National Institutions and Economic Insecurity*

Raphael Godefroy (Université de Montréal)
Joshua Lewis (Université de Montréal)

February 2018

Abstract

We study the impact of national institutions on local economic activity in Africa. Combining plausibly exogenous discontinuities in institutional arrangements at country borders with detailed geospatial data on a range of agricultural outcomes, we estimate stark differences in rural economic decision-making at country borders. Farmers in countries with worse national institutions engaged in more crop diversification, grew fewer cash crops, and grew more subsistence crops, although we find no impact on agricultural output. These patterns demonstrate how better functioning institutions can reduce reliance on self-insurance practices, supporting longstanding views on the essential link between formal institutions and economic insecurity. The results also show how institutions can shape economic decision-making in subtle ways that may not be captured by standard economic indicators.

*Contact: Raphael Godefroy: raphael.godefroy@umontreal.ca. Joshua Lewis: joshua.lewis@umontreal.ca.

1 Introduction

Economists have long recognized the potential for institutional structures to influence economic insecurity. In seminal work, Douglass North argues that national institutions impact the underlying level of uncertainty in the economy through both property rights enforcement and contract enforcement (North, 1981, 1990). While this view has gained prominence among economists and political scientists, it has never been directly tested.

There are several key challenges to assessing the link between national institutions and economic insecurity. The first stems from a lack of cross-country data on the underlying economic insecurity that agents face.[1] If the quality of national institutions varies systematically with underlying economic risk (e.g. Bloom and Sachs, 1998; Sachs, 2001), empirical assessments of its effects will be biased. Second, the uncertainty associated with poor functioning institutions may influence decision-making through a variety of subtle channels. Insecure property rights may alter individual labor supply and savings decisions (Deaton, 1990; Field, 2007); poor contract enforcement may limit access to financial markets (Glaeser, Johnson, and Schleifer, 2001); and economic uncertainty, more generally, may reduce firm dynamism and investment (Johnson, McMillan, and Woodruff, 2002; Bond, Bloom, and Van Reenen, 2007). An investigation of the effects of national institutions requires that we measure agent 'self insurance' practices systematically across countries, outcomes that are unlikely be captured by standard economic indicators such as GDP per capita. Third, despite a number of innovative research strategies (e.g., La Porta et al., 1997, 1998; Acemoglu, Johnson, and Robinson, 2001, 2002), it is inherently difficult to find credible sources of identification in cross-country analyses (e.g., Glaeser et al., 2004; La Porta et al., 2008; Nunn, 2012).

This paper investigates the impact of national institutions on economic decision-making among farmers in Africa. This context presents an exceptional opportunity to study the re-

[1] Recent work on the effects of uncertainty has typically focused on economic responses to discrete temporal shocks such as the Cuban Missile crisis and 9/11, rather than underlying regional differences in uncertainty (Bloom, 2009; Baker, Bloom, and Davis, 2016).

lationship between institutional quality and economic insecurity. Rural producers face high levels of underlying income risk due to volatile agroclimatic conditions (Carter, 1997). Detailed geospatial data on annual potential yields by crop allow us to measure this underlying agroclimatic risk, independently from decision-making over production inputs, crop choice, and the output prices that they face.[2] We use information on agricultural outcomes from GAEZ to develop several novel measures that capture how agents respond to economic insecurity through self-insurance. These measures include crop diversification (the number of different crops grown on a particular plot of land) and decisions over which crops to cultivate (high return / high risk versus low return / low risk). In addition, information on total agricultural output and output per hectare allow us to explore the extent to which self-insurance behavior affected total production.

Our empirical analysis combines detailed data on agricultural outcomes with differences in national institutions across country borders in Africa, taking advantage of the plausibly exogenous drawing of African borders during colonization. Our identification strategy compares outcomes in adjacent regions that belong to the same historical ethnic homeland but belong to different countries with different national institutions (see Michalopoulos and Papaioannou, 2013).[3] We identify the prevalence of crop diversification and self-insurance practices across farmers, who operated on nearby plots of land that faced similar underlying agroclimatic risk, shared historical ethnocultural farming practices and faced similar market conditions, but were exposed to different national institutions. We validate this approach with information on potential yields across twenty different crops. The results show no relationship between land quality and institutions throughout Africa, consistent with the historiography of country formation.[4]

[2]These data were compiled by the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). These potential yield estimates are constructed regardless of whether the crop is actually cultivated. On the vast majority of plots, potential crop yields follow a stochastic order process for the probability of harvest failure (zero yield), allowing us to rank crops from riskiest to safest without local information on output prices.

[3]This identification strategy was first implemented by Michalopoulos and Papaioannou (2013) to assess how national institutions affect economic development as measured by luminosity.

[4]In contrast, we find systematic differences in land productivity across European borders, consistent with its long history of disputed territory which we discuss in detail in Section 4.1.

The impacts of national institutions are interpreted using a model of agricultural production in which producers choose how to allocate land across a portfolio of crops with different risk profiles.[5] Poor functioning national institutions will increase crop diversification through both income and wealth expropriation, which raises the marginal cost of insecurity.[6] The increased risk of expropriation lowers the return to paid employment, which leads to a relative expansion in agricultural employment and cultivated farmland. Farmer decisions over which crops to cultivate will also be influenced by the quality of national institutions. Although institutional quality has no impact on the decision to cultivate high-return / high risk crops which are profitable everywhere, farmers facing higher expropriation risk will also be more likely to cultivate safer crops with lower average returns.

Our analysis of outcomes across country border reveals a number of empirical patterns consistent with a link between institutional quality and economic insecurity. Farmers in countries with worse formal institutions engaged in more crop diversification. Despite the common underlying agroclimatic risk, we estimate large and statistically significant differences in the number of different crops grown per hectare. The estimated differences are similar across a range of distance intervals from the border, including for farmers operating right at the country boundary, providing further confidence that the results do not reflect differences in land quality, knowledge of farming practices, or market access. We also estimate significant differences in cultivated area across countries. More land was devoted to farming in countries with worse formal institutions, consistent these farmers facing a lower return to paid employment.

To further explore the link between national institutions, agroclimatic risk, and farmer decision-making, we combine annual information on potential crop yields from 1990 to 2000 to develop measures of harvest failure risk for each crop on every plot of land. This measure reflects the probability that the crop yield will be zero in any particular year, and as a result, can be constructed without information on the local output prices that farmers face. Moreover,

[5]Guided by the data on potential crop yields, we model the probability of crop failure as a stochastic order process from the riskiest crop to the safest crop.

[6]This effect will be reinforced by a direct self-insurance channel, if low quality institutions restrict access to insurance and leave producers more vulnerable to output fluctuations.

because each crop responds differently to local agroclimatic conditions, we can order crops from safest to riskiest.

We find that national institutions influenced the choice of which crops to cultivate in ways consistent with self-insurance behavior. Poor national institutions significantly increased the likelihood of cultivating low return / low risk crops, but had no influence on the likelihood of cultivating high return / high risk crops. These patterns are consistent with these farmers having added less profitable but safer crops to their portfolio in an effort to self-insure. Consistent with this evidence, we find that farmers in countries with worse formal institutions devoted more land to cassava, millet, pulses, groundnut, and sorghum; crops that are traditionally valued for their drought-resistant properties (FAO, 1997).

Despite these systematic differences in farmer decision-making, we find no relationship between formal institutions and measures of local economic output. Differences in the quality of national institutions are unrelated to total plot output, output per acre, or agricultural revenue. The lack of systematic differences in agricultural production may partially be attributed to differences in labor inputs across countries and selection across sectors. These findings support previous research that finds no systematic link between institutional quality and local development in Africa (Michalopoulos and Papaioannou, 2013). Taken together, the results demonstrate that standard measures of economic development may not capture the full impact of institutions on economic activity.

We explore a variety of alternative mechanisms that might account for the relationship between national institutional quality and rural economic outcomes. The evidence rules out mechanisms based on cross-border differences in market access, consistent with the porous nature of African borders (Aker et al., 2012). We find no impact of national institution on the cultivation of bananas, coffee, sugar, or other traditional exported-oriented cash crops. Similarly, we find no significant difference in the cultivation of the revenue-maximizing crop across a range of different international agricultural prices. Finally, we estimate systematically *larger* effects of national institutions in regions where underlying land productivity was *lower*.

These findings contrast sharply with the predictions of a trade mechanism, in which the effects should be concentrated in areas with more productive land, where gains to specialization are greatest. Instead, the evidence suggests that self-insurance motivations were particularly acute in regions where underlying productivity was lower.

The observed land-use patterns and widespread cultivation of low yield crops in countries with worse institutions cannot be attributed cannot farmers sorting onto lower quality plots. We find no significant differences in the underlying productivity of cultivated land across borders. Similarly, the results cannot be attributed to differential selection of high productivity workers into non-agricultural sectors (e.g., Lagakos and Waugh, 2013), since those who remained in farming would have been relatively less productive in countries with *better* institutions.

Our results support the viewpoint of Douglass North and a number of economists and political scientists that the political and legal organization of a society are fundamental determinants of economic uncertainty and how agents respond to it. The literature has largely focused on the distortions arising through investment decisions (Banerjee, Gertler, and Ghatak, 2002; Jacobi, Li, and Rozelle, 2002; Goldstein and Udry, 2008; Donovan, 2018).[7] Our results highlight an alternative channel through which poor functioning institutions can influence economic outcomes: agents' willingness to undertake higher return but riskier ventures. These findings offer a new perspective on the low rates of entrepreneurship in many underdeveloped countries (McMillan, and Woodruff, 2002; Audretsch, Keilbach, and Leimann, 2006; World Bank, 2016).

More broadly, our paper contributes to a large body of research that links formal institutional arrangements to economic development. Since John Locke and Adam Smith, scholars have long recognized the potential for institutions to influence economic outcomes. Seminal empirical work by La Porta et al. (1998; 1999) and Acemoglu, Johnson, and Robinson (2001; 2002) demonstrate the impact of institutions arrangements established under European colonization for long-run development. In a recent contribution, Michalopoulos and Papaioannou (2013) exploit within-country variation in light intensity to demonstrate that, on average, differences

[7] Field (2007) emphasizes the impact of insecure property rights through labor supply distortions.

in formal institutions do not translate into differences in economic development in Africa, once differences in geographical and ethnic-specific characteristics are properly taken into account. Rather than focusing on a single measure of local development, we build this literature, using detailed plot-level data on a range of agricultural outcomes, to evaluate the broader effects of institutional arrangement on a range of local economic outcomes. We demonstrate that standard development measures – GDP per capita and light intensity – may fail to capture important changes in local economic activity.

The paper also contributes to the large literature on the role of insecurity in agriculture. In rural areas, where producers often face volatile climatic conditions and output prices (Carter, 1997; Shanahan et al., 2009), the role of institutions in mitigating the effects economic uncertainty may be especially pronounced. Researchers have identified a number of channels through which individual households cope with income uncertainty in the face of credit constraints, including precautionary savings (Deaton, 1990, 1991; Fafchamps et al., 1998), remittances from urban family members (Rapoport and Docquier, 2006; Gonzalez-Velosa, 2012, Yang and Choi, 2007), delayed technological adoption (Antle and Crissman, 1990, Dercon and Christiaensen, 2011, Donovan, 2014), and crop diversification (Kurosaki, and Fafchamps, 2002; Di Falco, and Chavas, 2009, Nicola, 2015). Despite this research, there is ongoing debate of the effectiveness of these strategies (Udry, 2016). The lack of consensus is, in part, a result of the fact that the populations studied differ widely in both the underlying risk and institutional quality. By comparing the behavior of producers in different countries who were exposed to similar climatic risk, we are able to isolate the impact of formal institutions on uncertainty in agriculture. Moreover, detailed data on agricultural outcomes allow us to explore these relationships across 76 border pairs in Africa, allowing us to assess their impact across a broad set of countries.

2 Data

2.1 Potential Yields, Agricultural Output, and Land Use

We obtain information on potential yields, agricultural output, and land use from the Global Agro-Ecological Zones (GAEZ) project, which is organized under the Food and Agriculture Organization's (FAO) and the Institute for Applied Systems Analysis.

The GAEZ project provides estimates of potential yields for twenty different crops in every 5-arc-minute grid cell worldwide (approximately ten kilometers by ten kilometers, measured at the equator), regardless of whether the crop is actually grown. These estimates are obtained by combining information on geographic and climatic conditions at the grid cell level with state-of-the art agronomic models of how each crop will respond to these conditions. Potential crop yields are calculated based on a large number of plot growing characteristics. Invariant characteristics include eight different soil types, elevation, and land gradient. Climatic variables include rainfall, temperature, humidity, wind speed, and sun exposure.[8] These plot characteristics are combined with a vector of hundreds of crop-specific parameters that capture how the output of each crop responds to environmental conditions.[9] Because output on a plot of land depends on farmers decisions over inputs, such as irrigation, fertilizers, machinery, and labor, the GAEZ project calculates different potential productivity estimates depending on the choice of inputs. In our analysis we focus on "high input" potential productivity from "rain-fed" water supply systems, which account for more than 85 percent of agricultural land.[10] Potential yields

[8]The GAEZ project is particularly careful in its treatment of weather conditions. Annual crop productivities are derived based on an aggregation of daily weather conditions, and the model captures how potential yields of each crop are affected by weather conditions throughout the growing cycle.

[9]The parameter estimates are taken from the agronomic literature, and are typically established through field experiments at agricultural research stations.

[10]The GAEZ project defines "high input" technologies as market oriented farm management that is fully mechanized with low labor intensity, adopts fallow and conservation measures, uses high yielding seed varieties, and uses optimum applications of nutrients and chemical pest, disease, and weed control. We also explore the sensitivity of the findings to production under "low input" technologies – traditional farm management that is largely subsistence with labor intensive techniques, and no use of chemicals for pest control and minimal soil conservations measures.

are measured as potential output (in tons) per hectare of land for each of the twenty staple crops in 2000.

The GAEZ project also provides information on cultivated land and actual yields for every crop in 2000, assembled at the 5 arc-minute level from several different sources. First, GAEZ combines several global studies of land cover to categorize land use at the grid cell level.[11] Second, local measures of agricultural crop production are compiled from Monfreda et al. (2008), a comprehensive dataset on output and land use by crop for 150 countries at sub-national political units, and 19,751 political units two levels below the country. Figure 1 presents the map of political units from Monfreda et al. (2008). High resolution statistics were widely available in areas of active agricultural production, whereas the larger political units typically covered regions that were unsuitable for cultivation.[12]

We use these data to construct several measures of local economic activity. To measure agricultural diversification, we construct a measure of the number of number of different crops grown per hectare of cultivated land.[13] To measure the impact of national institutions on farmers' crop choice decisions, we construct a series of crop-specific dummy variables equal to one if a particular crop is cultivated and zero otherwise. Finally, we construct measures of total crop output per hectare cultivated to explore whether national institutions influenced agricultural production.

[11]Land use is categorized as follows: i) rain-fed cultivation, ii) irrigated cultivation, iii) forest, iv) pasture and other vegetated land, v) barren and sparsely vegetated land, vi) water, and vii) urban and land required for housing and infrastructure.

[12]The GAEZ project combines the two data sources with downscaling methods to derive spatial distributions of agricultural activity that are consistent with both the local data from remote sensing and the agricultural production statistics. In the empirical analysis, we explore the sensitivity of the results to excluding countries for which agricultural data is available at higher levels of aggregation.

[13]Because GAEZ does not provide disaggregated information for certain crop categories (e.g., vegetables), our preferred measure relies only on uniquely identifiable crops. We also report results based on an alternate measure of diversification that increases linearly with the number of unique crops and crop groups. The results based on either measure of similar in magnitude.

2.2 National Institutions and Ethnic Partitions

Data on national institutions is from the World Bank's Governance Matters Database (Kauffman, Kraay, and Matruzzi, 2008). These data are based on an aggregate of various institutional quality measures that the World Bank categorizes based on principal components methods. Our main analysis is based on measures of 'rule of law', which ranges from -2.5 to +2.5. This variable reflects institutional factors, such as the quality of the judiciary and the level of property rights enforcement, that have been found to be particularly relevant for land development (Alston, Libecap, and Schneider, 1996; De Soto, 2000), and farmers' investment decisions (Goldstein and Udry, 2008; Banerjee, Gertler, and Ghatak, 2002; Jacoby, Li, and Rozelle, 2002). To limit concerns of reverse causality, we rely on this measure in 1996, the first period in which it was recorded.[14]

We also assemble information on historical ethnic homelands from Murdock's (1959) Tribal Map of Africa. Drawing on a variety of anthropological sources, Murdock (1959) identifies the spatial distribution of more than 500 ethnicities in Africa in the mid- to late nineteenth century. Combining these data with contemporary African country borders, we are able to identify 219 ethnic homelands that were partitioned into two different countries, with positive potential capacity. By comparing the role of impact of national institutions across regions that share a common historical ethnic homeland, we are able to control for persistent cultural factors that might continue to influence economic activity today.

2.3 Sample selection and Main Outcomes of Interest

Our empirical strategy examines how different formal institutions affect economic activity. Detailed data allow us to compare economic outcomes across plots of land that fall within a common historical ethnic boundary, but that lie in adjacent countries with different formal institutions. We assign each plot to one of 219 ethnic territories, linking to plot centroid

[14]We also explore the sensitivity of the findings to alternate years and different measures of institutional quality.

to information from Murdock (1959), which identifies the pre-colonial spatial distribution of ethnicities across Africa. Each plot is then linked to national institutions based on the country in which its centroid falls. To identify the effects of national institutions, holding constant geographic and climatic conditions, our main sample is restricted to plots located within 100 km of the country boundary.[15] We also drop plots located with 10 km of the national border to avoid measurement error in the assignment of institutions among plots spanning multiple countries. Finally, we omit approximately 30 percent of sample, plots primarily located in the Sahara, that had zero agricultural productivity for all crops. The resulting sample consists roughly 35,000 plots that fall within a common ethnic territory that spans one of 76 country borders.

To assess the impact of national institutions, we focus on rural economic outcomes, which accounted for 70 percent of the sub-Saharan population in 2000. We exploit the detailed GAEZ dataset to construct a range of different measures of rural economic outcomes including the number of different crops cultivated on a plot of land, indicators for whether particular crops were grown, and total agricultural output and output per hectare. This broad set of outcomes allow us to assess both the impact of national institutions on local development and their link to economic uncertainty.

3 Empirical Strategy

3.1 Ethnicity Fixed Effects Estimation Strategy

Our first estimation strategy compares economic activity across plots of land within 100 km of the border between neighbouring countries that fall within a common ethnic homeland. We estimate the following regression model:

[15] In the analysis, we explore heterogeneity in the effects at different distance intervals within and beyond the 100 km cutoff.

$$Y_{iebc} = \alpha_0 + \beta Rule\ of\ Law_{bc} + X'_{iebc}\gamma + \lambda_{eb} + \epsilon_{iebc}, \tag{1}$$

where Y_{iebc} denotes outcome on plot i that falls in the historical territory of ethnicity e, near border segment, b, in country c. The term X_{iebc} represents a vector plot-level characteristics including potential land productivity by crop, and distance to the border. The term λ_{eb} denotes an ethnic-border fixed effect, so that our estimates reflect differences in economic outcomes within the same ethnic homeland on either side of a country border. The variable of interest, $Rule\ of\ Law_{bc}$, is a dummy equal to one if institutional quality – the rule of law – in country c is higher than those in the bordering country. The coefficient of interest, β, captures the average difference in outcomes in countries with better and worse institutions across nearby plots that share a common border, have similar land productivity, and fall within the same ethnic homeland.[16]

We also report estimates of the (local) impact of national institutions at the border using a regression discontinuity (RD) approach (see Imbens and Lemieux, 2008; Lee and Lemieux, 2010). We adopt the following parametric specification:

$$Y_{iebc} = \alpha_0 + \delta Rule\ of\ Law_{bc} + F(Dist_{iebc}) + X'_{iebc}\gamma + \lambda_{eb} + \epsilon_{iebc}. \tag{2}$$

In addition to the baseline controls, equation (2) includes an RD-polynomial, $F(Dist_{iebc})$, that controls for a third-order polynomial in distance from each plot centroid to the border, whose coefficients are allowed to differ on either side of the national border. The coefficient of interest, δ, captures the impact of institutions on economic outcomes in the immediate vicinity of the national border.

To further assess treatment heterogeneity, we explore a generalized version of equation (1)

[16] We also explore the sensitivity of the results to alternative estimation strategies. For example, in some specifications, we replace the ethnic-border fixed effects with controls for border and country fixed effects (separately). Identification is based on the fact that certain countries have better institutions relative one neighbor and worse institutions relative to another, allowing us to identify within-country variation in farmers' response to these institutional differences.

based on the following specification:

$$Y_{iebc} = \alpha_0 + \sum_{j=1}^{4} \beta^j Rule\ of\ Law_{bc} \times I(Dist_{iebc} = j) + \sum_{j=1}^{4} \rho^j I(Dist_{iebc} = j) + X'_{iebc}\gamma + \lambda_{eb} + \epsilon_{iebc}, \tag{3}$$

where the term $j \in 1, 2, 3, 4$ identifies four distance bins (10-25km, 25-50km, 50-75km, and 75-100km) from the national border. The variable $I(Dist_{iebc} = j)$ is a dummy equal to one if the plot falls within a particular distance bin. This generalized version of equation (1) allows the impact of national institutions to vary according to producer proximity to the country border. For example, β^3, captures the impact of institutions on economic activity on plots 50 to 75km from the national border.

Each of the three regression models compare relative economic outcomes on plots of land in adjacent countries with different formal institutions. The identifying assumption is that economic activity would have been similar in neighbouring plots that belong to the same ethnic homeland in the absence of differences in institutional quality. For this assumption to be valid, we require that decisions over the siting of country borders were made independently of all factors related to the potential for local development. In the next section, we explore this question in detail.

Two other estimation details are worth noting. First, regressions for agricultural production and number of crops are weighted by plot farmland to estimate the average effect per acre of farmland. Second, standard errors are two-way clustered across both country and historical ethnic boundaries, following the approach of Cameron, Gelbach, and Miller (2011). This multi-way clustering allows for arbitrary spatial correlation along both dimensions.

3.2 Identification and Potential Yields across African Borders

The identifying assumption for the empirical analysis is that national borders were drawn independently of local conditions relevant for current agricultural outcomes. This assumption is supported by a large historical narrative documenting the various factors that shaped the formation of African borders in the late nineteenth century (see Asiwaju 1985; Wesseling 1996; Herbst 2000; Englebert 2009). The historiography indicates that European colonizers drew African borders with little knowledge of local geographic conditions, since at the time of colonization the continent was largely unexplored. Moreover, European powers partitioned land for the creation of colonies and protectorates, rather than future independent countries. These divisions remained even with the arrival of new information about local productivity, and the original borders were almost entirely unchanged by independence.

The historical narrative documenting the arbitrary partitioning of African countries is supported by growing body of empirical research. Michalopoulos and Papaioannou (2013, 2016) document the plausibly exogenous formation in African borders, and show that there are no systematic differences across a range of factors related to economic potential through country borders. The authors exploit this quasi-experimental variation to study the effects of national institutions on local development, as measured by luminosity. Alesina, Easterly, and Matuszeski (2011) similarly demonstrate that eighty percent of African country borders follow longitude and latitude lines, more than any other continent. The arbitrary assignment of African borders has also been used by a number of other researchers as a source of quasi-experimental variation (see Miguel 2004; Bubb 2012; Cogneau and Moradi 2011; Spielvogel 2012).

To further explore the validity of our approach, we use detailed information on potential productivity for twenty different crops reported in the GAEZ dataset. These potential productivity estimates, derived based on a myriad of agroclimatic factors, incorporate all location-specific factors that may influence agricultural production. As a result, any differences in geographic fundamentals that are relevant for agricultural decision-making should be reflected in these measures. We estimate the relationship between institutional quality and potential productivity by crop. This analysis allows us to evaluate whether local agroclimatic factors were systematically related to institutional quality, and thus to assess whether the drawing of national borders was influenced by underlying geographic features relevant for agricultural production.

Table 1 reports the results based on equation (1). We find no systematic relationship between national institutions and potential productivity. The coefficient estimates are all small and statistically insignificant. This broad pattern is robust to a range of alternative specifications. We find similar effects in RD regressions that identify the local treatment effect at the country border (Table A.1). The findings are also robust to alternate functional form assumptions: estimates based on indicators for positive potential crop yields are all small and statistically significant (Table A.2). The results are similar if we restrict the sample to cultivated plots only (Tables A.3 and A.4). The results are also similar when ethnic homeland fixed effects are excluded, suggesting that potential yield varied little across borders regardless of which ethnic groups historically cultivated the land (Tables A.5 for all plots and A.5 for cultivated plots). Results based on equation (4), where we replace the ethnic-border fixed effects with country fixed effects. are qualitatitively similar, regardless of whether estimated for all land plots (Table **??**, or cultivated plots (Table **??**). Together with the historical evidence on country formation, these results provide confidence that African borders were drawn independently from local characteristics, supporting the identifying assumptions.

3.3 Potential Yields across European Borders

Before turning to the main analysis, it is instructive to contrast the results for potential yields across African countries to other regions where borders were not arbitrarily drawn by outside colonizers. We estimate a version of equation (1) across country borders in Europe. Because we lack information on historical ethnic boundaries in these regions, however, these models do not control for ethnic-homeland fixed effects. Given the important role of cultural factors in influencing farming practices (Guiso, Sapienza, and Zingales, 2006), the estimates alone cannot establish the exogeneity of country borders. In particular, if countries were partitioned on the basis of historical ethnic homelands cultural norms might change discontinuously at national borders even if the underlying land productivity were smooth. The results are nevertheless illustrative as to the role of local geography in shaping country border formation.

Table 2 reports the results. We find stark differences in land characteristics near national borders, and countries with higher quality institutions had systematically less productive land. For each of the main staple European crops (wheat, maize, potato), we estimate large and statistically significant differences in potential yields. To understand the sources of these land quality differences, even within a narrow window of country borders, it is useful to describe the process by which modern European borders were establish. The borders of European nation-states have been contested for centuries (Abramson and Carter, 2016). In the centuries following the Wars of Religion, interstate disputes and wars continued, their primary objective being territorial acquisition with little regard for the cultural and ethnic characteristics of the local population. The rise of nationalism in the nineteenth century led to a shift in attitudes, creating the impetus for state formation on the basis of ethnically homogenous populations (Zacher, 2001; Korman, 1996).[17] This tradition continued into the twentieth century during major episodes of borders changes following both World Wars and the collapse of the Soviet Union.[18] Thus, modern European borders have been heavily influenced by the ethnic characteristics of indigenous populations. In addition, these ethnic groups have historically colocated on land with similar geographic characteristics (Michalopoulos, 2012). As a result, differences in land quality across historical ethnic groups are now reflected across country boundaries which tend to track these ethnic divisions. This process of European border formation differed fundamentally from the African experience during 19th century colonization, where little attention was paid to local characteristics or ethnic composition of the indigenous population. These distinct histories largely explain the differing findings in Tables 1 and 2.

The results on the land characteristics across European country borders also suggest that a more nuanced view of nation formation is required. Previous theoretical work has emphasized that the size of nations is governed by a tradeoff between the benefits of economics of scale and the costs of preference heterogeneity, so that individuals at the border are indifferent between either of the countries (e.g., Alesina and Spolaore, 2003). Instead, our findings are

[17]For example, Anderson (1996) argues that the growth of nationalism led to the "sacrilization of the homelands."

[18]Famously, prior to the 1919 Paris Peace Conference, Woodrow Wilson's "Fourteen Points" emphasized the importance of national self-determination for the redrawing of international boundaries.

more consistent with previous evidence on the negative consequences of ethnic partitioning (e.g., Michalopoulos and Papaioannou, 2015; Dippel, 2015), which led European national borders to form around indigenous ethnic territories. Understanding the link between ethnicity identity, intra- and international conflict, and their joint influence on nation formation is a potentially interesting avenue for future research.

Tables 3 and 4 report the corresponding results for Asia and the Americas. Broadly, the results are consistent with the African experience: we find little evidence of systematic differences in land quality at country borders. These findings support the narrative of country formation in these regions. There was less nation-state conflict, country formation was often influenced by external colonization, and ethnic nationalism played a limited role in national formation. Despite similarities in underlying land characteristics, unobservable ethnocultural factors may still change discontinuously at country borders, which could impact farm decision-making independently from national institutions. Since we are unable to control for these potential ethnic factors in these regions, our primary analysis focuses on discontinuity in rural outcomes in Africa.

4 Results

4.1 National Institutions and Economic Insecurity in Africa

4.1.1 Number of cultivated crops

We next explore the link between national institutions and economic insecurity in Africa. The African context offers at least four main advantages: 1) there are no systematic differences geographic fundamentals relevant for agricultural production, 2) the historiography suggests that decisions over the drawing of national borders were made independently of local geographic conditions and the characteristics of the indigenous ethnic groups, 3) data on historical ethnic homelands allow us to exploit within-ethnicity differences in institutional quality, holding constant historical ethnic cultural factors that might influence economic activity, 4) given its disproportionate share in output, effects on agricultural production are likely to capture nontrivial economic impacts.

Table 5 reports the results for the number of crops grown per hectare for various specifications. We report the results separately for the log number of crop groups (cols. 1-6) and the log individual crops per hectare (cols. 7-12). All models are estimated separately with and without controls for land productivity.[19] Columns 1-2, 7-8 report the results from equation (1) of the average difference across all plots within 100 km of the country border. The results suggest that better national institutions led rural producers to grow fewer crops. Across all specifications, the coefficient estimates are large, negative, and statistically significant. The estimates imply farmers in countries with worse institutions grew 32 to 37 percent more crops per hectare relative to nearby producers in countries with better institutions who faced similar underlying agroclimatic conditions and belonged to the same historical ethnic homeland (cols. 1-2, 7-8).

As a robustness check, since the last estimations use the subsample of cultivated plots only, we run regressions on potential yield similar to those of the previous section on that subsample, and weight each observation by the total size of its cultivated area. Tables A.3 and A.4 of the Appendix show no significant difference in potential yield between plots in countries with better institutions.

In Figure 1, we plot, for all borders, the average log of the number of crops cultivated on each side of a border, where observations for the country with the lower (resp. higher) Rule of Law score are on the left (resp. right). The figure shows a drop in the number of crops grown at the border, which may stem from the aggregation of the diversity of borders and ethnic homelands, or could reflect some systematic heterogeneity in the effects of institutions according to proximity to the border. To examine this point, columns 3-4, 9-10 report the results based on the RD specification from equation (2), and columns 5-6, 11-12 report the

[19]The capacity covariates denote a vector of crop-specific controls for potential yield for all crops.

results from equation (3) that allow the effects to vary based on plot proximity to the country border. Across all specifications, we find large and statistically significant effects of national institutions on the number of crops. In fact, the effects are slightly larger among producers operating in the immediate vicinity of the border, suggesting that cross-country trade did not mitigate the effects of poor institutions.

The impact of national institutions on the number of crops grown is robust to a range of alternative specifications. Table A.7 reports the effects based on models that exclude ethnic-homeland fixed effects. The effects are similar in both sign, significance, and magnitude. Table **??** reports the results models that replace the ethnic-border fixed effects with country fixed effects. The results, based on within-country variation in institutional differences, are similar to baseline findings.

The results show that institutional quality played an important role in farmers' decision-making. Because the estimates are derived based on a comparison of nearby land plots with similar geographic-ecological characteristics, we can rule out that they are driven by differences in underlying climate-related risk faced by farmers or divergence in optimal farm practices, such as crop-rotation. The effects also cannot be attributed to the legacy of historical ethnocultural factors that might have influenced farming decision-making.

The effects appear to reflect the role of national institutional quality on farmer's economic insecurity. In principle, formal institutions might also affect risk-taking through changes in market access (e.g., Allen and Atkin, 2016). In practice, however, farmers operating on either side of the border faced largely similar market conditions, and informal trade across African borders was ubiquitous (Aker et al., 2012). The result that farmers exposed to worse institutions engaged in more crop diversification is consistent with the viewpoint of a number of economists and political scientists that national institutions, such as property rights enforcement and inefficient legal systems are fundamental drivers of economic uncertainty and how agents respond to it.

The relationship between national institutions and local crop choices is consistent with high levels of insecurity among agricultural producers in developing countries, who are subject to recurrent droughts and fluctuations in output prices (Carter, 1997; Shanahan et al., 2009.[20] In countries with poor institutions, rural producers often lack access to credit and insurance markets (Binswanger and Rosenzweig, 1986) and there may be limited government-provided security. As a result, crop diversification, a widespread phenomenon in developing countries (Kurosaki, and Fafchamps, 2002; Di Falco, and Chavas, 2009, Nicola, 2015), may be more prevalent.

4.1.2 Crop choice

To further explore the link between national institutions and rural insecurity, we examine whether there were discontinuities in the types of crops grown at country borders. We estimate versions of equation (1) for the probability that particular crops were grown on a given plot of land. The models are estimated for 12 different crops. Tables **??** and **??** show that there were large differences in the types of crops grown across country borders. Farmers in countries with better national institutions were significantly less likely to grow cassava, millet, pulses, potatoes, groundnut, and cotton. In contrast, there were no significant difference in cultivation of bananas, coffee, sugarcane, olives, palm oil, or sunflower. (Tables **??** and **??** report the results of the specification of equation (2).)

The patterns in Tables **??** and **??** are consistent with national institutions having influenced the decision to engage in subsistence versus market-oriented farming. Agricultural production in Africa can broadly be classified into two broad categories: market oriented farming of cash crops and subsistence farming of basic food crops, although in practice, crops fall along a continuum between these two categories. Some crops are grown primarily on small-scale farms for home consumption. These include cassava, sorghum, millet, groundnut, and potatoes which are attractive to subsistence farmers both for their high calorie content and drought resistant

[20] For example, Kazianga and Udry (1996) find that the standard deviation of rainfall-induced income variation is more than half of average rural household income in Burkina Faso.

properties (McCann, 2005; FAO, 1997).[21] Other crops, including rice, maize, and certain fruits and vegetables, may be grown for home consumption but also yield surpluses that are sold on the market (Achterbosch et al., 2014). Finally, there are a number of cash crops, such as banana, cotton, coffee, sugarcane, and palm oil, that are grown almost exclusively for sale on domestic or foreign markets.[22]

Because our analysis compares plots with similar geographic and climatic conditions, the results do not reflect differences in underlying risk, but rather how producers in countries with different institutions respond to a given level of underlying uncertainty. Institutional quality affected not just the number of crops grown, but the composition of crops. These findings provide further evidence on the impact of national institutions on the decision-making of rural producers. In countries with lower quality institutions, there is less market-oriented agricultural production. The continued reliance on subsistence farming could be a result of the underlying risk facing farmers in these countries. As a result, they may optimally decide to allocate land to safer, albeit less productive crops.

4.1.3 Crop choice

The revenue obtained from a given crop is $p_c q_c(1 - t_c)$, where p_c, q_c, t_c are respectively the price, quantity produced and a discount factor on the production of crop c. These three variables may be uncertain: price variations affect p, agroclimatic risk affect q, and the risk of war or expropriation may affect t. If there was full insurance or no risk aversion, the optimal strategy would be to specialize in growing the crop with the largest expected revenue $E(p_c q_c(1 - t_c))$. Otherwise, it may be optimal to diversify. In this section, we aim to examine whether the risk of expropriation, which is the main risk considered in the literature on development, explains the differences in farming practices across borders. To do so, we use two approaches: the first

[21]Cassava is a major source of calories for nearly half of the African population. Because of its drought-resistant properties, cassava has historically played a key role in famine prevention particularly in Eastern and Southern Africa (FA0, 1997).

[22]References on cash cropping versus subsistence farming in Africa include FA0 (1997); Achterbosch et al. (2014); Braun and Kennedy (1986); Chauvin et al. (2012).

one estimates how likely crops more vulnerable to expropriation are to be cultivated, the second one is a direct test of the effect of agroclimatic risk on cultivation.

For the first approach, we use evidence on the type of crop more likely to be expropriated, focusing on expropriation in times of conflict. A major determinant of conflict risk in agriculture is the ease with which crops can be expropriated for food. For instance, Rockmore (2014) estimates the impact of conflict on crop choice in Northern Uganda. He finds the following crops less likely to be grown in areas prone to conflict: cassava, maize, groundnut, and beans. The argument for cassava is based on its long growing cycle and the importance of consistent weeding, without which can reduce yields by 90%. Similarly, he argues that beans require weeding and were also viewed as very nutritious by raiders. Meanwhile, millet is more likely to be cultivated, primarily because of the difficultly in harvesting. Kibriya et al. (2014) argue that maize, cassava and beans are expropriation prone because they can be readily consumed and are typically grown far from home. Expropriation-resistant crops are small-seeded grains (wheat, barley, oats, rye, millet), which require considerable investment in drying, threshing and winnowing before the crop can be consumed.

In Table 7, we report the estimation of Equation (1) for the three crops that are documented to be more resistant to expropriation. This table indeed shows that some of these crops are indeed more likely to be cultivated when the Rule of Law improves.

To examine further the specific role of expropriation on crop choice, we estimate the effect of conflict on crop choice directly. We measure conflict with an indicator equal to one for a plot if and only if civilian deaths due to conflict occurred in at least one of the plots of the same country, and located near the same border, over the period 1997-1999.[23] Columns 2, 5 and 8 report the results of these estimations, which are consistent with the anecdotal evidence of the studies mentioned before: all coefficients are positive. Columns 3, 6 and 9 show little change in the coefficients of the variables Rule of Law and Conflict when they are both included in the specification.

[23] The results are similar if we extend the period to years before 1997, or if we count violence anywhere in the country, or only on the same ethnic homeland to define the variable Conflict.

These results are consistent with the hypothesis that protection from the risk of expropriation may explain the cultivation of some crops. However, the coefficients are not large, and barely significant. In addition, crops that are more resistant to the risk of expropriation may also protect against other types of risk. To examine this point, we estimate, in Table 6, the estimation of Equation (1) for the four crops that are documented to be more vulnerable to expropriation. Columns 1, 4, 7 and 10, we find that improved rule of law *decreases* the probability that these crops are cultivated. In columns 2, 5, 8 and 11, we find that, indeed, the probability to grow any of these crops decreases when there is conflict, which is consistent with the studies mentioned before. Worse institutions, which we expect to be correlated with more war, actually increase the probability that crops vulnerable to expropriation are cultivated. The effect of institutions thus does not only go through risk of expropriation in time of war.

The second approach is to test the impact of agroclimatic risk on crop choice directly. Such a test is limited by the fact that we do not have information on local prices. To circumvent this limit, we estimate for every plot and every crop the share of years in which the attainable yield of the crop dropped to 0 over the period 1986 to 2000. We then estimate the median share across all plots of a given ethnic homeland and border country-pair. We then rank crops by median share and denote the crop with highest median share "high-risk", and the crop with lowest median share "low-risk". To count only crops that may be worthy enough to grow, we only rank crops that are cultivated on at least one plot of that ethnic homeland and border country-pair. We then estimate the effects of Rule of Law on the crop with the high and low risk.

Table 8 reports the results of these estimations. Columns 1 and 4 show that there significant effect of the Rule of Law on the probability to grow the low-variance crop only. This result is consistent with the hypothesis that better institutions provide some form of insurance against agroclimatic risk, whereas, when institutions are worse, farmers grow some crops with lower risk as safe assets.

In addition, Columns 2, 3, 5 and 6 show that, where the risk of violence is higher, there is no significant increase in the probability to grow a crop with lower agroclimatic risk. These results further suggest that agroclimatic and expropriation risks may be uncorrelated.

To summarize, the estimations of this section show that the risk of expropriation is not the only risk that explains crops choices. In fact, the results do not show evidence that this risk has a large impact on crop choice. Agroclimatic risk seems to be a better predictor of crop choices. This is consistent with the hypothesis that, where institutions are worse, farmers use crops with low variance as safe assets, and cannot insure themselves against that risk any other way.

4.2 National Institutions and Agricultural Output in Africa

Table 9 reports the impacts of national institutions on actual yield in tons by Ha, separately for each crop. These estimates are based on a comparison of economic outcomes across nearby plots of land in adjacent countries with different formal institutions regardless of ethnic origin. Despite the systematic differences in diversification and crop choice across country borders, we find limited effects on agricultural output. Across a range of specifications, the estimates are generally statistically insignificant, suggesting that better functioning institutions did not allow farmers to achieve higher yields.

Why did the stark different in rural decision-making not translate into observed output differences? One possibility is that farmers in countries with worse institutions were able to achieve similar output through increases in labor inputs, through either increased adult work hours or a greater use of family and child labor.[24] This scenario could arise if these farmers had less access to formal insurance mechanisms but also faced a binding subsistence food requirement. To reduce insecurity, they diverted land to less productive crops, to overcome the corresponding loss in output, they need to employ additional labor inputs. The differences may also be a result of cross-sector selection, in which the most productive workers leave agriculture

[24] In contrast, differences in intermediate inputs, such as farm technologies, would be expected to exacerbate output differences across countries (Donovan, 2016).

in countries with better institutions (e.g., Lagakos and Waugh, 2013). We leave it to future work to explore the relative importance of these two channels.

Broadly, the findings support the conclusions of Michalopoulos and Papaioannou (2013), who find no link between institutional quality and local development, as measured by luminosity. Although there were no systematic differences in agricultural production, farmers exposed to worse national institutions engaged in markedly different production practices: they specialized less and grew more crop varieties, they were less likely to engage in market-oriented agricultural, and they were more likely to devote land to subsistence crops. A key implication of these results is that standard measures of economic development, such as agricultural production and luminosity, may fail to capture the role of national institutions on rural wellbeing in less-developed countries. The benefits of higher institutional quality, through their role in reducing risk and mitigating its consequences, may be overlooked by focusing solely on measures of economic development.

4.3 National Institutions and Land Use in Africa

To conclude the empirical analysis we examine the extent to which national institutions affected agricultural activity at the extensive margin. Table 10 reports the effects for an indicator for whether a plot is cultivated, and the total area cultivated. There is a clear link between the quality of national institutions and the likelihood of cultivation. Across the various specifications, better institutions are associated with a 10 to 14 percentage point decrease in the likelihood that a plot is cultivated. To the extent that this decline was driven by the self-selection of the most productive workers out of agriculture (Lagakos and Waugh, 2013), these general equilibrium effects may partially account for the limited impact on agricultural production despite the clear effects of national institutions on crop choices. Conditional on cultivation, however, we estimate modest increases in the total area cultivated, although the effects are not generally statistically significant.

5 Conclusion

This paper studies the impact of national institutions on local economic activity in Africa. Drawing on detailed geospatial data on a range of agricultural outcomes, we compare outcomes across producers who faced similar agroclimatic conditions but were exposed to different formal institutions. We find that national institutions had little impact on agricultural output or productivity. Nevertheless, these results mask substantial differences in economic decision-making of rural producers: farmers exposed to worse institutions were more diversification and engaged in less market-oriented production.

Our findings show the interactions between formal institutions, underlying uncertainty, and economic decision-making. In particular, the efficiency gains from property rights and contract enforcement may be especially large among agricultural producers, who are exposed to high levels of agroclimatic risk. In the absence of formal mechanisms, these producers may engage in a variety of costly self-insuring activities, such as crop diversification and the diversion of land to subsistence agriculture.

Finally, our findings highlight how standard measures of development may fail to capture differences in economic activity at the local level. These limitations may soon be addressed, given newly available sources of geospatial data on a variety of local outcomes. Combining these various measures of economic activity (e.g., luminosity and agricultural land use) with machine-learning techniques may allow researchers to develop better proxies for local development.

References

Adamopoulos, T. (2011). Transportation costs, agricultural productivity, and cross-country income differences. *International Economic Review*, 52(2): 489-521.

Adamopoulos, T. and D. Restuccia. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6): 1667- 1697.

Adamopoulos, T. and D. Restuccia. (2015). Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data. *Mimeo.*

Binswanger, H.P., and J. McIntire. Behavioral and Material Determinants of Production Relations in Land-Abundant Tropical Agriculture. *Economic Development and Cultural Change*, 36(1), 73-99.

Binswanger, H.P., and M.R. Rosenzweig. Behavioral and Material Determinants of Production Relations in Agriculture. *Journal of Development Studies*, 22(3), 503-539.

Bloom, D.E., and J.D. Sachs. (1998). Geography, demography, and economic growth in Africa. *Brookings Papers of Economic Activity*, 2, 207-273.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of Economic Growth, Volume 1*, P. Aghion and S. Durlauf (eds.), Elsevier, Ch.9, 679-741.

Carter, M.R. (1997). Technology, and the Social Articulation of Risk in West African Agriculture. *Economic Development and Cultural Change*, 45(3), 557-590.

Carter, M.R., and F.J. Zimmerman. (2003). Asset Smoothing, Consumption Smoothing and the Reproduction of Inequality Under Risk and Subsistence Constraints. *Journal of Development Economics*, 71, 233-260.

Carter, M.R., and T.J. Lybbert. Consumption Versus Asset Smoothing: Testing the Implications of Poverty Trap Theory in Burkina Faso. *Journal of Development Economics*, 99, 255-264.

Deaton, A. (1990). On Risk, Insurance and Intra-Village Consumption Smoothing. *Research Program in Development Studies*, Princeton University, N.J.

Deaton, A. (1991). Savings and Liquidity Constraints. *Econometrica*, 59(5), 1221-1248.

Dercon, S. 2004. *Insurance Against Poverty*, WIDER Studies in Development Economics. Oxford University Press, Oxford, U.K.

Di Falco, S., and J.P. Chavas. (2009). On crop biodiversity, risk exposure and food security in the highlands of Ethiopia. *American Journal of Agricultural Economics*, 91, 599-611.

Donovan, K. (2014). Agricultural risk, intermediate inputs, and cross-country productivity differences. *Mimeo.*

Fafchamps, M. 1992. Cash crop production, food price volatility and rural market integration in the third world, *American Journal of Agricultural Economics*, 74, 90-99.

Fafchamps, M., and S. Lund. (2003). Risk-Sharing Networks in Rural Philippines. *Journal of Development Economics*, 72(2), 261-287.

Fafchamps, M., C. Udry, and K. Czukas. (1998). Drought and Saving in West Africa: Are Livestock a Buffer Stock? *Journal of Development Economics*, 55(2), 273-305.

Ingersoll, J.E. (1987). *Theory of Financial Decision Making*. Rowman & Littlefield Publishers.

Kazianga, H. and C. Udry. (2006). Consumption Smoothing? Livestock, Insurance and Drought in Rural Burkina Faso. *Journal of Development Economics*, 79, 413-446.

Lagakos, D. and M. E. Waugh. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2): 948-980.

Lange, S. and M. Reimers. 2014. Livestock as an Imperfect Buffer Stock in Poorly Integrated Markets. *Mimeo*.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 40, 561-572.

Michaloupouls., S., and E. Papaioannou. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, 81(1): 113-152.

Michaloupouls., S., and E. Papaioannou. (2014). National Institutions and Subnational Development in Africa. *Quarterly Journal of Economics*, 129(1): 151-213.

Morduch, J. (1995). Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives*, 9, 103-114.

Park, A. (2006). Risk and Household Grain Management in Developing Countries. *Economic Journal*, 116, 1088-1115.

Paxson, C.H. (1992). Using Weather Variability To Estimate the Response of Savings to Transitory Income in Thailand. *American Economic Review*, 81(1), 15-33.

Restuccia, D., D.T. Yang, and X. Zhu. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2): 234-250.

Rosenzweig, M.R., and K.I. Wolpin. (1993). 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*, 101(2), 223-244.

Sachs, J.D. (2001). Tropical Underdevelopment. *NBER Working Paper #8119*.

Schultz, T. W. (1953). *The Economic Organization of Agriculture*, New York: McGraw Hill.

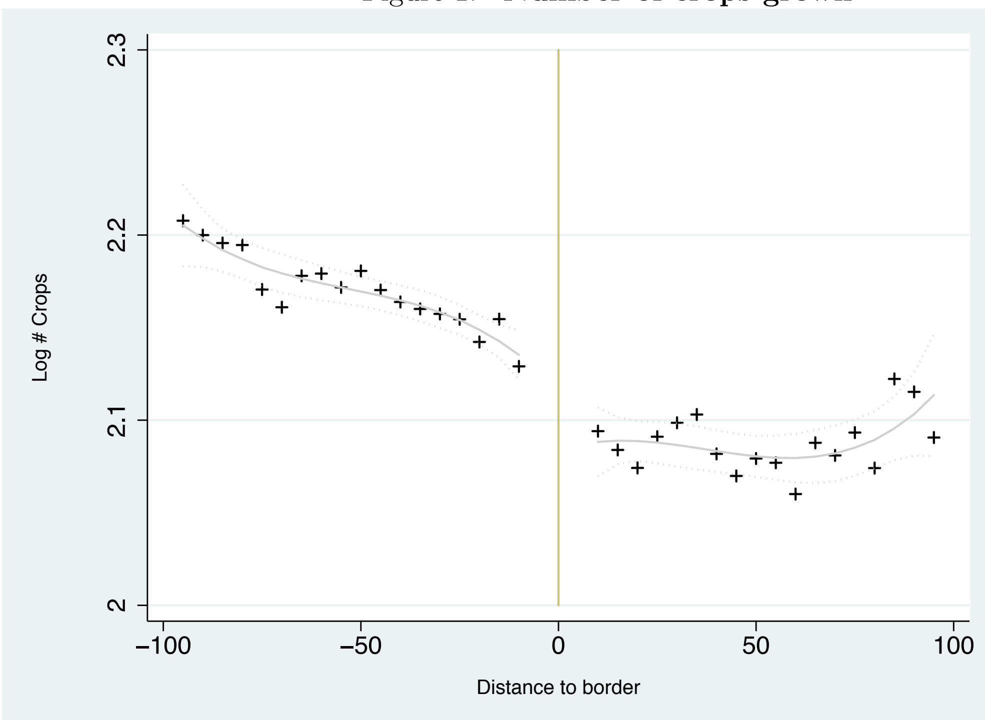
Tombe, T. (2014). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, Forthcoming.

Udry, C. (1995). Risk and Saving in Northern Nigeria. *American Economic Review*, 85(5), 1287-1300.

6 Figures and Tables

7 Figures and Tables

Figure 1: **Number of crops grown**



NOTES. This figure uses the sample of cultivated plots located between 10 and 100 km from a country border. For any border, the left side uses the subsample of those plots that are located in the country with the Rule of Law score smaller than its neighbour's at that border, whereas the right side uses the subsample of those plots that are located in the country with the Rule of Law score larger than its neighbour's at that border. For every bin x, it reports the average of the log of number of crops grown on all cultivated land located between x and x+10 km of the nearest border.

Table 1: **Africa: Potential yield at the border**

	Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.040	0.068	-0.149	-0.219	-0.082	-0.047	-0.248
	(0.075)	(0.121)	(0.200)	(0.174)	(0.118)	(0.070)	(0.421)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.707	0.731	0.680	0.697	0.678	0.702	0.676
Mean Dep. Var.	0.62	1.44	5.17	3.98	2.03	0.52	2.90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.403	-0.072	-0.051	-0.002	-0.044	-0.024	0.004
	(0.416)	(0.057)	(0.088)	(0.052)	(0.072)	(0.074)	(0.009)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.809	0.669	0.692	0.748	0.754	0.635	0.771
Mean Dep. Var.	0.64	1.60	2.17	0.54	1.25	1.87	0.04
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.183	-0.009	0.138	0.478	-0.133	-0.150	-0.078
	(0.417)	(0.017)	(0.458)	(1.490)	(0.295)	(0.192)	(0.092)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.692	0.820	0.807	0.846	0.752	0.900	0.879
Mean Dep. Var.	3.31	0.66	9.43	22.4	1.97	2.75	1.44
Polynomial	No	No	No	No	No	No	No
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	51280	51280	51280	51280	51280	51280	51280
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 2: **Europe: Potential yield at the border**

	Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.830***	-0.017	-0.762***	-0.409**	-0.186	-0.783***	-2.804***
	(0.295)	(0.017)	(0.256)	(0.180)	(0.135)	(0.287)	(0.981)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.463	0.141	0.547	0.672	0.555	0.460	0.457
Mean Dep. Var.	6.41	0.01	3.64	2.93	1.05	6.28	17.2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-3.631**	-0.287***	-0.324***	-0.342***	-0.360***	0.003	-0.156
	(1.391)	(0.088)	(0.110)	(0.112)	(0.115)	(0.003)	(0.110)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.500	0.490	0.543	0.496	0.492	0.386	0.379
Mean Dep. Var.	25.0	1.89	1.52	2.55	2.30	0.01	0.45
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-3.558***	0.001					
	(1.206)	(0.001)					
Crop	Cabbage	Cotton	.	.	.	.	.
R2	0.447	0.394	.	.	.	.	.
Mean Dep. Var.	25.7	0.00	.	.	.	.	.
Polynomial	No	No	No	No	No	No	No
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	32580	32580	32580	32580	32580	32580	32580
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. There are too few observations to estimate regressions for the follosing crops: cassava, sugarcane, palm oil, banana, and coffee. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and border, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 3: **Asia: Potential yield at the border**

	Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.055	0.055	0.301	0.121	-0.016	-0.033	-0.180
	(0.126)	(0.113)	(0.238)	(0.196)	(0.061)	(0.123)	(0.389)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.382	0.552	0.521	0.452	0.324	0.393	0.335
Mean Dep. Var.	0.76	0.46	1.12	1.02	0.26	0.75	2.27
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.399	0.033	0.039	-0.014	0.021	0.034	-0.039
	(0.393)	(0.064)	(0.080)	(0.054)	(0.062)	(0.078)	(0.088)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.288	0.518	0.527	0.389	0.406	0.471	0.414
Mean Dep. Var.	2.01	0.35	0.44	0.32	0.31	0.47	0.18
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.195	0.004	0.142	0.185	-0.218	-0.116	-0.115*
	(0.525)	(0.016)	(0.355)	(1.083)	(0.137)	(0.162)	(0.061)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.347	0.635	0.567	0.469	0.421	0.431	0.444
Mean Dep. Var.	3.16	0.20	1.94	6.12	0.38	0.78	0.24
Polynomial	No	No	No	No	No	No	No
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	76142	76142	76142	76142	76142	76142	76142
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and border, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 4: **America: Potential yield at the border**

	Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.213	-0.097	-0.262	-0.278	-0.134	-0.194	-0.452
	(0.208)	(0.136)	(0.246)	(0.254)	(0.165)	(0.197)	(0.859)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.503	0.531	0.548	0.517	0.460	0.494	0.553
Mean Dep. Var.	1.42	1.67	3.19	1.76	0.82	1.37	5.02
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-1.348	-0.091	-0.141	-0.118	-0.077	-0.036	0.058
	(1.184)	(0.084)	(0.134)	(0.107)	(0.112)	(0.073)	(0.044)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.554	0.448	0.484	0.534	0.473	0.642	0.595
Mean Dep. Var.	5.75	1.25	1.48	0.72	0.85	1.20	0.20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.841	0.009	0.078	0.879	0.443	0.234	0.106
	(0.866)	(0.013)	(0.482)	(1.734)	(0.338)	(0.343)	(0.100)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.534	0.779	0.710	0.686	0.763	0.720	0.731
Mean Dep. Var.	6.78	0.28	8.48	29.6	6.03	5.97	1.63
Polynomial	No	No	No	No	No	No	No
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	65373	65373	65373	65373	65373	65373	65373
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and border, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 5: **Number of crops grown**

	Ln # crop groups by Ha						*Ln # single crops by Ha*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Rule of Law	-0.363**	-0.377**	-0.464*	-0.446*			-0.321**	-0.324**	-0.466***	-0.455***		
	(0.171)	(0.172)	(0.242)	(0.239)			(0.147)	(0.147)	(0.170)	(0.167)		
Rule of Law, 25 km					-0.382*	-0.388**					-0.338**	-0.335**
					(0.193)	(0.192)					(0.156)	(0.155)
Rule of Law, 50 km					-0.354*	-0.373**					-0.313*	-0.319*
					(0.184)	(0.186)					(0.166)	(0.166)
Rule of Law, 75 km					-0.336**	-0.361**					-0.312**	-0.324**
					(0.150)	(0.150)					(0.149)	(0.147)
Rule of Law, 100 km					-0.298**	-0.325**					-0.267*	-0.275**
					(0.134)	(0.127)					(0.146)	(0.134)
50 km					0.037	0.043					0.019	0.027
					(0.052)	(0.050)					(0.037)	(0.036)
75 km					0.045	0.049					0.016	0.022
					(0.064)	(0.058)					(0.053)	(0.050)
100 km					0.041	0.032					-0.003	-0.002
					(0.103)	(0.086)					(0.102)	(0.086)
Polynomial	No	No	Yes	Yes	No	No	No	No	Yes	Yes	No	No
Potential yields covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	1	1	1	1	1	1	1	1	1	1	1	1
Mean Dep. Var.	3.26	3.26	3.26	3.26	3.26	3.26	3.15	3.15	3.15	3.15	3.15	3.15
# Observations	34925	34925	34925	34925	34925	34925	33970	33970	33970	33970	33970	33970
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots

NOTES. This table reports the effects of *Rule of Law* on the number of crops per hectare. Cols. 1-2, 7-8 report the results from equation (1), cols. 3-4, 9-10 report the results from equation (2), and cols. 5-6, 11-12 report the results from equation (3). The dependent variable is log number of groups per hectare (cols. 1-6), and log single crops per hectare (cols. 7-12). All regression models are weighted by total area cultivated, and from col. 7 to 12, we restrict to area cultivated by single crops. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. In addition, even-numbered columns control for the potential productivity by crop, for each 21 crop categories. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 6: **Probability of cultivating a crop vulnerable to expropriation - all plots**

	Crop grown											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Rule of Law	-0.105*		-0.088**	-0.104*		-0.093*	-0.123**		-0.114***	-0.153**		-0.131***
	(0.056)		(0.043)	(0.061)		(0.051)	(0.049)		(0.041)	(0.058)		(0.046)
Conflict 1996-1999		-0.203	-0.176		-0.137	-0.108		-0.133	-0.099		-0.275**	-0.235**
		(0.133)	(0.112)		(0.155)	(0.134)		(0.122)	(0.102)		(0.128)	(0.095)
Crop	Cassava	Cassava	Cassava	Maize	Maize	Maize	Groundnut	Groundnut	Groundnut	Pulses	Pulses	Pulses
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.486	0.487	0.493	0.405	0.401	0.407	0.414	0.407	0.416	0.443	0.444	0.456
Mean Dep. Var.	0.54	0.54	0.54	0.51	0.51	0.51	0.47	0.47	0.47	0.63	0.63	0.63
# Observations	51280	51280	51280	51280	51280	51280	51280	51280	51280	51280	51280	51280
Weights	No	No	No	No	No	No	No	No	No	No	No	No
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots

NOTES. This table reports the results of equation (1) for Africa, estimated separately for a subset of four crops. The dependent variable is an indicator if the particular crop is cultivated and zero otherwise. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 7: **Probability of cultivating a crop resistant to expropriation - all plots**

	Crop grown								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rule of Law	-0.009		-0.011	-0.089**		-0.097**	-0.047		-0.062
	(0.011)		(0.011)	(0.038)		(0.038)	(0.050)		(0.049)
Conflict 1996-1999		0.019	0.022		0.046	0.076*		0.141**	0.160***
		(0.017)	(0.017)		(0.048)	(0.042)		(0.061)	(0.059)
Crop	Wheat	Wheat	Wheat	Millet	Millet	Millet	Barley	Barley	Barley
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.536	0.536	0.537	0.485	0.480	0.486	0.625	0.629	0.632
Mean Dep. Var.	0.05	0.05	0.05	0.38	0.38	0.38	0.24	0.24	0.24
# Observations	51280	51280	51280	51280	51280	51280	51280	51280	51280
Weights	No	No	No	No	No	No	No	No	No
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots	All plots

NOTES. This table reports the results of equation (1) for Africa, estimated separately for a subset of four crops. The dependent variable is an indicator if the particular crop is cultivated and zero otherwise. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 8: **Probability of growing crop with highest vs. lowest variance - all plots**

	High-risk crop grown			Low-risk crop grown		
	(1)	(2)	(3)	(4)	(5)	(6)
Rule of Law	-0.060		-0.055	-0.153***		-0.141***
	(0.046)		(0.044)	(0.057)		(0.051)
Conflict 1996-1999		-0.069	-0.052		-0.168	-0.125
		(0.081)	(0.074)		(0.112)	(0.087)
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.487	0.485	0.487	0.471	0.460	0.474
Mean Dep. Var.	0.33	0.33	0.33	0.57	0.57	0.57
# Observations	51280	51280	51280	51280	51280	51280
Weights	No	No	No	No	No	No
Sample	All plots	All plots	All plots	All plots	All plots	All plots

NOTES. This table reports the results of equation (1) in Columns 1 and 5 (resp. equation (2) in Columns 2 to 4, and 6 to 8) for Africa. In columns 1 to 4 (resp. 5 to 8), the dependent variable is an indicator if the crop with the highest (resp. lowest) variance in attainable yield is cultivated and zero otherwise. All regression models are unweighted. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 9: **Africa: Actual productivity at the border - cultivated plots**

	Dependent variable: Actual yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.547***	0.028	-0.104	-0.204	-0.259	0.245	-0.832
	(0.154)	(0.195)	(0.185)	(0.172)	(0.163)	(0.164)	(0.845)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.877	0.644	0.784	0.731	0.839	0.866	0.902
Mean Dep. Var.	1.90	1.36	1.26	0.87	0.67	0.95	7.15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law		-0.033	-0.027		-0.455	0.025	-0.827
		(0.035)	(0.174)		(0.272)	(0.058)	(0.178)
Crop	Sugar beet	Pulses	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.000	0.827	0.916	0.000	0.913	0.759	0.819
Mean Dep. Var.	.	0.23	1.16	.	0.86	0.86	0.68
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.169	-0.049	1.418*	3.057	0.680	0.031	-0.166**
	(0.140)	(0.269)	(0.749)	(11.940)	(1.553)	(0.150)	(0.073)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Palm oil	Banana	Coffee
R2	0.797	0.487	0.736	0.911	0.925	0.799	0.941
Mean Dep. Var.	1.36	11.0	8.29	44.1	7.69	0.82	0.50
Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Potential yields covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	31187	17598	27907	8271	4099	8865	10853
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES. This table reports the results of equation (2) for Africa, estimated separately for crop-specific actual yields. The dependent variable is actual yield measured in tons per hectare of land where the crop if cultivated. All regression models are weighted by area cultivated with the crop. Each cell corresponds to a different crop. All models include controls for the interaction of border-country fixed effects. The sample comprises plots located between 10 and 100 km from the country border, where the crop is cultivated. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table 10: **Area cultivated**

	Plot Cultivated						*Log Area Cultivated*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Rule of Law	-0.125**	-0.107**	-0.146**	-0.142**			-0.205	-0.164	0.145	0.140		
	(0.057)	(0.053)	(0.063)	(0.065)			(0.161)	(0.152)	(0.191)	(0.185)		
Rule of Law, 25 km					-0.111**	-0.102**					-0.064	-0.048
					(0.051)	(0.049)					(0.161)	(0.155)
Rule of Law, 50 km					-0.115**	-0.098*					-0.199	-0.160
					(0.056)	(0.052)					(0.175)	(0.168)
Rule of Law, 75 km					-0.136**	-0.110*					-0.352**	-0.289*
					(0.063)	(0.059)					(0.168)	(0.160)
Rule of Law, 100 km					-0.168**	-0.140**					-0.354**	-0.298*
					(0.071)	(0.067)					(0.167)	(0.150)
50 km					0.013	0.008					0.093**	0.076**
					(0.012)	(0.011)					(0.036)	(0.036)
75 km					0.033*	0.022					0.158**	0.128*
					(0.017)	(0.016)					(0.067)	(0.068)
100 km					0.038*	0.028					0.160*	0.136
					(0.019)	(0.019)					(0.087)	(0.090)
Polynomial	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Potential yields covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.429	0.449	0.430	0.449	0.430	0.449	0.511	0.528	0.512	0.528	0.512	0.528
Mean Dep. Var.	0.68	0.68	0.68	0.68	0.68	0.68	-1.1	-1.1	-1.1	-1.1	-1.1	-1.1
# Observations	51280	51280	51280	51280	51280	51280	34925	34925	34925	34925	34925	34925
Weights	No	No	No	No	No	No	No	No	No	No	No	No
Sample	All plots	All plots	All plots	All plots	All plots	All plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots

NOTES. This table reports the effects of *Rule of Law* on the probability of cultivation and area cultivated. Cols. 1-2, 7-8 report the results from equation (1), cols. 3-4, 9-10 report the results from equation (2), and cols. 5-6, 11-12 report the results from equation (3). The dependent variables are an indicator for whether the plot is cultivated (cols. 1-6) and total area cultivated conditional on cultivation (cols. 7-12). All regressions are unweighted. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. In addition, even-numbered columns control for the potential productivity by crop, for each 21 crop categories. The sample comprises all plots located between 10 and 100 km from the country border (cols. 1-6), and cultivated plots between 10 and 100 km from the country border (cols. 7-12). Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

A Appendix. Additional Figures and Tables

A.1 Potential yield on all plots - Robustness checks

Table A.1: **Africa: Potential yield at the border - RD, all plots**

	Dependent variable: Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.041	-0.101	0.109	0.004	-0.023	0.021	0.070
	(0.053)	(0.112)	(0.237)	(0.174)	(0.097)	(0.041)	(0.253)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.708	0.731	0.680	0.697	0.678	0.702	0.676
Mean Dep. Var.	0.62	1.44	5.17	3.98	2.03	0.52	2.90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.150	0.025	-0.033	0.028	-0.074	-0.030	-0.040
	(0.100)	(0.070)	(0.089)	(0.048)	(0.066)	(0.077)	(0.040)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.809	0.669	0.693	0.749	0.756	0.635	0.771
Mean Dep. Var.	0.64	1.60	2.17	0.54	1.25	1.87	0.04
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.161	-0.017	-0.225	-0.862	-0.265	-0.227	0.020
	(0.343)	(0.013)	(0.431)	(1.265)	(0.286)	(0.170)	(0.081)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.693	0.820	0.808	0.846	0.752	0.901	0.879
Mean Dep. Var.	3.31	0.66	9.43	22.4	1.97	2.75	1.44
Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	51280	51280	51280	51280	51280	51280	51280
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (2) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table A.2: **Africa: Positive potential yield at the border - all plots**

	Dependent variable: Potential yield > 0						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.002	0.014	-0.023	-0.032	-0.012	-0.011	-0.006
	(0.014)	(0.011)	(0.021)	(0.026)	(0.039)	(0.017)	(0.015)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.798	0.905	0.667	0.693	0.691	0.778	0.767
Mean Dep. Var.	0.31	0.59	0.89	0.84	0.72	0.30	0.29
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.001	-0.016	-0.004	-0.000	0.007	0.006	0.000
	(0.002)	(0.021)	(0.020)	(0.014)	(0.019)	(0.019)	(0.000)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.906	0.643	0.728	0.828	0.801	0.662	0.817
Mean Dep. Var.	0.03	0.89	0.84	0.33	0.62	0.84	0.00
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.006	-0.021	0.003	0.039	-0.009	0.008	0.009
	(0.015)	(0.025)	(0.011)	(0.027)	(0.020)	(0.017)	(0.023)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.757	0.603	0.839	0.830	0.697	0.892	0.830
Mean Dep. Var.	0.27	0.87	0.67	0.51	0.14	0.27	0.40
Polynomial	No	No	No	No	No	No	No
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	51280	51280	51280	51280	51280	51280	51280
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is an indicator variable equal to 1 if and only if the potential yield is positive. All regression models are unweighted. Each cell corresponds to a different crop. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises all plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

A.2 Potential yield on cultivated plots - Robustness checks

Table A.3: **Africa: Potential yield at the border - cultivated plots**

	Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.145	0.135	-0.366	-0.218	-0.095	-0.141*	-0.846
	(0.088)	(0.084)	(0.351)	(0.270)	(0.117)	(0.084)	(0.521)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.731	0.773	0.648	0.640	0.630	0.736	0.699
Mean Dep. Var.	0.69	1.48	5.67	4.39	2.18	0.59	3.28
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.490	-0.080	-0.090	-0.078	-0.080	0.004	0.000
	(0.507)	(0.089)	(0.128)	(0.056)	(0.066)	(0.103)	(0.030)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.852	0.623	0.644	0.748	0.803	0.611	0.858
Mean Dep. Var.	0.94	1.74	2.34	0.58	1.33	2.00	0.06
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.759	-0.034	0.415	1.121	0.218	0.091	-0.032
	(0.496)	(0.029)	(0.363)	(1.130)	(0.256)	(0.222)	(0.082)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.714	0.783	0.800	0.828	0.798	0.861	0.858
Mean Dep. Var.	3.69	0.71	9.92	23.3	1.88	2.79	1.51
Polynomial	No	No	No	No	No	No	No
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	34925	34925	34925	34925	34925	34925	34925
Sample	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES. This table reports the results of equation (1) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are weighted by total area cultivated. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table A.4: **Africa: Potential yield at the border - RD, cultivated plots**

	Dependent variable: Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.164	-0.143	0.113	-0.000	0.035	0.141	1.034*
	(0.120)	(0.208)	(0.579)	(0.441)	(0.183)	(0.102)	(0.574)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.732	0.773	0.650	0.642	0.631	0.737	0.701
Mean Dep. Var.	0.69	1.48	5.67	4.39	2.18	0.59	3.28
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.240	0.077	-0.013	0.152**	-0.084	-0.044	-0.130
	(0.408)	(0.158)	(0.187)	(0.074)	(0.093)	(0.175)	(0.122)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.853	0.625	0.645	0.749	0.803	0.611	0.859
Mean Dep. Var.	0.94	1.74	2.34	0.58	1.33	2.00	0.06
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	1.170**	-0.035	-0.079	-1.576	-0.211	-0.057	-0.013
	(0.558)	(0.040)	(0.836)	(2.424)	(0.425)	(0.220)	(0.102)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.715	0.785	0.800	0.828	0.798	0.861	0.858
Mean Dep. Var.	3.69	0.71	9.92	23.3	1.88	2.79	1.51
Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border x Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	34925	34925	34925	34925	34925	34925	34925
Sample	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES. This table reports the results of equation (2) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are weighted by total area cultivated. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

A.3 Regressions with border FE only - Robustness checks

Table A.5: **Africa: Potential yield at the border - RD, all plots**

	Dependent variable: Capacity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.031	-0.097	0.014	-0.032	-0.041	0.013	0.074
	(0.037)	(0.088)	(0.190)	(0.146)	(0.075)	(0.031)	(0.198)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.528	0.561	0.580	0.581	0.546	0.521	0.507
Mean Dep. Var.	0.51	1.24	4.51	3.49	1.75	0.43	2.41
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	-0.187	-0.014	-0.051	0.028	-0.096*	-0.095	-0.027
	(0.141)	(0.051)	(0.069)	(0.028)	(0.051)	(0.069)	(0.031)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.632	0.589	0.598	0.580	0.646	0.550	0.150
Mean Dep. Var.	0.47	1.39	1.85	0.44	1.03	1.59	0.04
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.117	-0.027**	-0.562	-1.587	-0.292	-0.285*	-0.052
	(0.224)	(0.012)	(0.383)	(1.095)	(0.184)	(0.145)	(0.063)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.531	0.676	0.696	0.703	0.594	0.768	0.768
Mean Dep. Var.	2.71	0.59	7.96	18.9	1.65	2.32	1.23
Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	89299	89299	89299	89299	89299	89299	89299
# Clusters	81	81	81	81	81	81	81
Sample	All plots	All plots	All plots	All plots	All plots	All plots	All plots
Weights	No	No	No	No	No	No	No

NOTES. This table reports the results of equation (2) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are weighted by total area cultivated. All models include controls for border fixed effects. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and border, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table A.6: **Africa: Potential yield at the border - RD, cultivated plots**

	Dependent variable: Potential yield						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.119	-0.072	0.034	-0.017	-0.066	0.096	0.751
	(0.116)	(0.154)	(0.421)	(0.299)	(0.140)	(0.097)	(0.535)
Crop	Wheat	Rice	Maize	Sorghum	Millet	Barley	Potato
R2	0.581	0.535	0.458	0.459	0.447	0.589	0.549
Mean Dep. Var.	0.64	1.41	5.50	4.28	2.11	0.55	3.05
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.048	0.012	-0.033	0.113	-0.117	-0.050	-0.086
	(0.313)	(0.117)	(0.141)	(0.074)	(0.086)	(0.141)	(0.110)
Crop	Sugar beet	Phaseolus	Soybean	Rapeseed	Sunflower	Groundnut	Olive
R2	0.680	0.459	0.471	0.601	0.705	0.440	0.720
Mean Dep. Var.	0.75	1.68	2.21	0.53	1.19	1.88	0.06
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rule of Law	0.816	-0.024	-0.491	-2.484	-0.173	-0.403	-0.139
	(0.504)	(0.030)	(0.782)	(2.479)	(0.309)	(0.313)	(0.118)
Crop	Cabbage	Cotton	Cassava	Sugar cane	Oilpalm	Banana	Coffee
R2	0.569	0.575	0.651	0.670	0.621	0.672	0.680
Mean Dep. Var.	3.36	0.69	9.20	21.6	1.68	2.55	1.41
Polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	55618	55618	55618	55618	55618	55618	55618
Sample	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES. This table reports the results of equation (2) for Africa, estimated separately for crop-specific potential yields. The dependent variable is potential yield measured in tons per hectare. All regression models are weighted by total area cultivated. All models include controls for the interaction of border-country pair and ethnic homeland fixed effects. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and ethnic homeland, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

Table A.7: **Number of crops grown**

	Ln # crop groups by Ha						Ln # single crops by Ha					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Rule of Law	-0.296*	-0.313**	-0.463*	-0.460*			-0.249	-0.264*	-0.442**	-0.435***		
	(0.155)	(0.155)	(0.241)	(0.234)			(0.157)	(0.151)	(0.171)	(0.160)		
Rule of Law, 25 km					-0.346*	-0.354**					-0.290*	-0.292**
					(0.174)	(0.171)					(0.148)	(0.144)
Rule of Law, 50 km					-0.287*	-0.306*					-0.230	-0.248
					(0.168)	(0.169)					(0.173)	(0.169)
Rule of Law, 75 km					-0.255*	-0.281*					-0.223	-0.251
					(0.148)	(0.147)					(0.174)	(0.165)
Rule of Law, 100 km					-0.205	-0.236*					-0.193	-0.224
					(0.141)	(0.127)					(0.186)	(0.159)
50 km					0.021	0.019					-0.006	0.001
					(0.052)	(0.053)					(0.037)	(0.035)
75 km					0.017	0.012					-0.012	-0.005
					(0.062)	(0.058)					(0.051)	(0.046)
100 km					0.038	0.008					0.026	0.011
					(0.087)	(0.068)					(0.092)	(0.071)
Polynomial	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Capacity covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.544	0.565	0.546	0.566	0.545	0.565	0.530	0.556	0.531	0.556	0.530	0.556
Mean Dep. Var.	3.26	3.26	3.26	3.26	3.26	3.26	3.15	3.15	3.15	3.15	3.15	3.15
# Observations	34925	34925	34925	34925	34925	34925	33970	33970	33970	33970	33970	33970
# Clusters	75	75	75	75	75	75	75	75	75	75	75	75
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots	Cult'd plots

NOTES. This table reports the effects of *Rule of Law* on the number of crops per hectare. Cols. 1-2, 7-8 report the results from equation (1), cols. 3-4, 9-10 report the results from equation (2), and cols. 5-6, 11-12 report the results from equation (3). The dependent variable is log number of groups per hectare (cols. 1-6), and log single crops per hectare (cols. 7-12). All regression models are weighted by total area cultivated, and from col. 7 to 12, we restrict to area cultivated by single crops. All models include controls for border-country pair fixed effects. In addition, even-numbered columns control for the potential productivity by crop, for each 21 crop categories. The sample comprises cultivated plots located between 10 and 100 km from the country border. Standard errors are two-way clustered for both country and border, following the approach of Cameron, Gelbach, and Miller (2011). ***, **, * denote significance at the 1%, 5%, and 10% level.

B Appendix. Model

We consider a farmer's problem in which there are n crops. Our first assumption is that, for any k, l, with $k < l$, the probability of positive yield of crop k *zeroth-order stochastically dominates* the probability of positive yield of crop l, which means that, for any state of the world, crop l has a positive yield if and only if crop k has a positive yield as well. For simplicity, and without any loss of generality given the previous assumption, we assume that there are exactly $n + 1$ states of the world, numbered from 0 to n, and that crop k has a positive yield if and only if the state of the world is in the subset $\{k, k + 1, ..., n\}$. We denotep_k, the probability that state k occurs, with $\sum_{k=0}^{n} p_k = 1$.

Our second assumption is that, the value by unit of land of crop k is $q_k > 0$ in states k to n.

These assumptions reflect the fact that the yield of a crop depends on the rainfall, that crops are unevenly resistant to low rainfall, that more rainfall will weakly increase the yield of all crops. We show in Appendix ?? that these assumptions are consistent with the actual distribution of yields across crops and time.

B.1 Assumptions and notations

We consider an agent's portfolio allocation problem with m assets. The first two assumptions reflect the fact that the yield of a crop depends on the rainfall, that crops are unevenly resistant to low rainfall, and that more rainfall will weakly increase the yield of all assets. We show in Appendix ?? that these assumptions are consistent with the actual distribution of yields across crops and across years.

Assumption 1 For any k, l, with $k < l$, the probability of positive yield of asset k *statewise stochastically dominates* the probability of positive yield of asset l, which means that, for any state of the world, asset l has a positive yield if and only if asset k has a positive yield as well. For simplicity, we assume that there are exactly $m + 1$ states of the world, numbered from 0 to m, and that asset k has a positive yield if and only if the state of the world is in the subset

$\{k, k+1, ..., m\}$. We denotep_k, with $p_k > 0$ if $k > 0$, the probability that state k occurs, with $\sum_{k=0}^{m} p_k = 1$.

Assumption 2 The yield of asset k is $q_k > 0$ in states k to m. To avoid discussions of knife-edge cases, we also assume that all assets have different expected returns.

Assumption 3 Let $w > 0$ denote the agent's wealth. The agent's utility is the log of the sum of her wealth and the return from her portfolio.

Agent's problem The agent's problem is:

$$\underset{(\alpha_k)_{k=1}^m}{\text{Max}} \sum_{j=1}^{n} p_j \log\left(w + \sum_{i=1}^{j} \alpha_i q_i\right)$$
$$\textit{such that } 0 \leq \alpha_i \leq 1, \sum_{i=1}^{m} \alpha_i = 1$$

Notations We denote $V(w, q_1, ..., q_m, p_1, ..., p_m) \equiv \sum_{j=1}^{m} p_j \log\left(w + \sum_{i=1}^{j} \alpha_i^* q_i\right)$ the value of the objective function at the optimal solution of this problem.

We define $r_k \equiv q_k \sum_{i=k}^{m} p_i$, the expected return of asset k, and $w_{j,k} = \frac{\sum_{i=j} k-1 p_j q_j q_k}{r_k - r_j}$ if $1 \leq j < k \leq m$.

To solve the agent's problem, we analyze the interior solution first, and then the general solution.

B.2 Interior solution

Lemma 1. *For any $n, w,$ and $(q_k, p_k)_{k=1}^n$ such that there exists an interior solution to the agent's problem, V can be written as the sum of two terms $V(w, q_1, ..., q_n, p_1, ..., p_n) = \sum_{j=1}^{n} p_j \log(w + q_1) + G(q_1, ..., q_n, p_1, ..., p_n)$, where G is a function of $(q_1, ..., q_n, p_1, ..., p_n)$ and not of w.*

Proof. The lemma is trivially true if $n = 1$, in which case $V(w, q_1, p_1) = p_1 \log(w + q_1)$. Suppose $n = 2$. We have, assuming an interior solution exists:

$$\alpha_1^* = \frac{(p_1+p_2)wq_1 + p_1q_1q_2 - p_2q_2w}{(p_1+p_2)q_1(q_2-q_1)},$$

and hence,

$$V(w, q_1, q_2, p_1, p_2) = (p_1+p_2)\log(w+q_1) + G(q_1, q_2, p_1, p_2),$$

where G is a function of (q_1, q_2, p_1, p_2), not of w.

Suppose by induction that, for any interior solution with $n-1, w$, and $(q_k, p_k)_{k=1}^{n-1}$ for which there exists an interior solution to the agent's problem,

$$V(w, q_1, ..., q_{n-1}, p_1, ..., p_{n-1}) = \sum_{j=1}^{n-1} p_j \log(w+q_1) + G(q_1, ..., q_{n-1}, p_1, ..., p_{n-1}),$$

where G is a function of $(q_1, ..., q_{n-1}, p_1, ..., p_{n-1})$ and not of w.

The agent's problem for n assets,

$$\underset{(\alpha_k)_{k=1}^n}{\text{Max}} \quad \sum_{j=1}^{n} p_j \log\left(w + \sum_{i=1}^{j} \alpha_i q_i\right)$$
$$\textit{such that } 0 \leq \alpha_i \leq 1, \sum_{i=1}^{n} \alpha_i = 1$$

is equivalent to the following problem:

$$\begin{aligned}&\underset{\alpha,(\beta_k)_{k=2}^n}{\text{Max}} \quad p_1 \log(w+\alpha q_1) + \sum_{j=2}^n p_j \log\left(w+\alpha q_1 + (1-\alpha)\sum_{i=2}^j \beta_i q_i\right)\\ &\textit{such that } 0 \le \alpha \le 1, \sum_{i=2}^n \beta_i = 1\end{aligned}$$

We are interested in the optimal α of the last problem, which is the solution of the following problem:

$$\begin{aligned}&\underset{\alpha}{\text{Max}} \quad p_1 \log(w+\alpha q_1) + \sum_{j=2}^n p_j \log(1-\alpha) + V\left(\frac{w+\alpha q_1}{1-\alpha}, q_2, ..., q_n, p_2, ..., p_n\right)\\ &\textit{such that } 0 \le \alpha \le 1\end{aligned}$$

If this problem has an interior solution, then, necessarily, $0 < \alpha^* < 1$ and the agent's problem with $n-1$ assets and parameters $\frac{w+\alpha^* q_1}{1-\alpha}$ and $(q_k, p_k)_{k=2}^n$ has an interior solution. In addition, the induction hypothesis implies that there exists a function G such that, where defined,

$$V(\frac{w+\alpha q_1}{1-\alpha}, q_2, ..., q_n, p_2, ..., p_n) = \sum_{j=2}^n p_j \log(\frac{w+\alpha q_1}{1-\alpha} + q_2) + G(q_1, ..., q_n, p_1, ..., p_n),$$

so the last problem is equivalent to the following:

$$\begin{aligned}&\underset{\alpha}{\text{Max}} \quad p_1 \log(w+\alpha q_1) + \sum_{j=2}^n p_j \log\left(w+\alpha q_1 + (1-\alpha) q_2\right) + G(q_1, ..., q_n, p_1, ..., p_n)\\ &\textit{such that } 0 \le \alpha \le 1\end{aligned}$$

From the solution above, we thus have that

$$V(w, q_1, q_2, ..., q_n, p_1, p_2, ..., p_n) = \sum_{j=1}^{n} p_j \log(w + q_1) + G(q_1, ..., q_n, p_1, ..., p_n),$$

where G is a function of $(q_1, ..., q_n, p_1, ..., p_n)$, not of w, which proves the induction hypothesis for n, and the lemma. $\square$

Lemma 2. *The agent's problem can be rewritten as follows:*

$$\underset{(\beta_k)_{k=1}^n}{\text{Max}} \quad \sum_{j=1}^{n} p_j \log\left(w + \sum_{i=1}^{j} \beta_i q_i \Pi_{t=1}^{j-1}(1-\beta_t)\right)$$
$$\textit{such that } 0 \leq \beta_i \leq 1, \beta_n = 1$$

For any interior solution, we have with the notations defined above, for any $k = 1...n-1$, $\beta_k^* = \frac{(r_{k+1}-r_k)(w_{k,k+1}-w_{k-1,k})}{r_k(q_{k+1}-q_k)}$, *where* $w_{0,1} \equiv w$.

Proof. Necessarily, the optimal β_1 in the agent's problem must solve:

$$\underset{\beta_1}{\text{Max}} \quad p_1 \log(w_{0,1} + \beta_1 q_1) + \sum_{j=2}^{n} p_j \log(1-\beta_1) + V(\frac{w_{0,1} + \beta_1 q_1}{1-\beta_1}, q_2, ..., q_n, p_2, ..., p_n)$$
$$\textit{such that } 0 \leq \beta_1 \leq 1$$

where, by definition,

$$V(w, q_2, ..., q_n, p_2, ..., p_n) \equiv \underset{(\beta_k)_{k=2}^n}{\text{Max}} \quad \sum_{j=2}^{n} p_j \log\left(w + \sum_{i=2}^{j} \beta_i q_i \Pi_{t=2}^{j-1}(1-\beta_t)\right)$$
$$\textit{such that} \quad 0 \leq \beta_i \leq 1, \beta_n = 1$$

Suppose, by induction, that necessarily, for some $k \leq n-2$, the optimal β_k in the agent's problem must solve:

$$\begin{aligned}
\underset{\beta_k}{\text{Max}} \quad & p_1 \log(w_{0,1} + \beta_1^* q_1) + \sum_{j=2}^{n} p_j \log(1 - \beta_1^*) + ... \\
& + p_{k-1} \log(w_{k-2,k-1} + \beta_{k-1}^* q_{k-1}) + \sum_{j=k}^{n} p_j \log(1 - \beta_{k-1}^*) \\
& + p_k \log(w_{k-1,k} + \beta_k q_k) + \sum_{j=k+1}^{n} p_j \log(1 - \beta_k) \\
& + V(\frac{w_{k-1,k} + \beta_k q_k}{1 - \beta_k}, q_{k+1}, ..., q_n, p_{k+1}, ..., p_n) \\
\textit{such that } & 0 \leq \beta_k \leq 1
\end{aligned}$$

Then, using Lemma 1 to rewrite the last term of the objective function, this problem becomes:

$$\begin{aligned}
\underset{\beta_k}{\text{Max}} \quad & p_1 \log(w_{0,1} + \beta_1^* q_1) + \sum_{j=2}^{n} p_j \log(1 - \beta_1^*) + ... \\
& + p_{k-1} \log(w_{k-2,k-1} + \beta_{k-1}^* q_{k-1}) + \sum_{j=k}^{n} p_j \log(1 - \beta_{k-1}^*) \\
& + p_k \log(w_{k-1,k} + \beta_k q_k) + \sum_{j=k+1}^{n} p_j \log(1 - \beta_k) \\
& + \sum_{j=k}^{n} p_j \log\left(\frac{w_{k-1,k} + \beta_k q_k}{1 - \beta_k} + q_{k+1}\right) + G(q_{k+1}, ..., q_n, p_{k+1}, ..., p_n) \\
\textit{such that } & 0 \leq \beta_k \leq 1
\end{aligned}$$

where G does not depend on w_k or β_k.

The first order condition implies that:

$$\beta_k^* = \frac{(r_{k+1} - r_k)(w_{k,k+1} - w_{k-1,k})}{r_k(q_{k+1} - q_k)}$$

In addition, $\frac{w_{k-1,k}+\beta_k^* q_k}{1-\beta_k^*} = w_{k,k+1}$, so that, necessarily, the optimal β_{k+1} in the agent's problem must solve:

$$\begin{aligned}
\underset{\beta_{k+1}}{\text{Max}} \quad & p_1 \log(w_{0,1} + \beta_1^* q_1) + \sum_{j=2}^{n} p_j \log(1 - \beta_1^*) + ... \\
& + p_k \log(w_{k-1,k} + \beta_k^* q_k) + \sum_{j=k+1}^{n} p_j \log(1 - \beta_k^*) \\
& + p_{k+1} \log(w_{k,k+1} + \beta_{k+1} q_{k+1}) + \sum_{j=k+2}^{n} p_j \log(1 - \beta_{k+1}) \\
& + V(\frac{w_{k,k+1} + \beta_{k+1} q_{k+1}}{1 - \beta_{k+1}}, q_{k+2}, ..., q_n, p_{k+2}, ..., p_n) \\
\textit{such that } & 0 \leq \beta_{k+1} \leq 1
\end{aligned}$$

which proves the induction hypothesis. As can be seen, the induction hypothesis implies that, necessarily:

$$\beta_k^* = \frac{(r_{k+1} - r_k)(w_{k,k+1} - w_{k-1,k})}{r_k(q_{k+1} - q_k)},$$

which proves the lemma. □

Lemma 3. *(1) If the shares of all assets are positive at the optimum, necessarily $r_k < r_{k+1}$, which implies $q_k < q_{k+1}$, for any $1 \leq k \leq n-1$.*

(2) With the notations of Lemma 2, assuming that the shares of all other assets are positive at the optimum:

(2.1) the share of asset 1 is positive only if $w_{0,1} < w_{1,2}$,

(2.2) the share of asset k, for $1 < k < n$, is positive only if $r_{k-1} < r_k$ and $w_{k-1,k} < w_{k,k+1}$,

(2.3) the share of asset n is positive if and only if $r_{n-1} < r_n$.

Proof. (1) $r_k < r_{k+1}$ is a direct implication of the concavity of the objective function and the statewise stochastic order on the probabilities of positive yields of assets. Since $r_k = (p_k + \sum_{i=k+1}^{n})q_k < \sum_{i=k+1}^{n} q_k$, $r_k < r_{k+1}$ implies $q_k < q_{k+1}$, for any $1 \leq k \leq n-1$.

(2.1) From Lemma 2, assuming that the shares of all other assets are positive, the share of asset 1 is positive only if $\beta_1^* > 0$, which implies $w_{1,2} - w_{0,1} > 0$, since $r_2 > r_1$ from (1).

(2.2) From Lemma 2, assuming that the shares of all other assets are positive, the share of asset k is positive only if $\beta_k^* > 0$ and $\beta_{k-1}^* < 1$, i.e.:

$$(r_{k+1} - r_k)(w_{k,k+1} - w_{k-1,k}) > 0 \Leftrightarrow w_{k,k+1} - w_{k-1,k} > 0$$

since $r_k < r_{k+1}$ from (1), and

$$\begin{aligned} \frac{(r_k - r_{k-1})(w_{k-1,k} - w_{k-2,k-1})}{r_{k-1}(q_k - q_{k-1})} &< 1 \\ \Leftrightarrow (r_k - r_{k-1})(w_{k-1,k} - w_{k-2,k-1}) &< r_{k-1}(q_k - q_{k-1}), \end{aligned}$$

since $q_k > q_{k-1}$ from (1), and this last inequality is equivalent to the following ones:

$$\begin{aligned} r_{k-1}w_{k-2,k-1} - r_k w_{k-2,k-1} + p_{k-1}q_{k-1}q_k < r_{k-1}q_k - r_{k-1}q_{k-1} \\ \Leftrightarrow r_{k-1}(w_{k-2,k-1} + q_{k-1}) < r_k(w_{k-2,k-1} + q_{k-1}) \\ \Leftrightarrow r_{k-1} < r_k \end{aligned}$$

since $q_k > q_{k-1}$ from (1).

(2.3) From Lemma 2, assuming that the shares of all other assets are positive, the share of asset

n is positive if and only if $\beta^*_{n-1} < 1$, i.e.

$$\frac{(r_n - r_{n-1})(w_{n-1,n} - w_{n-2,n-1})}{r_{n-1}(q_n - q_{n-1})} < 1$$

which is equivalent to $r_{n-1} < r_n$, since $q_n > q_{n-1}$ from (1), using the same steps as in the proof of (2.2).

□

B.3 General solution

In general, some assets may not be used.

Lemma 4. *(1) Let $A = argmax_i\{r_i\}$.*

At the optimum, A has a positive share. No asset $a > A$ has a positive share.

(2) Fix $(p_k, q_k)_{k=1}^m$. For any asset a other than A, there exists $w(a)$ such that if $w \geq w(a)$, the optimal share of a is zero.

(3) Fix w and $(p_k, q_k)_{k=2}^m$. For any asset a, there exists $q(a)$ such that if $q_1 \geq q(a)$, the optimal share of a is zero.

Proof. (1) This is a corollary of Lemma 3 - (2.3).

(2) If $a > A$, it is a consequence of (1). If $a < A$, let $w(a) \equiv max_{(j,k), with\ 1 \leq j \leq a, j < k, k \leq A}(w_{j,k})$. Assume by contradiction that $w \geq w(a)$ and a has a positive share. Then, either a is the smallest such asset, in which case, from Lemma 3 - (2), necessarily there exists $k > a$ such that $w_{a,k} > w \geq w(a)$, which contradicts the definition of $w(a)$, or there exist j, k with $1 \leq j < k \leq a$, such that j is the smallest such asset, in which case, from Lemma 3 - (2), necessarily $w_{j,k} > w$, which contradicts the definition of $w(a)$.

(3) With $q(a) = q_a$, it is a consequence of Lemma 3 - (1). □