_{ig}(s))^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\right)^{1-\sigma}\right)^{-\frac{1}{\varepsilon_i}}\end{aligned} \tag{55}$$

$$\begin{aligned}\sum_{g\in\mathcal{G}}\alpha_{ig}(p_{ig}(s))^{1-\sigma} = &\left(\sum_{g=1}^{G}(\theta_{ig}A_{ig}(s))^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}} A_i(s)\theta_{ig}\right)^{-\frac{\sigma-1}{\varepsilon_i}}\\ &\times\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\left((\theta_{ig}A_{ig}(s))^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\right)^{1-\sigma}\right)^{\frac{\varepsilon_i+\sigma-1}{\varepsilon_i}},\end{aligned} \tag{56}$$

which together generalize equation (28) in the baseline model.

A.4.2 Equilibrium market prices

Equations (54), (55), and (56) together characterize the equilibrium village prices in a given state of the world, taking as given the prices in the regional markets. We proceed by calculating the equilibrium prices in the markets.

Consider a state of the world $s \in S$ whose realized yields and crop allocations result in a quantity $Q_{ig}(s)$ being produced of good $g \in \{1,...,G\}$ in village $i \in \mathcal{N}$. Let the price of a good in village be denoted by $p_{ig}(s)$, let the price of the good in market $m \in \mathcal{M}$ be denoted by $\bar{p}_{mg}(s)$, and let the price of the good in the central market be denoted by $p_g^*(s)$.

Suppose an arbitrage process – analogous in structure to the arbitrage process between villages and regional markets – occurs between each regional market and the central market, resulting in the following arbitrage relationship:

$$\bar{C}_{mg}(s) = \left(\frac{\bar{p}_{mg}(s)}{p_g^*(s)}\right)^{\varepsilon_m} \bar{Q}_{mg}(s), \tag{57}$$

where $\bar{C}_{mg}(s)$ is the quantity agents in regional market m consume of good $g \in \{1,...,G\}$ in state s and $\bar{Q}_{mg}(s)$ is the quantity of good g that arrives in market m from its constituent villages through the arbitrage process, which from equation (12) can be related to the quantities produced in these villages as:

$$\bar{Q}_{mg}(s) = \sum_{i \in \mathcal{N}_m} \left(1 - \left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i}\right) Q_{ig}(s), \tag{58}$$

where $\mathcal{N}_m \equiv \{i \in \mathcal{N} | m(i) = m\}$ is the set of villages that trade with regional market m.

The quantity agents demand in the regional market of good $g \in \{1,...,G\}$ for consumption is:

$$\bar{p}_g(s)\bar{C}_{mg}(s) = \beta_m \frac{\alpha_{mg}(\bar{p}_g(s))^{1-\sigma}}{\sum_{h=1}^G \alpha_{mh}(p_{ih}(s))^{1-\sigma}} \bar{Y}_m(s), \tag{59}$$

where $\bar{Y}_m(s)$ is the income earned by agents in the regional market both through the production of the numeraire good and through arbitrage:

$$\bar{Y}_m(s) = L_m + \sum_{g=1}^G \sum_{i \in \mathcal{N}_m} (\bar{p}_g(s) - p_{ig}(s)) \left(1 - \left(\frac{p_{ig}(s)}{\bar{p}_g(s)}\right)^{\varepsilon_i}\right) Q_{ig}(s), \tag{60}$$

where the arbitrage profits are as in equation (13). Combining the equations (57) and (59) yields:

$$\bar{p}_{mg}(s) = \left(p_g^*(s)\right)^{\frac{\varepsilon_m}{1+\varepsilon_m}} \left(\beta_m \frac{\alpha_{mg}(\bar{p}_{mg}(s))^{1-\sigma}}{\sum_{h=1}^G \alpha_{mh}(\bar{p}_{mh}(s))^{1-\sigma}} \frac{\bar{Y}_m(s)}{\bar{Q}_{mg}(s)}\right)^{\frac{1}{1+\varepsilon_m}}, \tag{61}$$

which – when combined with equations (58), (60) (for the quantity traded and income of the market) and (54), (55), and (56) (for the equilibrium prices in each village as a function of the market prices) – allow us to express the regional market prices only as a function of the central market price.

Note that equation (61) can be re-written as follows:

$$\ln \bar{p}_{mg}(s) = -\frac{1}{\sigma + \varepsilon_m} \ln \bar{Q}_{mg}(s) + \frac{\varepsilon_m}{\sigma + \varepsilon_m} \ln p_g^*(s) + \delta_{mg} + \delta_m(s), \tag{62}$$

where $\delta_{mg} \equiv \frac{1}{\sigma+\varepsilon_m} \ln \alpha_{mg}$ is a district-crop fixed effect and $\delta_m(s) \equiv \frac{1}{\sigma+\varepsilon_m}\left(\ln \beta_m \frac{\bar{Y}_m(s)}{\sum_{h=1}^G \alpha_{mh}(\bar{p}_{mh}(s))^{1-\sigma}}\right)$ is a market-state fixed effect. Hence, just as with the villages, we can identify the degree of openness of a regional market by regressing its equilibrium price on the quantity flowing into that market, after conditioning on the appropriate set of fixed effects and with the appropriate moment conditions.

The central market price, in turn, has the analogous quantity consumed as in equation (12):

$$C_g^*(s) = \sum_{m \in \mathcal{M}} \left(1 - \left(\frac{\bar{p}_{mg}(s)}{p_g^*(s)}\right)^{\varepsilon_m}\right) \bar{Q}_{mg}(s)$$

and total income:

$$Y^*(s) = L^* + \sum_{g=1}^G \sum_{m \in \mathcal{M}} \left(p_g^*(s) - \bar{p}_{mg}(s)\right) \left(1 - \left(\frac{\bar{p}_{mg}(s)}{p_g^*(s)}\right)^{\varepsilon_m}\right) \bar{Q}_{mg}(s)$$

Finally, given that traders residing in the central market also have the same CES demand, the following expression determines the equilibrium central market price:

$$p_g^*(s) = \left(\beta^* \frac{\alpha_g^*}{\sum_{h=1}^G \alpha_{ih} (p_h^*(s))^{1-\sigma}} \frac{Y^*(s)}{C_g^*(s)} \right)^{\frac{1}{\sigma}}, \tag{63}$$

which can be solved for simultaneously with equations (61) and (54) to determine the equilibrium prices in all markets and villages.

A.4.3 Distribution of real returns

As in the baseline model, we can also calculate the real income of farmers in the village:

$$Z_i^f(s) = \left((1-\beta_i)^{1-\beta_i} \beta_i^{\beta_i} \right) \frac{\sum_{g=1}^G \theta_{ig}^f A_{ig}(s) p_{ig}(s)}{\left(\sum_{g=1}^G \alpha_{ig} (p_{ig}(s))^{1-\sigma} \right)^{\frac{\beta_i}{1-\sigma}}}. \tag{64}$$

Substituting equations (54), (55), and (56) into (64) allows us to write the real returns of a farmer f in village $i \in S$ as a function of her crop allocation $\left\{ \theta_{ig}^f \right\}_{g \in \{1,..,G\}}$, realized yields, and the regional market price:

$$\begin{aligned} Z_i^f \left(s; \left\{ \theta_{ig}^f \right\}_{g \in \{1,...,G\}} \right) = & \left((1-\beta_i)^{1-\beta_i} \beta_i^{\beta_i} \right) \times \sum_{g=1}^G \theta_{ig}^f (A_{ig}(s))^{\frac{\varepsilon_i + \sigma - 1}{\sigma + \varepsilon_i}} (\theta_{ig})^{-\frac{1}{\sigma + \varepsilon_i}} \left(\bar{p}_{m(i)g}(s) \right)^{\frac{\varepsilon_i}{\sigma + \varepsilon_i}} (\beta_i \alpha_{ig})^{\frac{1}{\sigma + \varepsilon_i}} \\ & \times \left(\sum_{g=1}^G \theta_{ig} (A_{ig}(s))^{\frac{\varepsilon_i + \sigma - 1}{\sigma + \varepsilon_i}} (\theta_{ig})^{-\frac{1}{\sigma + \varepsilon_i}} \left(\bar{p}_{m(i)g}(s) \right)^{\frac{\varepsilon_i}{\sigma + \varepsilon_i}} (\beta_i \alpha_{ig})^{\frac{1}{\sigma + \varepsilon_i}} \right)^{\left(\frac{\varepsilon_i + \sigma}{\varepsilon_i} \right) \left(\frac{1-\beta_i}{\sigma + \varepsilon_i} \right)} \\ & \times \left(\sum_{g \in \mathcal{G}} \alpha_{ig} \left((\theta_{ig} A_{ig}(s))^{-\frac{1}{\sigma + \varepsilon_i}} \left(\bar{p}_{m(i)g}(s) \right)^{\frac{\varepsilon_i}{\sigma + \varepsilon_i}} (\beta_i \alpha_{ig})^{\frac{1}{\sigma + \varepsilon_i}} \right)^{1-\sigma} \right)^{\left(-\frac{\sigma + \varepsilon_i}{\varepsilon_i} \right) \left(\frac{1-\beta_i}{\sigma + \varepsilon_i} \right) + \frac{\beta_i}{\sigma - 1}}. \end{aligned} \tag{65}$$

The three summation terms capture the effect of the farmer's crop choice on her nominal income, the effect of the total production in the village on the relative price of manufacturing and agricultural goods, and the effect of the realized yields on the farmer's price index. Note that this expression collapses to equation (30) in the basic model when $\beta_i = 1$ and $\sigma \to 1$.

Given Assumption 2, an equivalent second-order approximation from above implying that the sum of log normal variables is itself approximately log normal, and a log-linearization of the equilibrium market prices around their means yields, we can then approximate the distribution of real returns across states of the world as itself log normal:

To characterize the distribution of real returns across states of the world, we begin by approximating the equilibrium market prices. Let $\ln \bar{\mathbf{p}}(s) \equiv [\ln \bar{p}_{mg}(s)]_{m \in \{1,..,M\}}^{g \in \{1,...,G\}}$ denote the $(M \times G) \times 1$ vector of equilibrium prices in the M markets with which villages trade directly in state of the world $s \in S$. A log-linear approximation of the equilibrium prices around the mean (log) yields in all locations yields the following result:

$$\ln \bar{\mathbf{p}}(s) \approx \ln \bar{\boldsymbol{p}} + \mathbf{B} \left(\ln \mathbf{A}(s) - \boldsymbol{\mu}^A \right), \tag{66}$$

where $\ln \bar{\mathbf{p}} \equiv [\ln \bar{p}_{mg}]_{m \in \{1,..,M\}}^{g \in \{1,...,G\}}$ are the equilibrium market prices when all villages realize the mean (log) productivity of all crops and $\mathbf{B} \equiv \left[\frac{\partial \ln \bar{p}_{mg}}{\partial \ln A_{ig}} |_{\ln \mathbf{A}(s) = \boldsymbol{\mu}^A} \right]$ is an $(M \times G) \times (N \times G)$ matrix of elasticities. Exponentiating allows us to write:

$$\bar{p}_{m(i)g}(s) \approx \bar{p}_{m(i)g} \prod_{g'=1}^{G} \prod_{j=1}^{N} \left(\frac{A_{g'j}(s)}{\mu_{g'j}^A} \right)^{B_{m(i)g,jg'}}$$

where $B_{mg,jg'} \equiv \frac{\partial \ln \bar{p}_{mg}}{\partial \ln A_{jg'}}|_{\ln \mathbf{A}(s) = \boldsymbol{\mu}^A}$.

We proceed by applying the an equivalent second-order approximation from above implying that the sum of log normal variables is itself approximately log normal. Taking logs of equation (65) yields:

$$\begin{aligned} \ln Z_i^f\left(s;\left\{\theta_{ig}^f\right\}_{g\in\{1,\dots,G\}}\right) =& \ln\left((1-\beta_i)^{1-\beta_i}\beta_i^{\beta_i}\right) + \ln\sum_{g=1}^{G}\theta_{ig}^f (A_{ig}(s))^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}} (\theta_{ig})^{-\frac{1}{\sigma+\varepsilon_i}} \left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}} (\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}} \\ &+ \left(\frac{1-\beta_i}{\varepsilon_i}\right) \ln\left(\sum_{g=1}^{G}\theta_{ig} (A_{ig}(s))^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}} (\theta_{ig})^{-\frac{1}{\sigma+\varepsilon_i}} \left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}} (\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\right) \\ &+ \left(-\frac{1-\beta_i}{\varepsilon_i} + \frac{\beta_i}{\sigma-1}\right) \ln\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\left((\theta_{ig}A_{ig}(s))^{-\frac{1}{\sigma+\varepsilon_i}} \left(\bar{p}_{m(i)g}(s)\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}} (\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\right)^{1-\sigma}\right) \end{aligned}$$

so that applying equation (66) combined with the second order approximation results in:

$$\ln Z_i^f\left(s;\left\{\theta_{ig}^f\right\}_{g\in\{1,\dots,G\}}\right) \approx \mu_i^Z + \sum_{g=1}^{G}\omega_{ig}^A\left(\ln A_{ig}(s) - \mu_{ig}^A\right) + \sum_{g=1}^{G}\omega_{ig}^B \sum_{g'=1}^{G}\sum_{j=1}^{N} B_{m(i)g,jg'}\left(\ln A_{g'j}(s) - \mu_{g'j}^A\right) \tag{67}$$

where:

$$\Sigma^Y \equiv \left(\left(\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{I} + \left(\frac{\varepsilon_i}{\sigma+\varepsilon_i}\right)\mathbf{B}\right)\Sigma^A\left(\left(\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{I} + \left(\frac{\varepsilon_i}{\sigma+\varepsilon_i}\right)\mathbf{B}\right)^T \tag{68}$$

$$\Sigma^P \equiv \left(\left(\frac{\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{I} + \frac{(1-\sigma)\varepsilon_i}{\sigma+\varepsilon_i}\mathbf{B}\right)\Sigma^A\left(\left(\frac{\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{I} + \frac{(1-\sigma)\varepsilon_i}{\sigma+\varepsilon_i}\mathbf{B}\right)^T \tag{69}$$

are the variance-covariance matrices of the income terms and price index terms, respectively, and where $\mathbf{I}$ is $(G\times N)\times(G\times N)$ identity matrix and $\mathbf{B} \equiv \left[B_{m(i)g,jh}\right]$ is an $(G\times N)\times(G\times N)$ matrix,

$$\omega_{ig}^A \equiv \left(\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}\right)\left(\theta_{ig}^f + \left(\frac{1-\beta_i}{\varepsilon_i}\right)\theta_{ig}\right) + \left(\frac{\beta_i}{\sigma+\varepsilon_i} - \left(\frac{1-\beta_i}{\varepsilon_i}\right)\left(\frac{\sigma-1}{\sigma+\varepsilon_i}\right)\right)\alpha_{ig}$$

$$\omega_{ig}^B \equiv \left(\frac{\varepsilon_i}{\sigma+\varepsilon_i}\right)\left(\theta_{ig}^f + \left(\frac{1-\beta_i}{\varepsilon_i}\right)\theta_{ig}\right) + \left(\frac{(1-\beta_i)(\sigma-1)}{\varepsilon_i} - \beta_i\right)\alpha_{ig}$$

are the weights placed on each of the local productivity shocks and productivity shocks throughout the world in location $i's$ real returns, respectively, and

$$
\begin{aligned}
\mu_i^Z \equiv & \ln\Big((1-\beta_i)^{1-\beta_i}\beta_i^{\beta_i}\Big) + \ln\sum_{g=1}^{G}\theta_{ig}^f\big(\exp\big(\mu_{ig}^A\big)\big)^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}}(\theta_{ig})^{-\frac{1}{\sigma+\varepsilon_i}}\big(\bar{p}_{m(i)g}\big)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}} \\
& -\frac{1}{2}\sum_{h=1}^{G}\sum_{g=1}^{G}\theta_{ig}^f\theta_{ih}^f\Sigma_{ig,ih}^Y + \frac{1}{2}\sum_{g=1}^{G}\theta_{ig}^f\Sigma_{ig,ig}^Y \\
& + \left(\frac{1-\beta_i}{\varepsilon_i}\right)\ln\left(\sum_{g=1}^{G}\theta_{ig}\big(\exp\big(\mu_{ig}^A\big)\big)^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}}(\theta_{ig})^{-\frac{1}{\sigma+\varepsilon_i}}\big(\bar{p}_{m(i)g}\big)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\right) \\
& -\frac{1}{2}\left(\frac{1-\beta_i}{\varepsilon_i}\right)^2\sum_{h=1}^{G}\sum_{g=1}^{G}\theta_{ig}\theta_{ih}\Sigma_{ig,ih}^Y + \frac{1}{2}\left(\frac{1-\beta_i}{\varepsilon_i}\right)^2\sum_{g=1}^{G}\theta_{ig}\Sigma_{ig,ig}^Y \\
& + \left(-\frac{1-\beta_i}{\varepsilon_i} + \frac{\beta_i}{\sigma-1}\right)\ln\left(\sum_{g\in\mathcal{G}}\alpha_{ig}\Big(\big(\theta_{ig}\exp\big(\mu_{ig}^A\big)\big)^{-\frac{1}{\sigma+\varepsilon_i}}\big(\bar{p}_{m(i)g}\big)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}(\beta_i\alpha_{ig})^{\frac{1}{\sigma+\varepsilon_i}}\Big)^{1-\sigma}\right) \\
& -\frac{1}{2}\left(-\frac{1-\beta_i}{\varepsilon_i} + \frac{\beta_i}{\sigma-1}\right)^2\sum_{h=1}^{G}\sum_{g=1}^{G}\alpha_{ig}\alpha_{ih}\Sigma_{ig,ih}^P + \frac{1}{2}\left(-\frac{1-\beta_i}{\varepsilon_i} + \frac{\beta_i}{\sigma-1}\right)^2\sum_{g=1}^{G}\alpha_{ig}\Sigma_{ig,ig}^P
\end{aligned} \tag{70}
$$

are the mean log returns. Equation (67) extends equation (17) to account for the fact that the market prices now are state dependent, depending on the realized yields of all crops in all locations.

From equation (67), we can characterize the approximate distribution of real returns in location i as follows:

$$
\ln Z_i^f\left(\left\{\theta_{ig}^f\right\}_{g\in\{1,\ldots,G\}}\right) \sim N\left(\mu_i^Z, \sigma_i^{2,Z}\right), \tag{71}
$$

where:

$$
\sigma_i^{2,Z} \equiv \left(\boldsymbol{\omega}_i^A + \boldsymbol{\omega}_i^B\mathbf{B}\right)\boldsymbol{\Sigma}^A\left(\boldsymbol{\omega}_i^A + \boldsymbol{\omega}_i^B\mathbf{B}\right)^T \tag{72}
$$

and $\boldsymbol{\omega}_i^A \equiv \left[0,\ldots,0,\omega_{i1}^A,\ldots,\omega_{ig}^A,\ldots,\omega_{iG}^A,0,\ldots,0\right]$ and $\boldsymbol{\omega}_i^B \equiv \left[0,\ldots,0,\omega_{i1}^B,\ldots,\omega_{ig}^B,\ldots,\omega_{iG}^B,0,\ldots,0\right]$ are $1\times(G\times N)$ matrices with zeros everywhere except in the portion of the matrix corresponding to location i which extends the equations (32) and equation (33) to incorporate the consumption of the manufacturing good and the correlation in real returns across locations that arise through market prices and the correlated productivity shocks.

A.4.4 Optimal crop choice

We now turn to farmers' crop choice. Given these updated definitions for μ_i^Z and $\sigma_i^{2,Z}$, the expression for farmers expected utility – equation (19) – remains unchanged, as does the farmers' crop choice problem:

$$
\max_{\left\{\theta_{ig}^f\right\}} \mu_i^Z + \frac{1}{2}(1-\rho_i)\sigma_i^{2,Z} \text{ s.t. } \Sigma_{g\in\mathcal{G}}\theta_{ig}^f = 1.
$$

The resulting first order conditions can be written as:

$$
\mu_{ig}^Z - \rho_i\sigma_{ig}^Z = \lambda_i
$$

where:

$$\mu_{ig}^{Z} \equiv \frac{\left(\exp\left(\mu_{ig}^{A}\right)\right)^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}}\left(\theta_{ig}\right)^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}\left(\alpha_{ig}\right)^{\frac{1}{\sigma+\varepsilon_i}}}{\sum_{g=1}^{G}\theta_{ig}^{f}\left(\exp\left(\mu_{ig}^{A}\right)\right)^{\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}}\left(\theta_{ig}\right)^{-\frac{1}{\sigma+\varepsilon_i}}\left(\bar{p}_{m(i)g}\right)^{\frac{\varepsilon_i}{\sigma+\varepsilon_i}}\left(\alpha_{ig}\right)^{\frac{1}{\sigma+\varepsilon_i}}} - \sum_{h=1}^{G}\theta_{ih}^{f}\Sigma_{i,gh}^{Y} + \frac{1}{2}\Sigma_{i,gg}^{Y} + \left(\boldsymbol{\omega}_i^{A} + \boldsymbol{\omega}_i^{B}\mathbf{B}\right)\boldsymbol{\Sigma}^{A}\left(\left(\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{1}_{ig} + \left(\frac{\varepsilon_i}{\sigma+\varepsilon_i}\right)\mathbf{1}_{ig}\mathbf{B}\right)^{T} \tag{73}$$

and:

$$\sigma_{ig}^{Z} = \left(\boldsymbol{\omega}_i^{A} + \boldsymbol{\omega}_i^{B}\mathbf{B}\right)\boldsymbol{\Sigma}^{A}\left(\left(\frac{\varepsilon_i+\sigma-1}{\sigma+\varepsilon_i}\right)\mathbf{1}_{ig} + \left(\frac{\varepsilon_i}{\sigma+\varepsilon_i}\right)\mathbf{1}_{ig}\mathbf{B}\right)^{T} \tag{74}$$

and $\mathbf{1}_{ig}$ is an $(G \times N) \times 1$ vector with zeros everywhere except in the element ig. Imposing symmetry and the land clearing constraint then allows us to write optimal crop choice as:

$$\theta_{ig} = \frac{\alpha_{ig} B_{ig}^{\varepsilon_i+\sigma-1}\bar{p}_{m(i)g}^{\varepsilon_i}}{\sum_{h=1}^{G}\alpha_{ih}B_{ih}^{\varepsilon_i+\sigma-1}\bar{p}_{m(i)h}^{\varepsilon_i}}, \tag{75}$$

where

$$B_{ig} \equiv \frac{\exp\left(\mu_{ig}^{A}\right)}{\left(\lambda_i - \left(\frac{1}{2}\Sigma_{i,gg}^{Y} - \sum_{h=1}^{G}\theta_{ih}^{f}\Sigma_{i,gh}^{Y} + \frac{1}{2}(1-\rho_i)\sigma_{ig}^{Z}\right)\right)^{\frac{\sigma+\varepsilon_i}{\varepsilon_i+\sigma-1}}}, \tag{76}$$

which is the generalization of equation (21).

A.5 Calculating the equilibrium of the quantitative model

Here we briefly describe how we calculate the equilibrium of the quantitative model to perform the counterfactual results in Section 5. The calculation is centered around two sub-routines - the price function and the crop choice function. The price function calculates the equilibrium price at the districts, states (regional markets) and central market given crop choice at the districts (not necessarily optimal) and yield realizations. The crop choice function calculates the farmers' optimal crop choices given state and central market preferences, trade costs, network linkages and crucially, moments of the yields distribution over the time period in question. We describe each in turn.

A.5.1 Price Function

In the interior price function routine, we take the realized quantities produced of all crops in all districts as given and calculate the market clearing prices. The basic structure of the algorithm as follows: given an initial guess of prices, we first hold constant the aggregate demand in all locations (so that the partial equilibrium excess demand function is assured to satisfy the gross substitutes property) and then use a bisection method to find the prices such that the excess demand is equal to zero in all locations. We then update the aggregate demand from these market clearing prices and iterate until convergence.

Algorithm 1 details the process, where we maintain the notation that X refers to a value of variable X at the district level, $\bar{X}$ refer to its corresponding value at the regional market level, and X^* refer to its value in the central market.

A.5.2 Crop Choice Function

The outer crop choice function computes the optimal crop choice of district farmers. The basic structure of the algorithm is as follows: given an initial guess of crop choice, we calculate the equilibrium prices and the elasticity of all market prices to yields of all crops in all districts. This allows us to calculate the (approximate) distribution of real returns for farmers in all locations. We then evaluate the first order conditions of farmers crop choice problem, and increase (decrease) their allocation of crops with higher (lower) marginal returns. We then iterate the procedure to convergence.

Algorithm 2 details the process.

A.5.3 The welfare effects of a counterfactual

Given the optimal crop allocations, we proceed by calculating the welfare. To do so, we calculate the equilibrium real returns $\left\{Z_i^f\left(s;\left\{\theta_{ig}^f\right\}_{g\in\{1,...,G\}}\right)\right\}$ from equation (65) given those optimal crop allocations and the actual observed yields for each year $t\in\{1970,1971,...,1979\}$, i.e. we determine what the realized returns for all agents would have been given a particular year's actual yield realizations and their (potentially counterfactual) crop choice. We then calculate the (log of the) mean and variance (of the log) of real returns by calculating the corresponding sample moments across the ten years within the decade, e.g.:

$$\mathbb{E}[Z_{id}(s)]\equiv\frac{1}{10}\sum_{t=1970}^{1979}Z_i^f\left(s_t,\{\theta_{igd}\}_{g\in\{1,...,G\}}\right)$$
$$\sigma_{id}^{2,Z}\equiv\frac{1}{10}\sum_{t=1970}^{1979}\left(\ln Z_i^f\left(s_t,\{\theta_{igd}\}_{g\in\{1,...,G\}}\right)-\frac{1}{10}\sum_{\tau=1970}^{1979}\ln Z_i^f\left(s_\tau,\{\theta_{igd}\}_{g\in\{1,...,G\}}\right)\right)^2,$$

where $\{\theta_{igd}\}_{g\in\{1,...,G\}}$ are the optimal crop allocations. Similarly, we can calculate the expected utility as the average of the utility across the ten years from equation (6):

$$\mathbb{E}[U_{id}(s)]\equiv\frac{1}{10}\sum_{t=1970}^{1979}\left(\frac{1}{1-\rho_0}\left(\left(Z_i^f(s_t)\right)^{1-\rho_{id}}-1\right)\right).$$

We remark that if the observed yields are distributed log normal and we apply the second order approximation used in equation (71) to determine the optimal crop choice, then we have:

$$\mathbb{E}[U_{id}(s)]\approx\frac{1}{1-\rho_0}\exp\left((1-\rho_0)\left(\ln\mathbb{E}\left[Z_i^f(s)\right]-\frac{1}{2}\rho_{id}\sigma_{id}^{2,Z}\right)\right),$$

which turns out to be an excellent approximation in the quantitative exercise (with a correlation across districts exceeding 0.999).

With these calculations in hand, we construct counterfactual results reported in Table 6. To do so, we first note that as the crop cultivation costs are calibrated to ensure the observed crop allocations are optimal, we can calculate the mean, volatility, and expected welfare for the actual 1970s by simply holding trade costs and bank access constant at their observed 1970s levels and proceeding with the calculations above, which provides a baseline set of parameters $\left\{\mathbb{E}[Z_{i0}(s)],\sigma_{i0}^{2,Z},\mathbb{E}[U_{i0}(s)]\right\}$. Columns 1 and 2 simply report the average difference across districts between the (log of the) mean real returns and variance (of the log) of real returns between the counterfactual and baseline values, i.e. $\ln\mathbb{E}[Z_{id}(s)]-\ln\mathbb{E}\left[Z_{i0}^f(s)\right]$ and $\sigma_{id}^{2,Z}-\sigma_{i0}^{2,Z}$. Because the expected utility is of course an ordinal measure, we instead calculate percentage increase in guaranteed income – holding constant everything else constant at the baseline values – that would yield the equivalent change in utility as the counterfactual being considered, which we refer to as the "certainty equivalent variation". Noting that demand is homothetic (i.e. the realized returns are homogeneous of degree one in income) and using the log-normal formulation of the expected utility from above, we can calculate CEV as follows:

$$\begin{aligned}\mathbb{E}[U_{id}(s)]-\mathbb{E}[U_{i0}(s)]=&\frac{1}{1-\rho_{i0}}\exp\left((1-\rho_{i0})\left(\ln\mathbb{E}[(\exp(CEV))Z_{i0}(s)]-\frac{1}{2}\rho_{i0}\sigma_{i0}^{2,Z}\right)\right)\\&-\frac{1}{1-\rho_{i0}}\exp\left((1-\rho_{i0})\left(\ln\mathbb{E}[Z_{i0}(s)]-\frac{1}{2}\rho_{i0}\sigma_{i0}^{2,Z}\right)\right)\iff\\CEV=&\left(\ln\mathbb{E}[Z_{id}(s)]-\frac{1}{2}\rho_{id}\sigma_{id}^{2,Z}\right)\\&-\left(\ln\mathbb{E}[Z_{i0}(s)]-\frac{1}{2}\rho_{i0}\sigma_{i0}^{2,Z}\right),\end{aligned}\tag{77}$$

i.e. the certainty equivalent variation is simply the difference between counterfactual and baseline in the combination of the (log of the) mean real returns and the variance (of the log) of the real returns, with the weight on the variance governed by the effective risk aversion parameter. (Note that this is consistent with the interpretation of the effective risk aversion parameter as a technological parameter governed by access to insurance, see Appendix A.3.2 for details.

A.6 Comparing the model to a traditional arbitrage model

In this subsection, we describe the methodology used to construct panel (c) of Figure 3 that compares the price arbitrage of our model to a traditional arbitrage model where iceberg trade costs are homogeneous. In both cases, consider a "village" (a district, in the data) whose autarkic relationship between prices and yields follows from CES preferences and the market clearing:

$$\log p_{ig}^{aut} = -\frac{1}{\sigma}\log A_{ig} + \frac{1}{\sigma}\log \beta_i \alpha_{ig} + \frac{1}{\sigma}\log \frac{\sum_{h=1}^{G} p_{ih}^{aut} L_i \theta_{ih} A_{ih}}{L_i \theta_{ig} \sum_{h=1}^{G} \alpha_{ih} (p_{ih}^{aut})^{1-\sigma}}, \tag{78}$$

where we omit the state of the world for readability. Suppose that the village is small in size relative to a market (a state, in the data) which has a price $\bar{p}_g$. Note that given estimates of β, α and σ from Section 5.3 and observed yields $\{A_{ig}\}$, allocations $\{\theta_{ig}\}$, and land areas $\{L_i\}$, there exists a unique (to-scale) set of autarkic prices p_{ig}^{aut} that satisfy equation (78).

A standard "kinked" model First consider a standard trade model, where the village is separated from the regional market by an iceberg trade costs $\tau_i > 1$. Then a standard no-arbitrage condition delivers the following relationship between the equilibrium local prices p_{ig}, the given market price, $\bar{p}_g$ and the autarkic local price p_{ig}^{aut}:

$$\log p_{ig} - \log \bar{p}_g = \begin{cases} \log \tau_i & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g > \log \tau_i \\ \log p_{ig}^{aut} - \log \bar{p}_g & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g \in [-\log \tau_i, \log \tau_i] \\ -\log \tau_i & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g < -\log \tau_i \end{cases} \tag{79}$$

The difference between the equilibrium local prices and the regional market prices then are a "kinked" function of the trade costs between the two (when trade occurs and the no-arbitrage equation holds) and the autarkic price p_{ig}^{aut} (when the trade costs are sufficiently high such that no trade occurs).

Our "smooth" model Now consider our framework, where from equation 54 equilibrium prices are:

$$\log p_{ig} = -\frac{1}{\sigma+\varepsilon_i}\log A_{ig} + \frac{\varepsilon_i}{\sigma+\varepsilon_i}\log \bar{p}_{m(i)g} + \frac{1}{\sigma+\varepsilon_i}\log \beta_i \alpha_{ig} + \frac{1}{\sigma+\varepsilon_i}\log \frac{\sum_{h=1}^{G} p_{ih} L_i \theta_{ih} A_{ih}}{L_i \theta_{ig} \sum_{h=1}^{G} \alpha_{ih} (p_{ih})^{1-\sigma}} \tag{80}$$

Combining equations (80) and (78) we can then write the difference between the equilibrium local price and the central market price

$$\log p_{ig} - \log \bar{p}_g = \frac{\sigma}{\sigma+\varepsilon_i}(\log p_{ig}^{aut} - \log \bar{p}_g) \tag{81}$$

Hence, unlike equation (79), equation (81) states that the local price relative to the market price should smoothly vary with the difference with the local autarkic price relative to the market price. Note that equations (79) and (81) coincide with each other under autarky ($\varepsilon_i = 0$, $\tau_i = \infty$) or free trade ($\varepsilon_i = \infty$, $\tau_i = 1$).

Empirical Strategy The basic idea is to compare the model fit of equations equations (79) and (81). In order to do so, we have to first solve a few implementation issues. First, as autarkic prices are only identified up to scale, we add a location specific constant c_i to both models, so that the standard "kinked" model becomes:

$$\log p_{ig} - \log \bar{p}_g = \begin{cases} \log \tau_i + c_i & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g > \log \tau_i + c_i \\ \log p_{ig}^{aut} - \log \bar{p}_g & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g \in [-\log \tau_i + c_i, \log \tau_i + c_i], \\ -\log \tau_i + c_i & \text{for } \log p_{ig}^{aut} - \log \bar{p}_g < -\log \tau_i + c_i \end{cases} \tag{82}$$

and our "smooth" model becomes:

$$\log p_{ig} - \log \bar{p}_g = \frac{\sigma}{\sigma+\varepsilon_i}(\log p_{ig}^{aut} - \log \bar{p}_g) + c_i \tag{83}$$

The advantage of the additional constant is that both models are now ensured to have an R-squared statistic between 0 and 1, which will be our statistic for goodness of fit. The second issue is how to measure prices.

Because of the potential endogeneity of yields to prices, as in Section 5.3, we use rainfall-predicted yields to construct a rainfall-predicted measure of autarkic prices from equation (78). Also as in Section 5.3, we measure the market price as the quantity weighted average price in all districts within a state except the one being examined to avoid mechanical correlations between market and local prices.

Estimation and Results As in Section 5.3, we allow the trade costs to vary by district-decade. To do so, we conduct the estimation of both the standard "kinked" model and our "smooth" model separately for each district-decade combination. The estimation for our smooth model is simply a linear regression of the log district price (relative to the state leave-one-out price) on the log rainfall-predicted autarkic price (again relative to the state leave-one-out price). The kinked model is similar, but uses a non-linear least squares routine to capture the kinks present in equation (82), where we constrain $\log \tau_i \geq 0$. Note that in both cases, we are estimating just two parameters using the same left hand side and right hand side variables: the constant c_i and a measure of trade costs ($\log\tau_i$ for the standard model and $\frac{\sigma}{\sigma+\varepsilon_i}$ for our model).

We compare the residual sum of squares from both models and normalize this by the variance of the dependent variable to create a comparable version of the R^2 for comparison. Panel (c) of Figure 3 plots the cumulative density of the fits for the two models. The smooth model has a better fit than the kinked model in nearly 71% of all district-decade combinations. The mean R^2 of the smooth and kinked model runs are 0.11 and 0.15 respectively.

Table A.1: Market Access Associations (district-decade level)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	Banks/capita	Mean(lnYield)	Var(lnYield)	HYV Prop	Irig. Prop	HYV Prop (fillin)	Irig. Prop (fillin)
State MA	-0.394	0.124	-0.014	-0.070	0.058	0.055	0.072
	(0.466)	(0.090)	(0.033)	(0.062)	(0.049)	(0.046)	(0.048)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.950	0.973	0.851	0.928	0.961	0.929	0.968
Observations	1,243	1,243	1,243	1,171	1,217	1,243	1,243

Notes: Various variables projected on within-state market access using the same specification as equation (4) and Table 3 in the main text. Variable names in column titles. Fillin columns replace missing HYV or irrigated area data with zeroes. All columns include district and state-decade fixed effects. Market access multiplied by 100,000. Each observation is a district-decade. Observations are weighted by district-decade total cropped area. Robust standard errors reported in parentheses. Stars indicate statistical significance: * $p<.10$ ** $p<.05$ *** $p<.01$.

Table A.2: Crop Choice and Openness: Robustness Part 1

Dependent variable:	IHS crop choice, no winsorizing			Crop choice			(Log) Crop choice		
	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV
Mean(log Yield)	0.004*	0.005*	0.003	0.004	0.005*	0.002	0.622***	0.643***	0.586***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.153)	(0.153)	(0.151)
Var(log Yield)	0.037***	0.024*	0.085***	0.030**	0.008	0.085***	5.817***	5.278***	6.898***
	(0.012)	(0.013)	(0.023)	(0.013)	(0.013)	(0.024)	(0.707)	(0.826)	(1.137)
Mean × State MA	0.009**	0.008**	0.011***	0.011***	0.010**	0.013***	0.709***	0.672***	0.768***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.184)	(0.186)	(0.182)
Var × State MA	-0.132***	-0.093**	-0.231***	-0.132***	-0.082**	-0.239***	-3.533**	-1.613	-7.149**
	(0.034)	(0.039)	(0.063)	(0.033)	(0.035)	(0.065)	(1.739)	(2.279)	(3.616)
Covar(log Yield)		0.017*			0.028***			0.657	
		(0.009)			(0.009)			(0.444)	
Covar × State MA		-0.057*			-0.074**			-2.698**	
		(0.029)			(0.031)			(1.362)	
Var × Bank			-11.271***			-14.023***			-250.208
			(3.441)			(3.961)			(170.556)
Var × State MA × Bank			21.734**			24.405***			772.486
			(8.507)			(8.770)			(550.242)
Crop-decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-crop FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	-0.005	-0.012	-0.011	-0.001	-0.014	-0.006	-0.097	-0.103	-0.100
Observations	18,626	18,626	18,626	18,626	18,626	18,626	18,626	18,626	18,626
First-Stage F Stat	13	22	6	117.1	37.7	78.6	117.1	37.7	78.6

Notes: This table replicates the regressions in Columns (2), (4) and (6) of Table 2 with alternate specifications. Crop choice regressed on the mean of log yields, the variance of log yields, and both terms interacted with within-state market access (i.e. access to districts in the same state). All columns include crop-decade, district-decade, and district-crop fixed effects. Columns (1)–(3) use non-winsorised variance of yields. Columns (4)–(6) use crop choice as a fraction of land planted. Columns (7)–(9) use (log) crop choice as a fraction of land planted (with zero-crop-share values replaced by the
1st percentile value of non-missing log crop shares, along with non-missing values below the 1st percentile). All columns instrument for mean log yields and the variance of log yields with the mean and variance of log predicted yields from a regression of log yield on local rainfall shocks for each month interacted with state-crop fixed effects and controlling for crop-decade, district-decade, and district-crop fixed effects. Interaction with market access instrumented with the predicted yield instruments interacted with market access. Market access variables multiplied by 100,000 and banks per capita multiplied by 1000. Each observation is a district-crop-decade. Observations are weighted by the district-decade total cropped area divided by the number of observations in a district decade. Standard errors clustered at the district-decade level reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table A.3: Crop Choice and Openness: Robustness Part 2

Dependent variable:	IHS crop choice, $\phi=1$			IHS crop choice, 1/4 highway speed			IHS crop choice
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IV	IV	IV	IV	IV	IV	IV
Mean(log Yield)	0.004	0.005*	0.002	0.003	0.004	0.002	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
Var(log Yield)	0.028**	0.006	0.080***	0.031**	0.005	0.085***	0.064**
	(0.012)	(0.012)	(0.023)	(0.013)	(0.013)	(0.024)	(0.028)
Mean × State MA	0.010***	0.010**	0.012***	0.009***	0.009**	0.011***	-0.004
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.007)
Var × State MA	-0.125***	-0.074**	-0.224***	-0.109***	-0.053	-0.190***	-0.183***
	(0.031)	(0.034)	(0.062)	(0.027)	(0.033)	(0.054)	(0.062)
Covar(log Yield)		0.028***			0.033***		
		(0.009)			(0.010)		
Covar × State MA		-0.076**			-0.079***		
		(0.030)			(0.029)		
Mean × National MA			-13.319***			-13.721***	0.024***
			(3.665)			(3.851)	(0.007)
Var × National MA			22.719***			18.484**	0.000
			(8.327)			(7.378)	(0.044)
Var × Bank							-9.522*
							(5.437)
Var × State MA × Bank							18.457**
							(8.584)
Var × National MA × Bank							-1.737
							(4.853)
Mean × Bank							1.121*
							(0.619)
Mean × State MA × Bank							1.080
							(0.810)
Mean × National MA × Bank							-0.640
							(0.701)
Crop-decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-crop FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	-0.001	-0.015	-0.006	-0.000	-0.016	-0.005	0.002
Observations	18,626	18,626	18,626	18,626	18,626	18,626	18,626
First-Stage F Stat	117.1	37.7	78.6	115.1	39.0	77.5	13.7

Notes: Columns (1)–(6) of this table replicate the regressions in Columns (2), (4) and (6) of Table 2 with alternate specifications. Crop choice regressed on the mean of log yields, the variance of log yields, and both terms interacted with within-state market access (i.e. access to districts in the same state). All columns include crop-decade, district-decade, and district-crop fixed effects. Columns (1)–(3) use $\phi = 1$ to calculate market access. Columns (4)–(6) use off highway speed of travel equal to 1/4 of the highway speed (instead of 1/3) when calculating market access. Columns (7) adds additional interactions between the mean of log yields, market access and banks. All columns instrument for mean log yields and the variance of log yields with the mean and variance of log predicted yields from a regression of log yield on local rainfall shocks for each month interacted with state-crop fixed effects and controlling for crop-decade, district-decade, and district-crop fixed effects. Interactions with market access instrumented with the predicted yield instruments interacted with market access. Market access variables multiplied by 100,000 and banks per capita multiplied by 1000. Each observation is a district-crop-decade. Observations are weighted by the district-decade total cropped area divided by the number of observations in a district decade. Standard errors clustered at the district-decade level reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table A.4: Crop Choice and Openness: Robustness Part 3

Dependent variable:	IHS fraction of land planted by crop					
Interaction variable:	HYV proportion				Years elapsed in decade	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean(log Yield)	-0.025***	-0.029***	0.001	0.003	0.003	0.002
	(0.006)	(0.007)	(0.003)	(0.003)	(0.002)	(0.002)
Var(log Yield)	0.058*	0.058*	0.043**	0.038**	0.029***	0.044***
	(0.031)	(0.032)	(0.017)	(0.016)	(0.011)	(0.012)
Mean × State MA	0.030***	0.038*	0.015***	0.001	0.011***	0.011***
	(0.010)	(0.020)	(0.005)	(0.005)	(0.003)	(0.003)
Var × State MA	-0.222**	-0.188*	-0.124***	-0.112***	-0.116***	-0.119***
	(0.106)	(0.107)	(0.038)	(0.036)	(0.027)	(0.028)
Main Effect of Variable	-0.014***	-0.019**	-0.007**	-0.100	0.000***	-0.001***
	(0.005)	(0.009)	(0.003)	(0.069)	(0.000)	(0.000)
Mean × Variable		0.002**		0.012		0.000***
		(0.001)		(0.009)		(0.000)
Var × Variable		-0.059		0.078		-0.003***
		(0.073)		(0.071)		(0.001)
Mean × State MA × Variable		-0.002		0.006***		-0.000***
		(0.004)		(0.002)		(0.000)
Var × State MA × Variable		-0.113		-0.444		0.001
		(0.425)		(0.392)		(0.002)
Crop-decade FE	Yes	Yes	Yes	Yes	Yes	Yes
District-decade FE	Yes	Yes	Yes	Yes	Yes	Yes
District-crop FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.003	-0.002	-0.001	0.006	-0.000	0.001
Observations	6,119	6,119	14,334	14,334	183,896	183,896
First-Stage F Stat	32.6	8.2	78.4	16.1	150.4	75.2

Notes: Crop choice regressed on the mean of log yields, the variance of log yields, both terms interacted with within-state market access (i.e. access to districts in the same state), plus an additional interaction with variable detailed in column header. All columns include crop-decade, district-decade, and district-crop fixed effects. All columns instrument for mean log yields and the variance of log yields with the mean and variance of log predicted yields from a regression of log yield on local rainfall shocks for each month interacted with state-crop fixed effects and controlling for crop-decade, district-decade, and district-crop fixed effects. Interactions with market access and further interaction terms instrumented with the predicted yield instruments interacted with market access and interaction term. Columns (2) and (4) include main effects and interactions of the proportion of area cropped planted with HYV varieties where columns (3) and (4) replace missing HYV information with zeroes. Columns (5) and (6) include additional interactions with years elapsed within a decade, and crop choice is at the annual rather than decadal level. Odd columns use the same sample as proceeding column but without the additional interactions. Market access variables multiplied by 100,000. Each observation is a district-crop-decade except for columns (5) and (6) that are at district-crop-year level. Observations are weighted by district-decade total cropped area divided by the number of observations in a district decade. Standard errors clustered at the district-decade level reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table A.5: Real Income and Roads: Robustness

Panel (a): Winsorized real income

Dependent variable:	Components of Real Income					
	(1)	(2)	(3)	(4)	(5)	(6)
	Var Log Nominal Y	Var Log P Index	Var Log Real Y	Var Log Nominal Y	Var Log P Index	Var Log Real Y
State Market Access	0.743	-0.531*	0.359	1.238*	-0.735**	0.719**
	(0.637)	(0.306)	(0.286)	(0.669)	(0.321)	(0.299)
National Market Access				-1.651**	0.681**	-1.199***
				(0.699)	(0.336)	(0.313)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.539	0.729	0.501	0.542	0.731	0.510
Observations	1210	1210	1210	1210	1210	1210

Panel (b): Alternative off-highway speed (1/4 instead of 1/3 highway speed)

Dependent variable:	Components of Real Income					
	(1)	(2)	(3)	(4)	(5)	(6)
	Var Log Nominal Y	Var Log P Index	Var Log Real Y	Var Log Nominal Y	Var Log P Index	Var Log Real Y
State Market Access	2.143***	-0.201	1.299***	2.204***	-0.205	1.329***
	(0.757)	(0.318)	(0.402)	(0.768)	(0.323)	(0.408)
National Market Access				-0.383	0.025	-0.192
				(0.809)	(0.340)	(0.430)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.543	0.754	0.490	0.543	0.754	0.491
Observations	1166	1166	1166	1166	1166	1166

Panel (c): ϕ=1 for Market Access

Dependent variable:	Components of Real Income					
	(1)	(2)	(3)	(4)	(5)	(6)
	Var Log Nominal Y	Var Log P Index	Var Log Real Y	Var Log Nominal Y	Var Log P Index	Var Log Real Y
State Market Access	1.429**	-0.192	0.859***	1.590**	-0.218	0.922***
	(0.568)	(0.239)	(0.302)	(0.618)	(0.259)	(0.329)
National Market Access				-0.255	0.041	-0.100
				(0.384)	(0.161)	(0.204)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
State-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.542	0.754	0.489	0.542	0.754	0.489
Observations	1166	1166	1166	1166	1166	1166

Notes: Regressions of the variance of the log of real income and its components on within-state market access (i.e. access to districts in the same state) multiplied by 100,000. Panel (a) uses winsorized real incomes and its components as the dependent variable. Panel (b) uses an alternative off-highway speed of travel equal to 1/4 the highway speed (as opposed to 1/3). Panel (c) uses $\phi=1$ for market access calculations. Columns (4)–(6) additionally include outside state market access (i.e. access to districts in other states) multiplied by 100,000. Each observation is a district-decade. Panels (b) and (c) only include observations with at least 25% of cropped area with observed prices. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table A.6: District Risk Sharing and Banks

Dependent variable:	Log(household per capita expenditure)						
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) IV	(6) IV	(7) Alt. IV
Log(Household crop income)	0.095***	0.096***	0.103***	0.107***	0.109***	0.110***	0.107***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log(District crop income)	0.144***	0.173***	-0.119***	-0.099***	-0.098***	-0.141***	-0.105***
	(0.012)	(0.013)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Banks per capita			1.820	2.732*	6.022***	4.534***	2.870*
			(1.620)	(1.630)	(1.627)	(1.634)	(1.631)
Log(Household income) × Banks			-0.297*	-0.406**	-0.384**	-0.529***	-0.434**
			(0.171)	(0.172)	(0.171)	(0.173)	(0.172)
Log(District income)× Banks			15.997***	16.093***	13.106***	16.701***	15.329***
			(1.119)	(1.155)	(1.140)	(1.156)	(1.122)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	No	No	No	No	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	No	Yes
R-squared	0.220	0.065	0.226	0.072	0.069	0.066	0.073
Observations	85,285	85,285	85,285	85,285	85,285	85,285	85,206
First-Stage F Stat	.	195455.4	.	102682.9	100914.7	115329.7	143542.4

Notes: Log monthly household per capita expenditure (mpce) regressed on household-level crop income and the district-average (for households observed in the same year) of household-level crop income. Columns (3)–(6) further include district banks per capita and banks per capita interacted with the crop income measures, where bank per capita is multiplied by 100,000. All columns include district fixed effects, column (5) additionally includes year fixed effects. Controls include differences between household and district-average crop shares for each of the 15 crops, as well as the difference between household and district-average total land cultivated. Sample constructed from rounds 43, 50, 55 and 61 of India's National Sample Surveys (1987/1988, 1993/1994, 1999/2000 and 2004/2005). Household cropping patterns extracted from whether household consumes crop out of home production in the consumption module (schedule 1.0) with area of land cultivated assumed to be equally allocated across all crops they produce. Total crop income calculated from (interpolated) yields and prices drawn from the VDSA data described in Section 2.3 of the paper. Crop incomes (and interactions) instrumented in columns (2) and (4)–(7) using predicted yields constructed from rainfall variation as in Section 3.2 of the paper (column (7) uses the predicted yields from Section 3.1). Regressions run only on farm households (those producing at least one of our 15 sample crops). To account for non farm income in the district average income measures, non farm households assigned income of 1 and farm household income is divided by the national average crop income for farmers multiplied by the ratio of average non farm to farm household mpce. Each observation is a household-year observation. Observations are weighted by NSS survey weights. Robust standard errors reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table A.7: More volatility can attenuate the gains from trade (Prop. 3, Part 3)

Case 1: Some volatility

	Village 1		Village 2	
	Crop A	Crop B	Crop A	Crop B
Mean yield	1	1	1	1
Variance of log yield	1	0	0	1
Autarkic crop allocation	0.5	0.5	0.5	0.5
Trade crop allocation	0.31	0.69	0.69	0.31
Gains from trade	0.124		0.124	

Case 2: More volatility

	Village 1		Village 2	
	Crop A	Crop B	Crop A	Crop B
Mean yield	1	1	1	1
Variance of log yield	1	1	1	1
Autarkic crop allocation	0.5	0.5	0.5	0.5
Trade crop allocation	0.5	0.5	0.5	0.5
Gains from trade	0		0	

Notes: This table provides an example of how increasing the volatility of yields may attenuate the gains from trade (Case 1 vs. Case 2). In Case 2, yield realizations are perfectly correlated between crops within village but uncorrelated across villages. In each example, both village types are the same size and have equal budget shares across the two crops. The numbers reported assume a risk aversion parameter $\rho = 2$ and gains from trade are calculated moving from autarky ($\epsilon_i = 0$) to costly trade ($\epsilon_i = 1$).

Table A.8: Agricultural expenditure shares

Panel (a): Crop-specific demand shifters (α_{ig})

Crop	25% percentile	Mean	75% percentile
Barley	0.000	0.003	0.000
Chickpea	0.008	0.022	0.035
Cotton	0.000	0.000	0.000
Finger Millet	0.000	0.012	0.000
Groundnut	0.001	0.047	0.091
Linseed	0.000	0.014	0.022
Maize	0.000	0.019	0.009
Pearl Millet	0.000	0.022	0.012
Pigeon pea	0.006	0.044	0.071
Rice	0.068	0.352	0.656
Rape and mustard seed	0.000	0.049	0.100
Sesame	0.000	0.014	0.022
Sorghum	0.000	0.054	0.051
Sugarcane (gur)	0.049	0.092	0.116
Wheat	0.046	0.256	0.433

Panel (b): Agricultural expenditure share (β_i)

	25% percentile	Mean	75% percentile
Ag. exp. share	0.330	0.381	0.444

Notes: This table provides summary statistics on the agricultural preferece shifters (α_{ig}) and total agricultural expenditure shares (β_i) used in the structural estimation. The preference shifters are normalized so that they sum to one across all crops. For each district, the parameters are calculated to match the observed district average expenditure shares from the Indian National Sample Survey Round 43.

Table A.9: Estimated crop costs and actual crop costs

Dependent variable:	Estimated Crop Costs (Log)	
	(1)	(2)
Observed Crop Costs (Log)	0.420**	0.420
	(0.197)	(0.359)
Decade FE	Yes	Yes
Crop FE	Yes	Yes
State-Decade-Crop Clustered SEs	No	Yes
R-squared	0.407	0.407
Observations	3030	3030

Notes: Regression of the estimated crop costs on the log of actual state-level crop costs, decade fixed effects and crop fixed effects. Each observation is a crop-district-decade triplet. Estimated crop costs come from a combination of fixed effects and residuals from regression (27) which are the unobserved crop costs that ensure that observed crop choices in the data are optimal crop choices in the model. As the crop costs are only identified up to scale within a district-decade, we normalize the cost of one crop (Barley) to zero in all district-decade pairs. Raw data on actual crop costs in Rupees/Hectare come from the Government publication *Cost of Cultivation of Principal Crops in India*. Data are annual at the state-crop level and cover 13 of our 15 crops between 1983-2008. To match with the crop-decade level estimated crop costs, actual costs are deflated by the all-India CPI and averaged over decades for each crop and state. Standard errors are reported in parentheses. As the actual crop costs are only at the State level, Column 2 clusters standard errors at the state-decade-crop level. Stars indicate statistical significance: * $p<.10$ ** $p<.05$ *** $p<.01$.

Table A.10: Correlation between actual and counterfactual crop choice

Dependent variable:	Observed (log) crop share			
	(1)	(2)	(3)	(4)
Predicted (log) crop share	0.897***	0.878***	0.669***	0.854***
	(0.008)	(0.265)	(0.199)	(0.209)
Crop-district FE	No	Yes	Yes	Yes
District-decade FE	No	No	Yes	Yes
Crop-decade FE	No	No	No	Yes
R-squared (within)	0.811	0.005	0.002	0.004
Observations	18660	18660	18660	18660

Notes: Each observation is a district-decade-crop triplet; there are 4 decades, 311 districts, and 15 crops. The dependent variable is the observed (log) crop share. The independent variable is the predicted equilibrium (log) crop share from the Indian highway expansion, holding all other parameter constant at their 1970s level. Standard errors clustered at the district level are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Algorithm 1: Calculate equilibrium prices given crop choice

Require: preference parameters $(\alpha,\bar{\alpha},\alpha^*,\beta,\bar{\beta},\beta^*,\sigma,\bar{\sigma},\sigma^*)$, populations $(L,\bar{L},L^*)$, trade openness parameters $(\epsilon,\bar{\epsilon})$, yield realization (A), trade network linkage from districts to states, **initial guesses of equilibrium price** $(\bar{p}_0,p_0^*)$, number of goods (G), number of districts (N), number of states (M)

Ensure: Δprice between consecutive iterations are small (specified $tol=10^{-8}$)

Initialize difference in price updates between iterations $\Delta P \leftarrow 1$
Initialize guess $\bar{p}_0 = J_{GM}, p_0^* = J_{G1}$
Initialise updating step size $=0.1$
while $\Delta P \geq tol$ **do**
Step 1: Calculate district market clearing prices (p) **using** $(\bar{p}_0,p_0^*)$ **from equation** 54 (using equations 55 and 56)

Step 2: Calculate regional market clearing prices $(\bar{p}_1)$ **consistent with** (p,p_0^*)

Initialise $\{\bar{p}_{ub},\bar{p}_{lb}\}$ as upper and lower bounds for $\bar{p}$
while $\bar{p}_{ub}-\bar{p}_{lb} \geq \overline{tol}$ **do**
$\bar{p}_1 = \frac{\bar{p}_{ub}+\bar{p}_{lb}}{2}$
Calculate supply to state from districts $(\bar{C}_{supply})$ using equation 58
Calculate state demand $(\bar{C}_{demand})$ using equation 59
if $\bar{C}_{demand} > \bar{C}_{supply}$ **then**
$\bar{p}_{lb} = \bar{p}_1$
else if $\bar{C}_{demand} < \bar{C}_{supply}$ **then**
$\bar{p}_{ub} = \bar{p}_1$
end if
end while

Step 3: Calculate central market clearing prices (p_1^*) **consistent with** $(p,\bar{p}_1)$

Repeat **Step 2** for central market using the corresponding supply to central market and central market demand equations

$$C^*_{g,supply}(s) = \sum_{m \in \mathcal{M}} \left(1-\left(\frac{\bar{p}_{mg}(s)}{p_g^*(s)}\right)^{\varepsilon_m}\right)\bar{Q}_{mg}(s)$$

$$C^*_{g,demand}(s) = \beta^* \frac{\alpha_g^* Y^*(s) p_g^*(s)^{-\sigma}}{\sum_{h=1}^{G} \alpha_h^* \left(p_h^*(s)\right)^{1-\sigma}}$$

Step 4: Update price guess $(\bar{p}_0,p_0^*)$

Update $\log\bar{p}_0 = update * \log\bar{p}_0 + (1-update) * \log\bar{p}_1$
Update $\log p_0^* = update * \log p_0^* + (1-update) * \log p_1^*$
Calculate $\Delta price$ between iterations $\Delta P = norm(\log p_0^* - \log p_1^*) + norm(\log\bar{p}_0 - \log\bar{p}_1)$
end while

Notes: This psuedo-code describes the interior algorithm used to calculate the equilibrium prices; see Appendix A.5 for details.

Algorithm 2: Calculate optimal crop choice

Require: preference parameters (α,$\bar{\alpha}$,α^*,β,$\bar{\beta}$,β^*,σ,$\bar{\sigma}$,σ^*), populations (L,$\bar{L}$,L^*), trade openness parameters (ϵ,$\bar{\epsilon}$), yield realisation (A), yield covariance matrix (Σ), utility costs of cultivation, risk aversion coefficient (ρ), trade network linkage from districts to states, **initial guesses of equilibrium price ($\bar{p}_0$,p_0^*)**, number of goods (G), number of districts (N), number of states (M)

Ensure: $\Delta\theta$ between consecutive iterations are small (specified $tol = 10^{-3}$)

Initialize θ_0 as initial guess (observed crop choices in decade)

Initialise $\Delta\theta \leftarrow 1$

while $\Delta\theta \geq tol$ **do**

Step 1: Compute prices at θ_0 using price function

Step 2: Calculate the regional market price elasticity of supply (perturbations in yield realisation) at θ_0 and the subsequent covariance matrices for income and price indices (equations 68 and 68)

Perturb the mean of (log) yields for a crop g in district i by a small difference (specified $d\ln\mu_{ig} = 0.001$)

Compute the new equilibrium regional market price $\bar{p}_{new}$ at the perturbed yield

The regional price elasticity matrix $B_{ig} = \frac{\log\bar{p}_{new} - \log\bar{p}}{d\ln\mu_{ig}}$

Step 3: Calculate how close θ_0 is to satisfying first order condition (FOC) and update

Calculate $\lambda_{ig} = \mu_{ig}^Z - \rho_i \sigma_{ig}^Z$ from equations 73 and 74

Calculate targeted shadow cost λ_i for each district as average of λ_{ig}

Calculate deviance from FOC for each district-crop as $\Delta = \lambda_{ig} - \lambda_i$ and update θ_0 as $\theta_1 = \theta_0 e^{\Delta}$

Normalise θ_1 to sum to 1 within each district

Step 4: Update crop choice guesses (θ_0)

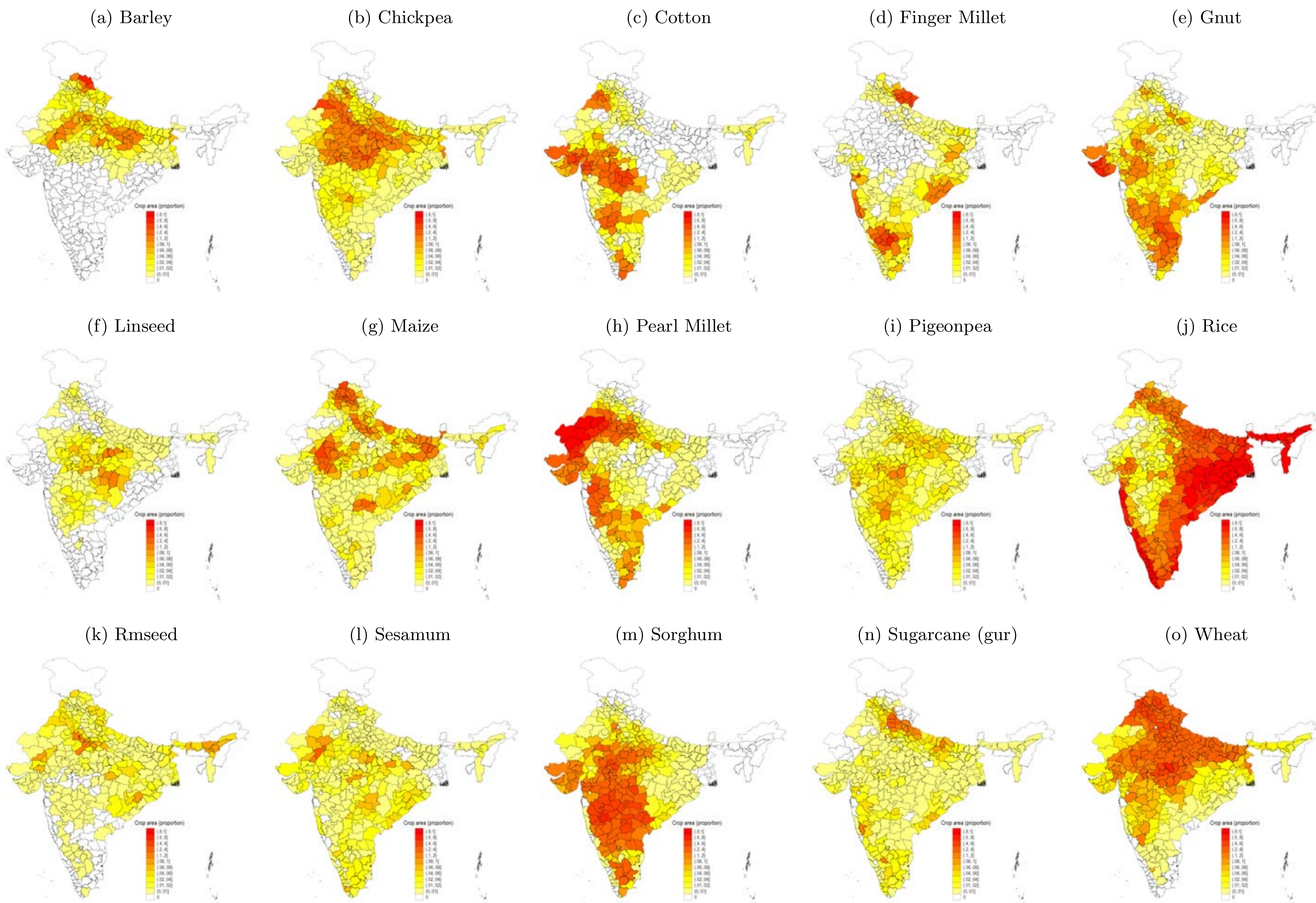
Update $\log\theta_0 = update * \log\theta_1 + (1 - update) * \log\theta_0$

Calculate $\Delta\theta$ as the difference between crop choice updates $\Delta\theta = norm(\theta_1 - \theta_0)$

end while

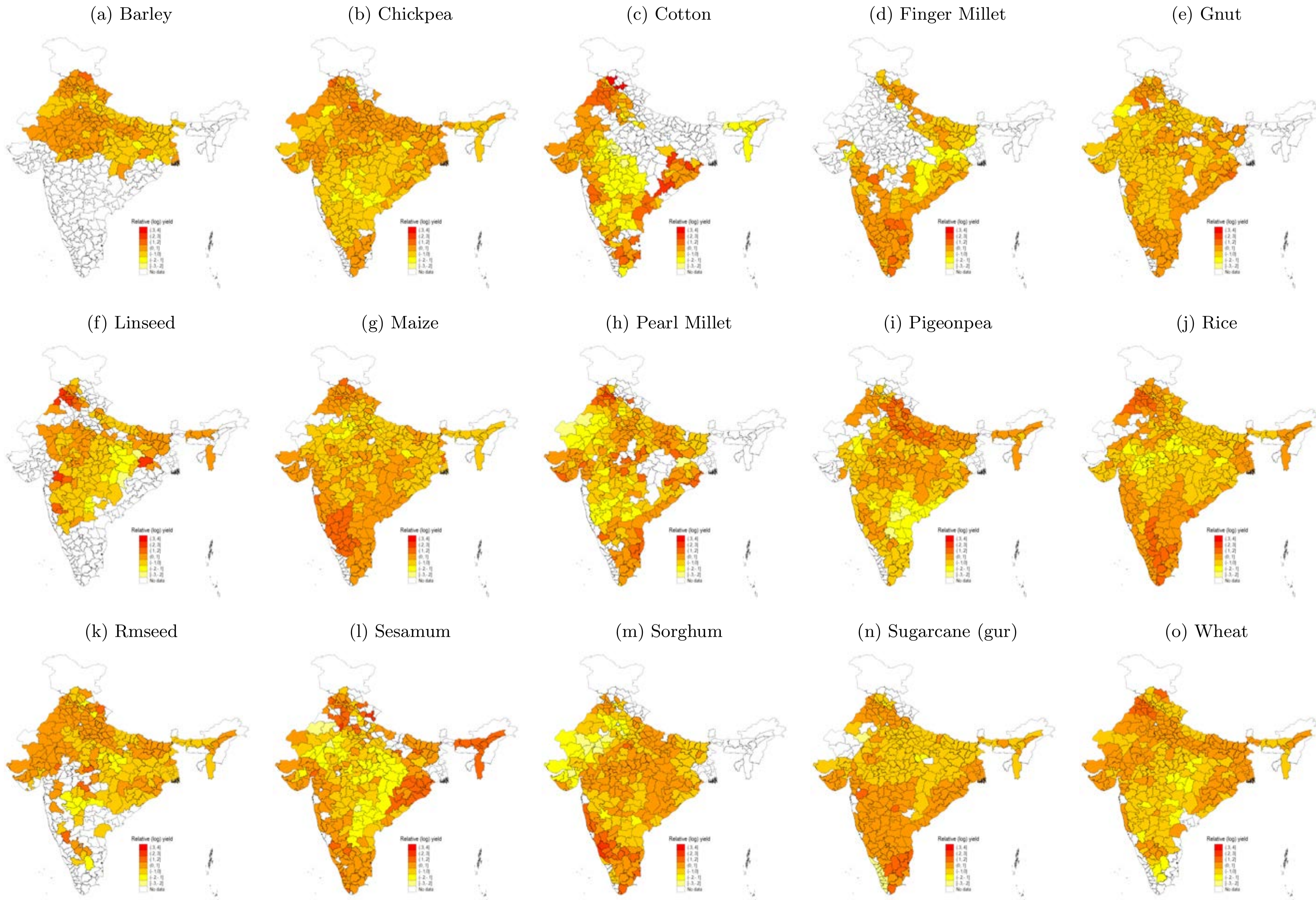
Notes: This psuedo-code describes the algorithm used to calculate the equilibrium crop choice; see Appendix A.5 for details.

Figure A.1: The 1970s crop allocations by district



Notes: This figure depicts the area of land allocated to each of the 15 crops across the 311 districts of India in our database for the 1970s decade. The shares are calculated by ensuring that the average annual allocation within the decade sums to one.

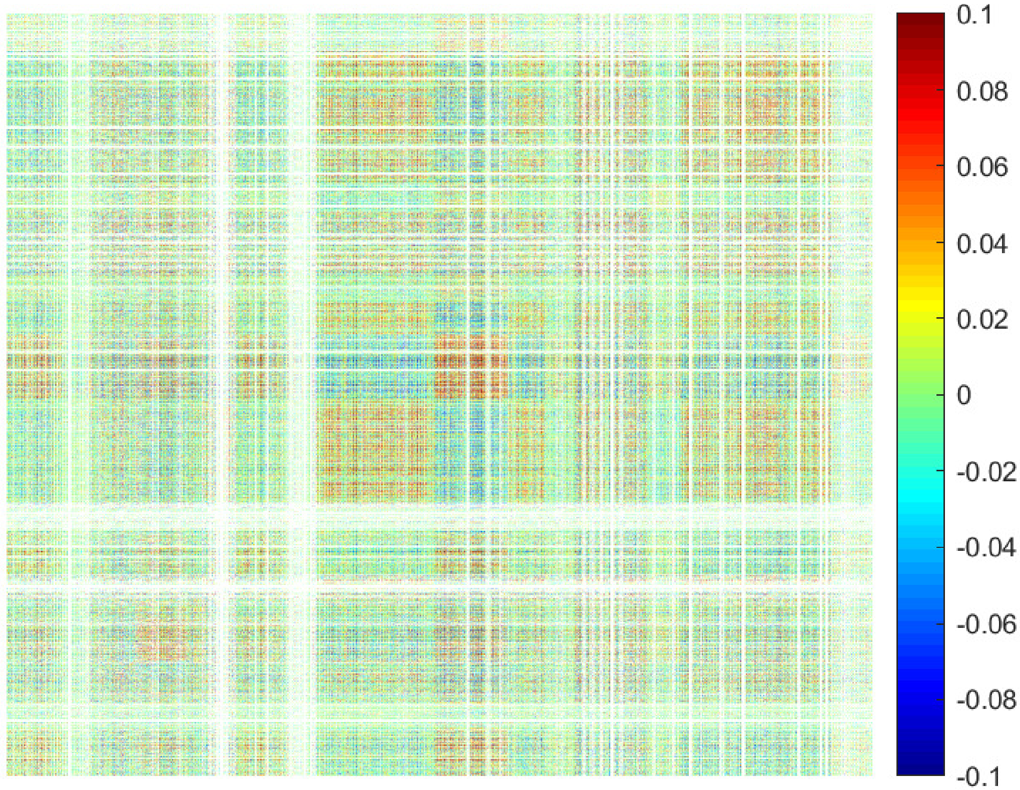
Figure A.2: The 1970s relative mean log yields by district



(a) Barley (b) Chickpea (c) Cotton (d) Finger Millet (e) Gnut (f) Linseed (g) Maize (h) Pearl Millet (i) Pigeonpea (j) Rice (k) Rmseed (l) Sesamum (m) Sorghum (n) Sugarcane (gur) (o) Wheat

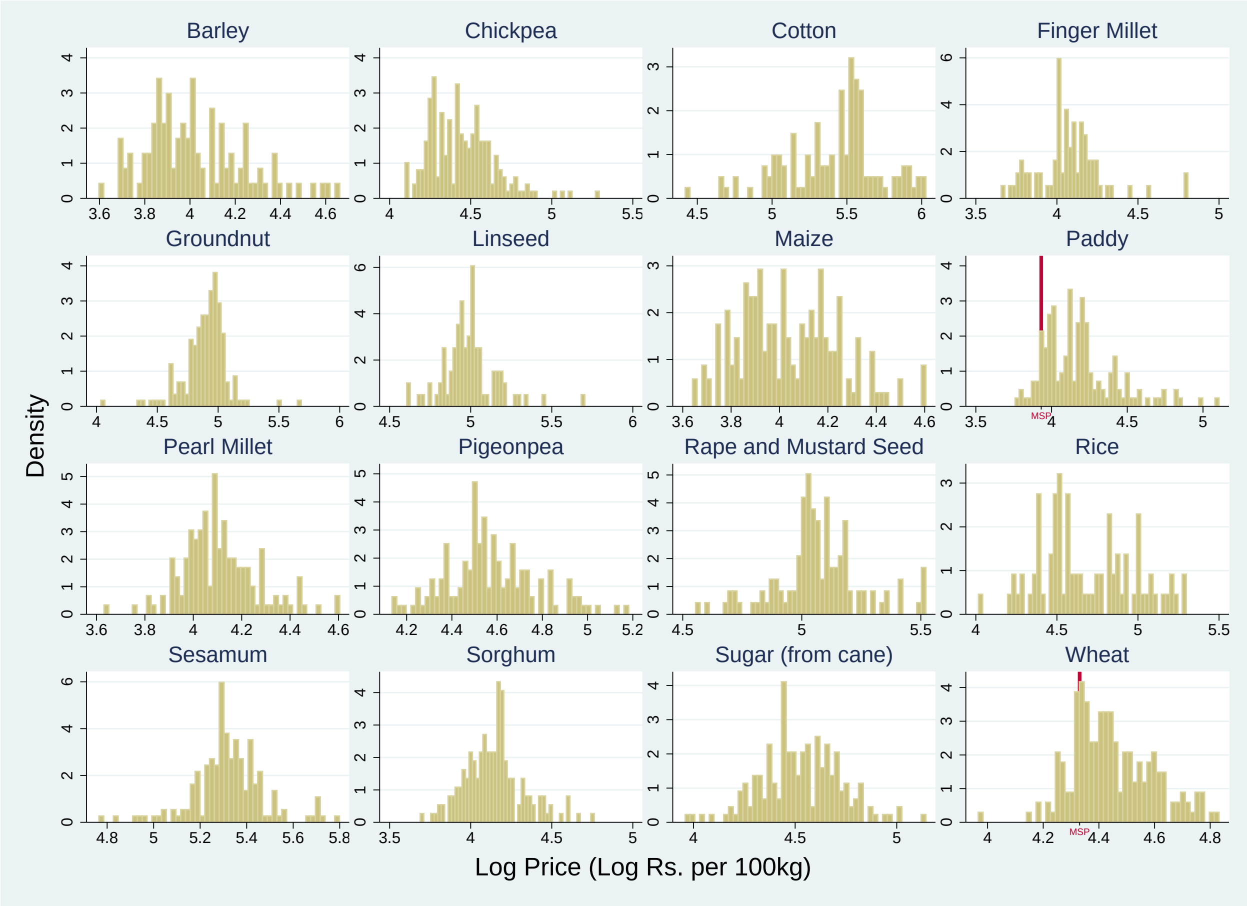
Notes: This figure depicts the mean (log) yields of crops in each of the 311 districts of India relative to the mean (log) yield for all of India for each of the 15 crops in our database and for the 1970s decade. The mean is taken across all years within the decade.

Figure A.3: The 1970s variance-covariance matrix of log yields



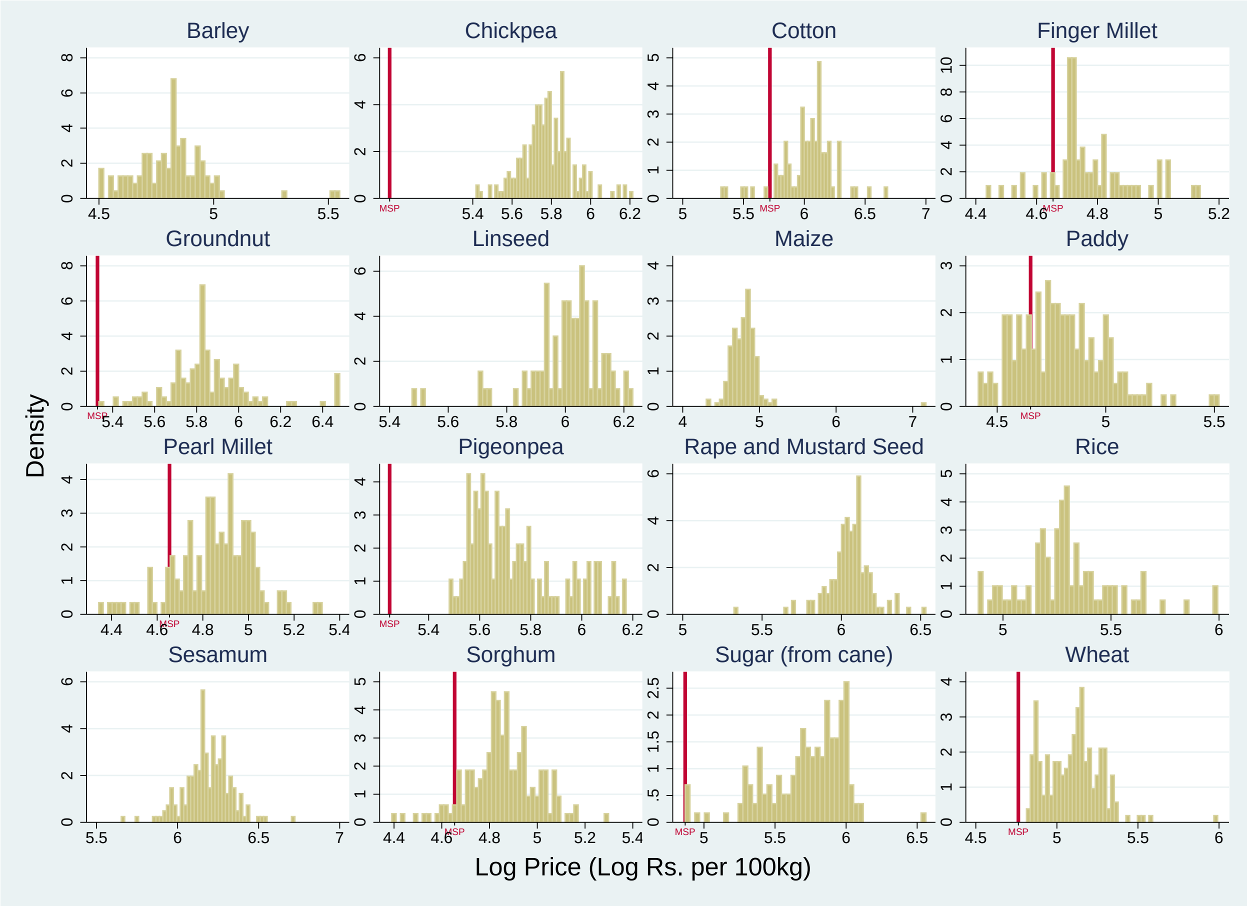
Notes: This figure depicts the variance-covariance matrix of (log) yields between all 311 districts across all 15 crops calculated across all ten years of the 1970s. The covariance is only calculated for crop-district pairs for which yields are observed for both crop-districts all ten years, otherwise it is treated as a zero (and appears in white). In total, there are 8,573,184 covariances calculated. For readability, approximately 5% of the the pixel colors are bottom/top coded at -0.1 and 0.1, respectively.

Figure A.4: Distribution of Prices and MSPs in 1970-71



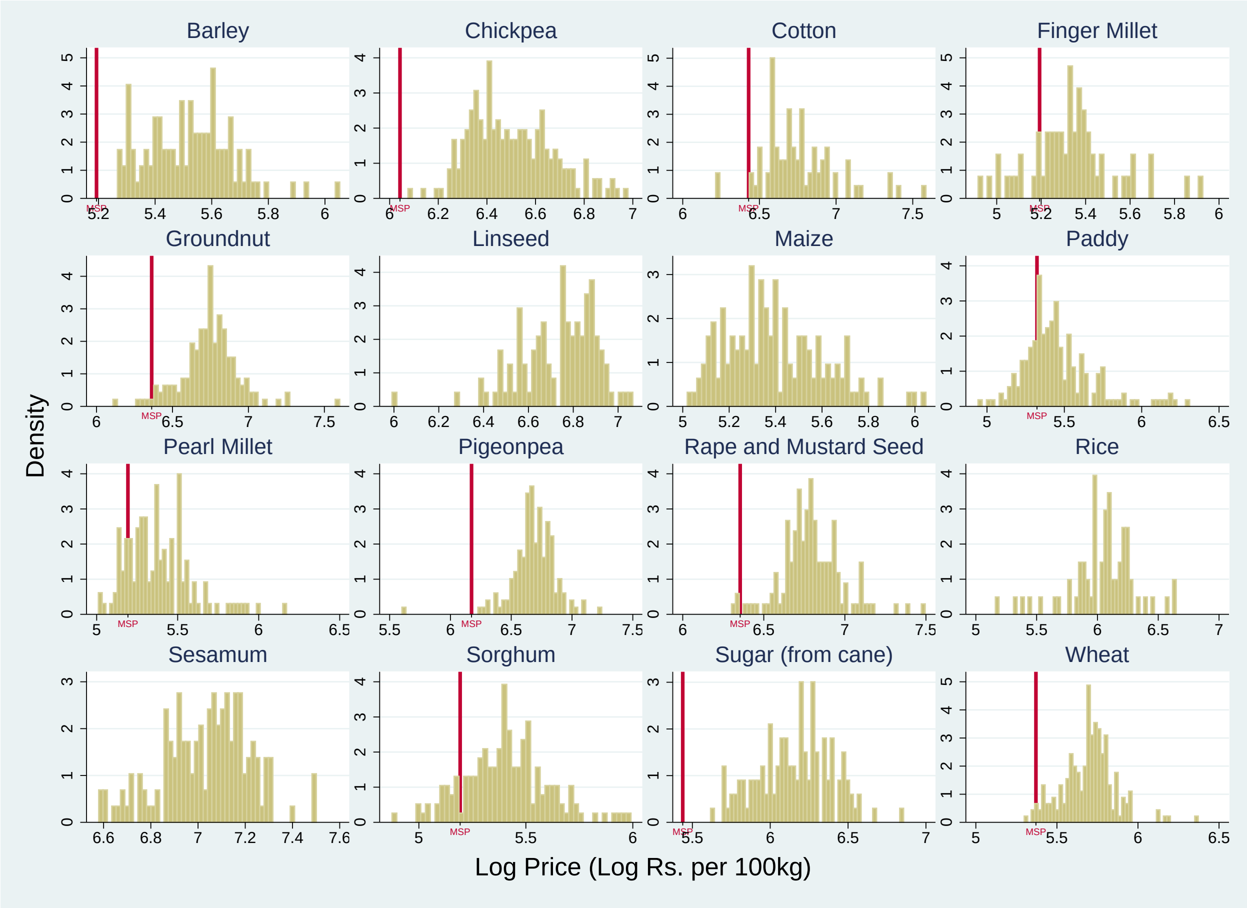
Notes: This figure plots the distribution of log prices across districts for our sample crops in the 1970-71 crop year. Vertical lines show minimum support prices (MSPs) for crops with MSPs in 1970-71.

Figure A.5: Distribution of Prices and MSPs in 1980-81



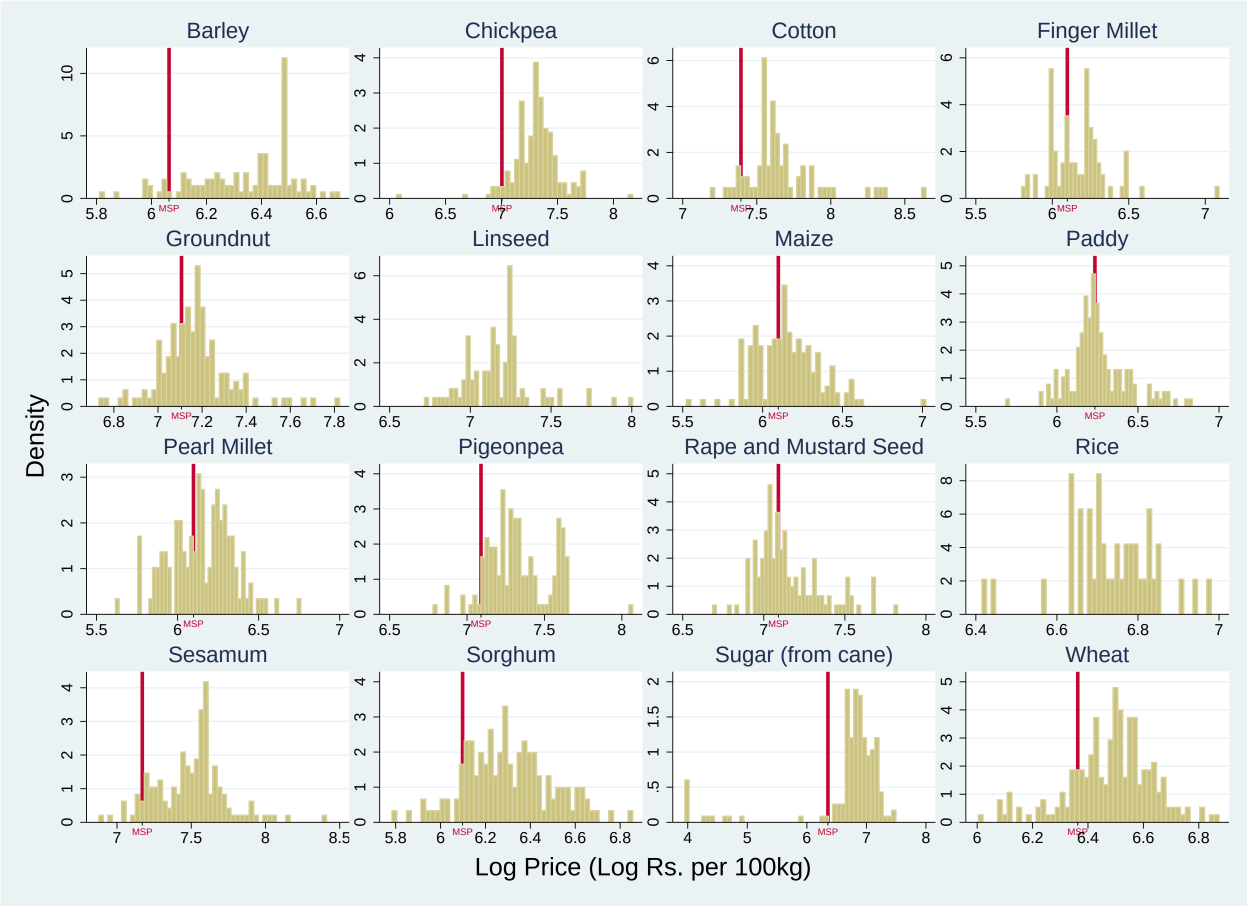
Notes: This figure plots the distribution of log prices across districts for our sample crops in the 1980-81 crop year. Vertical lines show minimum support prices (MSPs) for crops with MSPs in 1980-81.

Figure A.6: Distribution of Prices and MSPs in 1990-91



Notes: This figure plots the distribution of log prices across districts for our sample crops in the 1990-91 crop year. Vertical lines show minimum support prices (MSPs) for crops with MSPs in 1990-91.

Figure A.7: Distribution of Prices and MSPs in 2000-01



Notes: This figure plots the distribution of log prices across districts for our sample crops in the 2000-01 crop year. Vertical lines show minimum support prices (MSPs) for crops with MSPs in 2000-01.