STEG WORKING PAPER

THE FOOD PROBLEM AND STRUCTURAL CHANGE IN AN IMPERFECTLY OPEN ECONOMY

Lisa Martin

MARCH 2025
STEG WP091
PhD 1280

The food problem and structural change in an imperfectly open economy*

Lisa Martin†

September 2024

Abstract

Most models of structural transformation are based on closed economy frameworks in which the price mechanism leads the sectoral composition of production to match domestic consumption. An obvious criticism is that in an open economy framework, production will instead reflect the economy's comparative advantage with respect to the rest of the world. In particular, the "food problem" of Gollin et al. (2007) would not bind in a low-income economy with poor agricultural productivity; such an economy could simply import food. This paper shows that in an open economy with internal spatial frictions, the food problem can continue to bind. Domestic spatial frictions matter crucially for the sectoral and spatial allocation of economic activity, for the economy's pattern of international trade, and for spatial differences in food consumption.

Keywords— Structural Transformation, Spatial General Equilibrium, Trade, Transport Costs, Food Problem

*I am grateful to my PhD supervisors, Douglas Gollin and Simon Quinn, for their invaluable guidance and support. I also thank Christopher Adam, Channing Arndt, Banu Demir and seminar participants at Oxford's Centre for the Study of African Economies (CSAE) for helpful comments and suggestions. I gratefully acknowledge funding for this work from STEG through a PhD Research Grant.

†University of Oxford, Department of Economics

1 Introduction

In most low-income countries, a large share of the population remains engaged in subsistence agriculture. The persistence of low-productivity subsistence agriculture has several potential causes, but one widely invoked explanation is that, in a closed economy, low agricultural productivity combines with a minimum consumption requirement for food to imply that a high share of the economy's labour will be devoted to agriculture. This situation was described by Schultz (1953) as the "food problem," and it lies at the heart of more recent work on structural transformation, such as Gollin, Parente, et al. (2002) and Gollin, Parente, et al. (2007). By contrast, in a perfectly open economy that is integrated into world markets, we would not expect the food problem to bind in the same way: countries with low agricultural productivity might import food to meet subsistence needs. Countries would produce the goods in which they have a comparative advantage and would import those that they produce (relatively) inefficiently. Since few low-income economies are well-described as closed economies, this poses a puzzle. Why do so many people work in agriculture – a sector that appears to be low in both absolute and relative productivity?

In this paper, I show that the food problem can still bind in economies that are fully open at the border but that display spatial heterogeneity and face domestic spatial frictions. In these economies, some regions may be only loosely integrated with world markets, and subsistence needs may exercise considerable influence over the sectoral allocation of resources. This paper studies processes of structural change in such *imperfectly* open economies. I develop a stylised spatial general equilibrium model of a small open agricultural economy with spatial heterogeneity and high domestic trade frictions. Specifically, I analyse how domestic trade frictions, agricultural and non-agricultural productivity, and the country's terms of trade interact to shape processes of structural transformation such as labour reallocation across sectors and rural-urban migration.

Because of domestic trade frictions, world prices of goods – both consumption goods and intermediates – are passed through imperfectly into the domestic economy. The pass-through is most obviously mediated by the spatial structure of the economy, because of domestic trade costs. But it is also differentiated across sectors, because certain production activities tend to be concentrated in particular locations. For instance, the model locates the manufacturing sector in an urban area that is closely linked to the world market, so manufacturing faces world prices fairly directly. But agriculture takes place in both "near" and "far" rural locations, with different outputs and different input demands. The patterns of agricultural specialisation and production are thus shaped by characteristics of the domestic market as well as by world prices.

This paper links to a large body of literature on structural transformation (Matsuyama, 1992; Gollin, Parente, et al., 2002; Gollin and Rogerson, 2014; Ngai and Pissarides, 2007; Bustos et al., 2016; McCaig and Pavcnik, 2013; Fajgelbaum and Redding, 2022). Of these, only Matsuyama (1992) directly addresses the case of fully open economies. Matsuyama (1992) shows that the food problem fully disappears for an open economy and further argues that high agricultural productivity can in fact limit structural change in this case. Bustos et al. (2016) view local economies within Brazil as open, in the sense that they are price takers for agricultural and non-agricultural goods. But their framework does not allow mobility of labour between locations. This paper is perhaps closest to a set of papers that study the food problem in open economies (Tombe, 2015; Nath, 2020) and to two other papers that focus on the link between trade openness and food security (Burgess and Donaldson, 2012; Janssens et al., 2020).

More broadly, this paper relates to the growing literature that focuses on the role of intra-national trade frictions (Allen and Arkolakis, 2014; Atkin and Donaldson, 2015; Donaldson, 2018; Sotelo, 2020). The analysis in this paper is closer, however, to the simple but intuitive multi-sector multi-region general equilibrium model of Gollin and Rogerson (2014). To adapt the framework to my research questions, I extend that model in two important dimensions. First, I incorporate two additional sectors, a non-food (cash crop) agricultural sector and an urban food processing sector, which capture defining features of agricultural production and food consumption in many low-income countries.[1] Second, I open the economy to trade with the rest of the world. My model economy exports cash crops and imports food, manufactured goods, and intermediate inputs such as fertiliser.

The model economy has a stylised linear geography which features three domestic regions. Regions differ with respect to factor endowments (agricultural land), remoteness (transport costs) and production activities (agricultural, non-agricultural). The urban population lives in a city – which is imagined to occupy a coastal location that makes it the hub for all international trade. The rural population is divided into two regions, one "near" and the other "far". The near rural region and the far rural region specialise in agricultural production, while the urban region specialises in manufacturing and food processing. The urban region functions as the nexus between the economy's rural areas and the rest of the world, connecting the domestic economy to international markets. The economy is populated with a mass of identical individuals which is normalised to n. All individuals supply one unit of labour inelastically. Since there is no commuting, individuals work and consume in their region of residence. Goods traded across regions are subject to iceberg transport costs. Agricultural land is split between the far rural region and the near rural region, where it is used as an input in agricultural production. Total land is fixed within each region, and it is supplied for agriculture inelastically and with no fixed costs. There is no extensive margin and no cost of land clearing.

Non-homothetic preferences in food consumption and trade frictions, both at the border and domestically, are key features that allow application of the model to a range of different scenarios, depending on the exact parameterisation. Before focusing in more detail on a

[1] This non-food agricultural sector refers to the production of export cash crops that are not consumed domestically. Export crops differ from staple food crops in the sense that they do not have any nutritional value for domestic consumers.

specific parameterisation, I document three key properties of the model, relating to the role of agricultural productivity for the spatial distribution of population, the role of domestic transport costs for the sectoral allocation of economic activity, and the role of domestic transport costs for patterns of food consumption. First, agricultural productivity has a strong effect on the spatial distribution of population, but the direction and magnitude of the effect depend crucially on the degree of external openness and internal trade frictions. Second, spatial frictions matter for the sectoral allocation of economic activity and for the economy's pattern of international trade. Changes in spatial frictions (such as those that might be delivered by investments in transportation infrastructure) have the potential to alter both the domestic patterns of economic activity and the economy's patterns of trade. Third, domestic transport costs cause diverging food consumption patterns across regions and limit the penetration of imported goods to remote rural areas. Spatial frictions are a key determinant of food consumption patterns, as transport costs introduce a wedge between producer and consumer prices of food, and thus change the relative price of different food types in rural and urban regions. Rising transport costs drive overall food consumption closer to the subsistence level in all three regions in the economy, and they skew food consumption baskets to favor locally produced food, leading to diverging consumption patterns across regions.

After exploring the general properties of the model, I present a version of the model which is calibrated to Tanzanian data.[2] Tanzania is a small open economy with sizable domestic trade frictions. The economy is characterised by a large agricultural sector with relatively low productivity (Adam et al., 2018).[3] Although Tanzania registered robust economic growth over the past two decades, poverty remains widespread and a considerable part of the population continues to be food insecure (WFP, 2022).[4] For these reasons, Tanzania provides a relevant context in which to explore the research question. In addition, data availability for Tanzania is significantly better than for many other developing countries. I calibrate the model to the economy of Tanzania by estimating parameters from micro data, fitting parameters to aggregate data and targeting empirical moments such as regional population distributions and crop-specific land use patterns to set the remaining free parameters.

The calibrated model is then used to contrast comparative statics results for different productivity improvements in the economy. I consider productivity improvements in the transport sector, the agricultural sectors, and the non-agricultural sectors, and finally a combination of the three individual productivity improvements. All of these potential productivity im-

[2] While I choose to calibrate the model to Tanzanian data, the research questions in this paper are relevant for a broad set of low-income countries that exhibit similarly high domestic trade frictions, high agricultural employment shares, and low food imports. Empirical data from sub-Saharan African countries shows that low food import dependence is associated with low GDP per capita, high agricultural employment shares, and weak infrastructure (see Figure 6 in Appendix A).

[3] The latest World Bank data, which is based on an ILO estimate, reports agricultural employment in Tanzania as 65% of total employment in 2019.

[4] World Bank national accounts data reports annual GDP growth rates of 4-8% in 2000-2021, except in 2020 (2%).

provements would be expected to make the food problem less binding; they all lead to a higher urban population share, an expansion of the manufacturing sector, and a reduction in subsistence agriculture. However, these counterfactuals show that different productivity improvements have different effects on trade patterns, the spatial distribution of economic activity, and food consumption patterns.

The remainder of this paper is structured as follows: Section 2 briefly summarises related literature and highlights the contribution of this paper. Section 3 contains details of the modelling approach and Section 4 illustrates some of the model's key properties. Section 5 presents the calibration of the model to Tanzanian data, and Section 6 describes the baseline equilibrium at these calibration values. Section 7 presents results from comparative statics, contrasting the effects of different productivity improvements. Section 8 concludes and defines avenues for future research.

2 Related literature

This paper relates to three broad strands of literature. First, this paper links to the large empirical and theoretical literature on structural transformation, both in closed economies (Matsuyama, 1992; Gollin, Parente, et al., 2002; Gollin and Rogerson, 2014; Ngai and Pissarides, 2007; Emerick, 2018) and in open economies (McCaig and Pavcnik, 2013; Bustos et al., 2016; Fajgelbaum and Redding, 2022). In this structural transformation literature, different modelling approaches are used to generate structural change in the economy. One is a technological approach, in which differential sectoral Total Factor Productivity (TFP) growth drives shifts in the allocation of economic activity (Ngai and Pissarides, 2007). The other is a utility-based approach, in which non-homothetic preferences yield structural change (Matsuyama, 1992; Gollin, Parente, et al., 2002; Gollin and Rogerson, 2014). In both approaches, domestic output prices are fully endogenous, such that structural change is driven by changes in domestic relative prices.

In a small open economy, however, domestic prices are (at least partially) pinned down by exogenous world market prices, which can alter the effect of sectoral productivity improvements on the domestic economy. The utility-based approach to structural transformation suggests that in a closed economy in which households have minimum food consumption requirements, an increase in agricultural productivity induces growth as it releases labour from agriculture into the manufacturing sector. In an open economy, on the other hand, an increase in agricultural productivity may pull labour into agriculture, squeezing out the manufacturing sector. Similar to a 'Dutch disease' phenomenon, this can put the economy on a lower growth path if we assume that learning-by-doing occurs only in the manufacturing sector and that as such, only manufacturing promotes economic growth in the long run (Matsuyama, 1992).

The second strand of literature subsumes research on the role of trade in decoupling food consumption from domestic agricultural production. In a closed economy with low agricultural productivity, a large share of the population engages in subsistence agriculture to produce sufficient food to meet subsistence needs. In an open economy, however, we would not necessarily expect the food problem to bind in the same way, as economies may import food from global markets. However, this pattern is not observed empirically in trade data. Despite low productivity in agriculture, developing countries tend to import only a small share of domestic food consumption, while a large share of the population continues to engage in subsistence agriculture. There exists a small number of studies on the potential causes of this economic puzzle. Previous research points to the role of trade frictions at the border, such as tariffs and delays, in explaining low import shares for food (Tombe, 2015). This paper contributes to previous research in this area by showing that domestic transport costs can generate the same result as trade frictions at the border. The results suggest that, even as openness to trade is improved at the border, domestic transport costs have the potential to lock the economy in at a low level of food imports and a high employment share in agriculture.

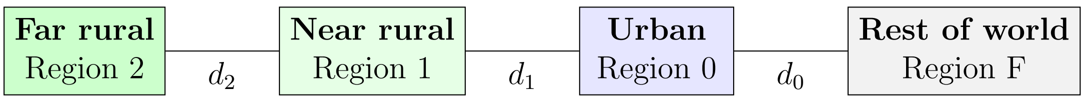
This finding is particularly relevant against the backdrop of accelerating climate change. Nath (2020) shows that, if trade costs remain high, climate change may exacerbate the food problem, such that developing countries with binding subsistence constraints intensify their specialisation in agriculture as increasing temperatures erode agricultural productivity. Model counterfactuals in Nath (2020) suggest that reducing trade costs could reduce the negative impact of climate change in these countries, as this would allow labour to shift from agriculture into non-agricultural sectors, where productivity is less vulnerable to climate change. Other studies focus on short-term volatility rather than long-term structural shifts and find that lower trade costs can help to stabilise food security by allowing for more diversified supply, thereby insuring against local production short-falls (Burgess and Donaldson, 2012; Janssens et al., 2020; Baptista et al., 2023).

Lastly, this paper relates more broadly to the growing spatial economics literature on the role of domestic trade frictions, which are particularly high in low-income countries (Allen and Arkolakis, 2014; Donaldson, 2018; Atkin and Donaldson, 2015; Sotelo, 2020; Farrokhi and Pellegrina, 2023). The models in this Quantitative Spatial General Equilibrium (QSGE) literature often feature very granular spatial differentiation and allow for spatial frictions to play an important role in determining patterns of production and trade. The granularity of the spatial structure in these models makes it more difficult to detect broader patterns in the ways that different types of regions relate to one another. In contrast to the QSGE literature, the model I develop in this paper has a simpler spatial structure that allows a more comprehensible and intuitive characterisation of the ways in which different regions interact. This stylised modelling approach provides a transparent illustration of the role of domestic trade frictions for persisting subsistence agriculture, processes of structural transformation, and trade patterns.

3 Model

I set up a stylised general equilibrium model of a small open economy called *Home*. *Home's* geography features three regions which differ with respect to remoteness (transport costs), production activities (agricultural, non-agricultural) and factor endowments (agricultural land). Figure 1 shows a schematic overview of *Home's* geography. Following Gollin and Rogerson (2014), I assume a simple linear geography. The urban population of this economy lives in a city on the coast; the rural population is divided into two regions, one 'near' and the other 'far'. The far rural region (region 2) and the near rural region (region 1) specialise in agricultural production, while the urban region (region 0) specialises in manufacturing and food processing. The urban region functions as the nexus between *Home's* rural areas and the rest of the world (region F), connecting the domestic economy to international markets. *Home* is a price taker, trading with the rest of the world at exogenous world market prices.

Figure 1: Schematic drawing of the model



Home is populated with a mass of identical individuals which is normalised to n. All individuals supply one unit of labour inelastically. There is no commuting, so individuals are assumed to work and consume in their region of residence. Migration from one region to another is costless and frictionless. By contrast, goods traded across regions are subject to iceberg transport costs. Agricultural land (l) is split between the far rural region (l_2) and the near rural region (l_1), where it is used as an input into agricultural production. Total land is fixed within each region, and it is supplied for agriculture inelastically and with no fixed costs. These assumptions remove, for tractability purposes, the possibility of expansion of agriculture on the extensive margin (e.g., through costly land clearing).

3.1 Consumption

All individuals in *Home* have the same utility function, irrespective of the region (r) in which they reside. Individuals derive utility from the consumption of food (f), for which there exists a subsistence requirement ($\bar{c}_f$), and a manufactured good (m). Utility is modeled using a two-tier nested structure, with households trading off consumption of food and the manufactured good at the top level

$$u(C_{r,f}, C_{r,m}) = \alpha \log\left(C_{r,f} - \bar{c}_f\right) + (1 - \alpha) \log(C_{r,m})$$

where α is a preference parameter, $C_{r,f}$ denotes consumption of food (with $\bar{c}_f$ the subsistence level) and $C_{r,m}$ denotes consumption of the manufactured good. On the lower tier, households choose consumption quantities of the staple food crop (a) and processed food (p)

$$C_{r,f} = \left[\lambda_f C_{r,a}^{\rho_f} + (1-\lambda_f) C_{r,p}^{\rho_f}\right]^{\frac{1}{\rho_f}}$$

where λ_f is a share parameter, ρ_f is a preference parameter, $C_{r,a}$ denotes consumption of the staple food crop and $C_{r,p}$ denotes consumption of processed food.

3.2 Production

There are four sectors of production in the economy: export cash crop agriculture (c), staple food crop agriculture (a), food processing (p), and manufacturing (m). Both rural regions, near and far, produce the two agricultural goods: a cash crop, which is a pure export good, and a staple food crop, which is only consumed domestically. The urban region specialises in manufacturing and food processing. The manufactured good is both consumed directly by households and used as an intermediate input in the production of other goods.

3.2.1 Export crop

Both rural regions (near and far) have the same technology for agricultural production of the export cash crop (c), using land, labour, and an intermediate input

$$Q_{r,c} = A_{r,c}\, l_{r,c}^{(1-\gamma_x-\gamma_n)}\, x_{r,c}^{\gamma_x}\, n_{r,c}^{\gamma_n}$$

where A is a region- and sector-specific productivity parameter, l is land, x is an intermediate input (e.g. fertiliser), n is labour, and γ denotes output elasticities. The cash crop is a pure export good, with *Home's* entire production being sold to *Foreign* at the international market price p_c. In the context of Tanzania's agricultural sector, examples of such export cash crops are coffee and cashew nuts.

3.2.2 Staple crop

Analogous to export crop production, both rural regions (near and far) use the same technology for production of the staple food crop (a)

$$Q_{r,a} = A_{r,a}\, l_{r,a}^{(1-\theta_x-\theta_n)}\, x_{r,a}^{\theta_x}\, n_{r,a}^{\theta_n}$$

where A is a region- and sector-specific productivity parameter, l is land, x is an intermediate input (e.g. fertiliser), n is labour, and θ denotes output elasticities. The staple crop is both consumed directly by households and used as an input into food processing. The staple food crop is neither exported nor imported, with the domestic market clearing perfectly. This simplifying assumption is motivated by the empirical observation that Tanzania is self-sufficient in the production of major staple food crops such as maize or cassava and exports little or none of these crops to other countries.

3.2.3 Processed food

The urban food processing sector turns the staple food crop into processed food by combining staple crop, intermediate inputs and labour

$$Q_{0,p} = A_{0,p}\, a_{0,p}^{(1-\delta_x-\delta_n)}\, x_{0,p}^{\delta_x}\, n_{0,p}^{\delta_n}$$

where A is region- and sector-specific total factor productivity, a is the staple crop which is being processed, x is an intermediate input (e.g. tin cans), n is labour, and δ denotes output elasticities. In addition to being produced domestically, processed food can also be imported from *Foreign* at the exogenous world market price p_p. The domestically produced variety is combined with the imported variety in the urban region to produce the composite processed food for consumption in *Home*

$$\bar{Q}_p = \bar{A}_p \Big[\lambda_p\, Q_{0,p}^{\rho_p} + (1-\lambda_p)\,\big((1-d_0)\,Q_{F,p}\big)^{\rho_p}\Big]^{\frac{1}{\rho_p}}$$

where $\bar{A}_p$ is a productivity parameter, $Q_{0,p}$ is the domestic production quantity, $Q_{F,p}$ is the imported quantity, d_0 is the transport cost between the urban region and the rest of the world, λ is the share parameter of the domestically produced good, and ρ is the elasticity parameter. The domestic and the imported variety are imperfect substitutes, with an elasticity of substitution of $\epsilon = \frac{1}{1-\rho}$. An example of such a processed food item in the context of Tanzania is vegetable cooking oil, varieties of which can be produced domestically using e.g. domestically grown sunflower seeds, or imported from other countries, mostly in the form of palm oil.

3.2.4 Manufactured good

In addition to processed food, the urban region also produces the manufactured good, using labour as the sole factor of production

$$Q_{0,m} = A_{0,m}\, n_{0,m}^{\omega_n}$$

Analogous to the processed food sector, there is also an imported variety of the manufactured good. This imported variety, which is an imperfect substitute for the domestically produced manufactured good, can be purchased from *Foreign* at price p_m. The two varieties, domestic and imported, are combined into a composite manufactured good ($\bar{Q}_m$) which can be consumed directly ($C_{r,m}$) or used as an intermediate input in the production of other goods ($x_{r,a}$, $x_{r,c}$, $x_{0,p}$):[5]

[5]The manufactured consumption good, m, and the intermediate production input, x, are assumed to be perfect substitutes here. Their world market prices are assumed to be identical ($p_m = p_x$). $Q_{F,m}$ and $Q_{F,x}$ are still included as separate quantities in the model to simplify calibration at later stages, when the model will be taken to data.

$$\bar{Q}_m = \bar{A}_m \Big[\lambda_m \, Q_{0,m}^{\rho_m} + (1 - \lambda_m) \left((1 - d_0) \, (Q_{F,m} + Q_{F,x} \right)^{\rho_m} \Big]^{\frac{1}{\rho_m}}$$

$\bar{A}_m$ is a productivity parameter, $Q_{0,m}$ is the domestic production quantity, $Q_{F,m}$ and $Q_{F,x}$ are the imported quantities, d_0 is the transport cost between the urban region and the rest of the world, λ is a share parameter, and ρ is the elasticity parameter. The elasticity of substitution between the two inputs is $\epsilon = \frac{1}{1-\rho}$. Examples of such imperfect substitutes could be imported cement and domestically produced bricks, or garments imported from China and garments from domestic textiles production.

3.3 Trade and transport costs

3.3.1 Domestic trade

The movement of goods between different regions in the domestic economy is subject to iceberg transport costs. The transport sector is efficient, in the sense that there are no rents or mark-ups, and transport costs represent technological trade frictions. If one unit of a good is transported between the urban region and the near rural region, only $(1 - d_1)$ units arrive at the destination. As goods can only be traded along the paths of the trade network, transport costs accumulate multiplicatively. That is, if one unit of a good is transported from the urban region to the far rural region, only $(1 - d_1)(1 - d_2)$ units of the good arrive in the far rural region. I assume for simplicity that transport costs are the same across all commodities and symmetric between regions (i.e., that the iceberg cost of shipping goods from the near rural region to the far rural region are the same as the cost of shipping goods from far to near).

3.3.2 International trade

Home is open to trade with the rest of the world. The entire domestic production of the cash crop (c) is exported to global markets. Foreign varieties of the manufactured good (x, m) and the processed food (p) are imported; as described above, these imported varieties are imperfect substitutes for domestic goods. *Home* is a price-taker in international markets, selling the export cash crop and purchasing the manufactured good and processed food at exogenous world market prices p_c, p_m (p_x), and p_p, respectively. International trade flows, like domestic trade flows, are subject to a symmetric iceberg transport cost (d_0), which reflects an exogenous technological trade friction. As the economy is otherwise assumed to be perfectly open to trade, there are no tariffs or duties on international trade. All traded goods have to pass through the urban centre, which is the only region that has a direct trade link with rest of the world. Note that for a certain parameterisation of this model, the domestic economy could conceivably be in autarky, as there exist domestically produced substitutes for all imported goods (p, m, x). In such an autarky equilibrium, rural regions would produce only the staple food crop, but not the export cash crop, since that would have no domestic market.

3.3.3 Direction of trade flows

Figure 2 shows the direction of feasible trade flows by sector. While any cross-regional trade in processed food and the manufactured good flows "from right to left", or *centre-to-periphery*, all traded volumes of the export cash crop flow the opposite way, "from left to right", or *periphery-to-centre*. Note that trade in the staple crop, on the other hand, need not be unidirectional, as traded quantities can originate in the far or the near rural region.

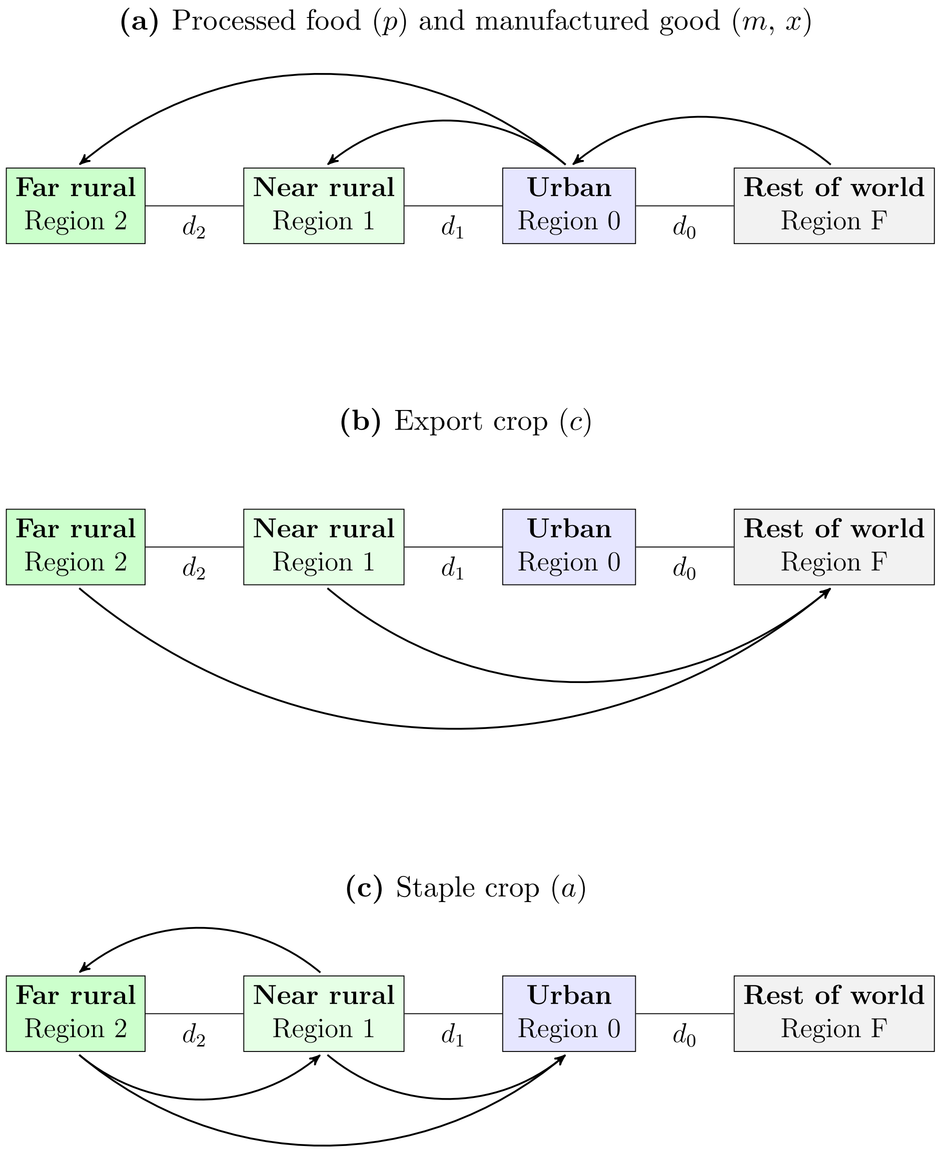


Figure 2: Direction of feasible trade flows by sector

There could in principle be an equilibrium in which the far rural region specialises in the production of the export cash crop and receives staple crop from the near region. In this case, traded volumes of the staple crop would be flowing both "from right to left", i.e. from near to far, and "from left to right", i.e. from near to urban.[6] This potential bidirectionality of staple crop trade flows, as opposed to the unidirectionality of flows in all other sectors, proves important for the formulation of market clearing conditions in Section 3.5 below.

3.4 Equilibrium

A good starting point for understanding the comparative statics of this model is to consider the social planner's problem, in which the utility of the individuals in the economy is maximised subject to feasibility constraints. The social planner maximises aggregate utility in *Home*, putting equal weight on each individual. Specifically, the social planner maximises the population-weighted sum of regional utilities,

$$\sum_{r=0}^{2} \frac{n_r}{n} \Big[\alpha \ \log \left(C_{r,f} - \bar{c}_f \right) + (1-\alpha) \ \log(C_{r,m}) \Big] \tag{1}$$

Note that, since all individuals in *Home* have the same preferences and all individuals in a given region face the same prices, consumption allocations will be identical for individuals residing in the same region. Equation (1) serves as the objective function in the numerical optimisation.

The utility maximisation in Equation (1) is subject to feasibility conditions with respect to market clearing, balanced trade, and factor utilisation. These linear and non-linear constraints require perfect market clearing in all four sectors of production, strictly balanced trade with *Foreign*, full employment in the economy and full utilisation of agricultural land in both rural region.

3.5 Commodity market clearing

Market clearing conditions require that supply must exactly equal demand in all four sectors. Depending on the sector, the supply side contains quantities from domestic production and imports from *Foreign*, while the demand side contains quantities that are consumed domestically, quantities used in domestic production and exports to *Foreign*.

Cash crop In order for the export crop market to clear, total export crop production from the far region ($Q_{2,c}$) and the near region ($Q_{1,c}$) has to be equal to the total amount exported to *Foreign*, accounting for transport costs between regions.

$$(1-d_1)Q_{1,c} + (1-d_1)(1-d_2)Q_{2,c} = \frac{C_{F,c}}{(1-d_0)} \tag{C.1}$$

[6] Note that, as staple crop produced in far and staple crop produced in near are perfect substitutes, there will never be bilateral trade in staple crop between the two rural regions. That is, we will never simultaneously observe staple crop trade flows from near to far and from far to near.

Staple crop Staple crop market clearing is slightly more complex than export crop market clearing, due to the potential bi-directionality of trade flows discussed in Section 3.3.3 above. The way in which transport costs enter the market clearing condition depends crucially on the direction of trade flows observed in equilibrium.

If we assume that the far rural region is either self-sufficient or a net exporter of the staple crop, i.e. if any traded volumes of the staple crop flow only from the far region to the near region and the urban region, the market clearing condition takes the following form

$$Q_{2,a} + \frac{Q_{1,a}}{(1-d_2)} = n_2 C_{2,a} + \frac{n_1 C_{1,a}}{(1-d_2)} + \frac{(n_0 C_{0,a} + a_{0,p})}{(1-d_1)(1-d_2)} \tag{C.2.a}$$

If on the other hand we assume that the far rural region is either self-sufficient or a net importer (that is, any staple crop trade flows between regions originate only from the near rural region, with some of its excess production flowing to urban and some of it flowing to far), the market clearing condition must be:

$$Q_{2,a} + (1-d_2)Q_{1,a} = n_2 C_{2,a} + (1-d_2) n_1 C_{1,a} + \frac{(1-d_2)(n_0 C_{0,a} + a_{0,p})}{(1-d_1)} \tag{C.2.b}$$

There are in principle two approaches to deal with this complexity in staple crop market clearing. The first approach is to formulate a clearing condition that nests both regimes. The second approach is to solve the model twice, one time adding a condition that ensures the far rural region is either self-sufficient or a net exporter of the staple crop, and a second time adding a condition that ensures the far rural region is either self-sufficient or a net importer of the staple crop. Note that both times nest the corner case in which the far rural region produces just enough staple crop to meet own consumption needs, i.e. the far rural region is self-sufficient. In solving for the baseline equilibrium in Section 6, I implement the second approach.

Processed food The processed food market clearing condition states that the total supply of processed food, which consists of both domestic production as well as imports, must equal the sum of aggregate consumption volumes in all three regions, accounting for transport costs along the way:

$$\bar{Q}_p = n_0 C_{0,p} + \frac{n_1 C_{1,p}}{(1-d_1)} + \frac{n_2 C_{2,p}}{(1-d_1)(1-d_2)} \tag{C.3}$$

Manufactured good Similarly, market clearing for the manufactured good requires that total supply, again containing both the domestically produced as well as the imported variety, must be equal to the sum of aggregate consumption volumes in all three regions, and the quantities used as inputs to production in the different sectors, accounting for transport costs along the way

$$\bar{Q}_m = n_0 C_{0,m} + x_{0,p} + \frac{n_1 C_{1,m} + x_{1,a} + x_{1,c}}{(1-d_1)} + \frac{n_2 C_{2,m} + x_{2,a} + x_{2,c}}{(1-d_1)(1-d_2)} \tag{C.4}$$

3.6 Balanced trade

International trade Trade must be strictly balanced in value terms. *Home* does not have a capital account, so the current account must also be balanced to achieve balance of payments. *Home* exports the cash crop at price p_c and imports processed food at price p_p, the manufactured consumption good at price p_m, and the intermediate production input at price p_x. T represents net unilateral transfers such as international aid and remittances. In the absence of international transfers ($T = 0$), *Home* must run a zero trade balance to ensure this feasibility condition holds:

$$\underbrace{p_m Q_{F,m} + p_x Q_{F,x} + p_p Q_{F,p}}_{\text{imports}} = \underbrace{p_c\left[(1-d_1)Q_{1,c} + (1-d_1)(1-d_2)Q_{2,c}\right]}_{\text{exports}} + \underbrace{T}_{\text{transfers}} \tag{C.5}$$

Domestic trade If the social planner's problem is to be implemented as a competitive equilibrium, then domestic trade must be strictly balanced in value terms, i.e. the social planner cannot make transfers between regions.[7] A region's aggregate expenditure on the three consumption goods (staple food, processed food, and the manufactured good) must be fully accounted for by the region's income from production. Prices are indexed by region (0, 1, 2) as the spatial frictions (d) lead to spatially differentiated prices.

$$\underbrace{n_0(p_{0,a}C_{0,a} + p_{0,p}C_{0,p} + p_{0,m}C_{0,m})}_{\text{expenditure in urban}} = \underbrace{p_{0,p}Q_{0,p} + p_{0,m}Q_{0,m}}_{\text{income in urban}} \tag{C.6}$$

$$\underbrace{n_1(p_{1,a}C_{1,a} + p_{1,p}C_{1,p} + p_{1,m}C_{1,m})}_{\text{expenditure in near rural}} = \underbrace{p_{1,c}Q_{1,c} + p_{1,a}Q_{1,a}}_{\text{income in near rural}} \tag{C.7}$$

$$\underbrace{n_2(p_{2,a}C_{2,a} + p_{2,p}C_{2,p} + p_{2,m}C_{2,m})}_{\text{expenditure in far rural}} = \underbrace{p_{c,2}Q_{2,c} + p_{2,a}Q_{2,a}}_{\text{income in far rural}} \tag{C.8}$$

3.7 Factor market clearing

Regional and sectoral factor endowments of land and labour must fully be utilised in production, such that labour and land markets clear.

Labour There must be full employment in *Home's* economy. Every individual must work in some sector and reside in some region, such that the sum of labour shares $n_{r,j}$ is equal to

[7] By contrast, if the social planner is unconstrained, or if the inhabitants of different regions are viewed as members of one household, able to make transfers costlessly across regions, then a different equilibrium might be attained. The presence of spatial frictions means that this is not a setting where the first and second welfare theorems hold unconditionally.

total population n

$$n_{2,a} + n_{2,c} + n_{1,a} + n_{,c} + n_{0,p} + n_{0,m} = n \tag{C.9}$$

Land Agricultural land in both the far rural region (l_2) and the near rural region (l_1) must be fully utilised. While the overall land endowment in each region is fixed ex-ante, the sectoral shares are choice variables in the optimisation

$$l_{2,a} + l_{2,c} = l_2 \tag{C.10}$$

$$l_{1,a} + l_{1,c} = l_1 \tag{C.11}$$

4 Characterising the model

This section illustrates the model's key properties to elucidate how different model features drive the qualitative predictions. By adjusting the parameterisation, the model can flexibly accommodate different scenarios, such as an open economy with low domestic transport costs or a closed economy with high domestic transport costs. I solve the model at different parameterisations to highlight the importance of, and interaction between, openness at the border and internal trade frictions in the model economy.[8] I present three key properties of the model below, relating to (1) the role of agricultural productivity for the spatial distribution of population, (2) the role of domestic transport costs for the sectoral allocation of economic activity, and (3) the role of domestic transport costs for patterns of food consumption.

Property 1. *The relationship between agricultural productivity levels and the spatial distribution of population depends on the degree of external trade openness and internal spatial frictions.*

In the model, agricultural productivity has a strong effect on the spatial distribution of population, but the direction and magnitude of the effect depend crucially on the degree of external openness and internal trade frictions. In a closed economy, an increase in agricultural productivity releases labour from subsistence agriculture into manufacturing, which implies a shift in population from rural regions to the urban region. This effect is well documented in the structural transformation literature, where most models depict a closed economy without domestic spatial frictions. However, as noted by Matsuyama (1992), the positive effect of rising agricultural productivity on urbanisation may be reversed if the economy is open to trade with the rest of the world. In an open economy, an increase in agricultural productivity provides an incentive to increase the production of agricultural exports. Labour moves into the agricultural sector to produce export cash crops, which implies a shift in population from the urban region to rural regions, as agricultural land is spatially dispersed and not mobile.

[8] In this section of the paper, I solve the unconstrained social planner's problem without enforcing regional budget constraints (Equations C.6-C.8) to allow the model to accommodate a larger parameter space.

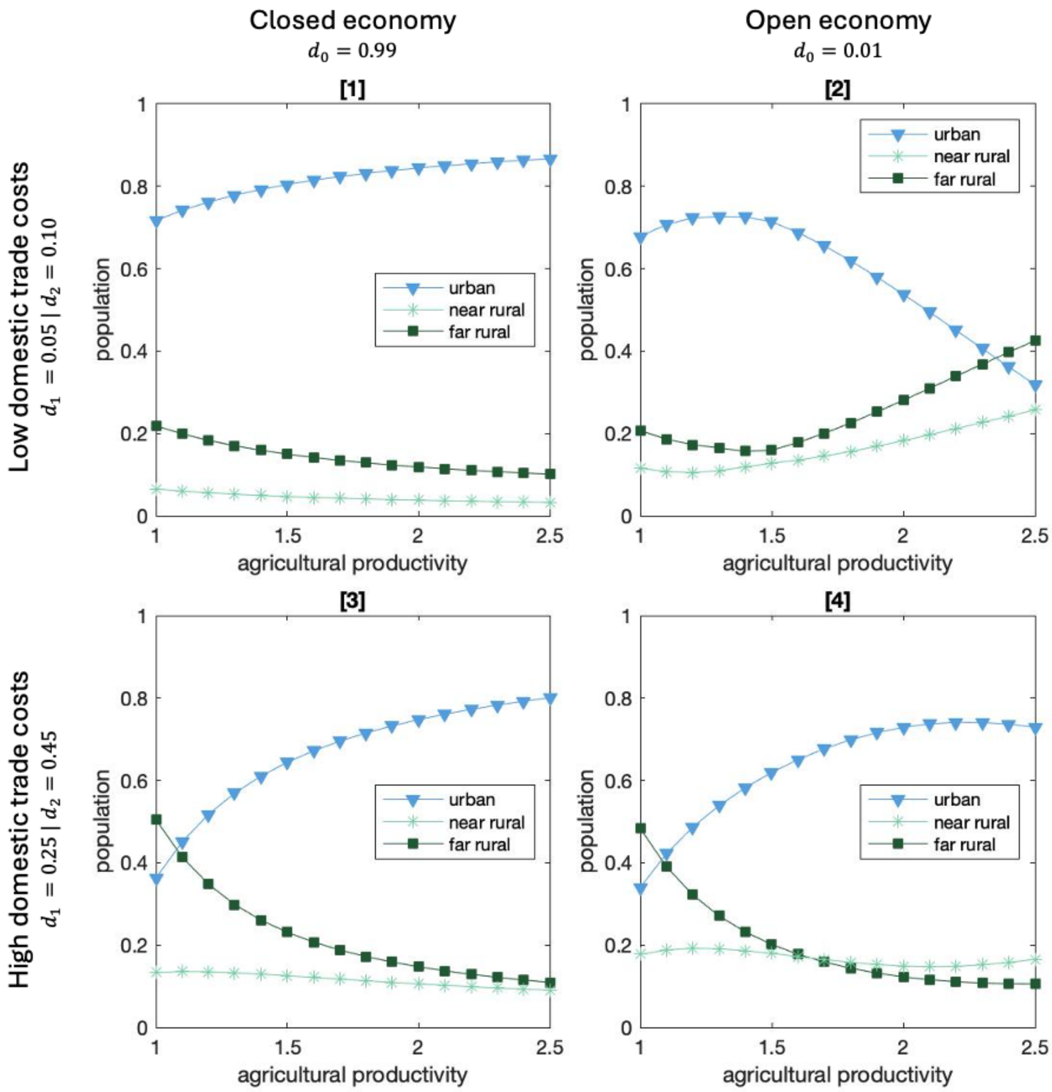
The structure of the model I develop here allows me to explore the interaction between agricultural productivity and trade openness in a setting with domestic transport frictions. High domestic transport frictions imply that the economy is effectively only 'partially' open – even if it is perfectly open to trade at the border – as domestic transport costs limit rural regions' integration into world markets. When domestic transport costs are sufficiently high, and the country's key export potential lies in agricultural goods which must be produced in rural regions, trade openness at the border plays a much smaller role for the way in which agricultural productivity affects urbanisation. High domestic transport costs limit the profitability of agricultural exports and thus reduce the incentive to produce a cash crop for export when the economy is open to trade. In this setting, agricultural productivity continues to drive structural transformation much in the same way as it does in an economy that is fully closed at the border, with an increase in agricultural productivity releasing labour into urban manufacturing. The closed economy result here is in line with the findings in Gollin and Rogerson (2014). While not explicitly modelling an open economy, Gollin and Rogerson (2014) presume that at relatively low levels of income, high domestic transport costs would still result in a large share of the population engaging in subsistence agriculture, even if the economy had access to imported food. The open economy model results presented here confirm this presumption.

Figure 3 plots regional population shares at different levels of agricultural productivity. The four panels in the figure correspond to four different parameterisations of the model, which differ with respect to trade openness (open versus closed at the border) and domestic trade frictions (low versus high domestic transport costs). The stark differences in the y-intercepts across panels highlight the importance of trade openness and domestic spatial frictions in determining spatial population patterns at a given level of agricultural productivity. Due to the non-homotheticity in food consumption, higher transport costs increase the share of the population living in rural regions at any given level of agricultural productivity. Panels [1] and [2] illustrate the results established by Matsuyama (1992): In a closed economy (panel [1]), an increase in agricultural productivity leads to a higher share of the population working in manufacturing, while in an open economy (panel [2]), an increase in agricultural productivity leads to a higher share of the population working in agriculture. When domestic transport costs are high, openness at the border has a much weaker effect. Panels [3] and [4] show that, when domestic transport costs are high, the impact of agricultural productivity on regional labour allocations is remarkably similar, irrespective of openness at the border. In both cases, an increase in agricultural productivity leads to a rise in the urban population, as labour is released from subsistence agriculture, and high transport costs make producing the export cash crop in rural regions less attractive than in panel [2], where transport costs are lower.

Property 2. *Spatial frictions matter for the sectoral allocation of economic activity and for the economy's pattern of international trade.*

Spatial frictions are an important determinant of the sectoral allocation of economic activity and the economy's pattern of international trade. Changes in spatial frictions (such as those

Figure 3: Agricultural productivity and the spatial distribution of population

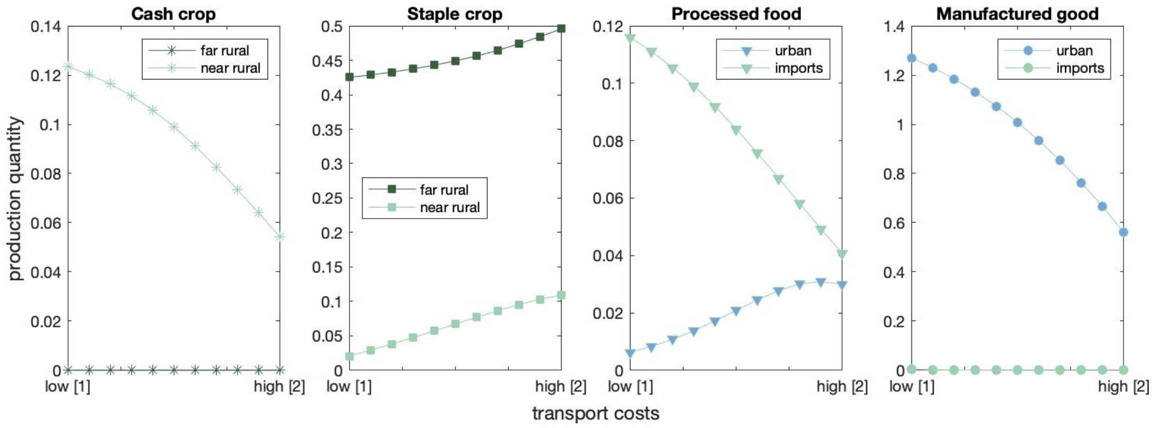


Note: This figure shows the effect of agricultural productivity on regional population allocations at different levels of trade openness and domestic trade costs. [1] Closed economy and low domestic trade costs; [2] Open economy and low domestic trade costs; [3] Closed economy and high domestic trade costs; [4] Open economy and high domestic trade costs. Agricultural productivity is assumed to be the same across all sectors (staple crop, cash crop) and regions (far rural, near rural). Other parameter values are held fixed across simulations: $A_{0p} = 1$, $A_{0m} = 2$, $\theta_x = 0.1$, $\theta_n = 0.3$, $\theta_l = 0.6$, $\gamma_x = 0.33$, $\gamma_n = 0.33$, $\gamma_l = 0.33$, $\delta_x = 0.33$, $\delta_n = 0.33$, $\delta_a = 0.33$, $\omega_n = 1$, $\lambda_p = 0.5$, $\rho_p = 0.9$, $\epsilon_p = 10$, $\lambda_m = 0.5$, $\rho_m = 0.9$, $\epsilon_m = 10$, $\lambda_f = 0.5$, $\rho_f = 0.4$, $\epsilon_f = 1.66$, $\alpha = 0.2$, $\bar{c}_f = 0.15$, $\bar{c}_m = 0$, $T = 0$, $p_m = p_x = 0.3$, $p_p = 1$, $p_c = 1$, $n = 1$, $l_2 = .8$, $l_1 = .2$

that might be delivered by investments in transportation infrastructure) have the potential to alter both the domestic patterns of economic activity and the economy's patterns of trade. Overall, an increase in domestic transport costs causes a spatial shift in economic activity from the urban region to rural regions. At the sectoral level, economic activity shifts out of the export cash crop sector and the manufacturing sector, and into the staple food crop sector and the urban food processing sector. These effects are driven crucially by two features of the model: the subsistence requirement in food consumption and the fact that the exports in this economy are purely agricultural.

Higher transport costs reduce the farm-gate price of the export cash crop, because *Home* is a price taker in international markets. At lower farm-gate prices, cash crop production declines, which results in a lower capacity to import processed food and manufactured goods from the rest of the world due to the balanced trade constraint. To satisfy subsistence needs for food consumption, production of the staple food crop and domestic processed food increase to compensate for lower food imports. The shift from cash crop into staple crop agriculture is particularly pronounced in the near rural region which, at low levels of domestic transport costs, specialises in the production of the export crop. The manufactured good is not subject to any subsistence consumption requirements, such that an increase in spatial frictions leads to a particularly pronounced decline in manufacturing production due to the decline in demand. As higher transport costs erode real incomes, households prioritise the consumption of food over the consumption of the manufactured good. In sectors where the manufactured good is used as an intermediate input, it is partially substituted with other factors of production (land and labour).

Figure 4: Domestic transport costs and the sectoral allocation of economic activity

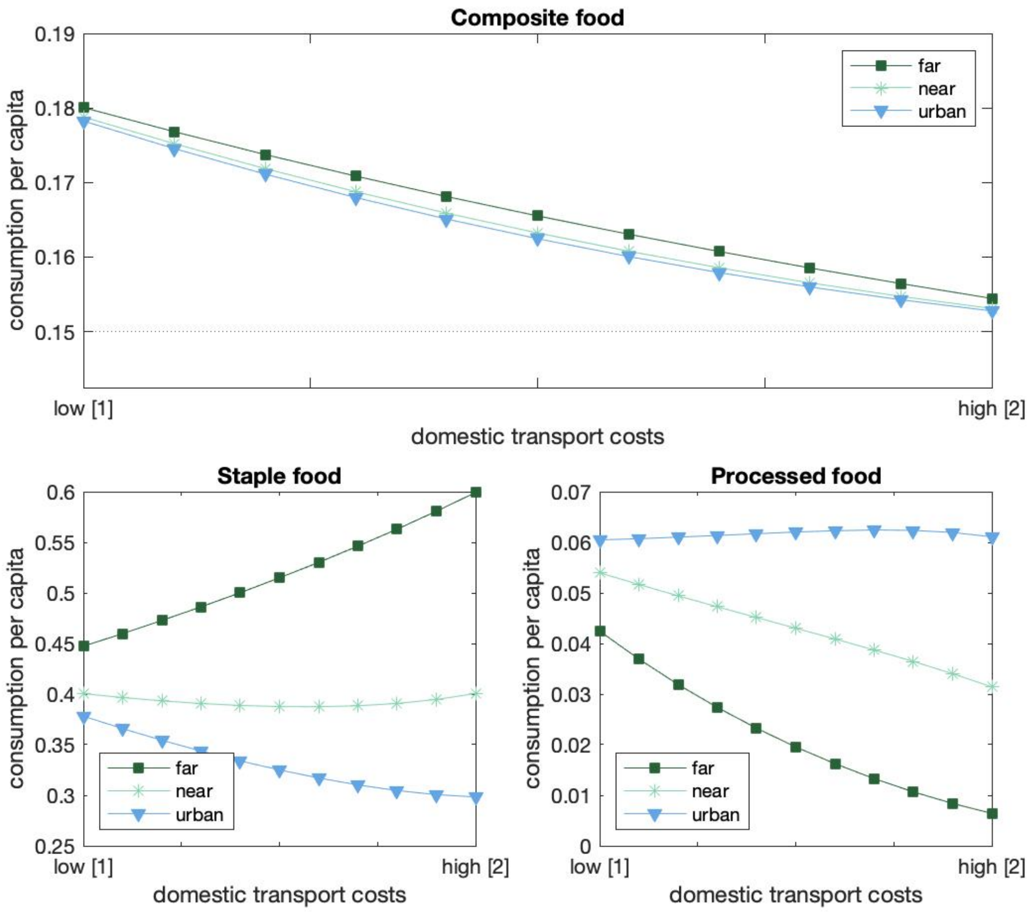


Note: This figure shows domestic production and import quantities by sector at different levels of domestic transport costs (d_1, d_2); [1] low domestic transport costs at the far left of the x-axis are $d_1 = 0.05$ and $d_2 = 0.1$; [2] high domestic transport costs at the far right of the x-axis are $d_1 = 0.25$ and $d_2 = 0.45$; the model is solved at 10pp increments between these low and high values.

Property 3. *Domestic transport costs cause diverging food consumption patterns across regions and limit the penetration of imported goods to remote rural areas.*

Spatial frictions are a key determinant of regional consumption patterns. This reflects the differing relative prices of food and other goods across regions. The effect is perhaps most pronounced with food. Transport costs introduce wedges between consumer prices of food in different regions as well as between producer prices, thus changing the relative price of different food types in rural and urban regions. Rising transport costs drive overall food consumption closer to the subsistence level in all three regions in the economy (see top panel in Figure 5). Food consumption allocations favor locally produced items, leading to diverging consumption patterns across regions (see lower panels in Figure 5). Two types of effects drive these results. First, there are direct effects that operate through the transport cost channel for non-local types of food. Rising transport costs increase the cost of moving staple food from rural regions to the urban region and processed food from the urban region to rural regions, leading to higher consumer prices of these goods in destination locations – and hence to reduced consumption. Second, there are indirect effects which operate through the input cost channel. High transport costs drive down the use of intermediates in agricultural production in the rural regions and drive up the cost of staple crop used in food processing in the urban region.

Figure 5: Food consumption patterns at different levels of transport costs



Note: This figure shows food consumption per capita in each of the three regions (far rural, near rural, urban) at different levels of domestic transport costs (d_1, d_2); [1] low domestic transport costs are $d_1 = 0.05$ and $d_2 = 0.1$; [2] high domestic transport costs are $d_1 = 0.25$ and $d_2 = 0.45$; simulations are run at simultaneous 10% increments between the low and high values. Composite food refers to the CES aggregate of the two types of food in this economy (staple food, which is domestically produced in the two rural regions, and processed food, which is produced domestically in urban and imported from RoW).

In the rural regions, the price of processed food rises due to costly transportation from the urban region, while the price of the staple food declines as labour shifts from cash crop to staple crop agriculture. Taken together, these opposing price changes lead to a surge in the relative price of processed food, which prompts households to heavily substitute for processed food with the locally produced staple food to meet subsistence needs. This effect is particularly strong in the far rural region, which faces substantially higher transport costs than the near rural region.

In the urban region, the price of the staple food rises as transportation from the rural regions becomes more costly. The increase in transport costs also has an indirect effect on the price of locally produced processed food, since the staple crop is a key input into the production of processed food. However, the price of locally produced processed food still falls overall, as urban labour shifts from manufacturing into food processing to increase production, more than offsetting the higher input costs for the staple crop.

While the economy is open at the border and can import food, the distribution of food imports to remote rural regions is subject to the same increase in transport cost as the distribution of domestic food. In the urban region, the price of imported processed food remains unchanged (since I assume here that changes in transport cost are restricted to the domestic economy and so do not affect the transport costs from world market to the urban region), but the economy's capacity to import food is limited by its production of the export cash crop, due to balance of trade constraints.

5 Calibration

The model properties described in Section 4 above underscore the importance of key parameter values for model predictions. In this section, I draw on various data sources to calibrate the model to the economy of Tanzania. I parameterise the model using a combination of methods which include (i) estimating parameters from micro data, (ii) fitting parameters to aggregate data, and (iii) targeting empirical moments to set the remaining free parameters. The resulting parameterisation allows me to explore quantitative model predictions for a range of potential productivity improvements in the comparative statics Section 7. A comprehensive overview of the model's parameterisation at the baseline equilibrium can be found in Table 11 in Appendix B.5.

5.1 Agricultural production

Estimation of the agricultural production functions draws on the detailed agriculture questionnaires in the LSMS-ISA 2019 National Panel Survey (NPS-SDD), which is the most recent LSMS data available for Tanzania. I estimate production functions separately for staple food crops and export cash crops.[9] Output (in log value terms) is regressed on land (in log acres),

[9] See Appendix B.1 for details on the distinction between staple food crops and export cash crops, and Appendix B.4 on details of data preparation and estimation.

labour (in log days) and intermediate inputs (in log value terms). In line with the constant returns to scale assumption, I normalise the estimated coefficients to sum to one and use these adjusted coefficients to parameterise agricultural production functions in the structural model. Table 1 presents raw and normalised coefficient estimates.[10]

Table 1: Agricultural production functions – Parameter estimates

	Staple crop		**Cash crop**	
	Raw	Normalised	Raw	Normalised
Sum of coefficients	0.73	1.00	1.27	1.00
Land	0.35	$\theta_l = 0.48$	0.58	$\gamma_l = 0.46$
Labor	0.19	$\theta_n = 0.26$	0.26	$\gamma_n = 0.20$
Intermediates	0.19	$\theta_x = 0.26$	0.43	$\gamma_x = 0.34$
Number of obs.	397	397	68	68

Notes: This table shows results of a regression of agricultural output (in log value terms) on land (in log acres), labour (in log days) and intermediate inputs (in log value terms). Data comes from the NPS-SDD 2019-2020 dataset collected by the Tanzanian National Bureau of Statistics. Production functions are estimated separately for staple food crops (43% of observations are Maize, 14% are beans, 9% are Paddy) and export cash crops (53% of observations are cashew nut, 45% are cotton, 2% are coffee). The regressions are run at the cluster level, pooling across all crops within crop type (staple crops / cash crops).

5.2 Food processing

The parameterisation of the production function for processed food is anchored in the 2015 input-output (I-O) table of the Tanzanian economy.[11]. I take industries C10 (*Manufacture of food products*) and C11 (*Manufacture of beverages*) in the I-O table to correspond to the food processing sector in the model. On the inputs side, I map industries A01-A03 to the staple food crop input and industries B05-C33 to intermediate inputs in the model.[12] For an estimate of labour inputs used in production, I sum *Compensation to employees* for the two food processing industries (C10 and C11) in the I-O table. To estimate output elasticity parameters for the processed food production in the model, I calculate normalised input cost shares over the three input types. This calculation yields an output elasticity of 0.7 for the staple food crop (δ_a), 0.1 for labour (δ_n), and 0.2 for intermediates (δ_x).

[10]This of course constitutes a fairly naive approach to estimating agricultural production functions, since inputs are endogenous. There are several papers (e.g. Gollin and Udry (2021), Suri (2011)) that tackle this endogeneity more fundamentally, but since I am only interested in the broad macroeconomic relationship between inputs and outputs, I keep the estimation more rudimentary.

[11]Source: Tanzania - Input-Output Tables 2015, https://www.nbs.go.tz/index.php/en/census-surveys/national-accounts-statistics/na-statistics-by-subject, accessed on 9th July 2022

[12]Other inputs (energy & construction, trade & transport, and services) are not included in the estimation.

To parameterise the CES aggregator which combines domestically produced processed food and imported processed food into the composite good for final consumption, I focus on vegetable oils as an example of a widely consumed processed food item with good trade data availability (FAOSTAT and IMF data for 2010-19). To approximate the elasticity of substitution between domestically produced vegetable oils and imported vegetable oils, I regress the change in the ratio of imports to total domestic supply on the change in the exchange rate. This allows me to assess how sensitive the relative consumption of imported vegetable oils is to a change in the relative price. The calculation yields an elasticity of substitution of 1.72, resulting in a value of ρ_p of 0.42. The remaining parameters in the CES aggregator for processed food are set to $\bar{A}_p = 1$ and $\lambda_p = 0.5$.

5.3 Manufacturing

For the manufacturing sector, I follow Gollin and Rogerson (2014) and set the output elasticity of labour (ω_n), the only input in production, equal to one. To parameterise the CES aggregator that governs the substitution between domestically produced and imported manufactured goods, I use a similar approach as for processed food. I approximate the elasticity of substitution between domestically produced and imported manufactured goods using production and trade data for textiles as a representative category of manufactured goods. The value of domestic production of textiles is approximated from World Bank data on value added in manufacturing and the share of value added in manufacturing that accrues to textiles. The value of textile imports is calculated from UN Comtrade data, aggregating import values for products with 2-digit HS codes 61-63. Regressing the change in the ratio of imports to total domestic supply on the change in the exchange rate yields an estimate of the elasticity of substitution of 2.27, resulting in a value of ρ_m of 0.56. Analogous to the food processing sector, the remaining parameters in the CES aggregator are set to $\bar{A}_m = 1$ and $\lambda_m = 0.5$.

5.4 Preferences

To parameterise the utility function, I estimate consumer preferences from budget shares, drawing on the LSMS-ISA 2019 National Panel Survey micro data from Tanzania. I map household location into the spatial model regions (urban, near rural and far rural) and calculate median budget shares for each region (see Table 2).[13] I use the urban overall food budget share as an estimate for the preference parameter on food ($\alpha = 0.31$); this presumes that the subsistence requirement on food consumption is likely less of a binding constraint in the urban region, where incomes tend to be higher.[14] I estimate the elasticity of substi-

[13]I use the GHSL Degrees of Urbanisation dataset to classify households as urban, near rural, or far rural. See Appendix B.3 for details.

[14]The classification of food items into staple foods and processed foods is shown in Table 9 in Appendix B.2. The classification broadly follows Regmi et al. (2005). The value of non-food consumption (consumables and durables) is directly reported in the data. Food consumption is reported in quantity terms, distinguished by source (purchases, gifts, and own production). To approximate consumption

tution between staple foods and processed foods (ϵ_f) and the CES share parameter λ_f from the spatial differences in median prices and consumption quantities for staple and processed foods, using the geo-referenced LSMS data. The subsistence level parameter ($\bar{c}_f$) is set using a method of moments approach.

Table 2: Consumption expenditure – Budget shares by category

	Urban region	**Near rural region**	**Far rural region**
Staple food (%)	17.3	29.8	35.5
Processed food (%)	13.2	13.8	14.3
Manufactured good (%)	67.4	52.2	45.4

Notes: This table shows median budget shares for staple food, processed food, and non-food consumption in the three regions as calculated from the National Panel Survey 2019-2020 data on Tanzania. Where food consumption was reported in quantity terms, median prices of the respective items were used to infer the value of consumption. Food consumption is aggregated over all different sources (own production, purchases, gifts). Classification of items into staple food and processed food loosely follows Regmi et al. (2005); see Table 9 for a detailed overview of items in the two categories. Non-food consumption includes both consumables and durables, the consumption values of which were directly reported in the data. The overall number of households in the sample for these calculations was $N = 628$ (urban = 190, near rural = 108, far rural = 330).

5.5 Agricultural land

I estimate the share of agricultural land by region from micro data, using the GHSL's Degrees of Urbanisation location classification and data on plot size from the LSMS-ISA 2019 National Panel Survey micro data for Tanzania. I classify households as living in either the urban, near rural, or far rural region and aggregate the area of all plots under cultivation by households living within a respective region to calculate total agricultural land. These calculations yield relative shares of land in the far and near rural region of 81.9% and 18.1%, respectively. I take these values from the LSMS sample as estimates of the overall distribution of land between regions in Tanzania and set the corresponding parameters in the structural model to $l_2 = 0.82$ and $l_1 = 0.18$.

5.6 Transport costs, subsistence consumption, and prices

The remaining model parameters on international prices, domestic transport costs, and the subsistence requirement for food consumption are not observed directly or indirectly. Instead, I calibrate these using a method of moments approach. I target moments on regional population shares, agricultural land use, and import shares for processed food and manufactured goods, setting the values of remaining parameters to minimise the sum of squared differences between (normalised) data moments and model moments. Table 3 below shows the model fit with respect to the targeted moments and reports the resulting parameter values for trans-

in value terms, I use median prices (calculated from data on food purchases) to infer the value of consumption from own production and gifts.

port costs (d_1, d_2), subsistence consumption ($\bar{c}_f$), and international prices (p_p, p_m, p_c). The price of the export cash crop (p_c) is normalised to unity and serves as the numéraire.

Table 3: Model fit at baseline

Moment	Data	Model	Parameter	Value
Population in urban region (n_0)	0.280	0.249	d_1	0.11
Population in near rural region (n_1)	0.160	0.180	d_2	0.23
Population in far rural region (n_2)	0.560	0.572	$\bar{c}_f$	0.15
Land used for cash crop in near rural (l_{1c})	0.022	0.048	p_p	12.0
Land used for cash crop in far rural (l_{2c})	0.088	0.053	p_m	0.23
Import share for processed food	0.560	0.386	p_c	1.00
Import share for manufactured good	0.480	0.744		

Note: Population and land use data moments are calculated from LSMS data. Population shares are a weighted average over LSMS waves 1-3 (2008-2012) and based on household size (approximated using adult equivalents) rather than the number of households, to account for the fact that households tend to be larger in more rural regions. Land use moments are calculated from LSMS waves 2-3 (2010-2012) as the share of land located in a given region multiplied by the share of land used to grow cash crops in that region. Import shares are calculated as the value of imports over the value of total domestic supply (domestic production - exports + imports). The import share for processed food is approximated using FAOSTAT production and trade data on vegetable oils (2010-2019). The import share for the manufactured good is approximated using Comtrade data on textiles imports and World Bank data on value added in the domestic textiles industry (2014-2019). Note that for reasons of data availability, data moments are not all calculated for the same year. The international price of the cash crop, p_c, is normalised to one.

6 Baseline equilibrium

Given the set of parameters described above, I solve the model numerically.[15] Aggregate utility (Equation 1) serves as the objective function, which is maximised subject to the feasibility conditions. These feasibility conditions include commodity market clearing, factor market clearing, balanced trade, and regional budget constraints (for details see Equations C.1-C.11). The remainder of this section presents results at the baseline equilibrium.

Table 4 presents labour and land allocations at baseline. The overall rural population share is approximately 75%, implying an urban population share of about 25%. The largest sector of employment in the economy is the staple crop agricultural sector in the far rural region, with a labour share of about 54% of the total population. This corresponds to 95% of the population in the far rural region engaging in staple crop agriculture. In the near rural region, where 18% of the overall population live, the share of workers engaged in staple crop agriculture is lower at, 77%. The near region, which is better connected to global markets than the far rural region, uses a relatively larger share of labour to produce the export cash crop. In the urban region, the majority of the workforce is employed in the manufacturing sector, while a smaller share of workers is employed in food processing. Land use patterns

[15] I solve the model in MATLAB, using the optimisation toolbox's function for constrained nonlinear optimisation, *fmincon*.

again reflect the near rural region's relative specialisation in cash crop production. In the near rural region, 26.7% of land is used to produce the export cash crop while in the far rural region this share is substantially lower at 6.4%.

Table 4: Labour and land allocation (% of total)

		Far rural	Near rural	Urban
Labour	*All sectors*	57.2	18.0	24.9
	Staple crop agriculture	54.2	13.9	–
	Cash crop agriculture	3.0	4.0	–
	Food processing	–	–	3.2
	Manufacturing	–	–	21.7
Land	*All sectors*	82.0	18.0	–
	Staple crop agriculture	76.7	13.2	–
	Cash crop agriculture	5.3	4.8	–

Note: Allocations are reported as the share of economy-wide total supply of labour and land. Regional population shares closely match the distribution of population observed in the data, as these shares were targeted as moments in the calibration. The regional land allocation was set based on microeconomic data, while the within-region sectoral splits were targeted moments in the calibration.

Table 5 shows consumption quantities per capita at the baseline equilibrium. Because the transport costs between regions in the model induce spatial price gradients, we would expect consumption of a good to be higher closer to its origin and to decline as we move away from the region of production. In line with this, consumption of the staple crop is highest in the far rural region, while consumption of processed food and the manufactured good are highest in the urban region. How quickly consumption of a good falls as we move away from the region of production depends on the level of transport costs between regions and on the degree of substitutability between different goods.

Table 5: Consumption per capita

	Far rural region	Near rural region	Urban region
Staple food	0.501	0.410	0.353
Processed food	0.020	0.032	0.039
Manufactured good	0.168	0.178	0.178

Note: Consumption is reported as consumption quantity per capita for each of the three representative households (far rural, near rural, and urban) across the three consumption goods (staple food, processed food, and the manufactured good).

The overall supply and demand structure of the economy at baseline is summarised in Table 6. For each of the four commodities, the export cash crop, the staple food crop, processed food, and manufactured goods, the table shows the distribution across different sources on the supply side and different uses on the demand side. The near region supplies 57% of the total domestic cash crop production, while the far region supplies 43%. The cash crop is not consumed domestically; 80% of production are exported, while iceberg transport frictions account for 20% of production.

The far rural region produces 80.6% of total supply of the staple food crop, with the remaining 19.4% produced in the near rural region. The majority of the far rural staple crop production (48.7 pp) remains in the far region, where it is consumed by households. The remaining supply of the staple crop is consumed by households in the urban region (15.0 pp) and near rural region (12.5 pp), used as an intermediate input in the urban food processing sector (13.1 pp), and accounted for by transport frictions (10.7 pp). The relatively low transport losses in the staple crop sector are due to the fact that consumption of the staple crop is more local than consumption of other commodities.

Table 6: Supply and demand (% of total by commodity)

	Cash crop	Staple crop	Processed food	Manufact. good
Supply [1]				
Far rural	43.0	80.6	–	–
Near rural	57.0	19.4	–	–
Urban	–	–	80.9	11.3
RoW (imports)	–	–	19.1	88.7
Demand [2]				
Final demand				
Urban household	–	15.0	29.4	6.4
Near rural household	–	12.5	17.8	4.6
Far rural household	–	48.7	34.7	13.9
RoW (exports)	80.2	–	–	–
Intermediate use				
Food processing	–	13.1	–	11.9
Staple crop – Near rural	–	–	–	10.2
Cash crop – Near rural	–	–	–	5.1
Staple crop – Far rural	–	–	–	24.6
Cash crop – Far rural	–	–	–	2.3
Transport				
Melt	19.8	10.7	18.1	21.0

Note: [1] Supply is reported as percent of total supply quantity by commodity. Total supply includes domestic production and imports. [2] Demand is reported as percent of total demand quantity by commodity. Total demand includes final consumption by households, exports, intermediate input use, and transport losses. For each commodity, supply from different origins sums to 100% and demand at different destinations also sums to 100%.

Imports account for around 19% of the total domestic supply of processed food. The much larger share, around 81%, are produced domestically in the urban region. Per capita consumption of processed food is highest in the urban region, where it is cheapest, such that 29.4% of total supply are consumed by urban households. Households in the near rural region consume 17.8%, and households in the far rural region 34.7% of total supply. For processed food, transport losses amount to 18.1% of total supply.

The manufactured good is largely imported (88.7%), with only 11.3% coming from domestic production in the urban region. The manufactured good is consumed directly by all households and used as an intermediate input in production across all other sectors in the economy. Only around 18% of total supply are consumed or used in the urban region. The majority of the manufactured goods is transported to the rural regions for use as an intermediate input into agricultural production and, to a lesser extent, for final consumption by households. As a result, the 'melt' share is highest for the manufactured good, because its consumption is less local than that of other commodities.

7 Counterfactual analysis

In this section, I consider four counterfactual experiments to explore the differential effects of alternative productivity improvements in the economy.[16] The experiments in this section are comparative statics, i.e. comparisons of general equilibrium outcomes at different model parameterisations. This does not take into account potential adjustment costs that could be associated with transitions between these equilibria. I consider productivity improvements in the transport sector, the agricultural sectors, and the non-agricultural sectors, and finally a combination of the three individual productivity improvements. All of these potential productivity improvements would be expected to make the food problem less binding for the economy.

7.1 Experiment design

[1] Increase in transport productivity (reduction in domestic trade costs)
A reduction in transport costs reduces spatial price gradients, which allows domestic agriculture to shift from subsistence farming to more market-oriented production. Lower transport costs imply higher farm-gate prices of the export cash crop, thus increasing the economy's scope for export earnings; they also lead to lower consumer prices of processed food in the rural regions, thus allowing for a larger share of subsistence needs to be met by food imports. In addition, a reduction in transport costs lowers the price of intermediate inputs used in agriculture and processed food production, which decreases the cost of growing and processing food.

[2] Increase in agricultural productivity across all sectors and regions
An increase in agricultural productivity directly affects the food problem by allowing the

[16]Note that this comparative analysis does not suggest that the different experiments would incur similar costs if policies to achieve these productivity improvements were implemented. Rather than a cost-benefit analysis or direct comparison of similarly costly policy counterfactuals, the analysis in this section should be understood in the same way as the experiments conducted in Gollin and Rogerson (2014), who state that their "motivation derives from the perspective (see, e.g., Hall and Jones, 1999) that differences in productivity are the dominant source of differences in living standards across countries, thereby making it of interest to compare the relative importance of different dimensions of productivity growth" (p. 45).

economy to maintain a given level of food production at lower levels of labour, land, and intermediate inputs, thus releasing some productive capacity into other sectors in the economy. Higher agricultural productivity implies that fewer people are required to grow food in the far rural region, where non-locally produced consumption goods are expensive, due to transport costs. This increases rural-to-urban migration. An improvement in agricultural productivity also increases trade with the rest of the world, as more efficient production of the cash crop implies higher export potential.

[3] Increase in non-agricultural productivity in urban areas
An increase in non-agricultural productivity affects the food problem through two channels. The first is a direct channel that operates via the food processing sector. A productivity increase in the food processing sector allows for more efficient production and reduces the price of processed food. This alleviates pressures on staple crop agriculture, as some staple food consumption can be substituted with processed food consumption. The second channel is more indirect and highlights the importance of intermediate inputs in production. A productivity increase in the manufacturing sector lowers the price of the domestic manufactured good. This reduced cost of intermediates can have significant effects on agricultural production of both the staple food crop (impacting the food problem directly through higher food production) and the export cash crop (allowing for more food and non-food imports due to higher export earnings).

[4] Combined increase in transport, agricultural & non-agricultural productivity
In the final experiment, I analyse the effects of combining the first three productivity improvements. Simulating simultaneous productivity improvements across the transport, agricultural, and non-agricultural sectors in the economy reveals potential interaction effects of these different changes in general equilibrium.

7.2 Results

Table 7 reports results for the four experiments described above, with effects reported as relative changes over baseline values. Column [1] shows the effects of a 10% reduction in domestic transport costs, which lowers the iceberg trade cost parameters to $d_1 = 0.1$ and $d_2 = 0.2$. The reduction in transport costs decreases spatial gradients between farm- and factory-gate prices, consumer prices, and input prices across the three regions. This implies a reduction in the cost of non-local consumption goods and intermediate inputs. Labour shifts from the rural regions to the urban region, leading to an increase in urban population by around 3%. Production in the urban sectors expands roughly in line with the growth of the urban workforce. The food processing sector grows by an additional percentage point (4.2%) as lower transport costs reduce the urban price of staple crop, which is used as an input in food processing. Staple crop production is flat overall, but some production relocates from the far to the near rural region to supply staple crop to the urban region. As the farm-gate price of the export cash crop increases in both rural regions, production expands by 3%

overall, with a significant shift in the location of production from the near to the far rural region.

Table 7: Counterfactual productivity improvements (change over baseline in %)

	Baseline value	[1] Trade costs d_1, d_2	[2] Ag. Prod. A_a, A_c	[3] Non-ag Prod. A_p, A_m	[4] Comb. d_1, d_2, A_a A_c, A_p, A_m
Population					
Urban region	0.249	3.1 %	7.5 %	8.9 %	8.9 %
Near rural region	0.180	-0.3 %	6.7 %	5.1 %	7.2 %
Far rural region	0.572	-1.3 %	-5.4 %	-5.5 %	-6.1 %
Trade					
Imports–Processed food	0.017	2.2 %	-6.8 %	7.4 %	10.1 %
Imports–Manufact. good	1.700	3.7 %	4.6 %	15.8 %	25.8 %
Exports–Cash crop	0.047	3.2 %	0.6 %	12.9 %	20.4 %
Production					
Food processing	0.072	4.2 %	4.9 %	15 %	27.5 %
Manufacturing	0.217	3.3 %	8.9 %	20.9 %	29.6 %
Staple crop–Overall	0.588	-0.1 %	7.6 %	-0.3 %	7.2 %
Staple crop–Near rural	0.114	3.5 %	19.1 %	11.8 %	41.2 %
Staple crop–Far rural	0.474	-14.9 %	4.9 %	-3.2 %	-1 %
Cash crop–Overall	0.073	3.1 %	3.1 %	17.9 %	29.2 %
Cash crop–Near rural	0.042	-0.9 %	-14.2 %	-17 %	-62.6 %
Cash crop–Far rural	0.031	26.9 %	26 %	64.1 %	150.8 %
Labour					
Food processing	0.032	1.5 %	-2.2 %	2.4 %	3 %
Manufacturing	0.217	3.3 %	8.9 %	9.9 %	17.8 %
Staple crop–Overall	0.681	-1.3 %	-2.3 %	-4.5 %	-6.7 %
Staple crop–Near rural	0.139	3.8 %	13.7 %	11.7 %	35.8 %
Staple crop–Far rural	0.542	-2.7 %	-6.4 %	-8.7 %	-9.1 %
Cash crop–Overall	0.070	2 %	-4.6 %	12.4 %	8.6 %
Cash crop–Near rural	0.040	-14.4 %	-17.6 %	-17.7 %	-64 %
Cash crop–Far rural	0.030	24.2 %	12.9 %	53.1 %	106.5 %
Land					
Staple crop–Overall	0.899	-0.8 %	0 %	-2.5 %	-4 %
Staple crop–Near rural	0.132	4.9 %	7.9 %	7.5 %	24.3 %
Staple crop–Far rural	0.767	-1.7 %	-1.3 %	-4.2 %	-8.9 %
Cash crop–Overall	0.101	6.9 %	-0.4 %	22 %	35.6 %
Cash crop–Near rural	0.048	-13.5 %	-21.8 %	-20.8 %	-67 %
Cash crop–Far rural	0.053	25.3 %	19 %	60.7 %	128.5 %
Intermediate inputs					
Food processing (a_{0p})	0.077	5.3 %	8.1 %	3.2 %	18.6 %
Food processing (x_{0p})	0.082	2 %	-2.4 %	10.8 %	13.4 %
Staple crop-Overall	0.241	1.9 %	-5.2 %	9.4 %	5.7 %
Staple crop–Near rural	0.071	0.7 %	3.7 %	20.4 %	28.6 %
Staple crop–Far rural	0.170	2.5 %	-9 %	4.8 %	-3.8 %
Cash crop–Overall	0.051	-2 %	-14 %	16 %	-0.9 %
Cash crop–Near rural	0.035	-16.9 %	-24.8 %	-11.3 %	-65.9 %
Cash crop–Far rural	0.016	30.7 %	9.8 %	75.7 %	141.1 %

–Continued on next page–

Table 7 – Continued from previous page

	Baseline value	[1] Trade costs d_1, d_2	[2] Ag. Prod. A_a, A_c	[3] Non-ag Prod. A_p, A_m	[4] Comb. d_1, d_2, A_a A_c, A_p, A_m
Consumption per capita					
Urban hh–Staple food	0.353	1.3 %	4.5 %	-2 %	2.7 %
Urban hh–Processed food	0.039	-0.3 %	-3.7 %	9.3 %	8.4 %
Urban hh–Manufact. good	0.178	13.8 %	29.9 %	31.9 %	78.2 %
Near hh–Staple food	0.410	1.8 %	8.8 %	-0.2 %	11.3 %
Near hh–Processed food	0.032	3.7 %	0.3 %	11.3 %	21.7 %
Near hh–Manufact. good	0.178	13.8 %	29.9 %	31.9 %	78.2 %
Far hh–Staple food	0.501	0 %	9.7 %	0.5 %	10.5 %
Far hh–Processed food	0.020	10.1 %	1.1 %	12.1 %	30.6 %
Far hh–Manufact. good	0.168	18.5 %	33.5 %	32.5 %	88.3 %

Note: This table shows the effects of counterfactual productivity improvements on population, production quantities, input use, and consumption per capita. Effects are reported as percent changes over baseline. [1] 10% reduction in domestic transport costs ($d_1 = 0.9d_1$ and $d_2 = 0.9d_2$); [2] 10% increase in agricultural productivity in all sectors and regions ($A_{1a}, A_{1c}, A_{2a}, A_{2c} = 1.1$); [3] 10% increase in non-agricultural productivity in food processing and manufacturing ($A_{0p}, A_{0m} = 1.1$); [4] 10% reduction in domestic transport costs ($d_1 = 0.9d_1$ and $d_2 = 0.9d_2$) in combination with 10% increase in agricultural productivity in all sectors and regions ($A_{1a}, A_{1c}, A_{2a}, A_{2c}, A_{0p}, A_{0m} = 1.1$).

Overall, agricultural labour moves out of production of the staple crop and into the cash crop sector, but labour movements flow in opposite directions in the two rural regions: In the near region, labour in staple crop production increases and labour in cash crop production declines, while in the far region, labour in cash crop production increases and labour in staple crop production declines. The same pattern can be observed for agricultural land. In the near region, land is reallocated from cash crop to staple crop production, while in the far region land is reallocated from staple crop to cash crop production. Overall, the land area under staple crop cultivation declines by 0.8% whereas the area under cash crop cultivation expands by 6.9%. Intermediate input use increases in urban food processing, particularly the amount of staple crop used in production, as prices fall with lower transport costs. In staple crop agriculture, intermediate input use increases in both rural regions individually and overall. In cash crop agriculture, intermediate input use increases strongly in the far rural region and declines significantly in the near rural region. Taking into account the reallocation of agricultural land, these changes imply more intensive use of intermediates in the far region (x_2/l_2) and less intensive use of intermediates in the near region (x_1/l_1). Lower transport costs allow households to increase their consumption, especially of the manufactured good and the non-locally produced type of food (processed food in rural regions and staple food in the urban region).

Column [2] shows the effects of a 10% increase in agricultural TFP across both rural regions and both crop sectors ($A_{1a}, A_{1c}, A_{2a}, A_{2c} = 1.1$). The primary effect of this productivity improvement is a shift of labour from subsistence agriculture into market-oriented agriculture and urban manufacturing. Population in the far rural region declines by 5.4%, whereas the near rural and urban region grow by 6.7% and 7.5%, respectively. Below these top-line population effects, the increase in agricultural productivity also causes a significant reallocation of labour across sectors within a given region. In the near rural region, labour shifts out of cash crop and into staple crop agriculture. In the far rural region, the increase in agricultural productivity has the opposite effect such that labour is released from staple crop agriculture and pulled into cash crop agriculture. The reallocation of agricultural land mirrors the labour movements across the agricultural sectors, with staple crop in far rural and cash crop in near rural declining, and staple crop in near rural and cash crop in far rural expanding. In the urban region, manufacturing employment increases substantially while food processing employment contracts slightly. Production in both urban sectors expands, which is driven by the labour inflow (for manufacturing) and the increased use of intermediates (for food processing). Imports of processed food decline as the agricultural productivity improvement allows for a higher share of households' food consumption to be met with domestically produced staple and processed food. Exports of the cash crop rise and imports of manufactured good increase to satisfy producers' increased demand for intermediate inputs and households' increased demand for non-food products.

Column [3] shows the effects of a 10% increase in non-agricultural TFP in the urban food processing and manufacturing sectors ($A_{0p}, A_{0m} = 1.1$). This productivity improvement pulls labour into the urban region. The magnitude of the population outflow from the far rural region is identical to the outflow following the agricultural productivity improvement in experiment [2]. However, a higher share of the labour leaving the far rural region reallocates to the urban region, rather than to agriculture in the near rural region. Urban production expands significantly, particularly in the manufacturing sector which absorbs the majority of the labour inflow. Total staple crop production is flat overall, with production shifting from the far rural region to the near rural region and becoming more intensive in intermediate inputs while employing less labour and using less land. Cash crop production falls in the near rural region but expands significantly in the far rural region, where the decreased cost for intermediate inputs and a significant reallocation of land and labour facilitate higher production volumes. The overall increase in export cash crop production allows for increased imports of processed food and manufactured goods from the rest of the world. Households increase their consumption of processed food and the manufactured good, the two sectors in which productivity increased, whereas their consumption of the staple crop is largely flat.

Column [4] shows the effects of combining the previous three experiments, i.e. a 10% reduction in transport costs and a 10% increase in agricultural and non-agricultural TFP across all regions and sectors. By comparing the effects in this joint experiment to the sum of the individual effects in the previous three experiments, the combined experiment allows to explore potential knock-on and interaction effects of the different productivity improvements in general equilibrium.

Where the individual productivity improvements cause changes in the same direction, experiment [4] shows whether the joint effect is larger (such that the different productivity improvements amplify each other) or smaller (such that they crowd each other out) than the sum of the individual effects. For example, population effects appear to largely crowd each other out. Though the reallocation of labour from far rural to near rural is somewhat stronger than in any of the three isolated experiments, the increase in the urban population in the joint experiment has the same magnitude as in the case of an isolated increase in non-agricultural productivity. By contrast, cash crop production in the far rural region constitutes an example of amplification of effects. All three productivity improvements individually lead to a significant increase in cash crop production, but the joint effect is nearly 50% stronger than the sum of individual effects.

Where the individual productivity improvements cause changes in opposite directions, the overall net effect shows which force outweighs the other in the combined experiment. For example, the transport cost reduction and non-agricultural TFP improvement cause a decline in staple crop production in far, whereas the agricultural TFP improvement leads to an expansion of this sector. In the joint experiment, the three combined productivity improvements essentially offset each other, leading to a small reduction in staple crop production in the far rural region. However, this reduction is much weaker than the sum of the three individual experiments would suggest.

8 Conclusion

In a closed economy, the sectoral composition of production matches domestic consumption. At low levels of agricultural productivity and with minimum consumption requirements for food, this requires a large share of the population to work in agriculture, producing food. In an open economy, food imports can in theory resolve this food problem. The qualitative model predictions in this paper show that, in an economy that is *imperfectly open*, the food problem can continue to bind. In an economy that is open at the border but which faces domestic trade frictions – such that parts of the economy are effectively not integrated into world markets – a large share of the labour force remains tied up in agriculture in rural regions.

The exact parameterisation of the model influences the extent to which this result holds. The strength of domestic trade frictions, the level of agricultural and non-agricultural productivity, and the economy's terms of trade all affect the extent to which the food problem continues to exert pressure in an economy that is open to trade. Openness at the border has a stronger effect on the economy if – *ceteris paribus* – domestic trade costs are lower, agricultural productivity is higher, or the terms of trade are more favorable (higher export prices, lower import prices). In addition, modelling assumptions around the subsistence requirement for food consumption and the degree of substitutability between different types of food (staple and processed, domestic and imported) affect model predictions on structural change in imperfectly open economies.

I calibrate the model to Tanzanian data to to set a sensible parameterisation for counterfactual analysis of different productivity improvements that have the potential to alleviate the food problem. Comparative statics results show that productivity improvements in the transport sector, agricultural sectors and non-agricultural sectors can all make the food problem less binding but differ with respect to their effect on trade patterns, the spatial distribution of economic activity, and food consumption patterns.

Interesting future extensions of the analysis in this paper would be to allow for different tradability of goods. For example, how would the qualitative model results change if there was an urban export good – either in addition to or instead of the agricultural export good? I presume that parameterisation would be key in this case. Depending on world market prices and the degree of substitutability between domestic and imported food, and under the assumption that urbanisation is not limited e.g. by congestion externalities or migration frictions, the existence of an urban export good could potentially resolve the food problem in an economy that is open to trade.

References

Adam, Christopher, David Bevan, and Douglas Gollin (2018). "Rural–Urban Linkages, Public Investment and Transport Costs: The Case of Tanzania". In: *World Development* 109, pp. 497–510.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *The Quarterly Journal of Economics* 129(3), pp. 1085–1139.

Atkin, David and Dave Donaldson (2015). "Who's getting globalized? The size and implications of intra-national trade costs". In: *NBER Working Paper 21439*.

Baptista, Diogo, John Spray, and D. Filiz Unsal (2023). "Coping with Climate Shocks: Food Security in a spatial framework". In: *Working Paper*.

Burgess, Robin and Dave Donaldson (2012). "Railroads and the Demise of Famine in Colonial India". In: *Working Paper*.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli (2016). "Agricultural Productivity and Structural Transformation: Evidence from Brazil". In: *American Economic Review* 106(6), pp. 1320–1365.

Donaldson, Dave (2018). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure". In: *American Economic Review* 108((4-5)), pp. 899–934.

Emerick, Kyle (2018). "Agricultural productivity and the sectoral reallocation of labor in rural India". In: *Journal of Development Economics* 135, pp. 488–503.

Fajgelbaum, Pablo and Stephen Redding (2022). "Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914". In: *Journal of Political Economy* 130(5), pp. 1249–1318.

Farrokhi, Farid and Heitor S. Pellegrina (2023). "Trade, Technology, and Agricultural Productivity". In: *Journal of Political Economy* 131(9), pp. 2289–2595.

Gollin, Douglas, Stephen Parente, and Richard Rogerson (2002). "The Role of Agriculture in Development". In: *American Economic Review (Papers and Proceedings)* 92(2), pp. 160–164.

Gollin, Douglas, Stephen Parente, and Richard Rogerson (2007). "The food problem and the evolution of international income levels". In: *Journal of Monetary Economics* 54, pp. 1230–1255.

Gollin, Douglas and Richard Rogerson (2014). "Productivity, transport costs and subsistence agriculture". In: *Journal of Development Economics* 107, pp. 38–48.

Gollin, Douglas and Christopher Udry (2021). "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture". In: *Journal of Political Economy* 129(1).

Janssens, Charlotte, Petr Havlík, Tamás Krisztin, Justin Baker, Stefan Frank, Tomoko Hasegawa, David Leclère, Sara Ohrel, Shaun Ragnauth, Erwin Schmid, Hugo Valin, Nicole Van Lipzig, and Miet Maertens (2020). "Global hunger and climate change adaptation through international trade". In: *Nature Climate Change* 10, pp. 829–835.

Matsuyama, Kiminori (1992). "Agricultural Productivity, Comparative Advantage, and Economic Growth". In: *Journal of Economic Theory* 58, pp. 317–334.

McCaig, Brian and Nina Pavcnik (2013). "Moving out of agriculture: Structural change in Vietnam". In: *NBER Working Paper* 19616.

Moszoro, Mariano and Mauricio Soto (2022). *Road Quality and Mean Speed Score*. IMF Working Papers 2022/095. Washington, DC: International Monetary Fund.

Nath, Ishan B. (2020). "The Food Problem and the Aggregate Productivity Consequences of Climate Change". In: *NBER Working Paper* 27297.

Ngai, Rachel and Christopher Pissarides (2007). "Structural Change in a Multisector Model of Growth". In: *American Economic Review* 97(1).

Regmi, A., M. Gehlhar, J. Wainio, T. Vollrath, P. Johnston, and N. Kathuria (2005). *Market Access for High-Value Foods*. Agricultural Economic Report No. AER-840. U.S. Department of Agriculture, Economic Research Service.

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

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

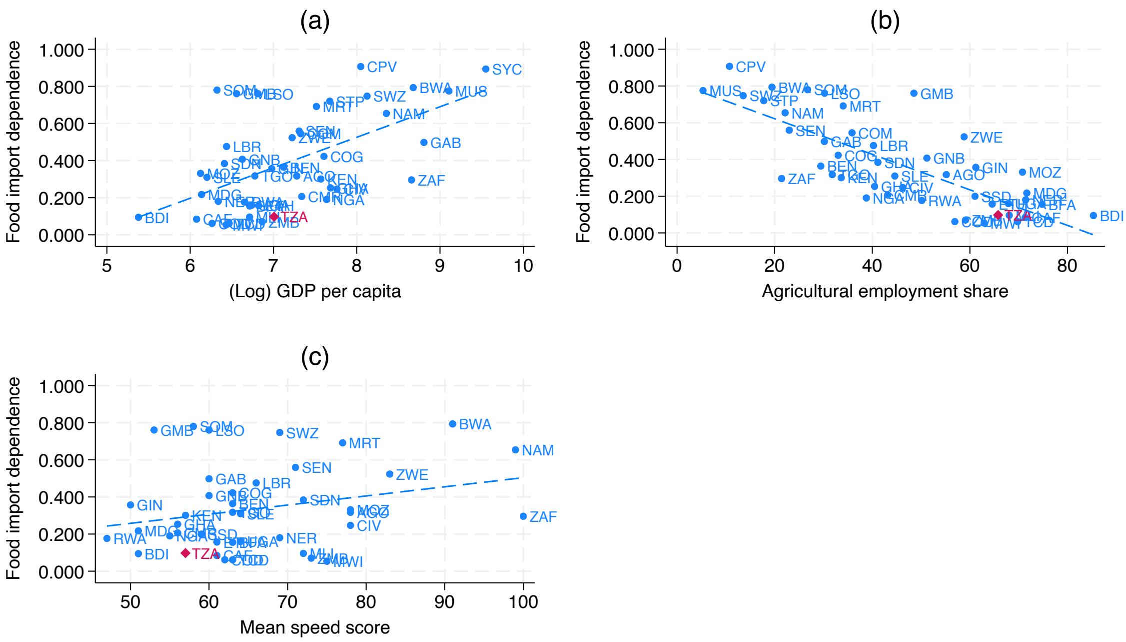
Suri, Tavneet (2011). "Selection and comparative advantage in technology adoption". In: *Econometrica* 79 (1), pp. 159–209.

Tombe, Trevor (2015). "The Missing Food Problem: Trade, Agriculture, and International Productivity Differences". In: *American Economic Journal: Macroeconomics* 7(3), pp. 226–258.

WFP (2022). *Tanzania Country Brief*. December 2022. World Food Program.

A Empirical appendix

Figure 6: Correlates of food import dependence



Note: This figure visualises bivariate correlations between food import dependence and GDP per capita, agricultural employment share, and mean speed score for 48 sub-Saharan African countries. Food import dependence is defined as the ratio of net import quantities to food quantities for all commodities in the FAO Food Balance Tables (data source: FAOSTAT, 2020). Quantities are weighted by the proportion of calories that each commodity provides in the overall calorie supply of the country. Since net imports of any given commodity can be negative or greater than unity, I impose a minimum value of zero (if the country is a net exporter) and a maximum value of one (if the country imports more than it consumes, either because of re-exports or processing). GDP per capita is reported in currect USD and shown in logs (data source: World Bank, 2020). Agricultural employment is measured as % of total employment (data source: World Bank, 2020). Mean speed score is a measure of road infrastructure quality, with higher scores indicating better infrastructure (data source: Moszoro and Soto (2022)).

B Calibration appendix

B.1 Classification of crops

To classify crops into export cash crops and staple food crops, I use data on agricultural production and trade from FAOSTAT to calculate the export share and import share for each crop. The export share indicates the proportion of domestic production that is exported to the rest of the world. The import share, on the other hand, gives imports relative to domestic production. Crops that have a high export share ($> 50\%$) and a low import share ($< 10\%$) are subsequently classified as export cash crops. Exceptions to this methodology are seed cotton, sugar cane and sisal, where raw production quantities cannot directly be linked to export and import quantities, as these crops are not traded in raw form. Instead, they are processed into e.g. cotton lint, refined sugar and fibre prior to being traded. I still classify cotton, sugar cane and sisal as export cash crops in the context of Tanzania, based on UN Comtrade statistics. The following ten crops are classified as export crops in the Tanzanian data:

Table 8: Export (cash) crops in Tanzania, 2010-20

	Item	Area	Production	Exports	Imports	Exports %	Imports %
1	Cashew nuts	575,581	182,137	174,114	1	96	0
2	Seed cotton	378,332	237,509	-	-	-	-
3	Coffee, green	187,685	55,664	52,033	7	93	0
4	Tobacco	123,100	93,826	69,622	2,414	74	3
5	Chick peas	83,121	67,664	46,785	121	69	0
6	Sugar cane	50,615	3,063,594	-	-	-	-
7	Sisal	50,254	34,308	-	-	-	-
8	Tea	19,197	39,840	25,805	252	65	1
9	Cocoa, beans	14,708	10,455	9,667	1	92	0
10	Pepper	956	443	256	33	58	7

Note: All values in this table are calculated from FAOSTAT data. Crops are ranked by Area (area harvested measured in hectares). Production, Exports and Imports are 2010-2020 mean quantities in tonnes. Exports (Imports) % shows the mean export (import) quantity as a percentage of the mean production quantity. All crops refer to unprocessed raw crops (for example, cashew nuts are in shell and tobacco is unmanufactured). Export and import quantities for seed cotton, sugar cane and sisal are missing because these crops are not traded in raw form. Instead, they are processed into e.g. cotton lint, refined sugar and fibre prior to being traded.

B.2 Classification of food items

Table 9: Classification of food items – Staple foods & processed foods

	Panel A: Staple foods	
bananas (cooking), plantains	groundnuts	rice (paddy)
bananas (ripe)	irish potatoes	seeds
cashew, almonds, other nuts	maize (flour)	starches (other)
cassava (dry/flour)	maize (grain)	sugarcane
cassava (fresh)	maize (green, cob)	sweet potatoes
cereals (other)	millet and sorghum (flour)	vegetables (green)
coconuts	millet and sorghum (grain)	vegetables (other)
fruits (citrus)	pulses	wheat (flour)
fruits (other)	rice (husked)	yams, cocoyams
	Panel B: Processed foods	
beer	fish and seafood (processed)	milk (fresh)
birds and insects (wild)	fish and seafood (fresh)	milk products
bread	fruits (canned)	salt
brews (local)	honey, syrups, jams	soft drinks
buns, cakes, biscuits	macaroni, spaghetti	spices (other)
butter, margarine, ghee	meat (beef)	sugar
cereal products (other)	meat (goat)	sweets
coffee and cocoa	meat (pork)	tea (dry)
cooking oil	meat (poultry)	tea, coffee (prepared)
eggs	meat products (other)	vegetables (processed)
fish (packaged)	milk (canned), milk powder	wine and spirits

Note: This table gives an overview of the classification of food items into either staple foods or processed foods for the purpose of calculating budget shares to parameterise the utility function. The classification is loosely based on Regmi et al. (2005). The overview is meant to be illustrative; Item codes are omitted and some items are summarised into broader categories to condense the information, for instance "fruits (other)".

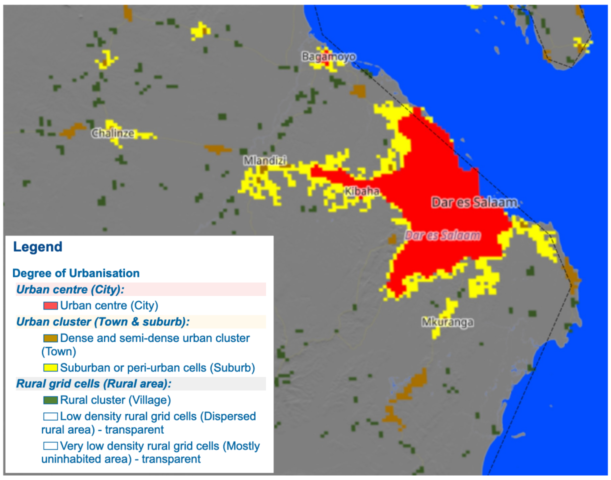
B.3 Mapping locations to model regions

I use the Global Human Settlement Layer's (GSHL) Degree of Urbanisation (DoU) classification to map households' geo-coordinates in the Living Standards Measurement Study (LSMS) data to regions in the stylised model. The LSMS data includes (modified) household coordinates, which can be mapped to one of the GHSL's seven DoU categories. I subsequently map these seven DoU categories to one of the three model regions to connect the geo-coded data to the stylised model.[17]

Table 10: Mapping Degrees of Urbanisation (DoU) clusters to model regions

Model Region	**DoU Grid Cell**
Urban region	Urban Centre (30)
	Dense Urban Cluster (23)
	Semi-Dense Urban Cluster (22)
	Suburban Or Peri-Urban (21)
Near rural region	Rural Cluster (13)
	Low Density Rural (12)
Far rural region	Very Low Density Rural (11)
N/A	Water (10)

Figure 7: Global Human Settlement Layer (GHSL) Degree of Urbanisation



[17] The mapping from household coordinates to DoU categories was conducted by Shraddha Mandi for a forthcoming working paper ("Urban-Rural Definitions and Spatial Income Gaps," by Douglas Gollin, Martina Kirchberger, David Lagakos, and Shraddha Mandi). I am grateful to the authors for sharing their data and code.

B.4 Estimation of agricultural production functions

Output. Output in value terms is calculated from output weight by crop and median crop prices in a given survey cluster. The use of median crop prices allows attaching a value to yields even of crops that are not sold in the market and for which value of output is not reported in the survey otherwise. To reflect the spatial heterogeneity in prices, I calculate median crop prices at the cluster level. Where median price at the cluster level is missing for given crop, so that the value of output cannot be estimated reliably, production quantities are not included in the subsequent estimation of production functions.

Land. Land use is calculated from GPS-measured plot size.[18] Calculating the amount of land used in production of a given crop requires a few key assumptions and follows different procedures for annual crops on the one hand, and permanent crops on the other hand. For annual crops, respondents report the share of a plot that is planted with a given crop.[19] Where the sum of these shares exceeds one, I renormalise them to sum exactly to one. For permanent crops (including fruit trees), the data only contains the number of plants per plot, but not the share of the plot area planted with a given crop. To avoid having to invoke plant size and spacing assumptions for the many different types of crops in the dataset, I proxy land use for permanent crops by the total area of the plot on which a given crop is planted.

Labour. To calculate labour used in production of a given crop, I combine the data on household labour and hired labour for a given plot. Harmonization of the data on household labour and hired labour requires some pre-processing of the household labour data. I aggregate hours spent working on a plot across activities and divide the number of hours by 10 to approximate the number of days. I then aggregate household labour and hired labour at the plot level to yield total labour used on a given plot. Where several different annual crops are planted on a plot, I use the adjusted land shares to attribute labour to the different crops.

Intermediates. Intermediates comprise seeds, fertiliser, pesticides, and herbicides. Seed use is reported at the crop-plot level while the use of chemicals (fertiliser, pesticides, herbicides) is reported only at the plot level. I aggregate the value of the four different types of inputs at the crop-plot level.

Estimation. I estimate the production function separately for staple food crops and export cash crops. Before estimation, I aggregate the data to the crop-cluster level, summing values for a given crop across all households within a cluster. This is to increase the number of observations, since many household do not use any intermediates in production, which would mean they would have to be excluded from the regression. Using the crop-cluster level data, I run the following log-log regression for staple food crops, pooled over all crops that are classified as staple foods

[18] Where GPS-measured plot size is missing, reported plot size is used.

[19] This measure is discretised into 25%, 50%, 75% or 100% of a plot being planted with a given crop.

$$\log(Q_{i,a}) = cons + \hat{\gamma_l}\ \log(l_{i,a}) + \hat{\gamma_n}\ \log(n_{i,a}) + \hat{\gamma_x}\ \log(x_{i,a}) + u_{i,a}$$

where $l_{i,a}$ is land (in acres), $n_{i,a}$ is labour (in days), and $x_{i,a}$ is intermediate inputs (in value terms). Regression results and corresponding parameter estimates are reported in Table 1.

B.5 Model parameterisation

Table 11: Overview of model parameters

Parameter	Value	Description
θ_x	0.26	Output elasticity of the intermediate input in staple crop production
θ_n	0.26	Output elasticity of labour in staple crop production
θ_l	0.48	Output elasticity of land in staple crop production
γ_x	0.34	Output elasticity of the intermediate input in cash crop production
γ_n	0.20	Output elasticity of labour in cash crop production
γ_l	0.46	Output elasticity of land in cash crop production
δ_x	0.20	Output elasticity of the intermediate input in processed food production
δ_n	0.10	Output elasticity of labour in processed food production
δ_a	0.70	Output elasticity of staple crop in processed food production
ω_n	1.00	Output elasticity of labour in manufactured good production
$A_{2,a}$	1.00	TFP in staple crop production in the far region
$A_{2,c}$	1.00	TFP in cash crop production in the far region
$A_{1,a}$	1.00	TFP in staple crop production in the near region
$A_{1,c}$	1.00	TFP in cash crop production in the near region
$A_{0,p}$	1.00	TFP in processed food production in the urban region
$A_{0,m}$	1.00	TFP in manufactured food production in the urban region
$\bar{A}_p$	1.00	Productivity parameter (processed food)
$\bar{A}_m$	1.00	Productivity parameter (manufactured good)
λ_p	0.44	Share parameter (processed food)
λ_m	0.49	Share parameter (manufactured good)
ρ_p	0.42	Elasticity parameter (processed food)
ρ_m	0.56	Elasticity parameter (manufactured good)
λ_f	0.57	Share parameter (composite food)
ρ_f	0.26	Elasticity parameter (composite food)
$\bar{c}_f$	0.15	Subsistence level consumption of composite food
α	0.31	Preference parameter for food consumption
n	1.00	Total population (available labour) in the economy
l_2	0.82	Land available for production in the far region
l_1	0.18	Land available for production in the near region
d_2	0.22	Transport cost between the far region and the near region
d_1	0.11	Transport cost between the near region and the urban region
d_0	0.20	Transport cost between the urban region and RoW
$ratio_{trade}$	0.10	Balancing ratio to account for unmodelled exports in external balance
p_x	0.23	International market price for the intermediate input
p_m	0.23	International market price for the manufactured good
p_p	12.00	International market price for processed food
p_c	1.00	International market price for the cash crop