The Globalization of Farmland

Rabah Arezki, Christian Bogmans, and Harris Selod

Working Paper Series

n° 346

January 2021

African Development Bank Group

Feed Africa

AFRICAN DEVELOPMENT BANK GROUP

Working Paper No. 346

Abstract

This paper proposes a model that sheds light on foreign direct investments in farmland. Countries can obtain food from other countries through international trade as well as by means of foreign land acquisition to offshore production. In equilibrium, bilateral trade and investment decisions are a function of cross-country differences in technology, land endowments, land governance, trade costs and domestic demands for differentiated food varieties. Using global data on transnational land deals, a test of the gravity equation for land investments shows evidence of bilateral patterns in line with the theory. In particular, the positive role of investor and host-country remoteness from markets in explaining bilateral investments is indicative of investor's food self-sufficiency motives. This contrasts with the negative role of host country remoteness in explaining platform-motivated FDI as is often the case with manufacturing. The paper also finds evidence that global financial centers in investor countries have facilitated transnational land deals.

This paper is the product of the Vice-Presidency for Economic Governance and Knowledge Management. It is part of a larger effort by the African Development Bank to promote knowledge and learning, share ideas, provide open access to its research, and make a contribution to development policy. The papers featured in the Working Paper Series (WPS) are considered to have a bearing on the mission of AfDB, its strategic objectives of Inclusive and Green Growth, and its High-5 priority areas: Power Africa, Feed Africa, Industrialize Africa, Integrate Africa, and Improve Living Conditions of Africans. The authors may be contacted at workingpaper@afdb.org.

Working Papers are available online at https://www.afdb.org/en/documents/publications/working-paper-series/

Produced by the Macroeconomics Policy, Forecasting, and Research Department

Coordinator
Adeleke O. Salami

Citation: Arezki R, C. Bogmans, and H. Selod. (2020), "The Globalization of Farmland" Working Paper Series N° 346, African Development Bank, Abidjan, Côte d'Ivoire.

The Globalization of Farmland*

Rabah Arezki†‡, Christian Bogmans§, and Harris Selod**

JEL Classification: F1, F2, Q1, Q17, Q3

Key words: Large-Scale Land Acquisitions, Food Self-sufficiency, Land Governance

* We are grateful to Vanessa Diaz Montelongo, Ziqi Li and Amjad Khan for excellent research assistance, to Brian Blankespoor for guidance on the construction and use of geographic data, and to Olivier Blanchard, Jeffrey Frankel, Davide Malacrino, Akito Matsumoto, Maurice Obstfeld, François de Soyres, and the participants of the EAERE 2017 Conference in Athens, the 2018 Land Governance and Poverty Conference held at the World Bank, the US International Trade Commission seminar and the IMF Research Department seminar for very useful comments on a previous version of this paper. The views expressed in this paper are those of the authors and do not necessarily represent those of the African Development Bank, the IMF or the World Bank, and their member states or policies.

† Chief Economist and Vice President for Economic Governance and Knowledge Management, African Development Bank, Avenue Joseph Anoma 01 BP 1387 Abidjan 01 Côte d'Ivoire. Email: r.arezki@afdb.org

‡ John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy St, Cambridge, MA 02138

§ Christian Bogmans (corresponding author): Research Department, International Monetary Fund, 1900 Pennsylvania Ave NW, Washington, DC 20431, USA. Email: cbogmans@imf.org.

** Harris Selod: Sustainability and Infrastructure team, Development Research Group, The World Bank, 1818 H Street NW, Washington, DC 20433, USA. Email: hselod@worldbank.org.

1 Introduction

In 2008, the world experienced a steep increase in food prices likely caused by a combination of factors, including a booming demand for biofuels and rising oil prices. Although it was not the first episode of sharp food price volatility, with the doubling of wheat and rice prices it was the largest in decades. This led to uncoordinated although familiar responses such as the enactment of export restrictions by food exporters. Interestingly, there was also a totally novel and unanticipated response with the sudden interest of many different stakeholders in directly acquiring agricultural land abroad, a phenomenon described by some observers as the "global land rush". A large fraction of such investments came from land-scarce economies–such as the GCC countries, in particular–and were directed at middle-income and developing countries with large endowments of land, a pattern which seems to have continued at a steady pace well into the 2010s. To a certain extent, the 2008 food price spike may have played the role of a tipping point in the organization of global agricultural markets. Although short-term volatility is unlikely to have been the trigger of long-term investments, it is possible that the 2008 price spike had shifted global expectations about long-term food prices and raised awareness that climate change and population and income growth will continue to increase the demand for food while the global land endowment will remain limited.

In a sense it seems that with the purchase of foreign agricultural land, investors opened up a new channel to secure food supply without exclusively relying on trade. This recent expansion in the globalization of farmland has led to a polarized debate between those welcoming foreign investments, hoping they will help raise agricultural land yields, foster economic development and alleviate poverty, and those who see the phenomenon as a "land grab" (Financial Times (2016) and Bloomberg (2017)). For some observers, the surge in land acquisitions also brings back memories from the colonial era, where advanced maritime powers (the Portuguese, Spanish, British, French and Dutch) conquered a large number of countries around the world and subsequently secured control over these countries' land resources. Those invasions led to a dramatic expansion of global trade in natural resources and agricultural products. In the present context, direct investments in land seem more to be driven by food self-sufficiency or profit-seeking motives of a variety of investors from many countries (private investors, including through multinational companies, or public investors, including through Sovereign Wealth Funds) rather than competition for economic clout between a handful of colonial powers dominating global commerce.[1] It also reflects that globalization has entered a new phase, one that is defined by the integration of pristine land in developing countries into the world economy (see Collier and Venables (2012)).

In this paper, we propose a multi-country model of trade and investment that helps us understand the different forces at play in potential host and investor countries driving bilateral investments

[1] It is noticeable for instance that a number of wealthy Gulf states, who are dependent on imports for 80 to 90 percent of their food consumption, have invested large sums of money over the last decade to buy cheap farmland in Tanzania, Ethiopia and other African nations.

in land, and derive a gravity equation for land investments which we estimate with transnational investment data. In our model, there are two sectors of production (agriculture and manufacturing) and two factors of production (labor and land). As in Ramondo and Rodríguez-Clare (2013), we assume that investors can use a home-country specific agricultural technology to engage in multinational agricultural production abroad. Agricultural varieties are both host-country *and* investor-country specific *à la* Armington (1969). In this setting, agricultural trade and multinational production are (imperfect) substitutes: a consumer country can import directly from native producers or instead offshore agricultural production using home-country technology. The rationale behind the double Armington assumption is that whereas differences in agro-ecological conditions lead to differentiated agricultural products around the globe, agricultural products also differ given the diversity of agricultural technologies used by different investors (e.g., GMO versus non GMO) or because of their different investor branding (e.g., fair trade versus conventional). In each country, there is an exogenous and fixed land price, which influences the quantity of land supplied to the commercial agricultural sector.

Using this framework, we derive a bilateral gravity-type specification of demand for land. To test the predictions of the model, we combine data from the Land Matrix (an online database of large-scale land acquisitions) with country-level data on land endowment, population size, yield gap, access to finance, land governance, agricultural productivity per worker and two measures of remoteness *à la* Anderson and Yotov (2016) and *à la* Baier and Bergstrand (2009). Following Silva and Tenreyro (2006), Fally (2015), and Anderson and Yotov (2016) and others, we estimate our gravity-type specification using Poisson Pseudo Maximum Likelihood (PPML) to deal with heteroskedasticity and a large proportion of zero observations.

As predicted by the model, bilateral investment in land is shown to depend on proxies for trade costs (negatively on distance and common border, positively on former colonial ties) as well as on host country characteristics (positively on land endowment, yield gap, and host country remoteness) and on investor country characteristics (positively on agriculture productivity, population size, global reach of the financial sector, and investor country remoteness). These effects reflect mechanisms that relate food needs and profitability of investments (from lower trade costs and greater technology and access to finance) to bilateral investments. In particular, we find that investor countries with international or global financial centers invest more than investor countries without those centers: The presence in the investor country of a global financial center in the top 15 (e.g. as Singapore or New York) as compared to having no financial center in the top 50 at all increases the number of deals originating from that country by 488 percent. As for the positive role of remoteness, it can be understood as follows: Pairs of investor and host countries that are both remote from the rest of the world will trade more land with one another. This is due to the fewer alternative investment opportunities for the investor (since the investor is remote from other potential host countries) and to the weaker interest of other investors to invest in the host country (since the host country is remote from other potential investor countries). This is a result that was obtained in the trade literature (Anderson and van Wincoop (2003)) but not yet shown for investments in land. This suggests that investments in land are more akin to trade than to standard FDI in which a negative effect of remoteness is possible (see Larch et al. (2017b)'s dynamic trade model in which host-country remoteness negatively affects FDI, but which, to our knowledge, was not yet empirically tested).

In quantitative terms, we find that when remoteness of the host country from investor country consumption markets increases by 1 percent (1/47th of a standard deviation increase from the mean), then the expected number of bilateral deals increases by 7.7 percent. Similarly, an increase of 1 percent in the remoteness of an investor country from host country producers (1/59th of a standard deviation increase from the mean) increases the number of deals by 2.79 percent. In fact, host-country remoteness appears to be a relevant determinant of investment, and its effect even exceeds that of agricultural land endowments in explaining large-scale land acquisitions. This finding is fully consistent with the food self-sufficiency motive underpinning foreign land investments. The fact that the elasticities of host-country and investor-country remoteness differ suggests an imbalance between investors and hosts at the disadvantage of hosts in capturing investments. To a certain extent, our empirical test confirms the intuition formulated by Collier and Venables (2012) in the case of Africa that the markets for such investments are "land-abundant, investor-scarce", with demand small relative to potential supply.

Our paper contributes to the literature in two important ways. First, to our knowledge, we are the first to provide a multi-country model with testable implications on the determinants of transnational agricultural land investments. The only other theoretical model that we know of is that of Rosete (2018), who formulated a game theoretical model of a single investor choosing between different locations. Rosete's finding supports the empirical result in Arezki et al. (2015) that weak recognition of existing users' property rights facilitates investments as the credible threath of eviction that hey face lowers the compensation that investors have to offer them to purchase the land. Contrary to our model, however, the nature of Rosete's model prevents it from shedding any light on the bilateral relationships in the data that has been collected since 2008. Second, our model introduces new mechanisms that were overlooked in previous analyses of large-scale land investments, namely the role of technology, of finance, and of remoteness to markets, which were found to be key drivers of those investments although they were unexplored in the previous empirical papers of Arezki et al. (2015) and Lay and Nolte (2017).

Our paper is also related to the literature on trade and deforestation (see Angelsen and Kaimowitz (1999) for an early review of the literature). Recently, Harstad (2020) formulated a dynamic North-South model that explains the two-way relationship between trade and natural resource depletion: Because trade liberalization raises the demand for beef and timber in the South, it encourages land conversion to agriculture in the South. Reciprocally, because the North is harmed by damages from land conversion in the South, it only benefits from trade when the remaining stock of the resource is low. Abman and Lundberg (2020) present empirical evidence of the first effect in Harstad (2020) whereby trade causes depletion by showing that regional trade agreements can foster deforestation. Somehow, our paper also confirms this first effect as we find that lower bilateral trade costs encourage foreign land acquisitions (and likely forest conversion). In a way, our work also relates to the second effect in Harstad (2020) whereby depletion causes trade. This is because the rush for land that we are studying is in part motivated by increasing global scarcity of land resources.

The remainder of this paper is as follows. Section 2 presents the multi-country model of trade and land acquisitions. Section 3 uses the model to derive a gravity style equation for domestic and transnational demand for land. Section 4 presents the econometric specifications for our cross-country analysis. Section 5 describes the data used. Section 6 estimates the determinants of transnational land deals using bilateral regressions, highlighting the role of remoteness. Section 7 concludes.

2 Modeling international trade and land acquisitions: the food self-sufficiency motive

2.1 The general setting

In this section we present a multi-country model that incorporates international investments in land and trade of agricultural products. The world consists of N countries. Each country produces an agricultural good (A) and a manufactured good (M) using labor (L) and land (T). In each country $i = 1...N$, factors of production are available in quantities L_i and T_i. Although labor can be used in both the manufacturing and agricultural sector, land is used for the production of agricultural goods only. Labor is a mobile factor of production that can move between the agricultural and the manufacturing sector within each country.

In-country and multinational production The manufacturing and the agricultural sectors produce goods that are differentiated by the country of origin. For the agricultural sector, this is consistent with trade being motivated by the consumption of different varieties of agricultural products (see Costinot et al. (2016)). Because agricultural potential varies across countries, producers grow different crop varieties of the same crop or different crops across countries. Our model, however, offers an extra layer of differentiation for agricultural goods as investors from a country i can also engage in production of an agricultural variety in another country l for the purpose of "re-importing" the agricultural good to its own domestic market only, something we refer to as *multinational production.* Differentiation according to both the country of production and the origin of the investor is a realistic feature of the model as consumers may for instance differentiate between coffee sold by a national producer of a Latin American country (i.e., a "fair trade" producer) and coffee produced in the same country by a multinational firm.

We will detail in the next subsection how these two types of differentiation affect consumer utility. For the time being, simply note that on the production side, the agricultural technologies Z_{li} used by investor country i differ for each host country l and are given by a vector of Total Factor Productivity (TFP) terms, $\boldsymbol{Z_i} \equiv \{Z_{li}\,,\, l = 1...N\}$. With these notations, Z_{ll} is the technology for own-country production of the agricultural variety by country l. For simplicity, we abstract from multiple crop choices by assuming that an investor country produces only one type of differentiated crop in a given country.[2]

Note that, in this setting, we assume that an investor must purchase land in a country l in order to sell agricultural goods to their home market.[3] This corresponds to an FDI type of investment, which has long been the norm in other commodity sectors such as energy and metal. We thus abstract from modeling contract farming, which would require the investor to buy other inputs instead of land.[4]

Also note that, as standard in the literature, trade is subject to an iceberg cost: in order to sell one unit of good to consumers in importing country n, a firm producing in country l must ship $t_{nl} \geq 1$ units of its product.[5]

Food self-sufficiency motivated investments in land Our representation of multinational production in the agricultural sector revolves around the transfer of technology from an investor country to a host country (see Ramondo and Rodríguez-Clare (2013)). As in the Ramondo and Rodríguez-Clare framework, investors can engage in multinational production and use a host-specific variant of their home-country technology to benefit from lower production costs in the host country. In the Ramondo and Rodríguez-Clare framework, investors also benefit from producing abroad if that location offers close proximity to other markets (thereby saving on trade costs t_{nl} as in Brainard (1997), Markusen and Venables (1998), Helpman et al. (2004) and others).

In our model, only the first feature of the Ramondo and Rodríguez-Clare framework is present as producers invest abroad with the purpose of re-exporting to their domestic market only, a pattern of investment to which we refer as "food self-sufficiency" (or vertical) FDI. Hence, investors are competing with other investors for (i) access to resources (i.e., labor and land) in the host countries. The model generates three types of production: (i) in-country production for home consumption, (ii) in-country production for export (i.e., international trade), and (iii) multinational production for export to oneself.

[2]We also abstract from considering multinational production in the manufacturing sector, an issue which is outside the scope of this paper.

[3]In a previous working paper we considered a second framework in which investors could also export to other countries, a pattern of investment to which we referred to as "*platform*" FDI (see Arezki et al. (2018)).

[4]This simplification is justified to the extent that contract farming constitutes only a relatively small share of investments (see Oya, 2012).

[5]For simplicity and without loss of generality, we do not differentiate the trade costs of the manufacturing and the agricultural goods. Introducing sector-specific trade costs would only make notations burdensome without changing the intuition of the model.

2.2 Preferences and consumption

Each country n has a representative agent who consumes all varieties of the agricultural and manufacturing goods. We assume that the upper-tier utility function is Cobb-Douglas:

$$u_n = \left(Q_n^A\right)^{\gamma} \left(Q_n^M\right)^{1-\gamma}, \tag{1}$$

where Q_n^A and Q_n^M are the quantities of the composite agricultural good and composite manufacturing good consumed in country n (see fromulas (2) and (3) below for the exact specifications), and $\gamma \in]0,1[$ and $1-\gamma$ represent the expenditure shares on agricultural and manufactured goods respectively.

To account for the idea that consumers differentiate between goods based on the country of origin, we specify agricultural consumption and manufacturing consumption in country n as CES aggregates over the N number of discrete varieties that are differentiated by country of production, with

$$Q_n^A = \left(\sum_{l=1}^{N} \left(q_{nl}^A\right)^{\frac{\varepsilon_A - 1}{\varepsilon_A}}\right)^{\frac{\varepsilon_A}{\varepsilon_A - 1}} \tag{2}$$

and

$$Q_n^M = \left(\sum_{l=1}^{N} \left(q_{nl}^M\right)^{\frac{\varepsilon_M - 1}{\varepsilon_M}}\right)^{\frac{\varepsilon_M}{\varepsilon_M - 1}}, \tag{3}$$

where q_{nl}^A (respectively q_{nl}^M) is the consumption of agricultural goods (respectively manufactured goods) produced in country l and consumed by the representative consumer in country n, and $\varepsilon_A > 1$ (respectively $\varepsilon_M > 1$) is the elasticity of substitution between agricultural varieties (respectively manufacturing varieties) produced in different countries.

Because our model allows for differentiated multinational production of the agricultural good, we further specify the consumption in country n of the agricultural goods produced in country l as a CES aggregate over two agricultural goods, one produced and exported by the host country l and one produced and "re-imported" by the consumer/investor country n,

$$q_{nl}^A = \left(\left(q_{nll}^A\right)^{\frac{\sigma-1}{\sigma}} + \left(q_{nln}^A\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{4}$$

where q_{nli}^A represent country n's consumption of the agricultural variety that is produced in country l by investors/producers from country i, and $\sigma > 1$ is the elasticity of substitution between agricultural varieties produced by different investors. Observe that if $\sigma \to \infty$, then the agricultural varieties are not differentiated by investor country. Formula (4) encapsulates the idea that agricultural trade and multinational production are (imperfect) substitutes: the consumer country can import directly from native producers or instead offshore agricultural production using home-country technology.

In this "double Armington" setting for the consumption of agricultural goods, observe that (2) accounts for product differentiation by country of origin (*where* the good is produced), whereas (4) accounts for product differentiation by investor country (*by whom* the good is produced).

2.3 Production

Firms in the manufacturing sector produce a manufacturing good under constant returns to scale using labor L_l^M only, such that

$$y_l^M = B_l L_l^M, \tag{5}$$

where we have assumed that in order to produce one unit of the manufactured good a producer in country l requires $1/B_l$ units of labor. Production of agricultural commodities by an investor from country i in host country l requires labor input L_{li}^A, and land input T_{li}^A (which is leased–or acquired–locally in host country l). The constant return to scale Cobb-Douglas production technology is

$$y_{li}^A = Z_{li} \left(\frac{L_{li}^A}{\alpha} \right)^{\alpha} \left(\frac{T_{li}^A}{1-\alpha} \right)^{1-\alpha}, \tag{6}$$

where α and $1 - \alpha$ are the output elasticities with respect to labor and land. To account for adjustment costs in the export of agricultural technology or adaptation to foreign contexts, we assume that agricultural TFP may decay with distance between the investor and the host country in a multiplicative way.[6] Sound institutions in the host country, such as secure property rights or an effective rule of law, are also expected to positively impact agricultural TFP. We thus have

$$Z_{li} = q_l Z_i F(t_{il}), \tag{7}$$

where Z_i is the intrinsic productivity of investors from country i, q_l measures institutional quality (including land governance) in the host country, and F is a decreasing function of transport costs t_{li}, with $F(t_{li}) = 1$ for $i = l$. It should be noted that whether multinational producers can fully leverage their domestic agricultural technology may also depend on other intrinsic factors. Chief among these factors is the quality of financial intermediation offered in the investor country. In this respect, Poelhekke (2015) finds evidence that large international banks have facilitated Dutch

[6]Our assumption is based on the ideas of Diamond (1997), who hypothesized that the diffusion of agricultural knowledge and technology is inhibited by geography.

FDI in foreign markets, in particular in host countries where institutions are weak. Hence, we broadly interpret Z_i to depend on both agricultural productivity as well as the quality of financial intermediation in the investor country.

2.4 Equilibrium

Assuming perfect competition for each variety, the producer price of each manufacturing variety equals its unit cost. Based on the manufacturing production function (5), this implies that the price for the manufactured good in country l is

$$p_l^M = \frac{1}{B_l} w_l, \tag{8}$$

where w_l is the return to labor in country l. Similarly, under perfect competition, the unit cost function associated with the agricultural production function (6) of investor i in country l is

$$c_{li}^A = \frac{(w_l)^\alpha (f_l)^{1-\alpha}}{Z_{li}}, \tag{9}$$

where f_l is the price of agricultural land in country l.

Consumers maximize utility by choosing how much to consume of each manufacturing variety and each agricultural variety. We can solve the utility maximization problem of consumers in three stages. First, at the top-tier, consumers maximize (1) by choosing agricultural and manufacturing consumption subject to the budget constraint, $I_n = P_n^A Q_n^A + P_n^M Q_n^M$, where I_n, P_n^A and P_n^M are defined as national income and the respective agricultural and the manufacturing price indices in country n. We obtain expenditures D_n^A and D_n^Mon agricultural and manufactured goods consumed in country n respectively:

$$D_n^A = \gamma I_n \tag{10}$$

and

$$D_n^M = (1-\gamma) I_n. \tag{11}$$

Second, consumers maximize agricultural utility (2) and manufacturing utility (3) by allocating total expenditures across the agricultural and manufacturing varieties from different countries. Let p_{nl}^A be the consumer price in country n of the agricultural varieties produced in country l. Then X_{nl}^A and X_{nl}^M, the respective consumptions by country n of the agricultural and manufacturing varieties produced in country l, are given by the following Marshallian demand functions:

$$X_{nl}^A = \left(\frac{p_{nl}^A}{P_n^A}\right)^{1-\varepsilon_A} \gamma I_n \tag{12}$$

and

$$X_{nl}^{M}=\left(\frac{t_{nl}p_{l}^{M}}{P_{n}^{M}}\right)^{1-\varepsilon_{M}}(1-\gamma)I_{n}, \tag{13}$$

where

$$P_{n}^{A}\equiv\left(\sum_{h=1}^{N}\left(p_{nh}^{A}\right)^{1-\varepsilon_{A}}\right)^{\frac{1}{1-\varepsilon_{A}}} \tag{14}$$

and

$$P_{n}^{M}\equiv\left(\sum_{h=1}^{N}\left(t_{nh}p_{h}^{M}\right)^{1-\varepsilon_{M}}\right)^{\frac{1}{1-\varepsilon_{M}}} \tag{15}$$

are the respective consumer price indexes of the composite agricultural good and the composite manufacturing good consumed in country n, also commonly referred to as the inward multilateral resistance indices (see Anderson and van Wincoop, 2003).

Third, consumers in country n maximize agricultural sub-utility (4) by choosing the level of consumption of the two investor-country differentiated varieties produced in country l, taking total expenditures on varieties from country l as given. The resulting Marshallian demand function, X_{nli}^{A}, that is, country n consumption of the variety produced in country l by investor country $i\in\{l,n\}$, is

$$X_{nli}^{A}=\left(\frac{t_{nl}c_{li}^{A}}{p_{nl}^{A}}\right)^{1-\sigma}\left(\frac{p_{nl}^{A}}{P_{n}^{A}}\right)^{1-\varepsilon_{A}}\gamma I_{n}, \tag{16}$$

where

$$p_{nl}^{A}=\left(\left(t_{nl}c_{ll}^{A}\right)^{1-\sigma}+\left(t_{nl}c_{ln}^{A}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{17}$$

is the CES price index of the two agricultural varieties produced in country l and consumed in country n. Observe that we can rewrite this price index as

$$p_{nl}^{A}=t_{nl}z_{ln}c_{ll}^{A} \tag{18}$$

where $z_{nl}\equiv\left(1+\left(\frac{Z_{ll}}{Z_{ln}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is a term that reflects the agricultural productivity of domestic investors relative to foreign investors (from country n) producing in country l. As can be seen from equations (16) and (13), imports decrease with distance from exporters and with producer price.

Next, let us define total revenues of sector $s\in\{A,M\}$ in producer country l as the sum of imports from all countries $n\in[1,..,N]$ of goods produced in l (including "own imports"), so that $V_{l}^{A}\equiv\sum_{n=1}^{N}X_{nl}^{A}$ and $V_{l}^{M}\equiv\sum_{n=1}^{N}X_{nl}^{M}$. With these notations, we also define global nominal outputs as $V^A \equiv \sum_{l=1}^{N} V_l^A$ and $V^M \equiv \sum_{l=1}^{N} V_l^M$ for the agricultural and manufacturing sectors respectively. Using (12) and (13), substituting for X_{nl}^A and X_{nl}^M in the formulas for V_l^A and V_l^M gives

$$V_l^A = (c_{ll}^A \Omega_l^A)^{1-\varepsilon_A} V^A \tag{19}$$

and

$$V_l^M = (p_l^M \Omega_l^M)^{1-\varepsilon_M} V^M, \tag{20}$$

where

$$\Omega_l^A \equiv \left[\sum_{n=1}^{n=N} \left(\frac{t_{nl} z_{ln}}{P_n^A} \right)^{1-\varepsilon_A} \frac{\gamma I_n}{V^A} \right]^{1/(1-\varepsilon_A)} \tag{21}$$

and

$$\Omega_l^M \equiv \left[\sum_{n=1}^{n=N} \left(\frac{t_{nl}}{P_n^M} \right)^{1-\varepsilon_M} \frac{(1-\gamma) I_n}{V^M} \right]^{1/(1-\varepsilon_M)} \tag{22}$$

are outward multilateral resistance indices in the agricultural and manufacturing sectors respectively. They are also the inverse of the agricultural and manufacturing market potentials of country l, that is, the weighted sum of the market sizes of all trade partners of country l.[7] Observe here that revenues (or nominal outputs) by sector decrease with "factory gate" prices and increase with market potential. Rearranging (19) to substitute for the power transform, $(c_{ll}^A)^{1-\varepsilon_A}$, into equation (14), we can write the inward multilateral resistance terms P_n^A as

$$P_n^A = \left(\sum_{h=1}^{h=N} \left(\frac{t_{nh} z_{hn}}{\Omega_h^A} \right)^{1-\varepsilon_A} \frac{V_h^A}{V^A} \right)^{1/(1-\varepsilon_A)}. \tag{23}$$

To close the model, we must now specify how the amount of agricultural land $T_l^A \equiv \sum_{i=1}^{i=N} T_{li}$ is determined. For simplicity, we assume that the government leases land to agricultural investors at a fixed price, $f_l = \overline{f_l}$, that is determined outside of our model, and supplies a total quantity of land that meets investors' demand at this price. Hence, in our framework f_l is exogenous and T_l^A is endogenous, and some land, in the amount of $T_l - T_l^A$, is not used for commercial agriculture. In practice, governments have indeed often leased land to investors at a flat rate, see Deininger et al. (2011).[8]

[7]Note that these measures of market potential attach more weight to nearby markets (smaller t_{nl}) and markets with higher prices (larger P_n^s).

[8]Note that it is straightforward to consider an alternative setting in which the government sets the supply of land to be allocated to agricultural use, $T_l^A = \overline{T_l^A}$, and lets the price f_l adjust. Alternatively, one could

Next, we note that national income in each country is given by the sum of labor earnings and government revenue from supplying land to the agricultural sector: [9]

$$I_l = w_l L_l + f_l T_l^A. \tag{24}$$

Before defining the equilibrium, let us now define aggregate demands for labor and land in each country. Cobb-Douglas production technology implies that land receives a fraction $1-\alpha$ of all agricultural revenues. Using equation (19), this property implies that the aggregate demand for land in host country l (the sum over all investors' demand for land in country l, $T_l^A = \sum_{n=1}^N T_{ln}^A$) can be written a $T_l^A = \frac{1-\alpha}{f_l}\left(c_{ll}^A \Omega_l^A\right)^{1-\varepsilon_A} V^A \leq T_l$.

Similarly, since a fraction α of agricultural revenues are paid out to agricultural labor and manufacturing only requires labor, the aggregate demand for labor (the sum over all agricultural investors' demand for labor in country l, $L_l^A = \sum_{i=1}^{i=N} L_{li}^A$, plus the demand for labor by manufacturing producers L_l^M) can be written as $L_l = \frac{\alpha}{w_l}\left(c_{ll}^A \Omega_l^A\right)^{1-\varepsilon_A} V^A + \frac{1}{w_l}(p_l^M \Omega_l^M)^{1-\varepsilon_M} V^M$.

Note that all terms on the right-hand side of the aggregate labor demand and the aggregate land demand equations can be written as functions of the vectors $\boldsymbol{w} = w_1, ..., w_N$ and $\boldsymbol{T^A} = T_1^A, ..., T_N^A$, such that we can think of these equations as a system of $2N$ equations in $\boldsymbol{w}$ and $\boldsymbol{T^A}$.[10] Solving this system of equations in the unknowns $\boldsymbol{w}$ and $\boldsymbol{T^A}$ can also be interpreted as finding the zeros of an excess demand system $G(\boldsymbol{w}, \boldsymbol{T^A}) = \left\langle G^L(\boldsymbol{w}, \boldsymbol{T^A}), G^T(\boldsymbol{w}, \boldsymbol{T^A})\right\rangle$, where $\boldsymbol{G^L} = \left(G_1^L, G_2^L, ..., G_N^L\right)$ and $\boldsymbol{G^T} = \left(G_1^T, G_2^T, ..., G_N^T\right)$ are respectively the N-dimensional excess labor demand and excess land demand vectors, with

$$G_l^L(\boldsymbol{w}, \boldsymbol{T^A}) = \frac{\alpha}{w_l}\left(c_{ll}^A \Omega_l^A\right)^{1-\varepsilon_A} V^A + \frac{1}{w_l}(p_l^M \Omega_l^M)^{1-\varepsilon_M} V^M - L_l \tag{25}$$

and

$$G_l^T(\boldsymbol{w}, \boldsymbol{T^A}) = \frac{1-\alpha}{T_l^A}\left(c_{ll}^A \Omega_l^A\right)^{1-\varepsilon_A} V^A - f_l. \tag{26}$$

Finally, note that we derive the country's balanced trade condition by adding up the excess demand for labor and land conditions and rearranging terms:

consider a setting in which both f_l and $T_l^A$are endogenous, resulting from an upward-sloping land supply curve. Under an upward sloping land supply curve, higher land prices would crowd out investment and act as a dispersal force. Here we choose to abstract from such a specification as our goal is not to produce a model that generates quantitative predictions, but rather to present a framework that delivers qualitative insights into the drivers of cross-border land acquisitions, and acts as a point of reference in setting up the empirical analysis that follows.

[9] Our setting with perfect competition in all markets implies zero profits.

[10] Our reasoning is as follows. First, note that unit costs, prices and national incomes can be written as functions of wages and agricultural land supplies. Unit costs c_{li}^A depend on w_l, so that prices p_l^A and p_{nl}^A depend on w_l, implying that P_n^A depends on $\boldsymbol{w}$. Similarly, p_l^M depends on w_l and P_l^M depends on $\boldsymbol{w}$. National income I_l depends on w_l and T_l^A. Second, it follows that V_l^A,V^A, V_l^M, V^M, Ω_l^A and Ω_l^M are functions of the vectors $\boldsymbol{w}$ and $\boldsymbol{T^A}$.

$$I_l = \left(c^A_{ll}\Omega^A_l\right)^{1-\varepsilon_A} V^A + \left(p^M_l\Omega^M_l\right)^{1-\varepsilon_M} V^M. \tag{27}$$

We can now define the model's equilibrium:

Definition 1. An equilibrium is a vector $\left\langle \boldsymbol{w}, \boldsymbol{T^A} \right\rangle \in \mathbb{R}^{2n}_{++}$ such that $G^L_l = 0$ and $G^T_l = 0$ for $l = 1, ..., N$, with c^A_{ll}, P^A_n and Ω^A_l satisfying equations (9), (14) and (21) respectively, and p^M_l, P^M_n and Ω^M_l satisfying equations (8), (15) and (22) respectively.

To prove that an equilibrium as in Definition 1 exists, we extend Theorem 2 in Alvarez and Lucas (2007)–which is an existence theorem for an exchange economy–to our Armington economy with two sectors and two factors of production. We assume that all countries own a sufficient quantity of land so that an interior equilibrium always exists, i.e., $T^A_l < T_l$ for all $l = 1, ..., N$.

Proposition 1. *There is at least one vector of wages and quantities of agricultural land,* $\left\langle \boldsymbol{w}, \boldsymbol{T^A} \right\rangle \in \mathbb{R}^{2n}_{++}$*, such that the markets for agricultural land and labor clear in every country:* $G = 0$*.*

Proof. See Appendix A. □

Now that we have established the foundations of our framework, we next derive testable hypotheses that can be used in conjunction with data on flows of cross-border land acquisitions, land availability and land governance, among other determinants.

2.5 Patterns of land investment

In this section, we use our theoretical framework to characterize patterns of land investment by deriving a gravity-style equation for transnational investment in land between any given investor-host country pair.

To do so, we substitute for p^A_{nl} from equation (18) into equation (16). Noting that $\frac{c^A_{ln}}{c^A_{ll}} = \frac{Z_{ll}}{Z_{ln}}$, we can write the agricultural revenues of investors from country n in country l as $V^A_{ln} = \frac{z^{1-\sigma}_{nl}-1}{z^{1-\sigma}_{nl}}\left(\frac{t_{nl}z_{nl}c^A_{ll}}{P^A_n}\right)^{1-\varepsilon_A}\gamma I_n$. Next, substituting for $c^A_{ll} = \left(\frac{1}{\Omega^A_l}\right)^{1-\varepsilon_A}\frac{V^A_l}{V^A}$ from equation (19) into the formula for V^A_{ln}, and noting that the land share of agricultural revenues is $1-\alpha$, we obtain a formula for land demand by agricultural investor n in country l:

$$T^A_{ln} = \left(\frac{t_{nl}\frac{Z_{ll}}{Z_{nl}}}{\Omega^A_l P^A_n}\right)^{1-\varepsilon_A}\frac{\gamma I_n T^A_l}{V^A} = \left(\frac{\frac{t_{nl}}{F(t_{nl})}\frac{Z_l}{Z_n}}{\Omega^A_l P^A_n}\right)^{1-\varepsilon_A}\frac{\gamma I_n T^A_l}{V^A}. \tag{28}$$

where, following Ramondo and Rodríguez-Clare (2013), we assume $\varepsilon_A = \sigma$ and finite, and where the second equality follows from substituting for $\frac{Z_{ll}}{Z_{ln}}$ using equation (7).[11] Observe that once we

[11] This simplifying assumption equates the elasticity of substitution between agricultural varieties produced in different countries and by different investors.

substitute out for factor costs and institutional quality (which drive production costs c_{ll}^A) in (28) by including remoteness, the formula does not explicitly link up factor costs and institutional quality to the number of land deals. However, factor costs and institutional quality are indirectly accounted for through their equilibrium relationship with host-country remoteness (see eq. 19).

We have the following proposition:

Proposition 2. *The quantity of land leased by investor country n in host country l is increasing in host-country remoteness from agricultural consumers Ω_l^A, investor-country remoteness from agricultural producers P_n^A, relative agricultural productivity of the investor-country $\frac{Z_{nl}}{Z_{ll}}$, expenditures on agricultural goods γI_n, the size of the host-country land market T_l^A, and decreasing in bilateral trade costs t_{nl}.*

Proof. The results follow directly from inspection of equation (28). □

Conditional on the amount of agricultural land T_l^A that is made available to investors in equilibrium, country pairs with remote hosts and remote investors feature more deals. This is similar to the role of importer-country and exporter-country remoteness in the determination of bilateral flows in trade models, as in Anderson and van Wincoop (2003). The intuition is as follows: Pairs of investor and host countries that are both remote from the rest of the world will trade more land with one another. This is due to the fewer alternative investment opportunities for the investor (since the investor is remote from other potential host countries / agricultural producers (P_n^A large)) and to the weaker interest of other investors to invest in the host country (since the host country is remote from other potential investor countries / consumer markets (Ω_l^A large)). [12]

3 Econometric Approach

3.1 Specification

The static model presented in the previous section shed light on the determinants of the long-term quantities of land put into cultivation and exchanged between countries. In line with our model, we now assess how these determinants affect the stock of FDI built up over the period of time for which data is available (see the data section 4.1). Specifically, we test the extent to which endowment of land, population size, agricultural productivity in host and investor countries, and host and investor country remoteness affect the stock of bilateral land deals over that period of time.

[12] Our prediction would capture for instance investments from Australia in Papua New Guinea which are both remote countries.

Based on the bilateral land investment equation (28), we propose a two-stage estimation procedure. The first stage aims to explain bilateral investment in land as a function of bilateral trade costs and serves to extract host-country and investor-country fixed effects used in our second stage. The second stage regresses (separeltey for host countries and for investor countries) the estimated fixed effects on all relevant unilateral determinants.

The empirical specification for this first stage is:

$$T_{li}^{A} = \beta_0 + \sum_{m=1}^{m=4} \beta_{1,m} ln(DIST_{li,m}) + \beta_2 COL_{li} + \beta_3 CONT_{li} + \mu_l + \lambda_i + \varepsilon_{li} \tag{29}$$

where T_{li}^{A} represents investments of country i in country l (i.e., the number of deals or total land covered under those deals). In this bilateral gravity equation, $DIST_{li,m}$ is the physical distance between investor country i and host country l if that distance falls in interval m, and zero otherwise. Following, Eaton and Kortum (2002) and Anderson et al. (2015), this construction allows us to account for non-linearities by estimating the effect of physical distance for four different intervals, $m = 1, 2, 3$ and 4, where the distance intervals, in km, are: [0,3000); [3000, 7000); [7000,10000) and [10000, maximum], COL_{li} is a dummy variable for historical colonial ties between investor country and host country, and $CONT_{li}$ is a dummy variable that equals 1 if investor country and host country share a common border and 0 otherwise. Observe that physical distance, colonial ties and common border are all proxies for trade costs between country i and country l. Finally, μ_l and λ_i represent host-country and investor-country fixed effects that absorb all country-level variation. Following Proposition 2 and Arezki et al. (2015), we expect distance between the investor and the host country to be an impediment to investment ($\beta_{1,m} < 0$ for all m), while colonial ties and a common border should encourage bilateral investment ($\beta_2 > 0, \beta_3 > 0$).

In the second stage, we then regress the estimates of the host-country and investor-country fixed effects on relevant unilateral investment drivers that stem from the gravity equation for land deals (28). The respective empirical specifications for the host-country and investor-country are as follows:

$$\lambda_i = \gamma_0 + \gamma_1 ln(L_i) + \gamma_2 ln(Z_i) + \gamma_3 FIN_i + \gamma_4 ln(P_i^A) + \varepsilon_i \tag{30}$$

and

$$\mu_l = \gamma_5 + \gamma_6 ln(T_l) + \gamma_7 GAP_l + \gamma_8 ln(\Omega_l^A) + \gamma_9 ln(w_l) + \gamma_{10} ln(q_l) + \varepsilon_l \tag{31}$$

where L_i is investor-country population size (proxying for expenditures on agricultural consumption), Z_i is investor-country agriculture value-added per worker (proxying for intrinsic agricultural TFP), FIN_i measures the quality-adjusted number of international or global financial centers in the investor country (proxying for the quality of financial intermediation), P_i^A is a measure of investor country remoteness, T_l is a measure of agricultural potential on host-country land that is both suitable and "available", GAP_l is a measure of the gap between actual and potential yields for major crops (proxying for the host-country productivity Z_{ll}), Ω_l^A is a measure of host country remoteness, w_l is host-country GDP per capita (proxying for labor costs), q_l is the host-country index that measures the recognition of preexisting land rights and associated level of tenure security from Arezki et al. (2015), and ε_l and ε_i represent the unexplained components of the host-country and investor-country fixed effects respectively.

In light of Proposition 2, our expectations are as follows: Investor countries acquire more land if they have a larger population ($\gamma_1 > 0$) as in Arezki et al. (2015), if they are agriculturally more productive ($\gamma_2 > 0$), if they have more global financial centers ($\gamma_3 > 0$), and if they are remote ($\gamma_4 > 0$). Host countries tend to attract more investments if they have abundant suitable agricultural land ($\gamma_6 > 0$) as in Arezki et al. (2015) and Lay and Nolte (2017), if their yield gap is large ($\gamma_7 > 0$) as in Arezki et al. (2015), and if the host country is remote ($\gamma_8 > 0$). Note that although including factor costs and institutional quality is redundant because of the equilibrium relationship with remoteness, we still include these terms as controls in the host-country regression. We expect host countries to attract more investment if labor is cheap ($\gamma_9 < 0$), and following Arezki et al. (2015) if land governance is weak ($\gamma_{10} < 0$).

[13]

3.2 Estimation of remoteness terms

We describe below the estimation procedure for our theory-inspired proxies for the host-country and investor-country remoteness terms. (summary statistics will be provided

in subsection (4.1). For robustness, we rely on two different approaches from the trade literature on the topic.

We estimate the first-stage equation (29) using Poisson Pseudo-Maximum Likelihood (PPML), which provides a consistent estimator for our bilateral gravity equation that is robust to measurement error, heteroskedasticity and the inclusion of zeros (see Silva and Tenreyro (2006)). The second-stage equations (30) and (31) are estimated using OLS. An important challenge in estimating our gravity-style equations is how to properly construct measures of remoteness. Anderson and van Wincoop (2003) have shown that if gravity equations do not account for these terms, the other coefficients of interest (such as e.g. trade costs) may become biased (Baldwin and Taglioni, 2007). To address this issue, we alternatively consider two different approaches to measure or account for remoteness. We present them sequentially below.

[13] The term $\left(\Omega_l^A\right)^{1-\varepsilon_A} = \sum_{n=1}^{n=N}\left(\frac{t_{nl} z_{ln}}{P_n^A}\right)^{1-\varepsilon_A}\frac{\gamma I_n}{V^A}$ is known in the economic geography literature as an index of market potential (see (Head and Mayer, 2015)). Since it is inversely related to remoteness Ω_l^A, there will be a negative relationship between bilateral investment and market potential in host-country and in investor-country.

First, we use linear approximations to the remoteness terms *à la* Baier and Bergstrand (2009).[14] Namely, we account for host-country l's remoteness (in logs) with the GDP-weighted average log distance of that country to all other countries, which accounts for agricultural export potential. We measure investor country i's remoteness (in logs) as the agricultural GDP-weighted average log distance to all host countries, which accounts for agricultural import potential. Hence, we use

$$log\left(\Omega_l^A\right) = \prod_{n=1,n\neq l}^{N} DIST_{nl}^{\theta_n} = \sum_{n=1,n\neq l}^{N} \theta_n log\left(DIST_{nl}\right) \equiv MRDIST_l(log) \tag{32}$$

and

$$log\left(P_i^A\right) = \prod_{l=1,l\neq i}^{N} DIST_{il}^{\xi_l} = \sum_{l=1,l\neq i}^{N} \xi_l log\left(DIST_{nl}\right) \equiv MRDIST_i(log), \tag{33}$$

where θ_n is the importer's share of global GDP, ξ_l is the exporter's share of global agricultural production, and MRDIST is a mnemonic for "multilateral resistance (or remoteness) based on physical distance". Following Baier and Bergstrand (2009), we also calculate remoteness measures based on the other proxies for trade costs, that is, former colonial relationship, common border and common language, to which we refer as MRCOL, MRCONTIG and MRCOMLANG respectively. The calculations for these additional remoteness measures are similar to eqs. (32)-(33), except that we replace $log\left(DIST_{nl}\right)$ with the relevant dummy variable, for example COL_{nl} (for a colonial relationship between countries n and l) in the case of MRCOL.

Note that the remoteness measures of eqs. (32)-(33) eliminate the first-order endogeneity between GDP and bilateral investment by removing the internal distance component $\theta_l log\left(DIST_{ll}\right)$. Like our first proxy, the linear approximations suggested by Baier and Bergstrand (2009) are a particularly attractive solution to the puzzle of measuring host-country and investor-country remoteness, as it allows us to calculate the remoteness measures directly using observable information on distance, GDP and agricultural GDP. A disadvantage of this approach, however, is that one would preferably use a direct measure of trade cost t_{nl} instead of physical distance to calculate these remoteness terms.

Second, we also resort to another approach which improves upon the proxy suggested by Baier and Bergstrand (2009) by using estimates of bilateral agricultural trade costs instead of distance. To come up with trade-cost estimates, we use a two-stage structural gravity approach recently proposed by Anderson and Yotov (2016) and Anderson et al. (2015). In the first step of the two-stage procedure we estimate a dynamic panel version of the bilateral agricultural trade equation (12) in multiplicative form using PPML. This estimation allows us to obtain estimates of the bilateral fixed effects $\hat{\mu}_{nl}$ for the country pairs for which agricultural imports are observed:

[14] Based on Monte Carlo analysis, Baier and Bergstrand (2009) show that their reduced form approach to remoteness (which we use in our paper) results in regression coefficients that are virtually identical to those estimated in the structural approach of Anderson and van Wincoop (2003).

$$X^A_{nl,t} = exp\left[\pi_{l,t} + \chi_{n,t} + \mu_{nl} + \beta_1 RTA_{nl,t}\right] \times \varepsilon_{nl,t}, \tag{34}$$

where $\pi_{l,t}$, $\chi_{n,t}$ and μ_{nl} respectively represent the producer/exporter time-varying fixed effect, the consumer/importer time-varying fixed effect, and the pair fixed effect, and $RTA_{nl,t}$ is a dummy variable that indicates the presence of a regional trade agreement between countries l and n at time t–the inclusion of which is necessary to obtain estimates of the bilateral fixed costs–and $\varepsilon_{nl,t}$ is a remainder error term. Since we do not observe agricultural trade flows for all possible country-pair combinations, we add an additional step to our procedure. In that second stage, we use the estimates of the pair fixed effects $\hat{\mu}_{nl}$ as a dependent variable and regress them on a set of standard determinants of trade costs and importer and exporter fixed effects. This can be written:

$$exp\left[\hat{\mu}_{nl}\right] = exp\left[\pi_l + \chi_n + \sum_{m=1}^{m=4} \eta_{1,m} ln(DIST_{ln,m}) + \eta_2 CONTIG_{nl} + \eta_3 COMLANG_{nl} + \eta_4 COL_{nl}\right] \times \varepsilon_{nl}, \tag{35}$$

where $CONTIG_{nl}$ and $COMLANG_{nl}$ are respectively dummy variables for a common border and for a common language between countries n and l. The predicted values from this second stage regression, that is,

$$\hat{t}_{nl}^{1-\varepsilon_A} = exp\left[\hat{\pi}_l + \hat{\chi}_n + \hat{\eta}_1 ln\left(DIST_{nl}\right) + \hat{\eta}_2 CONTIG_{nl} + \hat{\eta}_3 COMLANG_{nl} + \hat{\eta}_4 COL_{nl}\right],$$

are then used to complete the full bilateral trade cost matrix by substituting the predicted values for the country-pairs that are missing.[15]

To recover the trade cost estimate $\hat{t}_{nl}$ from the exponentiated term $\hat{t}_{nl}^{1-\varepsilon_A}$, we need to pick a value for the agricultural trade elasticity of substitution ε_A. We follow Tombe (2015) who finds an agricultural trade elasticity of substitution of 4.06. This value sits almost exactly in between the value of 5.4 and 2.82 that Costinot et al. (2016) obtain for respectively the trade elasticity of substitution at the crop level and the elasticity of substitution across different crops, and thus appears to be a reasonable pick. Armed with a bilateral trade cost matrix, we then calculate the agricultural multilateral resistance indices, Ω_l^A and P_i^A. As before, we rely on the approximations of Baier and Bergstrand (2009), eqs. (32)-(33), but use our trade cost estimates instead of physical distance to substitute for bilateral trade cost terms.

[15] Note that to estimate the two-stage regression model expressed by eqs. (34)-(35), we rely on the Stata command "ppml_panel_sg", developed by Thomas Zylkin, which is the only PPML command in Stata that allows for fast estimation of panel specifications with a large number of fixed effects (see Zylkin et al. (2017) for an application and a technical companion to this application).

4 Data

4.1 Sources

To measure bilateral investments in land, we use the Land Matrix (as of June 2016), which is an online database that contains extensive information on land deals that have undergone ground verification by NGOs affiliated with the International Land Coalition (Anseeuw et al. (2012)) and provides the most complete dataset available.[16] The 2016 version that we use is much improved compared to the 2011 version previously used in Arezki et al. (2015). The data includes information on 2,152 transnational deals negotiated between 2000 and 2016, including the origin country of the investor, and the total area covered by the investment. For the few land deals that involve multiple locations or multiple investor countries, we split them into several subprojects, which leaves us with a sample of 2,601 bilateral land deals. Removing the projects outside agriculture or biofuels, such as those associated with tourism, industry and other renewable energy, we are left with a total of 2,122 bilateral deals. With this information we calculate the cumulative number of projects (and cumulative quantity of land associated with these deals) by investor-country / host-country pair. We sum all deals by country pair over an unrestricted period (2000-2016) and a restricted period (2006-2013). Our main empirical analysis focuses on the restricted period. The reason for this is twofold. First, monitoring of large-scale land deals did not consistently take place before the onset of the global economic crisis of 2008. As retroactive information on deals in early years may be hard to come by, it is therefore possible that the selective inclusion of early years increases the likelihood of mismeasurement. Second, there is reason to believe that the intense scrutiny of these deals by civil society was most intense in the immediate aftermath of the 2008 global crisis so that the inclusion of later years also increases the possibility of mismeasurement. As we discuss in section (4.6), results are nevertheless similar across the restricted and unrestricted periods.

To measure the determinants of land acquisitions at the national level, we compile data from a variety of sources, including: the physical distance between countries and a dummy variable for a former colonial relationship from the GeoDist database (Mayer and Zignago, 2011); a land governance index at the national level which measures the level of recognition and associated tenure security of preexisting rights in rural areas (see Arezki et al. (2015));[17] a dummy variable for regional trade agreements for the period 1995-2004 that is borrowed from Mario Larch's Regional Trade Agreements Database from Egger and Larch (2008) and which includes all 468

[16] In this dataset, a deal is defined as an intended, concluded, or failed attempt to acquire land through purchase, lease, or concession that meets the following criteria: It (1) entails a transfer or rights to use, control, or ownership of land through sale, lease, or concession; (2) occurred after the year 2000; (3) covers an area of 200 hectares or more; and (4) implies the potential conversion of land from smallholder production, local community use, or important ecosystem service provision to commercial use. The vast majority of the deals in the database (80%) are listed based on two or more independent sources, e.g., research papers, policy reports and governments sources, and only 6% of the deals are based on media reports only (Nolte et al. 2016).

[17] The index is the first component of a principal component analysis on land governance variables contained in the 2009 Institutional Profiles Database (de Crombrugghe et al. 2009).

multilateral and bilateral trade agreements as notified to the World Trade Organization for the last 66 years from 1950 to 2015; and a set of variables from the 2009 World Development Indicators database (World Bank) that includes agricultural value added per worker (in constant 2010 US $), GDP per capita (in current US $) and population size.

To proxy for the quality of financial intermediation, we use data from Z/Yen Group (2008), which ranks global and international financial centers by their competitiveness, which is measured by a large number of instrumental factors that are based on both third party data as well as questionnaire responses. To construct our raw variable, each country receives 4 points for a financial center city ranked 1-15, 2 points for a financial center city ranked 16-30 and 1 point for a financial center city ranked 31-50. All other cities do not count. For example, the United Kingdom scores 8 points: the sum of the score for London (4), Glasglow (2) and Edinburgh (2), while the Netherlands scores 1 point for Amsterdam.

Furthermore, for gross food production (current US $) and bilateral food trade (current US $), we use the CEPII TradeProd database (a trade and production dataset originally constructed by de Sousa et al. (2012) that covers 26 sectors using the International Standard Industrial Classification (ISIC) revision 2, and 151 importing and exporting countries). For our purpose of estimating trade costs as explained in the previous section, we sum the trade flow data of the categories food (ISIC 311) and beverages (ISIC 313) for each year in the period 1995-2004.[18] Because proper estimation of (34)-(35) requires the availability of *intra-national* trade flow data, we use the CEPII TradeProd dataset which already contains these flows for many countries and impute missing values by using the Anderson and Yotov (2010) method.[19]

Finally, we construct measures of available suitable land, using the Global Agro-Ecological Zones (GAEZ) data jointly developed by the Food and Agriculture Organization (FAO) and the International Institute for Applied System Analysis (IIASA) (http://gaez.fao.org/). For each of five crops (wheat, maize, oil palm, sugarcane, soybean) we calculate the maximum of the suitability indexes (which has been re-scaled and is comprised between 0 and 100) under rainfed and baseline climate (1961-1990) conditions in each 5-arc minute (approximately 10km by 10km) grid cell. We then calculate the total area of all the cells for which this maximum for at least one crop is greater than 70 (which corresponds to high or very high suitability under rainfed conditions), after having excluded already cultivated land, forests, and protected land. The variable is expressed in 1,000 hectares.[20]

Summary statistics for our data are presented in Table (1).

[18] We also experimented with using agricultural trade flow data from the UN Comtrade database and found comparable estimates of trade costs.

[19] To do this, we first determine for each country the ratio of aggregate intranational trade over total goods expenditures for all sectors for which the data is available. Second, we multiply total expenditures on food and beverages with this ratio and take this as our estimate of intranational trade for those countries for which the data is missing.

[20] Using this approach, we estimate that the total quantity of suitable available land (defined here as uncultivated, non-forest, non-protected land suitable for agriculture outside urbanized areas) is about 446 million hectares for the whole world.

Table 1: Descriptive Statistics - Large-Scale Land Acquisitions Dataset

Variables	(1) N	(2) mean	(3) sd	(4) min	(5) max
Bilateral variables					
Project (2006-2013)	34,272	0.0475	0.763	0	54
Project (2000-2016)	34,272	0.0599	0.933	0	58
Distance (log), [0,3000) km	34,272	0.945	2.460	0	8.006
Distance (log), [3000,7000) km	34,272	2.578	3.914	0	8.854
Distance (log), [7000,10000) km	34,272	2.069	3.797	0	9.210
Distance (log), $\geq$ 10000 km	34,272	3.214	4.496	0	9.901
Former colonial relationship	34,272	0.0106	0.103	0	1
Common land border	34,272	0.0155	0.124	0	1
Common language	34,272	0.170	0.376	0	1
Host-country variables					
GDP per capita (log)	85	8.684	1.271	6.364	10.77
Max potential outp. forest and non-forest (log)	85	9.294	1.926	1.431	13.46
Max potential outp. forest and non-forest	85	41,338	89,914	4.184	701,295
Yield gap	85	0.551	0.263	0	0.910
Land governance	85	-0.00309	0.967	-1.924	1.849
MRDIST host (log)	85	8.508	0.623	6.205	9.264
MRDIST host	85	5,782	2,729	495.0	10,547
MRDIST host (log)(AY)	85	1.342	0.213	0.752	1.830
MRDIST host (AY)	85	3.915	0.852	2.121	6.235
MRCOL host (log)	85	0.0436	0.0594	0	0.351
MRCONTIG host (log)	85	0.0329	0.0498	0	0.244
MRCOMLANG host (log)	85	0.114	0.141	0	0.453
MRAGRIVA host (log)	85	-1.401	1.579	-4.160	1.089
Investor-country variables					
Total population (log)	111	16.35	1.756	11.07	21.00
Total population	111	5.446e+07	1.670e+08	64,000	1.318e+09
Agri value added per worker (log)	111	8.394	1.651	5.510	11.20
Agri value added per worker	111	13,911	19,631	247.1	73,187
No. of global financial centers	111	0.775	2.169	0	16
MRDIST investor (log)	111	7.742	2.326	2.279	9.519
MRDIST investor	111	5,600	3,295	9.762	13,616
MRDIST investor (log)(AY)	111	1.747	0.675	0.291	2.880
MRDIST investor (AY)	111	6.976	3.997	1.338	17.81
MRCOL investor (log)	111	0.0376	0.0557	0	0.356
MRCONTIG investor (log)	111	0.0259	0.0412	0	0.244
MRCOMLANG investor (log)	111	0.128	0.157	0	0.468
MRGAP investor (log)	111	-0.160	0.237	-0.551	0.328

Notes: *Project (2006-2013)* (*Project (2000-2016)*) measures the number of deals concluded per investor-host country pair between 2006-2013 (2000-2016), *Distance* measures physical distance (in km) between investor and host , *Former colonial relationship* is a dummy that equals 1 if investor and host country have been in a colonial relationship and equals 0 if not, *Max potential outp. forest and non-forest* represents the potential agricultural output of quantity on all non-forest, non-cultivated land (in 1,000 ha), *Land governance* is an index of tenure security enjoyed by existing land users, *Agri value added per worker* is value added per worker in the agricultural sector of the investor country (in constant 2010 USD), *Total population* is population size, *No. of global financial centers* is a score of the quality-adjusted number of global financial centers in the investor country constructed from Z/Yen Group (2008), *GDP per capita* is GDP per capita (in current US dollar), *MRDIST, MRCOL, MRCONTIG, MRCOMLANG, MRAGRIVA* and *MRGAP* are the remoteness terms based on physical distance, former colonial relationship, common border, common language, agricultural value added and the yield gap. All terms are based on the methodology by Baier and Bergstrand (2009), except those with AY in parentheses which are based on the methodology of Anderson and Yotov (2016) (see section 3.2).

4.2 Descriptive statistics

Table (2) provides an overview of land that is available and suitable for agricultural production for different regions in the world. It reveals that little land remains in the Middle East and North Africa and South and East Asia, while Sub-Saharan Africa and Latin America still have significant endowments of unexploited land. Forests constitute almost 63 percent of all available, non-protected land. Note that in our regressions, we use a definition of suitable available land that excludes forests and protected areas.

Table 2: Availability of Suitable Land for Agriculture (in mn ha), by Region.

	All	MENA	SSA	LAC	ECA	ESA	RoW
Suitable uncultivated, non-forest, non-protected land	446	3.04	202	123	52	14	51
Suitable uncultivated, forest, non-protected land	775	0.21	163	291	140	46	135
Total suitable uncultivated, non-protected land	1221	3.25	365	414	192	61	186

Source: Deininger et al., 2011. Notes: The data are for areas where population density is below 25 inhabitants per km2. Land is determined suitable if its suitability index exceeds 60 percent for any of the following five rainfed crops: wheat, oil palm, sugarcane, soybean, maize (see Fisher and Shah, 2010). Region abbreviations refer to Middle East and North Africa (MENA), Sub-Saharan Africa (SSA), Latin America and the Caribbean (LAC), Europe and Central Asia (ECA), East and South Asia (ESA), and Rest of World (RoW).

Table (3) presents basic characteristics of deals using data from the Land Matrix. All projects in the Land Matrix database were initiated between 2000 and 2016, although almost 80% of the deals were negotiated between 2006 and 2013. As of June 2016, the Land Matrix has information on 2,152 transnational deals, which together cover 58.4 million hectares, affecting 88 host countries worldwide. This roughly corresponds to an area the size of France or Ukraine, or 13 times the size of the Netherlands.

As one can tell from Table (3), Sub-Saharan Africa (884 deals) and East-Asia (611 deals) have been the most important target regions for investment, followed by Latin-America (368 deals), with only a few deals recorded outside of these regions. Figure (1) visualizes this uneven geographical distribution. In addition, the size distribution of these land deals is quite uneven: the 52 largest deals represent a staggering 40.9 percent of all acquired land. Approximately 66 percent of all deals are smaller than 10,000 hectares (or 100 km2). Deal implementation has been slow, both at the extensive and intensive margin: In 2016, 51.9 percent of all projects have been implemented, covering a total of 43.54 million hectares, but only 37.6 percent of all the land associated with those implemented projects have been confirmed under production. We note that the fraction of deals not implemented is significantly smaller in Sub-Saharan Africa (37 percent).

Table 3: Key Characteristics of Land Projects by Region, 2000 - 2016, by Region.

	All	MENA	SSA	LAC	ECA	EAP	SA
Size of projects							
Host countries (#)	88	6	38	18	9	13	4
Total area (mn ha)	58.9	0.9	23.7	10.2	10.6	13.4	0.1
Projects >= 1 mn ha (#)	3	0	2	0	0	1	0
Projects >= 250k ha and < 1mn ha (#)	49	1	18	9	15	6	0
Projects >=10k ha and <250k ha (#)	675	4	284	144	54	188	1
Projects<10k ha (#)	1,425	16	580	215	97	416	101
Total # projects	2,152	21	884	368	166	611	102
Intended use (percent)							
Agriculture	62.3	96.2	59.8	55.8	89.9	65.9	13.9
Biofuels	14.2	0.0	19.3	12.3	2.1	13.0	7.0
Forestry, Industry and Other	14.3	3.8	8.5	17.1	6.3	16.7	73.0
"Green"	7.3	0.0	10.8	9.6	1.7	3.1	6.1
Unknown	1.9	0.0	1.6	5.2	0.0	1.3	0.0

Notes: Region abbreviations refer to Middle East and North Africa (MENA), Sub-Saharan Africa (SSA), Latin America and the Caribbean (LAC), East Asia and the Pacific (EAP), Europe and Central Asia (ECA) and South Asia (SA). The Land Matrix does not include any deals hosted in North-America.

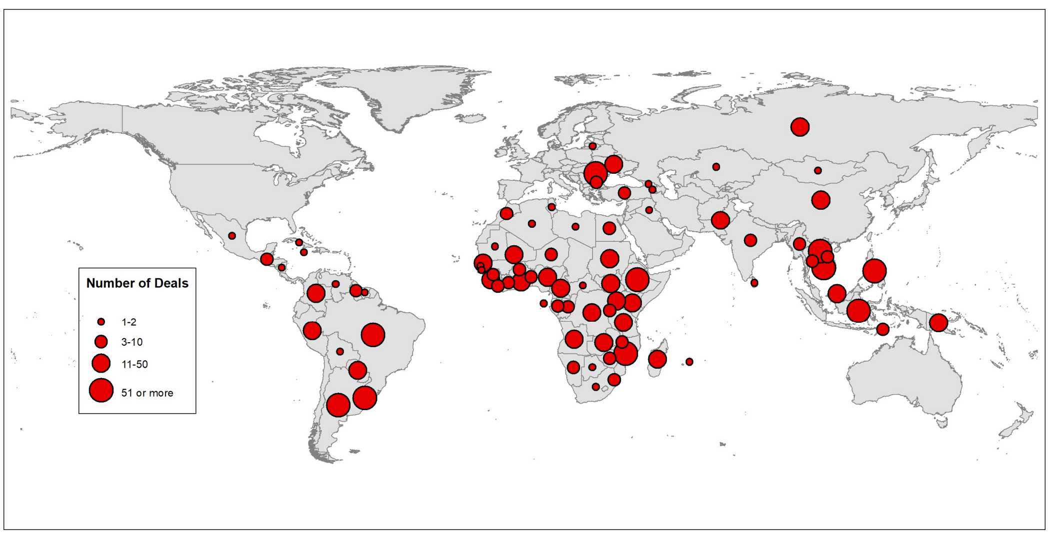

Figure 1: Number of deals by host country

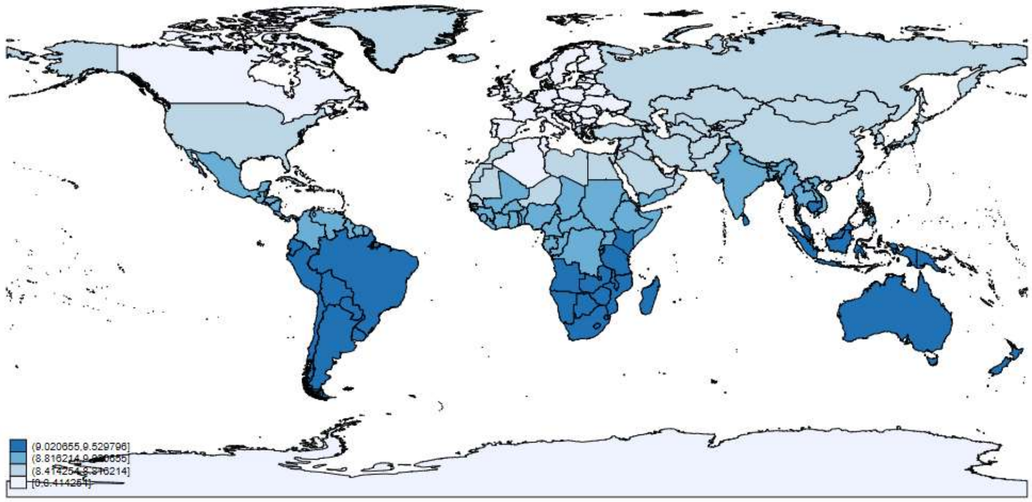

Figure 2: Remoteness a la Baier and Bergstrand (2009)

According to our calculations, 76.5 percent of all recorded deals have been linked to agricultural and biofuel related projects, see Table (3). This fact fits the narrative that large-scale land acquisitions have been largely motivated by private sector expectations of higher food and (bio)fuel prices and government concerns over food independence.

Finally, in figure (2) we provide an overview of the global distribution of remoteness, as measured by the Baier and Bergstrand (2009) methodology. Countries in Sub-Saharan Africa, those in East and Southern Africa in particular, tend to be more remote than Europe, the United States and Japan. A number of countries in Latin-America and South-East Asia likewise appear to be more remote from the rest of the world, as measured by the average distance to other markets.

4.3 Main results

Table (5) presents the first-stage results from our bilateral regression of the number of deals on bilateral trade costs variables (physical distance, colonial ties, and common border) and host and investor country fixed effects (as specified in (29)). With PPML, the coefficient estimates of regressors that are expressed in logs and levels can be interpreted as elasticities and semi-elasticities respectively.

In line with our model, we find that physical distance impedes bilateral deals. We observe that the impact of physical distance–our main proxy for bilateral trade costs–is rather constant across the various distance intervals, with the distance coefficients on the four intervals not being statistically different from one another. This contrasts with the impact of distance on bilateral trade flows as estimated in the literature, see e.g., Eaton and Kortum (2002) and Anderson and

Yotov (2016), who find that long distances impede trade more strongly than short distances. The absolute magnitude of each of the four distance coefficients is substantially larger than 1, which means that distance is a stronger deterrent of cross-border land acquisitions than it is of trade in goods because, for trade, scholars have often reported absolute coefficient values close to 1. In light of our model, one potential reason for this finding is that physical distance not only proxies for bilateral (agricultural) trade costs but also for impediments to transferring technology from investor to host country. In accordance with the literature on trade, we also find that colonial ties encourage land deals. An investor-host country pair with colonial ties is expected to have 131 percent more deals than a country pair with no colonial ties. However, a common border discourages them, decreasing the expected number of deals by 45 percent. The latter effect is consistent with the nature of land investments whereby developed countries invest in developing countries with whom they typically do not share a common border. Finally, contrary to the standard finding for trade, we find no evidence of common language affecting land deals. [21]

[21] Although we do not show the results without country fixed effects, note that their inclusion increases the coefficients of physical distance. This implies that not accounting for individual country characteristics would introduce a downwards bias as shown for international trade ((Head and Mayer, 2015; Anderson and van Wincoop, 2003)).

Table 4: First stage bilateral regression of the number of deals. The role of trade costs.

VARIABLES	(1) project_b
distance (log), [0,3000) km	-1.952***
	(0.148)
distance (log), [3000,7000) km	-1.948***
	(0.123)
distance (log), [7000,10000) km	-1.907***
	(0.116)
distance (log), ≥10000 km	-1.863***
	(0.113)
Former colonial relationship	0.837***
	(0.253)
Common land border	-0.610**
	(0.286)
Common language	0.167
	(0.202)
Constant	15.56***
	(0.987)
N	34272
r2_p	0.740
RESET	0.507

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, we recover the fixed effects from the first-stage regression and regress them on the relevant unilateral country variables for host countries and investor countries separately (as specified in (31) and (30) respectively). Table (5) reports the results of these unilateral regressions.

4.4 Host country regressions

Let us first present the host country results. In column (1) we present the regression of host country fixed effects on the yield gap (which inversely proxies for host-country productivity) and on the potential output value of uncultivated suitable land (capturing the potential size of the land market), controlling for GDP per capita and the land governance index (which together capture labor and land factor costs). In column (2-5), we add various remoteness terms to the specification: In column (2), we introduce a remoteness proxy for physical distance *à la* Baier and Bergstrand (2009). In column (3), we use instead the remoteness proxy for physical distance *à la* Anderson and Yotov (2016).[22] Column (4) extends column (2) by also including remoteness proxies for colonial ties, common borders and common language à la Baier and Bergstrand (2009). Column (5) adds another remoteness proxy for "technological distance".

In line with our model, we find that countries that attract more investments have a greater yield gap (which provides an opportunity for more technologically advanced foreign investors to bridge that gap). This effect is robust across all specifications and highly significant (at the one percent level). Similarly, the physical distance remoteness terms are also highly significant (at the one percent level) in all specifications. As for the other remoteness proxies in columns (4) and (5), only the border remoteness proxy is weakly significant (at the ten percent level). The endowment of of uncultivated suitable land has a positive and significant impact on the number of investments in regression (2). As for the controls, we see that, as expected, countries that attract investments are poorer (i.e., have lower labor costs) and have weaker tenure security as in Arezki et al. (2015) but the significance of these effects is not robust across specifications. The fact that tenure security is actually a deterrent of investments is consistent with the "land grab" narrative of many observers regarding these deals (Grain, 2008, 2016). The lack of robust significance may be due to the fact that "land grabs" have become more difficult in the recent period given the more intense civil society monitoring and increased transparency from initiatives such as the Land Matrix. Since the regression with the Baier and Bergstrand (2009) remoteness proxy (column 2) has a better fit than the regression with the Anderson and Yotov (2016) remoteness proxy (column 3) and since the additional remoteness proxies in columns (4) and (5) do not add much explanatory power, the parsimonious specification presented in column (2) is our preferred empirical model. We discuss its results below in more detail.

All things equal, we find that the number of bilateral deals increases by .84 percent for a 1 percent increase in the maximum potential agricultural output of available land (non-cultivated, non-forested and non-protected) in the host country. A country like Angola, which has 216.77 percent more agricultural potential on uncultivated land than the sample average, would attract 91.5 percent more land acquisitions than the average host.[23] Similarly, we find that an increase in the yield gap by 1 percentage point would increase the number of bilateral deals attracted by 0.092 percent.

We find that if the host country is more remote from agricultural consumers in the rest of

[22]In Appendix (B), we report the estimation results of regression (35) needed to obtain the Anderson and Yotov (2016) structural gravity remoteness measures, and confirm that the bilateral trade costs estimates from our first-stage fixed effects regression are partially explained by the usual set of geographical variables proxying for trade costs.

[23]For right hand-side variables that enter the regression specification in logs (such as potential agricultural output), an increase of one standard deviation (sdX) from the mean (X_{mean}) leads to increase in the number of deals of $\left(\left(\frac{X_{mean}+sdX}{X_{mean}}\right)^{\gamma_i}-1\right)*100$ percent. For variables that enter the estimated equation in levels (such as the output gap), a one standard deviation increase from the mean leads to an increase in the number of deals of $((sdX)^{\gamma_i}-1)*100$ percent.

the world–implying a weaker agricultural market potential–the expected number of bilateral deals increases. In quantitative terms, the coefficient on host-country remoteness implies that if distance-based remoteness to investor country markets increases by 1 percent, then the expected number of bilateral deals increase by 7.7 percent. Next, consider the following example: Because the average country in our dataset and Angola are located at a distance-based remoteness of 5,782 km and 8,554 km respectively from importing countries, Angola is 48 percent more remote than the average country. Applying the coefficient on the Baier and Bergstrand (2009) remoteness proxy from column (2) of Table (5), this would translate into 1,938 percent more bilateral deals for any given investor-host country pair in which Angola features as the host country than one in which the average country plays this role. While this quantitative result may seem excessively large, it is not as it represents slightly more than an one standard deviation increase in the quantity of bilateral deals.[24]

To shed more light on the relative importance of cross-country variation in host-country land endowments, yield gap and remoteness in explaining bilateral land deals, we consider a more general thought experiment. A standard deviation increase in the potential output of available agricultural land, which represents an increase of 318 percent from the mean, increases the number of deals by 164 percent. Similarly, a standard deviation increase of the yield gap increases the expected number of deals by 1012 percent, while a standard deviation increase of distance-based remoteness à la Baier and Bergstrand (2009) increases the expected number of deals by 1,854 percent. Using remoteness based on agricultural trade costs à la Anderson and Yotov (2016), the latter number changes to 548 percent.

In sum, comparing effects of a one standard deviation increase in independent variables, we find that host-country remoteness is a relevant determinant of bilateral investment, and its importance is comparable to that of the yield gap and easily exceeds the role of agricultural land endowments in explaining large-scale land acquisitions.

4.5 Investor country regressions

Let us now focus on the investor country regressions (columns(6)-(10) in Table 6). In column (6) we regress the investor country fixed effects on population size (proxying for expenditures on agricultural consumption), value-added per worker (proxying for intrinsic agricultural TFP), the quality-adjusted number of international or global financial centers (proxying for the quality of financial intermediation). In column (7-10), we add the investor country remoteness terms, replicating the same sequence as in the presentation of host country regressions: a remoteness proxy for physical distance à la Baier and Bergstrand (2009) in column (7) and a remoteness proxy for physical distance à la Anderson and Yotov (2016) in column (8). Column (9) extends

[24]Because of the distribution in the number of deals (with many country pairs not engaging in an investor/host relationship), a 1 standard deviation in the number of deals (0.76) represents a 1,600 percent increase from the mean (0.048).

column (7) with remoteness proxies à la Baier and Bergstrand (2009) for colonial ties, common borders and common language. Column (10) adds another proxy for "technological distance".

Consistently with our theoretical model, we find that countries with a larger population (and thus larger food consumption needs) tend to engage in more deals. This effect is robust across all specifications and highly significant (at the one percent level). Similarly, countries with higher valued added per worker in the agricultural sector (i.e., more productive countries) also engage in more deals, an effect that is also highly significant (at the one percent level) in all regressions. Countries that have more financial centers also engage in more deals. The effect is highly significant (at the one percent level) in all regressions that include remoteness terms. As predicted by the model, the coefficient for these remoteness terms is positive and significant (at the five percent or one percent level) in all regressions when remoteness is measured in terms of physical distance, common colonizer, and common border. The coefficient for remoteness measured in terms of common language is also significant, but it has the wrong sign. As for the host country regression, we see that the regression with the Baier and Bergstrand (2009) remoteness term has a better fit than the regression with the Anderson and Yotov (2016) remoteness term, and that adding remoteness terms based on other trade costs proxies such as former colonial relationships does not increase the explanatory power much. For these reasons, our preferred specification is the parsimonious one presented in column (7).

In line with our model, the results from our preferred specification show that increasing an investor-country's population size by 1 percent would raise the expected number of deals by roughly 0.55 percent. Similarly, increasing agricultural value added per worker in an investor country by 1 percent increases that country's expected number of deals by about .95 percent. The presence of a global financial center in the top 15 (e.g. as Singapore or New York) as compared to having no financial center in the top 50 at all (which implies a financial sector score increase from 0 to 4) would amount to an expected increase in the number of deals of the investor country by 488 percent. Remoteness of the investor country also plays a large role as we find that an increase of 1 percent in the remoteness proxy results in an expected increase in the number of deals by 2.79 percent.

As before, a better understanding of the relative importance of the various investor-country (push) factors can be attained by comparing the effects of a one standard deviation increase from the mean. Based on the regression coefficients in Table (5) and the descriptive statistics from Table (1), we find that a one standard deviation increase in population size from the population mean increases the number of expected deals by 116 percent. Likewise, a one standard deviation increase from the mean of agricultural value added, distance-based remoteness and the number of global financial centers increases the number of bilateral deals by respectively 131, 264 and 67 percent. Hence, remoteness and agricultural productivity – two factors playing a prominent role in our gravity equation for land, – represent the two most important push drivers in quantitative terms.

Table 5: Unilateral host-country and investory-country regressions. The role of endowments, technology, institutions and remoteness.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	host FE	host FE	host FE	host FE	host FE	Investor FE	Investor FE	Investor FE	Investor FE	Investor FE
GDP per capita (log)	-2.316**	-0.928	-2.287**	-1.034	-0.854					
	(0.972)	(0.883)	(0.989)	(0.913)	(1.190)					
Max potential outp. forest and non-forest (log)	1.253**	0.843*	0.743	0.778	0.775					
	(0.580)	(0.486)	(0.535)	(0.491)	(0.492)					
Yield gap	12.219***	9.192**	13.852***	9.869***	9.868***					
	(3.950)	(3.747)	(4.014)	(3.466)	(3.484)					
Land governance	-1.934	-0.896	-2.092*	-1.042	-1.070					
	(1.307)	(1.005)	(1.191)	(0.999)	(1.016)					
MRDIST host (log)		7.689***		8.196***	8.167***					
		(1.594)		(1.693)	(1.741)					
MRDIST host (log) (AY)			9.491***							
			(3.061)							
MRCOL host (log)				7.637	7.654					
				(7.145)	(7.120)					
MRCONTIG host (log)				19.849**	19.665**					
				(9.576)	(9.761)					
MRCOMLANG host (log)				-6.694*	-6.606*					
				(3.458)	(3.415)					
MRAGRIVA host (log)					-0.148					
					(0.702)					
Total population (log)						1.505***	0.548***	1.101***	0.448***	0.454***
						(0.421)	(0.119)	(0.263)	(0.120)	(0.115)
Agri value added per worker (log)						1.006***	0.949***	1.136***	0.779***	0.513***
						(0.376)	(0.117)	(0.257)	(0.130)	(0.151)
No. of global financial centers						0.345	0.386***	0.759***	0.390***	0.384***
						(0.273)	(0.074)	(0.195)	(0.069)	(0.073)
MRDIST investor (log)							2.790***		2.840***	2.918***
							(0.056)		(0.058)	(0.053)
MRDIST investor (log) (AY)								7.967***		
								(0.653)		
MRCOL investor (log)									3.396*	2.744
									(1.931)	(1.738)
MRCONTIG investor (log)									11.075***	9.969***
									(3.794)	(3.556)
MRCOMLANG investor (log)									-2.659**	-2.479**
									(1.103)	(1.113)
MRGAP investor (log)										2.819***
										(1.020)
R-squared	0.690	0.807	0.746	0.822	0.822	0.215	0.951	0.731	0.957	0.960
N	86	85	85	85	85	111	111	111	111	111

Notes: Estimates are obtained with OLS. A constant is included throughout but not reported. Dependent variables represent the country fixed effects (host and investor) from a first-stage bilateral land acquisition regression on physical distance, common colony and common border. All MR terms are based on the Baier and Bergstrand (2009) methodoloy, except those with AY in parentheses which are based on the Anderson and Yotov (2016) methodology (see section 3.2). Robust standard errors are reported in parentheses. *** p<0.01; ** p<0.05; * p<0.1

4.6 Robustness Checks

To verify the robustness of our results, we ran several alternative regressions. First, one may object that the relationship between land acquisitions and target-country remoteness is purely driven by an "Africa effect" as more than 41 percent of all land deals target countries in Sub-Saharan Africa (see Table 3 and Figure (1)). To control for this, we extended the regressions in Table (5) with an African continent dummy variable. In the benchmark specifications of both the host-country and investor-country fixed effects regressions, we find that the dummy variable is highly significant, negative, and leaves all other results qualitatively unaffected. Once we control for the (large) yield gap and (large) suitable land endowments of African countries, among other factors, we find, in fact, that countries in Sub-Saharan Africa host 95 percent fewer deals, and also initiate 74 percent fewer deals compared to countries in the rest of the world. A potential explanation for this counterintuitive finding is the generally weak institutional environment prevailing in Sub-Saharan Africa.[25] The inclusion of the dummy variable does reduce the magnitude of the remoteness regression coefficients slightly in some cases, but this is to be expected: the majority of African countries is substantially above-average in terms of remoteness, so the Africa dummy variable is expected to pick up this effect.[26]

Second, we re-estimated the first stage bilateral regressions of Table (4) by constraining the dependent variable, that is, the number of land deals by country pair, to strictly positive values. Because the majority of observations consists of zeros, we are left with a total of 409 bilateral flows.[27] Running our second stage regressions again after restricting the sample in the first stage, Table (A.1) shows that the distance-based host-country remoteness proxy remains statistically significant at the 5% level, whereas a few other regressors are no longer significant. For the subsample of country pairs with non-zero observations the presence of global financial centers, for example, does not appear to predict the total number of deals by country pair. This is consistent with the idea that financial centers with a global reach are a necessary condition to acquire land abroad but that they do not drive the number of deals. Furthermore we note that the elasticity on host-country remoteness in column (2) is smaller in magnitude than in the unconstrained regression. This makes sense as one could interpret the unconstrained elasticity as the combination of an extensive margin effect (i.e., whether to acquire land) and intensive margin effect (i.e., how many deals). As our results are by and large qualitatively unaffected and because gravity type regressions are typically not run on so few observations, we view the results of this robustness exercise as reaffirming the paper's main messages.

Third, we ran the first-stage bilateral regression with the quantity of land under all contracts as dependent variable. Recall that this variable measures the cumulative or total contract size associated with all deals per investor-country host-country pair. From columns 2 and 7 of Table (A.2), we learn that all regressors except the yield gap remain statistically significant. Hence the yield gap influences the decision whether to acquire land in another country or not (extensive margin), but it does not appear critical in determining total deal size (intensive margin).

Fourth, expanding our horizon to all land deals between 2000 and 2016 instead of 2006 and 2013 leaves our results unaffected too. In Table (A.3) we see again that our distance-based proxies for host-country and investor-country remoteness are positive and statistically significant, as in Table (5), while the magnitude of the coefficients changes little. The most notable differences are that host-country potential agricultural output is no longer a significant driver over the longer

[25]Indeed, once we introduce the "Rule of Law" variable from the World Bank Governance indicators, the significance of the Africa dummy variable in the host-country regressions disappears.

[26]For space considerations we do not include the regressions with the African dummy variable. The full set of results is available upon request from the authors.

[27]Note that the PPML estimator is generally well behaved even if the proportion of zeros in the dataset is very large (Silva and Tenreyo, 2011).

time period, whereas the coefficient on host-country GDP per capita becomes significant. These results confirm that, even across a larger time span starting well before the 2008 food crisis, the observed pattern of large-scale land acquisitions is consistent with the notion that investors aimed to secure land with the desire to re-export the produce to their home countries.

Fifth and finally, we re-ran the bilateral PPML regressions from Arezki et al. (2015) and added regressors, suggested by the gravity equation for bilateral land acquisitions, that were previously excluded (i.e., remoteness, global financial centers, agricultural value added per worker, and income per capita). The inclusion of this new set of regressors in table (A.4) leaves the results from Arezki et al. (2015) mostly unchanged, and all of the new regressors tend to be statistically significant at the 1 percent level. Arezki et al. (2015) analyzed three older land acquisition datasets, one of which was an older version of the Land Matrix dataset used in this paper. Hence, the results found in this paper are thus also robust to restricting our analysis to older data samples. The work by Arezki et al. (2015) also differentiated between deals that were already implemented and the total number of deals, the latter of which includes deals that are not (yet) implemented. Comparing columns 8 and 9 of table (A.4), we find that global financial centers are positively related to the number of started deals, but unrelated to the total number of deals. This suggests that financial intermediaries with a global reach are necessary to finance the implementation of transnational land deals.

5 Conclusion

The increased interest of international investors to acquire farmland is part of a broader set of developments that are changing the nature of commercial agriculture at the global scale. These include the increased importance of foreign direct investment and a more prominent role for global value chains in expanding food supply (Maertens and Swinnen (2015). In this paper, we presented the first multi-country trade and investment model that sheds light on the drivers of transnational land acquisitions and provide a testable empirical specification. The key innovation of the model is to account for two important features of bilateral investments in land. First, countries can resort to both trade and food production offshoring to obtain food from abroad. Second, investors can use their home-country agricultural technology when they produce agricultural goods abroad. The framework makes it possible to explain how cross-country differences in technology, endowments, land governance, demand for food, financial reach and bilateral trade costs drive transnational land acquisitions. Similar to the canonical gravity equation for trade in goods, our bilateral gravity specification for land investments predicts that host country remoteness (from agricultural consumers) and investor country remoteness (from agricultural producers) increase the quantity of acquired land.

All in all, the evidence presented in the paper sketches a clear picture of (i) large-scale land-investors targeting countries that have abundant quantities of available suitable land and that are not yet well integrated into the global food trading system, as observed by their significant degree of remoteness from existing food importers, and of (ii) agriculturally productive, populous investor countries with global financial centers being the driver behind the rush for land. The positive relationship between agricultural productivity and population size in investor countries and investment activity suggests that productive investors from existing centers of agricultural production in Western-Europe, North-America and South-East Asia are looking to accommodate the growth in demand for food by investing away from these centers of demand in new emerging centers of food production.

Transnational acquisitions of land raise important questions regarding food security and economic development. While agricultural prices have declined in recent years, they remain elevated compared to pre-crisis levels, with the future of food prices being dependent on, among other factors, the relative pace of population growth versus technological progress Baldos and Hertel (2016). More recently, the global COVID-19 pandemic has brought back price volatility in agricultural markets, raising the question as to whether this can lead to another rush for land. Expanding populations and a growing global middle class could continue to fuel global interest in large scale land acquisitions in the coming decades. The debate on the benefits, costs and risks of large-scale land acquisitions in developing countries in a context where farmland markets have become globalized seems far from settled.

References

Abman, R., and C. Lundberg (2020). "Does Free Trade Increase Deforestation? The Effects of Regional Trade Agreements." *Journal of the Association of Environmental and Resource Economists,* vol. 7 (1), pp. 35-72. 1

Angelsen, A., and D. Kaimowitz (1999). "Rethinking the causes of deforestation: lessons from economic models." *The world bank research observer,* vol. 14 (1), pp. 73-98. 1

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers*, 16(1), 159-178. 1

Alvarez, F., & Lucas Jr, R. E. (2007). General equilibrium analysis of the Eaton–Kortum model of international trade. *Journal of Monetary Economics*, 54(6), 1726-1768. 2.4

Anderson, J. and Y. Yotov (2010), "Specialization: Pro- and Anti-globalizing, 1990-2002,' NBER Working Papers 16301, National Bureau of Economic Research, Inc. 4.1

Anderson, J. and E. van Wincoop (2003), "Gravity with Gravitas: a solution to the border puzzle", *The American Economic Review*, vol. 69, pp.106-116. 1, 2.4, 2.5, 3.2, 21

Anderson, J., Larch, M. and Y. Yotov (2015), "Growth and Trade with Frictions: A Structural Estimation Framework", NBER Working Papers 21377, National Bureau of Economic Research, Inc. 3.1, 3.2

Anderson, J. and Y. Yotov (2016), "Terms of trade and global efficiency effects of free trade agreements, 1990–2002," *Journal of International Economics*, Volume 99, pp. 279-298, ISSN 0022-1996, https://doi.org/10.1016/j.jinteco.2015.10.006. (http://www.sciencedirect.com/science/article/pii/S0022199615001531) 1, 3.2, 4.3, 4.4, 4.5

Antras, P. and R. Staiger. (2012), "Offshoring and the Role of Trade Agreements", *American Economic Review*, vol. 102(7), pp. 3140-83.

Anseeuw, W., Alden Wily, L., Cotula, L. and M. Taylor (2012). Land Rights and the Rush for Land: Findings of the Global Commercial Pressures on Land Research Project. Rome: International Land Coalition. 4.1

Arezki, R., Deininger, K. and H. Selod (2015), "What Drives the Global "Land Rush"?", *World Bank Economic Review*, vol. 29 (2): 207-233 first published online October 21, 2013 doi:10.1093/wber/lht034 1, 3.1, 3.1, 4.1, 4.4, 4.6

Arezki, R., Bogmans, C., and H. Selod (2018). The Globalization of Farmland: Theory and Empirical Evidence. Policy Research Working Paper;No. 8456. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29885 License: CC BY 3.0 IGO 3

Baier, S. and J. Bergstrand (2009), "Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation," *Journal of International Economics*, Elsevier, vol. 77(1), pp.77-85, February. 1, 3.2, 3.2, 14, 3.2, 4.2, 4.4, 4.5

Baldwin, R. and D. Taglioni (2007). Trade effects of the euro: A comparison of estimators. *Journal of Economic Integration*, pp. 780-818. 3.2

Baldwin, R. and J. Harrigan (2011). "Zeros, quality, and space: Trade theory and trade evidence." *American Economic Journal: Microeconomics,* vol. 3 (2), pp. 60-88.

Baldos, U. and T. Hertel (2016), "Debunking the 'new normal': Why world food prices are expected to resume their long run downward trend", *Global Food Security*, vol. 8, pp.27-38. 5

Bloomberg (2017), "Farming the World: China's Epic Race to Avoid a Food Crisis", available at: https://www.bloomberg.com/graphics/2017-feeding-china/ 1

Brainard, S. (1997), "An empirical assessment of the proximity-concentration trade-off between multinational sales and trade", *The American Economic Review*, vol. 87 (4), pp. 520-544. 2.1

Collier, P. and A. Venables (2012), "Land Deals in Africa: Pioneers and Speculators", *Journal of Globalization and Development*, vol. 3(1), article 3, DOI: 10.1515/1948-1837.1228 1

Costinot, A., Donaldson, D. and C. Smith (2016), "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world", *Journal of Political Economy*, vol. 124(1), 205-248. 2.1, 3.2

Cotula, L. (2011) Land deals in Africa: What is in the contracts?, IIED, London, 49 pages.

de Crombrugghe, D., Farla, K., Meisel, N., de Neubourg, C., Ould Aoudia, J. and A. Szirmai (2009), "Institutional Profiles Database III. Presentation of the Institutional Profiles Database 2009 (IPD 2009)." Documents de Travail de la DGTPE, No.2009/14. Treasury Directorate General of the French Ministry of the Economy, Industry and Employment, Paris.

Deininger, K., Byerlee, D., Lindsay, J., Norton, A., Selod, H. and M. Stickler (2011) Rising Global Interest in Farmland, Washington, D.C.: The World Bank. 2.4

Diamond, J. (1997). Guns, germs, and steel: the fates of human societies. NY: WW Norton & Company, 14. 6

Dixit, A. and J. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, vol. 67(3), pp 297-308.

di Giovanni, J., Levchenko, A. and J. Zhang (2014), "The Global Welfare Impact of China: Trade Integration and Technological Change", *American Economic Journal: Macroeconomics*, vol. 6:3, pp.153-183.

Eaton, J. and S. Kortum (2002), "Technology, Geography and Trade", *Econometrica*, Vol.70(5), 1741-79. 3.1, 4.3

Economist Intelligence Unit (2016). Global Food Security Index. Accessed date, May 7, 2016. http://foodsecurityindex.eiu.com/Downloads

Egger, P. and M. Larch (2008), "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis", *Journal of International Economics,* vol. 76(2), pp. 384-399. 4.1

Ekholm, K., Forslid, R. and J. Markusen (2007), "Export-platform foreign direct investment", *Journal of the European Economic Association,* vol. 5 (4), pp. 776-795.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, 97(1), 76-85. 1

FAO (Food and Agriculture Organization of the United Nations) (2019), FAOSTAT Trade indices, http://www.fao.org/faostat/en/#data/TI/, accessed on August 27, 2019

Feenstra, R., (2004), "Advanced International Trade: Theory and Evidence". Princeton University Press, Princeton, New Jersey.

Financial Times (2016), The Great Land Rush, http://www.ft.com/great-land-rush , accessed on June 30, 2016. 1

Fischer, G., and M. Shah. "Farmland investments and food security." Rep. Prep. by World Bank Int. Inst. Appl. Syst. Anal.(IIASA), Luxemb. Austria (2010).

Frankel, J. and D. Romer (1999), "Does trade cause growth?", *American Economic Review*, pp. 379-399.

Grain (2008), "Seized: The 2008 land grab for food and financial security." October. http://www. grain. org/briefings_files/landgrab-2008-en. pdf (2008). 4.4

Grain (2016), "The global farmland grab in 2016: how big, how bad?" June. https://www.grain.org/article/entries/5492-the-global-farmland-grab-in-2016-how-big-how-bad 4.4

Harstad, Bård (2020. "Trade and Trees: How Trade Agreements Can Motivate Conservation Instead of Depletion". CESifo Working Papers 8569. 1

Head, K. and T. Mayer (2015), "Gravity Equations: Workhorse, Toolkit, and Cookbook". In: Elhanan Helpman, Kenneth Rogoff and Gita Gopinath, editor, Handbook of International Economics, Vol 4, Oxford: BV, pp. 131-195. 13, 21

Helpman, E., Melitz, M. and and S. Yeaple (2004), "Export versus FDI with heterogeneous firms", *American Economic Review*, vol 94 (1), pp. 300-316. 2.1

Land Matrix Global Observatory (2016), http://landmatrix.org/en/get-the-detail/, Accessed Date, May 7, 2016.

Larch, M., Wanner, J., Yotov, Y. and T. Zylkin (2017), "The Currency Union Effect: A PPML Re-assessment with High-dimensional Fixed Effects", Drexel University School of Economics Working Paper 2017-07.

Larch, Mario, James E. Anderson, and Yoto V. Yotov (2017), "Trade Liberalization, Growth, and FDI: A Structural Estimation Framework." 1

Lay, J. and K. Nolte (2017), "Determinants of foreign land acquisitions in low- and middle-income countries", *Journal of Economic Geography*, lbx011. doi: 10.1093/jeg/lbx011 1, 3.1

Levchenko, A. (2007), "Institutional Quality and International Trade", *Review of Economic Studies*, vol. 74 (3), pp. 791-819.

Markusen, J. and A. Venables (1998), "Multinational firms and the new trade theory", *Journal of International Economics,* vol. 46 (2), pp.183-203. 2.1

Maertens, M., and J. Swinnen. (2015). Agricultural trade and development: a value chain perspective (No. ERSD-2015-04). WTO Staff Working Paper. 5

Martin, P., Mayer, T., and M. Thoenig, M. (2008). Make trade not war? *The Review of Economic Studies*, vol.75(3), pp.865-900.

Mayer, T. and S. Zignago (2011), "Notes on CEPII's distances measures: the GeoDist Database", CEPII Working Paper 2011-25. 4.1

Matsuyama, K. (2007), "Beyond icebergs: towards a theory of biased globalization", *The Review of Economic Studies*, vol. 74(1), pp.237-253.

Nolte, K., Chamberlain, W. and M. Giger (2016). International Land Deals for Agriculture. Fresh insights from the Land Matrix: Analytical Report II. Bern, Montpellier, Hamburg, Pretoria: Centre for Development and Environment, University of Bern; Centre de coopération internationale en recherche agronomique pour le développement; German Institute of Global and Area Studies; University of Pretoria; Bern Open Publishing.

Oxfam (2012) "'Our Land, Our Lives' Time out on the Global Land Rush", Oxfam Briefing Note, October, 26 pages. https://www.oxfam.org/sites/www.oxfam.org/files/bn-land-lives-freeze-041012-en_1.pdf

Oya, C. (2012), "Contract farming in sub-Saharan Africa: A survey of approaches, debates and issues", *Journal of Agrarian Change*, 12(1), 1-33. 4

Poelhekke, S. (2015). Do global banks facilitate foreign direct investment? *European Economic Review*, vol. 76, pp. 25-46. 2.3

Ramondo, N. and A. Rodríguez-Clare (2013), "Trade, Multinational Production, and the Gains from Openness", *Journal of Political Economy*, vol. 121 (2), pp. 273 - 322. 1, 2.1, 2.5

Rosete, A. R. (2018). Understanding Investor Behavior in the Recent Global Land Rush. *Journal of Globalization and Development*, 9(1). 1

Redding, S., and Venables, A. J. (2004), "Economic geography and international inequality", *Journal of international Economics*, 62(1), 53-82.

Silva, J. and S. Tenreyro. (2006), "The Log of Gravity", *Review of Economics and Statistics,* vol. 88 (4), pp. 641–58. 1, 3.2

Silva, J. S. and S. Tenreyro (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2), 220-222. 27

Song, Z., Storesletten, K. and Zilibotti, F. (2011), "Growing Like China", *American Economic Review*, vol. 101(1), pp. 196–233.

de Sousa, J., Mayer, T. and S. Zignago (2012), "Market access in global and regional trade", *Regional Science and Urban Economics*, vol. 42 (6), pp.1037-1052. 4.1

Tombe, T. (2015), "The missing food problem: Trade, agriculture, and international productivity differences", *American Economic Journal: Macroeconomics*, vol. 7(3), pp.226-258. 3.2

Z/Yen Group (2008), "The Global Financial Centers Index 4", available on-line at http://www. longfinance. net/

The World Bank. 2019. World Development Indicators. Washington, D.C.: The World Bank. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators. Accessed August 27, 2019.

Appendix A. Technical Appendix.

Proposition 1.

Proof. We assume f_l is fixed in each country and the quantity of land T_l^A adjusts to clear the market for land. To prove existence of at least one vector $\langle \boldsymbol{w}, \boldsymbol{T^A}\rangle \in \mathbb{R}^{2n}_{++}$, we verify that $G(\boldsymbol{w}, \boldsymbol{T^A})$ has the following five properties:

(i) $G\left(\boldsymbol{w}, \boldsymbol{T^A}\right)$ is continuous,

(ii) $G\left(\boldsymbol{w}, \boldsymbol{T^A}\right)$ is homogenous of degree zero,

(iii) $\langle \boldsymbol{w}, \boldsymbol{T^A}\rangle \cdot G\left(\boldsymbol{w}, \boldsymbol{T^A}\right) = 0$ for all $\langle \boldsymbol{w}, \boldsymbol{T^A}\rangle \in \mathbb{R}^{2n}_{++}$ (Walras' law),

(iv) for $k = max\left(max_h L_h, max_h f_h\right)$, $G_l^L\left(\boldsymbol{w}, \boldsymbol{T^A}\right) > -k$ and $G_l^T\left(\boldsymbol{w}, \boldsymbol{T^A}\right) > -k$ for all $l = 1, ..., N$ and $\langle \boldsymbol{w}, \boldsymbol{T^A}\rangle \in \mathbb{R}^{2n}_{++}$, and

(v) if $\langle \boldsymbol{w}, \boldsymbol{T^A}\rangle \to \langle \boldsymbol{w^0}, \boldsymbol{T^{A,0}}\rangle$, where $\left\langle \boldsymbol{w^0}, \boldsymbol{T^{A,0}}\right\rangle \neq 0$ and $w_l^0 = 0$ and/or $T_l^{A,0} = 0$ for some l, then $max_h G_h(\boldsymbol{w}, \boldsymbol{T^A}) \to \infty$.

The existence result then follows from Proposition 17.C.1. of Mas-Collel et al. (1995, p. 585). The latter is an application of Kakutani's Fixed-Point Theorem.

(i) Since all the prices c_{ll}^A, P_n^A, p_l^M and P_n^M are continuous, Ω_l^A and Ω_l^M are continuous too, and so is $\boldsymbol{G}$.

(ii) The unit variety price c_{ll}^A ise homogenous of degree α in $\left\langle \boldsymbol{w}, \boldsymbol{T^A}\right\rangle$, while p_l^M is homogenous of degree one in $\left\langle \boldsymbol{w}, \boldsymbol{T^A}\right\rangle$. It is then immediate from (21), (22), (14) and (15) that $\frac{1}{T_l^A}\left(c_{ll}^A\Omega_l^A\right)^{1-\varepsilon_A} V^A$, $\frac{1}{w_l}\left(c_{ll}^A\Omega_l^A\right)^{1-\varepsilon_A} V^A$ and $\frac{1}{w_l}\left(p_l^M\Omega_l^M\right)^{1-\varepsilon_M} V^M$ are homogenous of degree zero, and so $\boldsymbol{G}$ has the same property.

(iii) We note that the requirement of $\left\langle \boldsymbol{w}, \boldsymbol{T^A}\right\rangle \cdot G\left(\boldsymbol{w}, \boldsymbol{T^A}\right) = \sum_{l=1}^{l=N} w_l G_l^L\left(\boldsymbol{w}, \boldsymbol{T^A}\right) + \sum_{l=1}^{l=N} T_l^A G_l^T\left(\boldsymbol{w}, \boldsymbol{T^A}\right) = 0$ corresponds to balanced trade in each country:

$$\sum_{l=1}^{l=N} w_l G_l^L(\boldsymbol{w}, \boldsymbol{T^A}) + \sum_{l=1}^{l=N} T_l^A G_l^T(\boldsymbol{w}, \boldsymbol{T^A}) = \sum_{l=1}^{l=N}\left(\sum_{n=1}^{n=N}\left(\frac{t_{nl} z_{ln} c_{ll}^A}{P_n^A}\right)^{1-\varepsilon_A}\gamma I_n\right)$$
$$+\sum_{l=1}^{l=N}\left(\sum_{n=1}^{n=N}\left(\frac{t_{nl} p_l^M}{P_n^M}\right)^{1-\varepsilon_M}(1-\gamma)I_n\right) - \sum_{l=1}^{l=N} I_l,$$

where we substituted for Ω_l^A and Ω_l^M into the excess demand functions. We observe that $\left\langle \boldsymbol{w},\, \boldsymbol{T^A} \right\rangle \cdot G\left(\boldsymbol{w},\, \boldsymbol{T^A}\right) = 0$ follows from the fact that in each country consumers spend all their income on the consumption of agricultural and manufacturing varieties, that is, $\left(\sum_{n=1}^{n=N} \left(\frac{t_{nl} z_{ln} c_{ll}^A}{P_n^A}\right)^{\frac{1}{1-\varepsilon_A}} \gamma I_n\right) + \left(\sum_{n=1}^{n=N} \left(\frac{t_{nl} p_l^M}{P_n^M}\right)^{1-\varepsilon_M} (1-\gamma) I_n\right) - I_l = 0$.

(iv) Inspection of equations (25) and (26) shows that each excess demand function has a lower bound that is equal to the inelastically supplied amount of labor or the exogenous price of land, that is, $G_l^L(\boldsymbol{w},\, \boldsymbol{T^A}) > -L_l$ and $G_l^T(\boldsymbol{w},\, \boldsymbol{T^A}) > -f_l$ for all $\left\langle \boldsymbol{w},\, \boldsymbol{T^A} \right\rangle \in \mathbb{R}^{2n}_{++}$. Hence, it holds that $G_l^L(\boldsymbol{w},\, \boldsymbol{T^A}) > -max\left(max_h L_h, max_h f_h\right)$ and $G_l^T(\boldsymbol{w},\, \boldsymbol{T^A}) > -max\left(max_h L_h, max_h f_h\right)$ for all $l = 1, ..., N$.

(v) Suppose $\left\langle \boldsymbol{w^m},\, \boldsymbol{T^{A,m}} \right\rangle \to \left\langle \boldsymbol{w^0},\, \boldsymbol{T^{A,0}} \right\rangle$, where $\left\langle \boldsymbol{w^0},\, \boldsymbol{T^{A,0}} \right\rangle \neq 0$ and $w_l^0 = 0$ or $T_l^{A,0} = 0$ for some l. For any $\left\langle \boldsymbol{w}, \boldsymbol{T^A} \right\rangle \in \mathbb{R}^{2n}_{++}$ we have

$$max_k G_k^L \geq max_k \left(\sum_{n=1}^{n=N} \left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{w_k} + \sum_{n=1}^{n=N} \left(\frac{p_k^M t_{nk}}{P_n^M}\right)^{1-\varepsilon_M} \frac{(1-\gamma) I_n}{w_k}\right) - max_k L_k$$

$$\geq max_{k,n} \left(\left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{w_k} + \left(\frac{p_k^M t_{nk}}{P_n^M}\right)^{1-\varepsilon_M} \frac{(1-\gamma) I_n}{w_k}\right) - max_k L_k$$

and

$$max_k G_k^T \geq max_k \sum_{n=1}^{n=N} \left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{T_k^A} - max_k f_k \geq max_{k,n} \left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{T_k^A} - max_k f_k.$$

Then $max_h G_h(\boldsymbol{w},\, \boldsymbol{T^A}) \to \infty$ will be proved if we can show that either (i) $max_{k,n} \left(\frac{p_k^M t_{nk}}{P_n^M}\right)^{1-\varepsilon_M} \frac{(1-\gamma) I_n}{w_k} \to \infty$, (ii) $max_{k,n} \left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{w_k} \to \infty$ or (iii) $max_{k,n} \left(\frac{t_{nk} z_{kn} c_{kk}^A}{P_n^A}\right)^{1-\varepsilon_A} \frac{\gamma I_n}{T_k^A} \to \infty$ for the factor price sequence $\left\langle \boldsymbol{w^m},\, \boldsymbol{T^{A,m}} \right\rangle$. This is straightforward since (i) and (ii) are verified by (a) $max_n w_n^m \to max_n w_n^0 > 0$ and $min_k w_k^m \to min_k w_k^0 = 0$, while (iii) is verified by (b) $max_n T_n^{A,m} \to max_n T_n^{A,0} > 0$ and $min_k T_k^{A,m} \to min_k T_k^{A,0} = 0$. Since we had assumed that either (a) or (b) holds, this verifies (v). Finally, it is straightforward to show that all five properties hold for the case where countries sell a fixed quantity of land to agricultural producers, $T_l^A = \overline{T}_l^A$, with the solution to the excess demand system given by a vector $\langle \boldsymbol{w},\, \boldsymbol{f} \rangle \in \mathbb{R}^{2n}_{++}$. This completes the proof. □

Appendix B. Trade cost regression

Agricultural bilateral trade costs are obtained by estimating equation (34) using data on agricultural imports between 1995-2004. For those country-pairs for which data is missing, we use the estimates of the pair fixed effects $\hat{\mu}_{nl}$ as a dependent variable and regress them on a set of standard trade cost related covariates and importer and exporter fixed effects. The results from this estimation are reported below in the estimated equation:

$$exp\left[\hat{\mu}_{nl}\right] = exp[\underset{(0.050)}{-0.764} lnDIST_{nl,1} - \underset{(0.041)}{0.781} lnDIST_{nl,2} - \underset{(0.040)}{0.784} lnDIST_{nl,3} - \underset{(0.038)}{0.755} lnDIST_{nl,4} \tag{A.1}$$

$$- \underset{(0.116)}{0.024} CONTIG_{nl} + \underset{(0.068)}{0.670} COMLANG_{nl} + \underset{(0.133)}{1.134} COL_{nl}]. \tag{A.2}$$

The predicted values that we obtain from this regression are then used to complete the full bilateral trade cost matrix. With this trade cost matrix we are then able to compute the multilateral resistance terms. We note that all the regression estimates have the expected sign, reasonable magnitude and are statistically significant; distance impedes trade with elasticities that sit around 0.77, while common language and colonial ties between importer and exporter promote trade with elasticities of respectively 0.67 and 1.13. The effect of a contiguous border between importer and exporter is found to be slightly negative, but the result is statistically insignificant.

Appendix C. Additional regression tables

Table A.1: Unilateral host-country and investory-country regressions. First stage dependent variable only includes host-investor pairs with a positive nummber of deals.

VARIABLES	(1) host FE	(2) host FE	(3) host FE	(4) host FE	(5) host FE	(6) Investor FE	(7) Investor FE	(8) Investor FE	(9) Investor FE	(10) Investor FE
GDP per capita (log)	-0.130	-0.015	-0.127	-0.021	-0.284					
	(0.155)	(0.151)	(0.158)	(0.168)	(0.179)					
Max potential outp. forest and non-forest (log)	0.196**	0.167*	0.200**	0.171*	0.175**					
	(0.090)	(0.092)	(0.095)	(0.088)	(0.085)					
Yield gap	-0.372	-0.316	-0.416	-0.581	-0.821					
	(0.944)	(0.949)	(1.062)	(0.954)	(0.994)					
Land governance	0.017	-0.046	0.016	0.010	0.020					
	(0.262)	(0.209)	(0.267)	(0.236)	(0.233)					
MRDIST host (log)		1.079**		0.942*	0.944*					
		(0.438)		(0.487)	(0.500)					
MRDIST host (log) (AY)			-0.077							
			(0.578)							
MRCOL host (log))				-1.544	-1.233					
				(2.357)	(2.291)					
MRCONTIG host (log))				-3.062	-2.725					
				(4.588)	(4.571)					
MRCOMLANG host (log))				0.429	0.311					
				(0.818)	(0.776)					
MRAGRIVA host (log)					0.215					
					(0.156)					
Total population (log)						0.311***	0.361***	0.322***	0.341***	0.338***
						(0.081)	(0.081)	(0.086)	(0.079)	(0.074)
Agri value added per worker (log)						0.066	0.104	0.079	0.057	-0.030
						(0.070)	(0.086)	(0.070)	(0.091)	(0.099)
No. of global financial centers						0.202	0.350	0.304	0.421*	0.130
						(0.286)	(0.240)	(0.292)	(0.250)	(0.358)
MRDIST investor (log)							0.917*		1.110**	0.800
							(0.476)		(0.526)	(0.590)
MRDIST investor (log) (AY)								0.309		
								(0.284)		
MRCOL investor (log)									5.152**	4.303**
									(1.917)	(2.018)
MRCONTIG investor (log)									2.253	0.444
									(4.528)	(4.932)
MRCOMLANG investor (log)									-1.237	-1.140
									(0.742)	(0.726)
MRGAP investor (log)										1.277
										(0.820)
R-squared	0.184	0.287	0.184	0.302	0.332	0.325	0.371	0.342	0.436	0.472
N	44	44	44	44	44	53	53	53	53	53

Notes: Estimates are obtained with OLS. A constant is included throughout but not reported. Dependent variables represent the country fixed effects (host and investor) from a first-stage bilateral land acquisition regression on physical distance, common colony and common border. All MR terms are based on the Baier and Bergstrand (2009) methodoloy, except those with AY in parentheses which are based on the Anderson and Yotov (2016) methodology (see section 3.2). Robust standard errors are reported in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$

Table A.2: Unilateral host-country and investory-country regressions. First stage dependent variable is the cumulative size of all deals by host-investor pair.

VARIABLES	(1) host FE	(2) host FE	(3) host FE	(4) host FE	(5) host FE	(6) Investor FE	(7) Investor FE	(8) Investor FE	(9) Investor FE	(10) Investor FE
GDP per capita (log)	-3.097***	-1.758*	-3.108***	-1.804*	-1.697					
	(1.095)	(0.999)	(1.028)	(1.050)	(1.355)					
Max potential outp. forest and non-forest (log)	1.481**	1.092**	0.924	1.063*	1.061*					
	(0.621)	(0.541)	(0.575)	(0.552)	(0.555)					
Yield gap	8.092*	5.110	9.928**	5.416	5.416					
	(4.176)	(4.046)	(3.773)	(4.079)	(4.099)					
Land governance	-0.998	0.062	-1.082	-0.007	-0.024					
	(1.330)	(0.999)	(1.171)	(1.017)	(1.036)					
MRDIST host (log)		7.530***		7.764***	7.747***					
		(1.606)		(1.702)	(1.718)					
MRDIST host (log) (AY)			10.901***							
			(2.940)							
MRCOL host (log)				3.691	3.701					
				(6.207)	(6.212)					
MRCONTIG host (log)				8.898	8.789					
				(10.846)	(11.102)					
MRCOMLANG host (log)				-3.105	-3.053					
				(3.776)	(3.686)					
MRAGRIVA host (log)					-0.088					
					(0.824)					
Total population (log)						1.613***	0.886***	1.317***	0.869***	0.872***
						(0.379)	(0.264)	(0.305)	(0.295)	(0.294)
Agri value added per worker (log)						1.203***	1.160***	1.299***	1.101***	0.950**
						(0.394)	(0.256)	(0.304)	(0.318)	(0.441)
No. of global financial centers						0.306	0.337***	0.609***	0.358**	0.355**
						(0.252)	(0.127)	(0.203)	(0.140)	(0.137)
MRDIST investor (log)							2.119***		2.149***	2.193***
							(0.207)		(0.230)	(0.256)
MRDIST investor (log) (AY)								5.845***		
								(0.894)		
MRCOL investor (log)									-5.001	-5.368
									(4.920)	(4.974)
MRCONTIG investor (log)									6.557	5.934
									(10.684)	(10.918)
MRCOMLANG investor (log)									-0.303	-0.201
									(3.055)	(3.049)
MRGAP investor (log)										1.588
										(3.329)
R-squared	0.624	0.734	0.686	0.737	0.737	0.245	0.659	0.516	0.661	0.662
N	86	85	85	85	85	111	111	111	111	111

Notes: Estimates are obtained with OLS. A constant is included throughout but not reported. Dependent variables represent the country fixed effects (host and investor) from a first-stage bilateral land acquisition regression on physical distance, common colony and common border. All MR terms are based on the Baier and Bergstrand (2009) methodoloy, except those with AY in parentheses which are based on the Anderson and Yotov (2016) methodology (see section 3.2). Robust standard errors are reported in parentheses. *** p<0.01; ** p<0.05; * p<0.1

Table A.3: Unilateral host-country and investory-country regressions. First stage includes all deals between 2000 and 2016.

VARIABLES	(1) host FE	(2) host FE	(3) host FE	(4) host FE	(5) host FE	(6) Investor FE	(7) Investor FE	(8) Investor FE	(9) Investor FE	(10) Investor FE
GDP per capita (log)	-3.065***	-1.633**	-3.052***	-1.801**	-1.429					
	(0.838)	(0.777)	(0.851)	(0.747)	(1.100)					
Max potential outp. forest and non-forest (log)	1.161**	0.739	0.587	0.709	0.704					
	(0.561)	(0.470)	(0.503)	(0.473)	(0.472)					
Yield gap	11.162***	8.029**	13.026***	8.448**	8.446**					
	(3.623)	(3.624)	(3.708)	(3.207)	(3.219)					
Land governance	-1.245	-0.164	-1.382	-0.285	-0.343					
	(1.318)	(0.939)	(1.145)	(0.942)	(0.956)					
MRDIST host (log)		7.951***		8.404***	8.344***					
		(1.548)		(1.650)	(1.692)					
MRDIST host (log) (AY)			10.935***							
			(3.055)							
MRCOL host (log)				6.097	6.131					
				(6.994)	(6.903)					
MRCONTIG host (log)				17.389**	17.010*					
				(8.638)	(8.797)					
MRCOMLANG host (log)				-8.319***	-8.138**					
				(3.137)	(3.096)					
MRAGRIVA host (log)					-0.306					
					(0.688)					
Total population (log)						1.637***	0.804***	1.280***	0.739***	0.744***
						(0.397)	(0.212)	(0.263)	(0.241)	(0.239)
Agri value added per worker (log)						1.056***	1.006***	1.172***	0.888***	0.649***
						(0.367)	(0.131)	(0.246)	(0.170)	(0.197)
No. of global financial centers						0.254	0.289***	0.620***	0.289***	0.284***
						(0.250)	(0.098)	(0.191)	(0.099)	(0.100)
MRDIST investor (log)							2.428***		2.461***	2.531***
							(0.228)		(0.246)	(0.250)
MRDIST investor (log) (AY)								7.055***		
								(0.752)		
MRCOL investor (log)									1.044	0.460
									(2.291)	(2.127)
MRCONTIG investor (log)									9.098*	8.108
									(5.296)	(5.222)
MRCOMLANG investor (log)									-1.535	-1.372
									(1.366)	(1.363)
MRGAP investor (log)										2.525**
										(1.080)
R-squared	0.698	0.826	0.767	0.844	0.844	0.252	0.862	0.695	0.866	0.869
N	86	85	85	85	85	111	111	111	111	111

Notes: Estimates are obtained with OLS. A constant is included throughout but not reported. Dependent variables represent the country fixed effects (host and investor) from a first-stage bilateral land acquisition regression on physical distance, common colony and common border. All MR terms are based on the Baier and Bergstrand (2009) methodoloy, except those with AY in parentheses which are based on the Anderson and Yotov (2016) methodology (see section 3.2). Robust standard errors are reported in parentheses. *** p<0.01; ** p<0.05; * p<0.1

Table A.4: Results from bilateral regressions of the number of deals according to three different databases

VARIABLES	(1) All (Grain)	(2) Started (Grain)	(3) All (A&C)	(4) Started (A&C)	(5) All (Land Matrix)	(6) All (Grain)	(7) Started (Grain)	(8) All (A&C)	(9) Started (A&C)	(10) All (Land Matrix)
Distance	-0.676***	-0.595***	-0.202	-0.411***	-0.916***	-0.888***	-0.806***	-0.627***	-0.711***	-1.377***
	(0.074)	(0.064)	(0.159)	(0.078)	(0.097)	(0.078)	(0.067)	(0.155)	(0.073)	(0.093)
Former colonial relationship	0.575*	0.964***	1.257**	1.461***	1.535***	0.224	0.596***	0.774	0.896***	1.338***
	(0.336)	(0.231)	(0.571)	(0.277)	(0.370)	(0.331)	(0.206)	(0.519)	(0.244)	(0.287)
Value net food imports (log)	2.800***	3.338***	3.015***	1.104	0.388	2.063***	2.590***	2.181***	1.014***	0.509
	(0.599)	(0.398)	(0.563)	(0.682)	(1.012)	(0.580)	(0.444)	(0.517)	(0.329)	(0.460)
Total population (log)	0.704***	0.767***	0.836***	0.688***	0.686***	0.708***	0.772***	0.869***	0.669***	0.693***
	(0.069)	(0.052)	(0.068)	(0.052)	(0.063)	(0.096)	(0.078)	(0.104)	(0.094)	(0.100)
Agri value added per worker (log)						0.640***	0.559***	0.615***	0.684***	0.687***
						(0.074)	(0.062)	(0.097)	(0.076)	(0.079)
No. of global financial centers						0.050*	0.038	0.027	0.085***	0.051**
						(0.029)	(0.025)	(0.035)	(0.019)	(0.025)
MRDIST investor (log) (BB)						0.530***	0.471***	0.315***	0.403***	0.322***
						(0.083)	(0.095)	(0.106)	(0.081)	(0.057)
Landlocked	-0.401*	-0.389**	-1.075***	0.028	0.213	-0.646**	-0.630***	-1.313***	-0.103	-0.010
	(0.229)	(0.197)	(0.366)	(0.231)	(0.223)	(0.268)	(0.224)	(0.391)	(0.267)	(0.252)
Max potential outp. non-forest (log)	0.674***	0.659***	0.897***	0.635***	0.572***	0.638***	0.605***	0.879***	0.614***	0.441**
	(0.108)	(0.092)	(0.152)	(0.090)	(0.162)	(0.126)	(0.101)	(0.190)	(0.095)	(0.194)
Max potential outp. forest (log)	-0.164**	-0.153**	-0.268***	0.059	0.060	-0.172	-0.166*	-0.298**	0.065	0.189*
	(0.071)	(0.060)	(0.087)	(0.058)	(0.086)	(0.107)	(0.089)	(0.131)	(0.072)	(0.109)
Yield gap	0.786	0.757	0.588	1.692**	1.193*	-0.448	-0.310	-2.525***	0.914	-0.773
	(0.704)	(0.555)	(0.777)	(0.690)	(0.676)	(0.784)	(0.611)	(0.971)	(0.640)	(0.889)
Value food exports (log)	0.176***	0.104***	0.001	0.122**	0.208***	0.227***	0.119**	0.024	0.156**	0.325***
	(0.046)	(0.038)	(0.050)	(0.050)	(0.061)	(0.072)	(0.054)	(0.082)	(0.065)	(0.113)
Land governance	-0.507***	-0.457***	-0.426**	-0.320**	-0.760***	-0.050	-0.109	0.192	0.058	0.265*
	(0.138)	(0.114)	(0.183)	(0.133)	(0.142)	(0.178)	(0.152)	(0.224)	(0.165)	(0.160)
Law and order	-0.109	-0.081	-0.031	0.221**	0.004	0.180*	0.181**	0.365**	0.517***	0.335**
	(0.111)	(0.093)	(0.167)	(0.110)	(0.142)	(0.104)	(0.080)	(0.161)	(0.102)	(0.149)
Weak investor protection	-0.054	-0.114	-0.246*	0.049	-0.080	0.025	-0.013	0.033	0.116	0.151
	(0.098)	(0.081)	(0.129)	(0.087)	(0.128)	(0.108)	(0.085)	(0.124)	(0.088)	(0.129)
GDP Per Capita (log)						-0.532***	-0.299**	-0.768***	-0.235*	-1.337***
						(0.146)	(0.117)	(0.194)	(0.136)	(0.177)
MRDIST host (log) (BB)						1.435***	1.573***	2.626***	1.404***	2.377***
						(0.353)	(0.305)	(0.614)	(0.235)	(0.463)
Observations	18,333	18,333	18,333	18,333	18,333	12,057	12,057	12,057	12,057	12,057
Pseudo R-squared	0.243	0.276	0.275	0.276	0.350	0.361	0.377	0.392	0.425	0.529

Notes: Estimates are obtained with the PPML estimator. A constant is included throughout but not reported. Columns 1-5 reproduce Table 7 from Arezki et al. 2015. Columns 6-10 extend columns 1-5 with additional regressors as discussed in main text. Robust standard errors clustered by country-pair are reported in parentheses. *** p<0.01; ** p<0.05; * p<0.1