Essays on Macro Development, Global Price Shocks and Food Security

Lisa Martin

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

September 2024

Essays on Macro Development, Global Price Shocks and Food Security

Lisa Martin

Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics

September 2024

In many low-income African countries, agriculture is the dominant sector in terms of employment and output. What are the key features of the agricultural sector? Two points stand out. First, agriculture is the sector most centrally involved in the production of food – and food accounts for the largest share of expenditures for low-income consumers. Indeed, many people live close to a subsistence level, and food availability is a central concern. Second, and less obviously, agriculture is the sector that most fundamentally uses land as an input. Consequently, production is necessarily dispersed across geographic space. The movement of inputs and outputs – often bulky and/or perishable – is an unavoidable feature of agricultural economies. This in turn implies that spatial frictions play an important role in agricultural economies. Moreover, transport costs are extremely high in many low-income economies – by one estimate, four or five times higher than in the United States. Faced with high spatial frictions, many people in these economies live in quasi-autarky, producing most of the food that they consume.

In this thesis, I consider the ways in which low-income economies in Africa are shaped by food dependence and by spatial frictions – and by the interplay between these two characteristics. The thesis consists of three stand-alone essays that address different aspects of development, focusing on spatial frictions and food security.

Chapter 1 analyses the role of domestic trade frictions for processes of structural change in an open economy. In a closed economy, low agricultural productivity leads a large share of the population to work in agriculture to produce sufficient food. By contrast, a perfectly open economy with low agricultural productivity may import food to meet subsistence needs, thus releasing labour into other more productive sectors in the economy. I develop a stylised spatial general equilibrium model of a small open economy with spatial heterogeneity and high domestic trade frictions. I find that in such an imperfectly open economy, 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.

Chapter 2 (co-authored with Christopher Adam and Douglas Gollin) investigates the impact of global food, fuel and fertiliser price shocks on welfare in low-income countries – where these shocks are again mediated by the spatial structure of the economy. We develop a spatial general equilibrium model of a small open agricultural economy, calibrated to Tanzanian data, to study the transmission of global price shocks to the consumption patterns of heterogeneous households located in different regions in the domestic economy. Our model simulations show strong spatial heterogeneity in the effects of import price shocks. While urban households' consumption baskets are more exposed to global food price shocks, remote rural households' production and consumption are heavily dependent on fuel and fertiliser imports. We use the calibrated model to compare potential mitigation policies and different financing arrangements.

Chapter 3 focuses on the prediction of household food insecurity. Using machine learning, I seek to predict household food insecurity using data from household surveys, combined with readily available additional data on weather, prices, and geography. I train and test tree-based ensemble models for the prediction of households' food security status. The models achieve overall accuracy of 54-65% at a true positive rate of 70%. The approach I develop provides a way to assess household food security without the need for costly and time-consuming primary data collection. In this paper, too, spatial frictions operate in subtle but important ways. In an open economy with well-integrated markets, household food security would essentially be determined by world prices and household income levels. But that is not the case for Tanzania. Markets are poorly integrated. Households and communities face various shocks that shift local (and individual) prices and wages. In short, spatial frictions shape the patterns of food insecurity.

The three thesis chapters thus share a common set of issues. All three chapters also focus on Tanzania, a country that is fairly typical of low-income coastal economies in sub-Saharan Africa. The chapters use different methodologies and draw on different tools. Taken together, they seek to contribute to an understanding of development issues in low-income agricultural economies.

Supervisors This thesis was supervised by Douglas Gollin and Simon Quinn.

Total words This thesis contains approximately 60,216 words. The word count was calculated as the number of words on a representative page (386 words, page 3) multiplied by the total number of pages, excluding the cover, table of contents, and references.

Statement of authorship Chapter 1 and Chapter 3 are single-authored. Chapter 2 is joint work with Christopher Adam and Douglas Gollin.

Acknowledgement of funding I gratefully acknowledge funding from the Oxford Department of Economics and the German Academic Scholarship Foundation. I am thankful for generous financial support from the Structural Transformation and Economic Growth (STEG) programme and the Centre for the Study of African Economies (CSAE).

Acknowledgements

I want to start by thanking the people who were most influential for my journey on the DPhil – my supervisors, Douglas Gollin and Simon Quinn. I deeply appreciate your mentorship and support over these years. To Doug – I cannot thank you enough for how generous you have been with your time, for the words of wisdom, your patience and kindness. I have learned so much from you about economics and about being a good economist. To Simon – thank you for your guidance and attention to detail, for sifting through code with me, and also for giving such brilliant and inspiring lectures during my MSc and PRS years.

To my co-author, Christopher Adam – Thank you for all you taught me about macroeconomic modelling, and for making my DPhil research so much more enjoyable. To the broader academic community at the Centre for the Study of African Economies (CSAE) – Sanjay Jain, Christopher Woodruff, Julien Labonne, Natalie Quinn and Banu Demir Pakel; I am grateful for all the valuable feedback I have received from this inspiring network of researchers.

I want to thank the broader community at the Department of Ecnomics for providing an academic home during my DPhil years. Particular gratitude also goes to Hamish Low and Abi Adams-Prassl for being so supportive of the Women in Economics initiatives. To Keble College – thank you being my a social home in Oxford since 2017. To Nicola Fuchs-Schündeln and her team at Goethe University Frankfurt: Thank you for giving me an office space and – more importantly – a community during the final months of my DPhil that I spent working in Frankfurt. I am particularly grateful to Celina Proffen for the companionship and stimulating conversations.

To friends in the DPhil community – Verena, Sanghamitra, Hannah Z., Luke, Wenzhang, Amelie, Lukas, Tom, Vatsal, Shihang, Binta, Giulio, Rocco, Manos, Alex. Thank you for stimulating discussions, valuable feedback, and for cheering me up on lunch and coffee breaks. Thank you to Hannah R. and Maria for keeping the Women in Economics mentoring programme and alumnae network alive after I took a step back to finish the DPhil. Thank you to Jonas for countless hours of co-working sessions on Teams – these sessions not only gave structure to long days spent working remotely, but also made the experience a lot less lonely.

I am incredibly grateful to Verena and Raphael for giving me the opportunity to join them on their research visits in Kenya – the weeks we spent together in Nairobi and Kakuma were some of the most insightful and joyful times during my DPhil. Thank you to Raphael and Rocco for making me feel at home at Marston St whenever I needed a place to stay in Oxford during the final stretch of my DPhil.

To Tatjana and Diala – your friendship, loving support and words of encouragement mean so much to me – you are an inspiration. To Vanessa and Cécile, I am very thankful for our deep conversations, shared laughter, and for being able to learn from you. To Raphael, I am so glad we finally met at Keble. Your humor and friendship brought much lightness into my DPhil years. Also, thank you for your bravery in teaching me to scull.

To my Oxford MSc friends – Vicky, Sophie and Ishleen, thank you for making even exam preparations on the MSc course such a happy time. Your friendship and support during the MSc encouraged me to return to Oxford for the DPhil. Thank you to Gideon, another fellow MSc alumnus, for hosting me in Nairobi during my research visit there.

To my community of friends outside of Oxford – Samy, Laura, Florian, Maximilian, Rosalie, Benny and Huong. Thank you for taking me out of the bubble, for bearing with my stories about the struggles of doing a DPhil, and for making me laugh so much and so heartily.

To my family – Karen, Helmut, Tanja, Hannah and Marie. Your encouragement and unwavering support over the years meant more than I can say. You shared in my joy and lifted me up when I felt down, and I could not have done it without you.

And last, but not least, to Filip – Thank you for always believing in me and for supporting me wholeheartedly. Whenever I was unsure if I was up to the challenge, you were there to tell me I could do it. I am eternally grateful to you for cheering me up with so many flowers, favorite meals and morning coffees. Thank you for travelling this road with me, now and always.

Finally, I gratefully acknowledge funding from the Oxford Department of Economics, the Structural Transformation and Economic Growth (STEG) programme, the Centre for the Study of African Economies (CSAE) and the German Academic Scholarship Foundation.

Lisa Martin
Oxford, April 2024

Contents

List of Figures

List of Tables

Introduction

In many low-income countries, the share of employment in agriculture is higher than agriculture's share in GDP. This implies that labour productivity in agriculture is lower than in other economic sectors, and new research suggests that agricultural productivity in several sub-Saharan African countries has recently even been declining (Wollburg et al., 2024). The agricultural productivity gap poses an economic puzzle. Why do so many people work in such an unproductive sector? And why do so many people continue to be food insecure in countries where such a large share of the population produces food?

This thesis consists of three stand-alone essays that focus on the interaction between food as a central production and consumption good, and the importance of domestic trade frictions in low-income countries. When food consumption is close to subsistence levels, food plays a central role in shaping an economy's overall production and consumption structure. Food has idiosyncractic properties that distinguish it from other consumption goods. Importantly, food is subject to minimum consumption requirements which cause demand to be highly inelastic. This non-homotheticity has important implications for processes of structural transformation and economic growth. In addition, food production differs from the production of other goods in important ways that warrant spatial analysis. Land, a key input in agriculture, is spatially dispersed and immobile, and the distribution of food is complicated by its bulkiness and perishability.

Domestic trade frictions, which are particularly high in low-income countries (Atkin & Donaldson, 2015), are a key reason why food remains so central. High domestic trade frictions make it economically optimal for households to locate in places where food is produced. Previous research has shown that high domestic trade costs can hinder structural transformation by limiting the use of intermediate inputs in agricultural production and by driving up the cost of food distribution (Gollin & Rogerson, 2014). Regions' differential integration into domestic and global markets matters for the way in which food is produced and consumed. High transport costs imply that often there is not a single domestic market for food. Transport costs create substantial spatial price gradients and limit the penetration of imports to remote regions. The resulting spatial heterogeneity in food consumption patterns imply spatially differentiated exposure to global and local shocks. Domestic trade costs can thus have profound implications for processes of structural change, for the pass-through of global price shocks on different regions in the economy, and for the effects of local agricultural production shocks on food security outcomes.

While all three chapters draw on data from the Tanzanian economy, the qualitative findings in this thesis are relevant more broadly beyond the context of Tanzania. Tanzania typifies the kind of economy that is central to my research questions on the intersection of two features of low-income economies: the importance of food as both a production and consumption good, and the salience of domestic trade frictions in settings where markets are poorly integrated. Tanzania shares several key features with many other low-income sub-Saharan African economies. The economy is open to world trade at the border, but domestic trade frictions are substantial. Its main city and trading hub is located on the coast, with an adjacent agricultural hinterland. Large spatial areas of the country are dedicated to agriculture, with the non-agricultural economy spatially concentrated in cities. A large fraction of the population works in agriculture, which has lower labour productivity than other sectors in the economy. Although Tanzania registered robust economic growth over the past two decades,[3] poverty remains widespread and a considerable part of the population continues to be food insecure (WFP, 2022).

In addition to the research questions' relevance in the case of Tanzania, data availability is significantly better than for many other developing countries. Other sub-Saharan African countries that have a broadly similar spatial and economic structure, such as Mozambique, Ghana, Ivory Coast, Senegal, Angola, Kenya

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

(though its major city is not coastal), or Nigeria (although bigger and more diverse), do not have the same quality and richness of data as Tanzania. Further, political institutions in Tanzania are relatively stable, so data are not greatly distorted by effects of conflict. Detailed household survey data from the Living Standards Measurement Study (LSMS) and macroeconomic data from the Social Accounting Matrix (SAM) form the cornerstones of the empirical analysis in this thesis. In addition, I draw on various other data sources, including domestic market price data provided by the World Food Programme (WFP), data on global food prices, agricultural production and land use from the Food and Agriculture Organization (FAO), and UN Comtrade trade statistics.

Chapter 1 explores the role of domestic trade frictions for structural transformation in an economy that is open at the border. Previous work on the "food problem" has shown that in a closed economy with low agricultural productivity and high domestic trade frictions, minimum consumption requirements for food create the need for a large share of the labour force to work in agriculture (Gollin & Rogerson, 2014). Existing research does not, however, investigate whether the same mechanism applies in an open economy that can import food to meet subsistence needs. While macroeconomic models typically conceptualise trade openness as a binary attribute, many economies are effectively *imperfectly* open, i.e. they are open at the border but varying degrees of domestic trade frictions imply that some regions are only loosely integrated with world markets. In this chapter of the thesis, I develop a stylised spatial general equilibrium model of a small open agricultural economy with spatial heterogeneity and high domestic trade frictions to explore processes of structural change in imperfectly open economies.

In imperfectly open economies, subsistence needs may still exercise considerable influence over the sectoral allocation of resources, despite trade openness at the border. I find that, depending on the parameterisation of the model, the food problem can continue to bind in imperfectly open economies. I document three key properties of the model. First, agricultural productivity directly affects the spatial distribution of population, but the direction and magnitude of this effect depend on the degree of external openness and internal trade frictions. Second, domestic trade frictions exert significant influence over the spatial and sectoral composition of production and the economy's pattern of trade with the rest of the world. Third, an increase in domestic trade frictions causes divergence in food consumption patterns across regions, limiting the penetration of imported goods to remote rural areas. Higher transport costs lead overall food consumption to approach the subsistence level in all regions in the economy, and food consumption baskets are skewed to favor locally produced food which are less affected by the increase in transport costs.

Chapter 1 contributes to the existing literature on structural transformation (Bustos et al., 2016; Fajgelbaum & Redding, 2022; Gollin & Rogerson, 2014; Gollin et al., 2002; Matsuyama, 1992; McCaig & Pavcnik, 2013; Ngai & Pissarides, 2007) by proposing a stylised framework to study structural change in an (imperfectly) open economy. The analysis relates closely to a set of papers that study the food problem in open economies (Nath, 2020; Tombe, 2015) and to two other papers that focus on the link between trade openness and food security (Burgess & Donaldson, 2012; Janssens et al., 2020). More broadly, this chapter also relates to the growing literature that studies the role of domestic trade frictions (Allen & Arkolakis, 2014; Atkin & Donaldson, 2015; Donaldson, 2018; Sotelo, 2020).

The analysis in Chapter 1 suggests that, due to domestic trade frictions, changes in world prices of imports and exports pass through imperfectly to the domestic economy. The pass-through is mediated both by domestic trade frictions and by the spatial specialisation of production, i.e. the fact that certain economic activities are concentrated in particular regions. The spatial structure directly affects the pass-through of import prices to consumers in different locations, implying differential welfare effects. In addition, the spatial structure also mediates the pass-through of global prices to domestic firms that use imported intermediate inputs in production or that export products to global markets.

Chapter 2 (co-authored with Christopher Adam and Douglas Gollin) investigates this spatially differentiated pass-through of world market price shocks. We analyse the resulting heterogeneous welfare effects and distributional consequences of potential fiscal mitigation policies and alternative financing arrangements. The analysis in this chapter is motivated in particular by the increased volatility in world market prices of food, fuel, and fertiliser in recent years, which was driven by the decompression of the post-pandemic global economy, and sharply accelerated by Russia's invasion of Ukraine in February 2022. The effects of such shocks are often felt most strongly in low-income countries, due to their dependence on imports, widespread poverty, and limited scope for offsetting fiscal policies. Consequently, global food, fuel, and fertiliser price shocks can have far-reaching implications for welfare, particularly with respect to food security, in low-income countries.

We develop a spatial multi-sector general equilibrium model of a small open agricultural economy to study the short- to medium-term impact of global price shocks and to consider alternative fiscal policy responses to these shocks. Previous work on the impact of price shocks focuses largely on aggregate country-level analyses, assessing vulnerability to shocks based on national trade statistics, average consumption baskets, and aggregate production data (Arndt et al., 2023; Bekkers et al., 2017; Meyimdjui & Combes, 2021; Okou et al., 2022). Our modelling approach contributes to this literature a framework that facilitates the analysis of spatially heterogeneous effects.

For calibration, we follow Adam et al. (2018) in fitting the model to the economy of Tanzania based on the 2018 Social Accounting Matrix (SAM). As the SAM in itself does not have a spatial dimension, we microfound modelling assumptions around the spatial distribution of production and consumption using geo-coded household survey data from the LSMS. We use the calibrated model to examine heterogeneous pass-through of shocks from a positive perspective and to conduct granular normative welfare analysis, across locations and household types. Specifically, we compare the short-term distributional consequences of different fiscal mitigation measures, such as import price subsidies and net household transfers under different financing arrangements.

We find that an isolated price shock for food imports has the most adverse effect on urban wage-earning households, who are net consumers of food and who rely more heavily food imports. In the case of a simultaneous shock to food, fuel, and fertiliser prices, remote agricultural households are equally negatively affected, due to the transport-cost intensity of consumption and the reliance on imported fertiliser in rural regions. Without a mitigating policy response, a simultaneous shock to import prices of food, fuel, and fertiliser of +20% decreases households' aggregate real income in the domestic economy by 7.3%.

Comparative static analysis of counterfactual policies shows that import price subsidies which fully offset the global price shocks can stabilise households' real incomes. When import price subsidies are financed domestically through a higher tax on the capitalist's income, average real income per capita falls by 0.3% for labour-supplying households, compared to a decline of 8.6% without policy mitigation. In the case of aid-financed policy mitigation, we find that import price subsidies exhibit an urban bias, with urban households benefiting disproportionately. Rural households' income, on the other hand, can be shielded more effectively through direct per capita transfer payments.

Accurate targeting of any such policy interventions to protect vulnerable households from the effects of adverse shocks – be they global price shocks or domestic production shortfalls – requires correct identification of the households most at risk. In Chapter 3 of this thesis, I explore whether machine learning methods can be used to guide targeting of food policy. Food security is a multi-dimensional phenomenon that is not directly observable. Many different indicators for measuring food security exist, all of which require costly and time-consuming primary data collection. In this chapter, I draw on existing detailed household survey data from the LSMS and readily available secondary data to train machine learning models for the prediction of household food insecurity.

Direct comparison of different experience-based, consumption-based, and monetary indicators for the same set of households confirms previous findings in the literature that these indicators are only weakly correlated and appear to capture independent dimensions of food security. In the prediction section, I thus focus on two different indicators, one experience-based (the reduced Coping Strategies Index, or rCSI) and one consumption-based (the Food Consumption Score, or FCS), which allow to classify household food security status based on pre-defined categorical bounds.

I use tree-based ensemble methods and readily available data on weather, prices, geography, and household assets to predict households' current food security status according to rCSI and FCS. In that, the approach is similar in spirit to other studies that use machine learning methods to bridge data gaps in low-income countries for improved policy targeting (Aiken et al., 2022; Blumenstock et al., 2015; Jean et al., 2016). This paper relates closely to recent studies that employ machine learning models and secondary data to predict the prevalence of food insecurity at various geographical scales (Lentz et al., 2019; Martini et al., 2022; Zhou et al., 2021).

Motivated by the findings in Chapter 1 and Chapter 2 that domestic trade frictions can create spatial heterogeneity in food consumption patterns and pass-through of global shocks, I include variables that capture households' remoteness in the predictor set for the machine learning models. These remoteness indicators, such as distance to the nearest major road and distance to the nearest agricultural market, reflect households' market access, which can mediate how different drivers affect food security outcomes. In addition, I include other predictors that have a spatial dimension, such as rainfall in the household's location and staple food prices in the nearest market.

Comparing predictive performance with and without certain categories of predictors reveals their differential importance. Whereas household demographics add relatively little predictive power, basic household assets (roof type and cellphone ownership, data on which can conceivable be accessed without household surveys) contribute substantially to model performance. While the remoteness indicators add relatively little predictive performance, I still include them in the predictor set of the main models, due to their ease of accessibility (compared to e.g. household demographics).

Under the assumption that policy makers care more about failing to identify truly food insecure households than about misclassifying a secure household as insecure, I prioritise avoiding errors of exclusion. To establish an upper bound on exclusion errors when comparing predictive performance of different models, I fix the recall rate to 70%, which implies that a model is able to detect 7 out of 10 food insecure households. At this level of recall, models predicting FCS achieve higher overall out-of-sample accuracy than models predicting rCSI (65% vs. 54%). The results suggest that inferring households' food security status using machine learning methods can provide a complementary approach for policy targeting that does not rely on costly and time-consuming primary data collection.

CHAPTER 1

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

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.

1.1 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 et al. (2002) and Gollin 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 (Bustos et al., 2016; Fajgelbaum & Redding, 2022; Gollin & Rogerson, 2014; Gollin et al., 2002; Matsuyama, 1992; McCaig & Pavcnik, 2013; Ngai & Pissarides, 2007). 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 (Nath, 2020; Tombe, 2015) and to two other papers that focus on the link between trade openness and food security (Burgess & 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 & Arkolakis, 2014; Atkin & 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.[2] Second, I open the economy to trade

[2]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.

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 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.[3] 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).[4] 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).[5] 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 productiv-

[3]While I choose to calibrate the model to Tanzanian data, the research questions in this chapter 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 1.6 in Appendix 1.A).

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

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

ity 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; 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 1.2 briefly summarises related literature and highlights the contribution of this paper. Section 1.3 contains details of the modelling approach and Section 1.4 illustrates some of the model's key properties. Section 1.5 presents the calibration of the model to Tanzanian data, and Section 1.6 describes the baseline equilibrium at these calibration values. Section 1.7 presents results from comparative statics, contrasting the effects of different productivity improvements. Section 1.8 concludes and defines avenues for future research.

1.2 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 (Emerick, 2018; Gollin & Rogerson, 2014; Gollin et al., 2002; Matsuyama, 1992; Ngai & Pissarides, 2007) and in open economies (Bustos et al., 2016; Fajgelbaum & Redding, 2022; McCaig & Pavcnik, 2013). 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 & Pissarides, 2007). The other is a utility-based approach, in which non-homothetic preferences yield structural change (Gollin & Rogerson, 2014; Gollin et al., 2002; Matsuyama, 1992). 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.

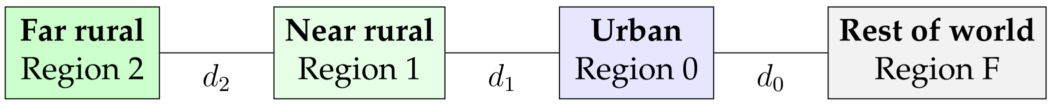
This finding is particularly relevant against the backdrop of accelerating climate change. Nath (2020) shows that, if trade costs remain high, climate change may exacerbate the food problem, such that developing countries with binding subsistence constraints intensify their specialisation in agriculture 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 (Baptista et al., 2023; Burgess & Donaldson, 2012; Janssens et al., 2020).

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 & Arkolakis, 2014; Atkin & Donaldson, 2015; Donaldson, 2018; Farrokhi & Pellegrina, 2023; Sotelo, 2020). 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.

1.3 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 1.1 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 1.1: 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).

1.3.1 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.

1.3.2 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.

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.

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.

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.

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 ($x_{r,a}$, $x_{r,c}$, $x_{0,p}$):[6]

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

$\bar{A}_m$ 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.

1.3.3 Trade and transport costs

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

[6]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.

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).

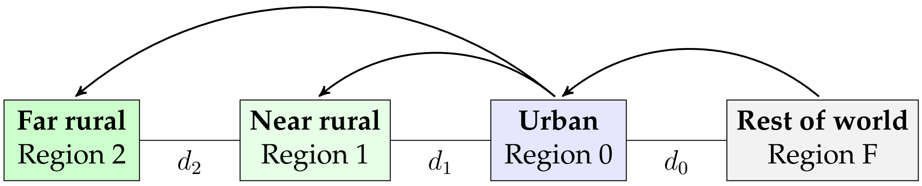
International trade

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.

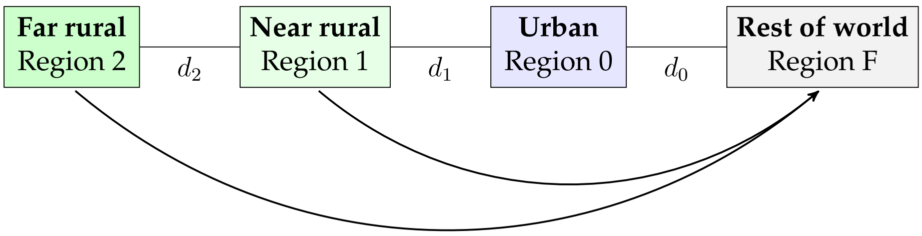
Direction of trade flows

Figure 1.2 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.

(a) Processed food (p) and manufactured good (m, x)



(b) Export crop (c)



(c) Staple crop (a)

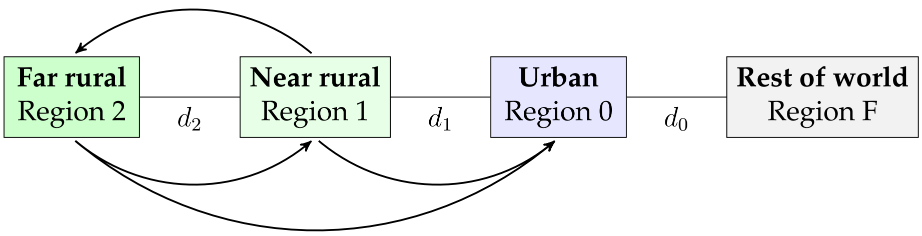


Figure 1.2: Direction of feasible trade flows by sector

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.[7] 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 1.3.5 below.

1.3.4 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,

$$\sum_{r=0}^{2} \frac{n_r}{n} \Big[\alpha \, \log \big(C_{r,f} - \bar{c}_f\big) + (1-\alpha) \, \log(C_{r,m})\Big] \tag{1.1}$$

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 (1.1) serves as the objective function in the numerical optimisation.

The utility maximisation in Equation (1.1) 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.

1.3.5 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*.

[7] 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.

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.

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

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 1.3.3 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 1.6, 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

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

1.3.6 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.[8] A region's aggregate

[8]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.

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.

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

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

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

1.3.7 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 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

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

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

1.4 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.[9] 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.*

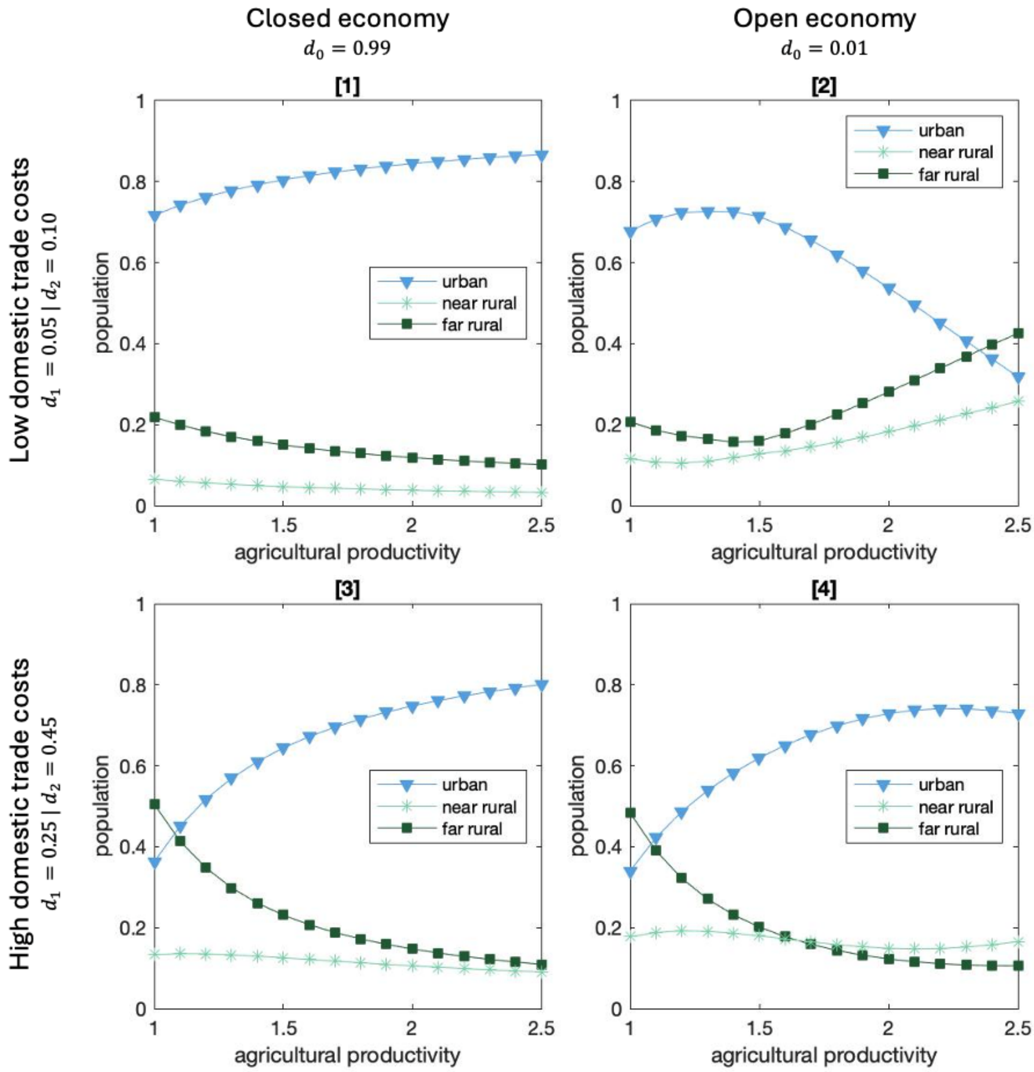
In the model, agricultural productivity has a strong effect on the spatial distribution of population, but the direction and magnitude of the effect depend crucially on the degree of external openness and internal trade frictions. In a closed economy, an increase in agricultural productivity releases 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.

[9]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 1.3 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 econ-

Figure 1.3: 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: $A_{0p} = 1$, $A_{0m} = 2$, $\theta_x = 0.1$, $\theta_n = 0.3$, $\theta_l = 0.6$, $\gamma_x = 0.33$, $\gamma_n = 0.33$, $\gamma_l = 0.33$, $\delta_x = 0.33$, $\delta_n = 0.33$, $\delta_a = 0.33$, $\omega_n = 1$, $\lambda_p = 0.5$, $\rho_p = 0.9$, $\epsilon_p = 10$, $\lambda_m = 0.5$, $\rho_m = 0.9$, $\epsilon_m = 10$, $\lambda_f = 0.5$, $\rho_f = 0.4$, $\epsilon_f = 1.66$, $\alpha = 0.2$, $\bar{c}_f = 0.15$, $\bar{c}_m = 0$, $T = 0$, $p_m = p_x = 0.3$, $p_p = 1$, $p_c = 1$, $n = 1$, $l_2 = .8$, $l_1 = .2$

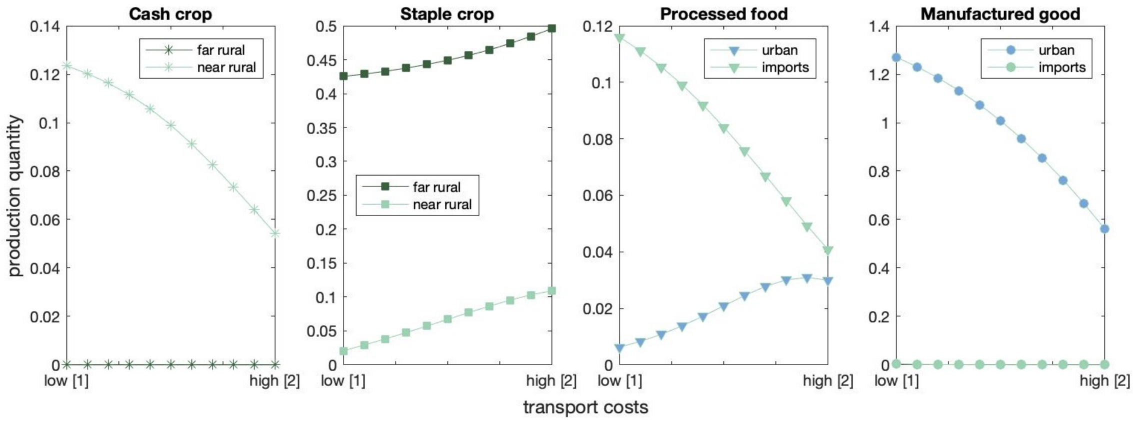
omy (panel [2]), an increase in agricultural productivity leads to a higher share of the population working in agriculture. When domestic transport costs are high, openness at the border has a much weaker effect. Panels [3] and [4] show that, when domestic transport costs are high, the impact of agricultural 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 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 1.4: Domestic transport costs and the sectoral allocation of economic activity

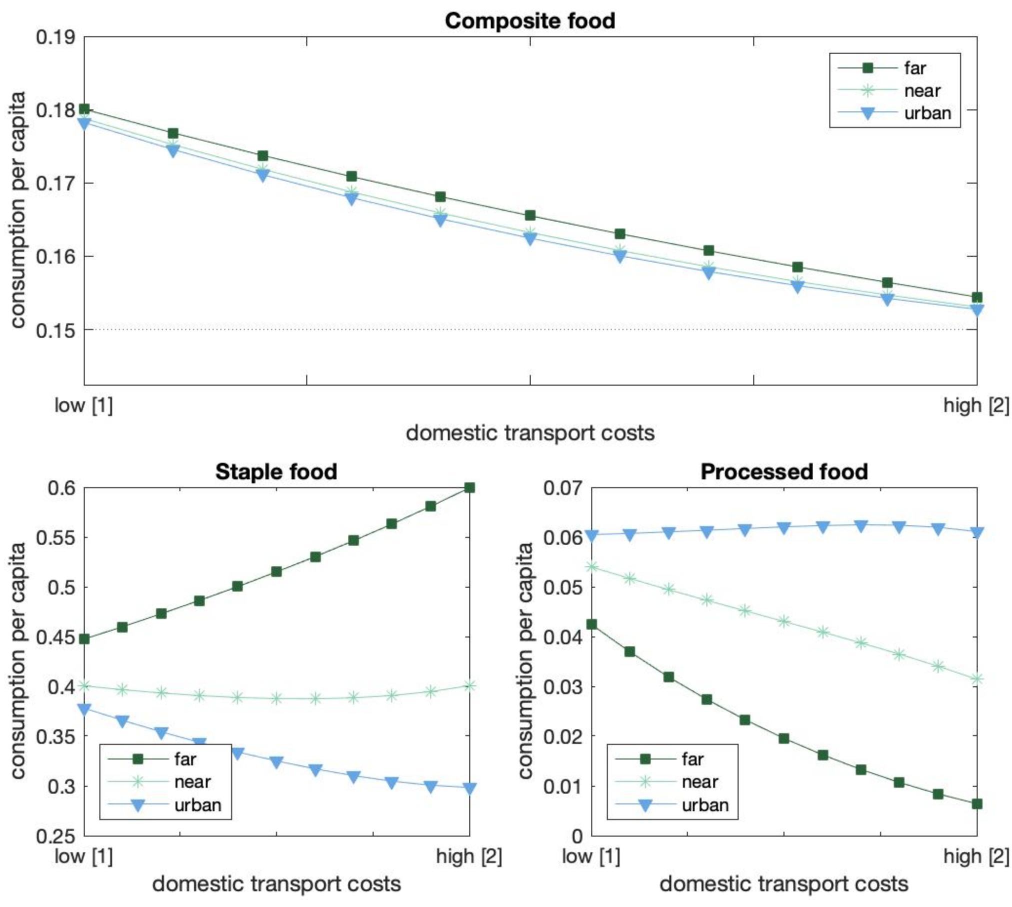


Note: This figure shows domestic production and import quantities by sector at different levels of domestic transport costs (d_1, d_2); [1] low domestic transport costs at the far left of the x-axis are $d_1 = 0.05$ and $d_2 = 0.1$; [2] high domestic transport costs at the far right of the x-axis are $d_1 = 0.25$ 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 1.5). Food consumption allocations favor locally produced items, leading to diverging consumption patterns across regions (see lower panels in Figure 1.5). 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 1.5: 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.

1.5 Calibration

The model properties described in Section 1.4 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 1.7. A comprehensive overview of the model's parameterisation at the baseline equilibrium can be found in Table 1.11 in Appendix 1.B.5.

1.5.1 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.[10] Output (in log value terms) is regressed on land (in log acres), 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 1.1 presents raw and normalised coefficient estimates.[11]

Table 1.1: 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).

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

[11] 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.

1.5.2 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.[12]. 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.[13] 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).

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 $\lambda_p = 0.5$.

1.5.3 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

[12]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

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

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 $\lambda_m = 0.5$.

1.5.4 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 1.2).[14] 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.[15] I estimate the elasticity of substitution 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.

[14] I use the GHSL Degrees of Urbanisation dataset to classify households as urban, near rural, or far rural. See Appendix 1.B.3 for details.

[15] The classification of food items into staple foods and processed foods is shown in Table 1.9 in Appendix 1.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 in value terms, I use median prices (calculated from data on food purchases) to infer the value of consumption from own production and gifts.

Table 1.2: 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 1.9 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 $= 330$).

1.5.5 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 $l_1 = 0.18$.

1.5.6 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 1.3 below shows the model fit with respect to the targeted moments and reports the resulting parameter values for transport 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 1.3: 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.

1.6 Baseline equilibrium

Given the set of parameters described above, I solve the model numerically.[16] Aggregate utility (Equation 1.1) 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 1.4 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

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

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 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 1.4: 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 1.5 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 1.5: 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 1.6. 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.

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

Table 1.6: 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%.

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.

1.7 Counterfactual analysis

In this section, I consider four counterfactual experiments to explore the differential effects of alternative productivity improvements in the economy.[17] 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.

1.7.1 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 economy to maintain a given level of food production at lower levels

[17]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).

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.

1.7.2 Results

Table 1.7 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 1.7: 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 %

–Continued on next page–

Table 1.7 – 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
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 %
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.

1.8 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.

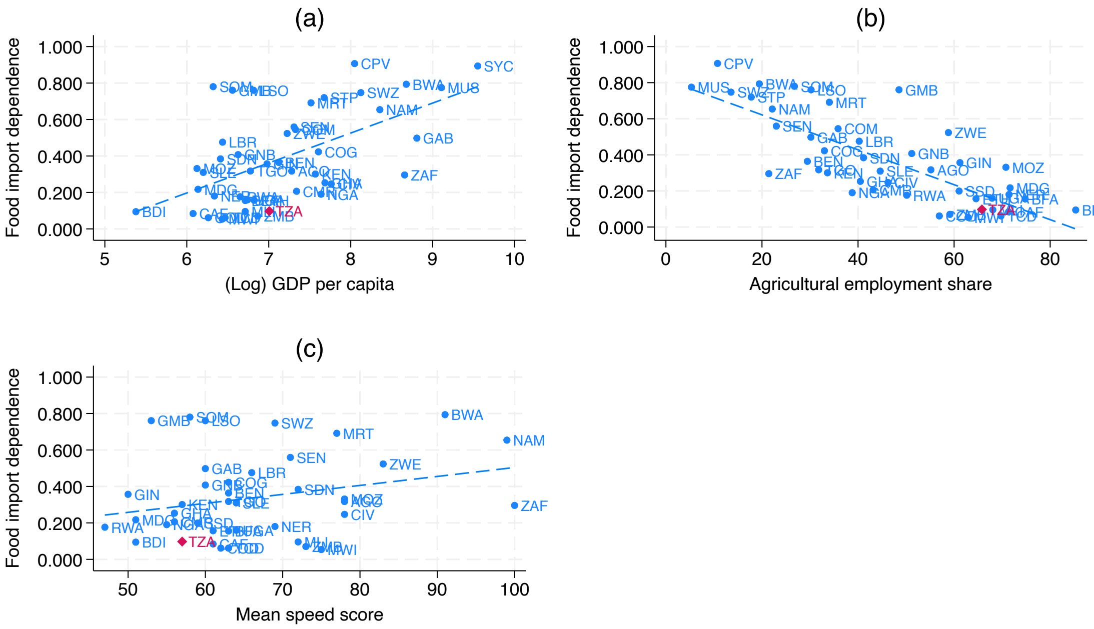
Interesting future extensions of the analysis in this paper would be to allow for different tradability of goods. For example, how would the qualitative model results change if there was an urban export good – either in addition to or instead of the agricultural export good? I presume that parameterisation 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.

Although it is not the focus of this paper, a model with these features would also allow us to analyse the impact of various short-term shocks, ranging from a world price shock to a local shock to agricultural productivity. In contrast to many widely used models, in which world prices are transmitted evenly and fully through the domestic market, this framework suggests a less complete pass-through. A spike in the world price of food, for instance, would pass through more or less fully to urban consumers in the model economy, but the effects would be muted for the near and remote rural areas. By contrast, a production shock in a remote rural area (e.g., from a drought or flood) would have a muted effect on urban consumers, who have access to alternative sources of food through the world market. In a sense, the model used here begins to provide a framework for thinking about the exposure of a low-income African economy to different types of shocks. The spatial and sectoral segmentation of the economy is a central feature of an economy like that of Tanzania. Models that reflect this segmentation can help in understanding the ways in which shocks are propagated and dampened as they are transmitted across space and through the input-output structure of the economy. In short, a model with these features serves as a useful tool for thinking about long-run structural change but also about shorter-run phenomena. Those shorter-run phenomena are the subject of the next chapter of my thesis.

Appendices of Chapter 1

Appendix 1.A Empirical appendix

Figure 1.6: 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)).

Appendix 1.B Calibration appendix

1.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 1.8: 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.

1.B.2 Classification of food items

Table 1.9: 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)".

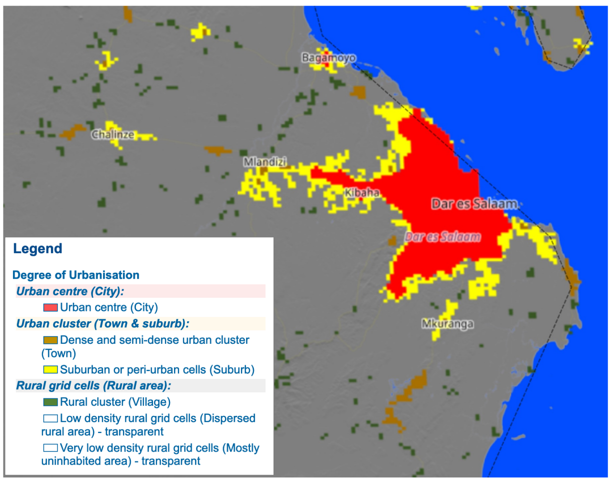
1.B.3 Mapping locations to model regions

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.[18]

Table 1.10: 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 1.7: Global Human Settlement Layer (GHSL) Degree of Urbanisation



[18]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.

1.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.[19] 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.[20] 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 (fer-

[19]Where GPS-measured plot size is missing, reported plot size is used.
[20]This measure is discretised into 25%, 50%, 75% or 100% of a plot being planted with a given crop.

tiliser, 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

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

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 1.1.

1.B.5 Model parameterisation

Table 1.11: 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

CHAPTER 2

Global price shocks and welfare in a spatial model of low-income countries*

with **Christopher Adam** and **Douglas Gollin**

Abstract We develop a spatial dynamic general equilibrium model of a small open agricultural economy, calibrated to Tanzanian data, to study the transmission of global food, fuel and fertiliser price shocks to the consumption patterns of heterogeneous households located in different regions. The model allows us to consider fiscal policy measures, such as direct price subsidies and net household transfers, which can be implemented to mitigate the shocks' impact on household consumption, domestic production, internal trade and labour movements. Our model simulations show strong spatial heterogeneity in the effects of import price shocks. While urban households' consumption baskets are more exposed to global food price shocks, remote rural households' production and consumption are heavily dependent on fuel and fertiliser imports.

*We are grateful for support for this work from the IMF-FCDO Macroeconomic Program on Low-Income Countries. We thank seminar participants at the IMF SPR Climate and Development seminar for helpful comments. All remaining mistakes herein are ours.

2.1 Introduction

Recent years have seen increased volatility in world prices for food, energy and fertiliser. Driven by the decompression of the post-pandemic global economy, and sharply accelerated by Russia's invasion of Ukraine in February 2022, real energy and fertiliser prices almost doubled while the composite tradable food price index rose by around 30 percent between 2020 and 2022 (Figure 2.1). Low-income countries (LICs) tend to be particularly strongly affected by such shocks, due to their dependence on imports, widespread poverty, and limited scope for offsetting fiscal policies. As a consequence, food, fuel and fertiliser price shocks have potentially large welfare implications for LICs, particularly with a view to food security.

When assessing vulnerabilities to import price shocks and considering the distributional consequences of fiscal policy responses, within-country spatial heterogeneity is an important dimension. In LICs in particular, weak infrastructure and thin markets generate considerable domestic spatial frictions, with connectedness to world markets varying significantly across regions.[2] These spatial frictions result in strong spatial differences in the level and composition of production, consumption, and trade which imply differential vulnerability to global price shocks. Policy makers must therefore take into account the spatial heterogeneity both in the impact of shocks and in the distributional consequences of policies aimed at shielding vulnerable populations.

In this paper, we develop a calibrated spatial multi-sector general equilibrium model to study the impact of global food, fuel and fertiliser price shocks on a small open agricultural economy and to consider alternative fiscal policy responses to these shocks. We follow Adam et al. (2018) in calibrating the model to the economy of Tanzania. Our version draws on a newer (2018) version of the Social Accounting Matrix (SAM) for Tanzania. We microfound modelling assumptions around the spatial distribution of production and consumption using household survey data from the World Bank's Living Standards Measurement Study (LSMS).[3] Given this structure, we can examine the heterogeneous pass-through of global price shocks

[2]Domestic trade costs are particularly high in sub-Saharan Africa. For example, Atkin and Donaldson (2015) estimate that the effect of log distance on trade costs within Ethiopia or Nigeria is four to five times larger than in the US.

[3]We use data from waves 1-4 (2008-2014) and complement this with the extended panel from 2019.

from a positive perspective. However, the model also lends itself to a granular normative welfare analysis, across locations and household types. In this spirit, we use the calibrated model to examine a set of alternative policy responses to shocks. Specifically, we consider the short-term distributional consequences of different fiscal mitigation measures, such as import price subsidies, net household transfers, and tariff or tax adjustments.[4]

We find that, consistent with their higher net food consumption, urban wage-earning households are more adversely affected than agricultural households by global food price shocks. In the case of a simultaneous shock to food, fuel and fertiliser prices, however, remote rural households are equally negatively affected, because of the high transport-cost intensity of their consumption and because of the effects on their ability to purchase imported fertiliser. In the absence of a mitigating policy response, a joint food, fuel and fertiliser price shock of +20% in global markets reduces households' aggregate real income in the domestic economy by 7.3%. The relative decline in real factor incomes is strongest for households in remote rural regions and for the 'capitalist' household in the urban region, whose income from transport rents declines as households' consumption becomes more local and domestic trade flows decrease.

In counterfactual policy experiments, we contrast an increase in import price subsidies with an increase in direct net household transfers, financed either externally through grant aid or domestically through an increase in the capital-owning household's income tax. We find that increasing import price subsidies to fully offset the price shock can shield households' real incomes. In the case where the additional expenditure on subsidies is financed domestically through a higher tax on the capitalist's income, average real income per capita of all other households falls by only 0.3%, compared to 8.6% in the case of an unmitigated shock. However, we find that aid-financed import price subsidies have an urban bias, as the policy

[4]The time horizon we consider extends beyond immediate short-term equilibrium impacts (where adjustment occurs only on the intensive margin) to include more medium-term impacts such as inter-sectoral labour movements and the implementation of fiscal policy. We still label this time horizon 'short-term' to reflect that we limit medium- to long-term factor adjustment (i.e. capital reallocation and migration), equivalent to a 'Ricardo-Viner' fixed-factor structure. Consequently, we do not speak to questions around rural-urban migration or structural change (e.g. productivity growth or fundamental changes on the supply side of the economy, such as the possibility to promote domestic fertiliser production).

disproportionately benefits the urban population. For rural households, real income can be protected more effectively by distributing transfer payments directly to households on a per capita basis.[5]

Previous work on the impact of price shocks focuses largely on aggregate country-level analyses, assessing vulnerability to shocks based on national trade statistics, average consumption baskets, and aggregate production data (Arndt et al., 2023; Bekkers et al., 2017; Meyimdjui & Combes, 2021; Okou et al., 2022). In recent empirical work, Okou et al. (2022) explore the determinants of domestic retail staple food prices in sub-Saharan Africa, highlighting the importance of import dependence for the extent of price pass-through.[6] In an analysis of 19 different countries, Arndt et al. (2023) use national economy-wide models to measure the near-term impacts on poverty and food insecurity of the food and fertiliser price shocks induced by the war in Ukraine. Such cross-country analyses are important in rapidly assessing which countries are likely to face the largest challenges as a result of global price shocks. At the same time, previous analyses of within-country heterogeneity are typically limited to income differences between households. However, questions around within-country spatial heterogeneity, which are largely missing from existing literature, are key for policy considerations in LICs.[7]

Our work also connects to the growing literature that focuses on the importance

[5]It should be noted, however, that these results are derived from comparative statics for a one-off shock without adjustment costs. To construct more realistic policy scenarios, we are currently working on a dynamic model to consider the case where price shocks are transitory, private agents face adjustment costs, and the government's fiscal measures include access to domestic debt financing.

[6]Okou et al. (2022) estimate the drivers of prices for five staple crops (maize, rice, cassava, wheat and palm oil) using a panel of monthly data across 68 market locations in 15 African countries for the period 2012-2021. Amongst other results they find "...pass-through from global to local staple prices is estimated to be a near one-to-one (0.97) in countries which import at least three quarters of their consumption…" (p. 23), although the estimate is significantly lower at the median (0.19).

[7]While some studies implicitly cover the spatial dimension by mentioning broad urban-rural differences, mostly in relation to differences in income, to the best of our knowledge, thorough analysis of spatial heterogeneity with respect to the impact of shocks and the consequences of policies is missing from existing literature. Notably, Arndt et al. (2008) analyse the effects of fuel and food price shocks, and possible policy responses, by calibrating a CGE model to the Mozambican Social Accounting Matrix. While maintaining the SAM's rural-urban distinction for households, their model does not, however, feature spatial heterogeneity in production or explicit spatial frictions. Minot and Dewina (2015) highlight the importance of considering second-order effects of food price shocks, but they do not conduct a spatial analysis of these shocks. Cudjoe et al. (2010) conduct an empirical analysis of spatially differentiated price transmission in Ghana, but they do not develop a theoretical framework that would allow to conduct policy experiments.

of spatial frictions within countries, both with a particular focus on transport costs (Atkin & Donaldson, 2015; Donaldson, 2018; Gollin & Rogerson, 2014; Sotelo, 2020) and market integration more broadly (Allen, 2014; Burke et al., 2019; Moser et al., 2009; Van Campenhout, 2007). To the best of our knowledge, there exists little research that examines the role of market integration for shock transmission and food insecurity.[8]

We contribute to the existing literature by presenting a quantitative assessment of the impact of food and energy price shocks, both individually and jointly, across locations and households. Our model allows us to identify key pass-through mechanisms from external shocks to domestic production, consumption and household incomes.[9] Additional extensions, which are work in progress, shed closer light on alternative fiscal mitigation measures and financing instruments, contrasting targeted subsidies on food, fuel and fertiliser with cash transfers to households, and comparing effects under external aid financing and domestic tax financing. As such, our framework constitutes a novel modelling approach that integrates a multi-sector spatial structure with macroeconomic models widely used in the policy arena, such as the workhorse DIG models in use at the International Monetary Fund.

The modelling framework presented in this paper lets us distinguish between shocks that superficially look the same but which can have quite different impacts and therefore may require quite different policy measures to mitigate the effect of the shock. Food price shocks that result from a global price shock, for instance, may look similar to food price shocks that are caused by a domestic supply shock (e.g. climate-related productivity shocks as in Baptista et al. (2023)). In addition, the model we develop allows us to consider aggregate and distributional welfare impacts of alternative fiscal and financing responses.

The remainder of this paper is organised as follows: Section 2.2 discusses potential mechanisms through which global price shocks affect consumer welfare in

[8]Exceptions are Burgess and Donaldson (2012), Gao and Lei (2021), Janssens et al. (2020), and Simola et al. (2022) though they focus primarily on the transmission of supply shocks in local agricultural production rather than the spatial transmission of import price shocks.

[9]While our focus in this paper is on external price shocks, this framework can be used to examine other sources of price and productivity volatility. Notably, recent related work by Baptista et al. (2022) examines the spatially differentiated effects of climate-induced shocks to agricultural productivity.

LICs and provides an overview of possible policy instruments. Section 2.3 describes the empirical setting of this study, briefly summarizing Tanzania's exposure to global price shocks and spatial patterns in the domestic economy. Section 2.4 presents our modelling framework, with full details contained in Appendix 2.B. Section 2.5 describes the calibration procedures and the baseline equilibrium. Section 2.6 presents results of the different policy experiments in comparative statics. Section 2.7 concludes.

2.2 Background

2.2.1 Global price shocks and their local effects in LICs

Global commodity price shocks are a recurring phenomenon that can impose severe terms of trade effects on importing countries. Figure 2.1 plots price indices for food, fuel, and fertilisers over time. The left panel shows price developments since 2006, including the three 'global food price crises' that have occurred this century; that is, the price spikes in 2007, 2011, and 2022. The right panel zooms in on price developments since 2019, showing the pronounced simultaneous increase in the global prices of food, fuel and fertiliser observed in 2021 and 2022. This increase was first driven by recovering aggregate demand as economies reopened after the COVID-19 shock. In early 2022, price increases accelerated drastically due to the effects of the war in Ukraine. The war triggered a strong negative global supply shock, as both Ukraine and Russia were important exporters of agricultural products, energy commodities, and raw materials used in fertiliser production.

In principle, rising global prices for food, fuel, and fertilisers exert pressure on the balance of payments of all net importers of these commodities. When passed through to domestic markets, commodity price spikes can have considerable welfare implications for consumers. However, the effects are typically felt most strongly in LICs as many of them have limited domestic production of key commodities and thus high import dependence, low capacity to implement offsetting macroeconomic policies, and a high degree of vulnerability due to widespread poverty. As a result, simultaneous price shocks in food, fuel, and fertilisers can pose a threat to food security in LICs through both direct and indirect channels.

Figure 2.1: Global prices of food, energy and fertiliser

Source: World Bank Commodity Price Data (retrieved February 17, 2023). All series are scaled by US CPI deflator (FRED, all consumers). Series in right panel are in logs with scale representing percentage change relative to January 2019.

First, an increase in the world market price of imported food has a direct effect on households' food expenditure share, all else equal. The magnitude of this price effect depends on households' import dependence, income level, and net food production. Households whose food consumption basket contains a relatively high share of imported foods – which is typically true for many urban households – are affected more strongly. Unlike other consumption goods, food consumption is highly inelastic due to binding subsistence requirements. Poorer households, for whom food expenditure necessarily constitutes a large share of overall expenditure, are hit hardest as there is limited room for offsetting demand adjustment. Net food-producing households in rural regions may be positively affected by higher global food prices if these (partially) pass through to markets for domestically produced food, and if the revenue increase from sales offsets the cost increase for (processed) foods that these households purchase in the market for own consumption.

Second, higher fuel prices drive up the cost of transporting and distributing food, both internationally and domestically.[10] In many LICs, which typically face high domestic trade frictions, transport costs constitute a significant share of final consumption prices for imported foods in remote rural regions.[11] In addition, higher transport costs put negative pressure on export earnings of cash crop producers in these rural regions, further lowering their ability to purchase food.

Third, fertiliser prices have a lagged effect on the price of food, as increased input prices are passed through to final output prices after harvest. When farmers have limited access to finance, such that they reduce fertiliser use in response to higher price, fertiliser price shocks may result not only in lower access to food (due to higher prices), but also in lower availability (due to weaker harvests).

2.2.2 Policy measures in response to price shocks

Governments can implement a range of policy measures to protect vulnerable populations from the adverse effects of global commodity price shocks. We can distinguish between short-term 'reactive' measures aimed at mitigating the imminent effects of realised shocks, and longer-term 'precautionary' measures aimed at reducing vulnerability to future shocks. Such longer-term measures consist mostly of structural reforms in agriculture, infrastructure and trade or financial sector reforms that strengthen e.g. insurance markets.[12] The following overview, which is informed by a recent IMF publication (Amaglobeli et al., 2022), is limited to more short-term policy measures.

Targeted policies, i.e. measures aimed at directly supporting vulnerable households while letting prices largely pass through to the domestic economy, are typically the *first best* course of action. Targeted direct and indirect transfers, including cash transfers, in-kind support, food vouchers, or utility bill cuts, constitute the most efficient policy response. However, such direct assistance can be difficult to implement if social safety nets are imperfectly targeted and institutional capacity

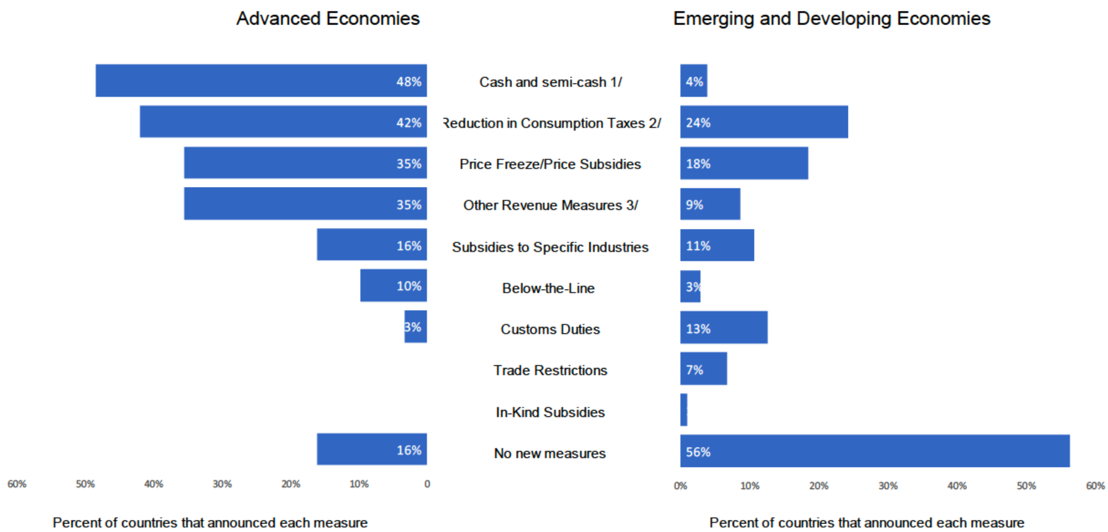
[10] Where agriculture is mechanised, higher fuel prices also influence the cost of food production. Mechanisation in agriculture is typically lower in low-income countries, but this effect on food production elsewhere can have considerable second-round effects on the price of imported food.

[11] In a study on the role of domestic trade costs in LICs, Atkin and Donaldson (2015) find that that "the effect of log distance on trade costs within Ethiopia or Nigeria is four to five times larger than in the US".

[12] For a comprehensive discussion of such reforms in the context of climate change and chronic food insecurity in sub-Saharan Africa see Baptista et al. (2022).

is low. In addition, targeted policies that are limited only to certain groups are politically less favorable than untargeted policies that benefit a larger share of the population. Untargeted policies generally aim to limit the pass-through of global price spikes to domestic markets. These policies include tax measures such as cuts in income taxes or import tariffs, and spending measures such as price subsidies. Untargeted policies constitute a *second best* approach, since limiting the price pass-through distorts price signals and corresponding supply and demand adjustments and reduces incentives for market responses from distributors and producers. In addition, untargeted policies are expensive and may potentially crowd out other productive public spending.

Figure 2.2: Policy measures announced in response to price spikes since 2022



Note: "No new measures" refers to countries that responded to the survey and have not announced any measures. 1) includes cash transfers and semi-cash, such as vouchers and utility bill discounts. 2) includes value-added and excise taxes. 3) includes changes to income taxes and other revenue measures. Source: Amaglobeli et al. (2022). IMF staff estimate based on an IMF survey of 134 countries (31 advanced economies and 103 emerging and developing economies) conducted in March 2022 on the measures taken by governments since 2022 in response to rising food and commodity prices.

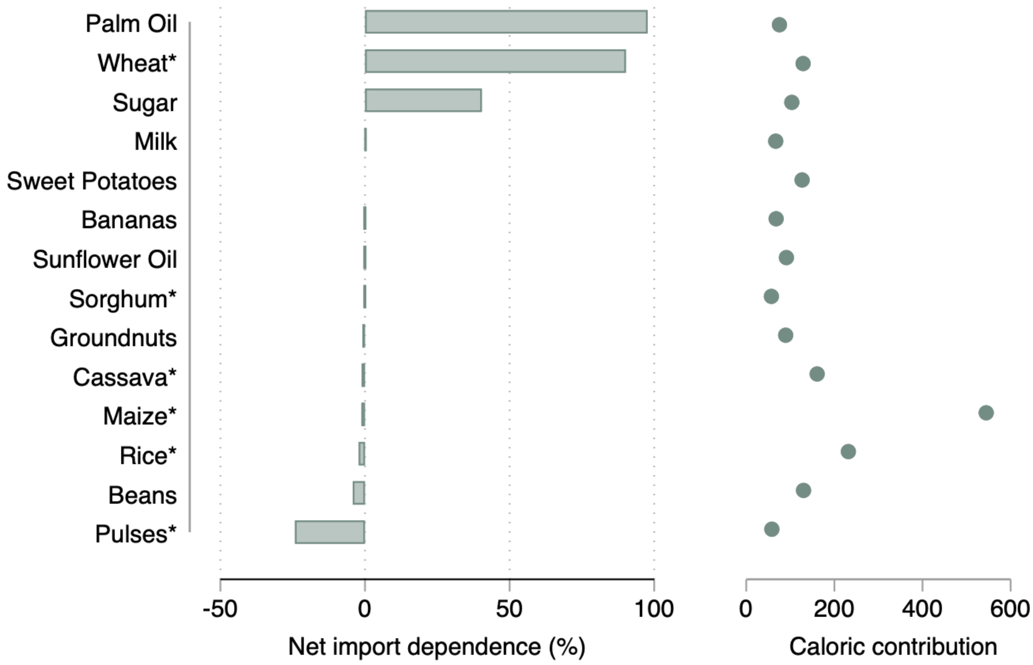
In response to the recent price spikes in 2022, many advanced economies as well as emerging market and LICs put in place various fiscal and quasi-fiscal measures to help insulate households and firms from the full effect of these price shocks. Figure 2.2 provides an overview of the prevalence of different measures in advanced economies and emerging and developing economies. The majority of advanced economies announced a host of different measures, the most widely used being cash and semi-cash transfers, with half of the countries (48%) putting these in place. A large share of advanced economies also introduced untargeted tax and spending measures, with 42% of countries reducing consumption taxes, and 35% providing price freezes and subsidies. At 56%, the share of countries that announced no new measures in response to the recent price spikes was much higher among the group of emerging and developing economies as they tend to have less fiscal scope. While targeted measures would constitute the most efficient use of these countries' limited funds, cash and semi-cash transfers typically require existing social safety nets as an effective means for distribution, which few low-income countries have. Instead, emerging and developing countries mostly resorted to cutting taxes, introducing price freezes and subsidies, and targeting trade (through customs duties and trade restrictions).

2.3 Empirical setting

The empirical part of this paper is focused on the United Republic of Tanzania. The setting in Tanzania is characterised by a large and relatively unproductive agricultural sector, high domestic trade frictions and low market integration, and strong spatial heterogeneity in production and consumption (Adam et al., 2018). While Tanzania has a balanced net food production position overall (Baptista et al., 2022, p. 4, Figure 2), import dependence for some (mostly processed foods) is very high (see Figure 2.3). In addition, the country is almost entirely dependent on imports for fuel and fertiliser. These features make Tanzania somewhat typical of low-income sub-Saharan economies. Our choice of Tanzania reflects its 'typical' characteristics – but also relates to the availability of data. Compared to other countries in the region, Tanzania has high-quality household survey data, along with a detailed social accounting matrix and input-output tables. These domestic data sources are supplemented by data from international sources such as UN Comtrade and FAOSTAT.

Import dependence and exposure to world markets. Tanzania is fully dependent on imports of fuel (petroleum) and synthetic fertilisers. As such, fuel and fertiliser are non-competitive imports and Tanzania is a price taker in international

Figure 2.3: Food – Net import dependence and caloric contribution

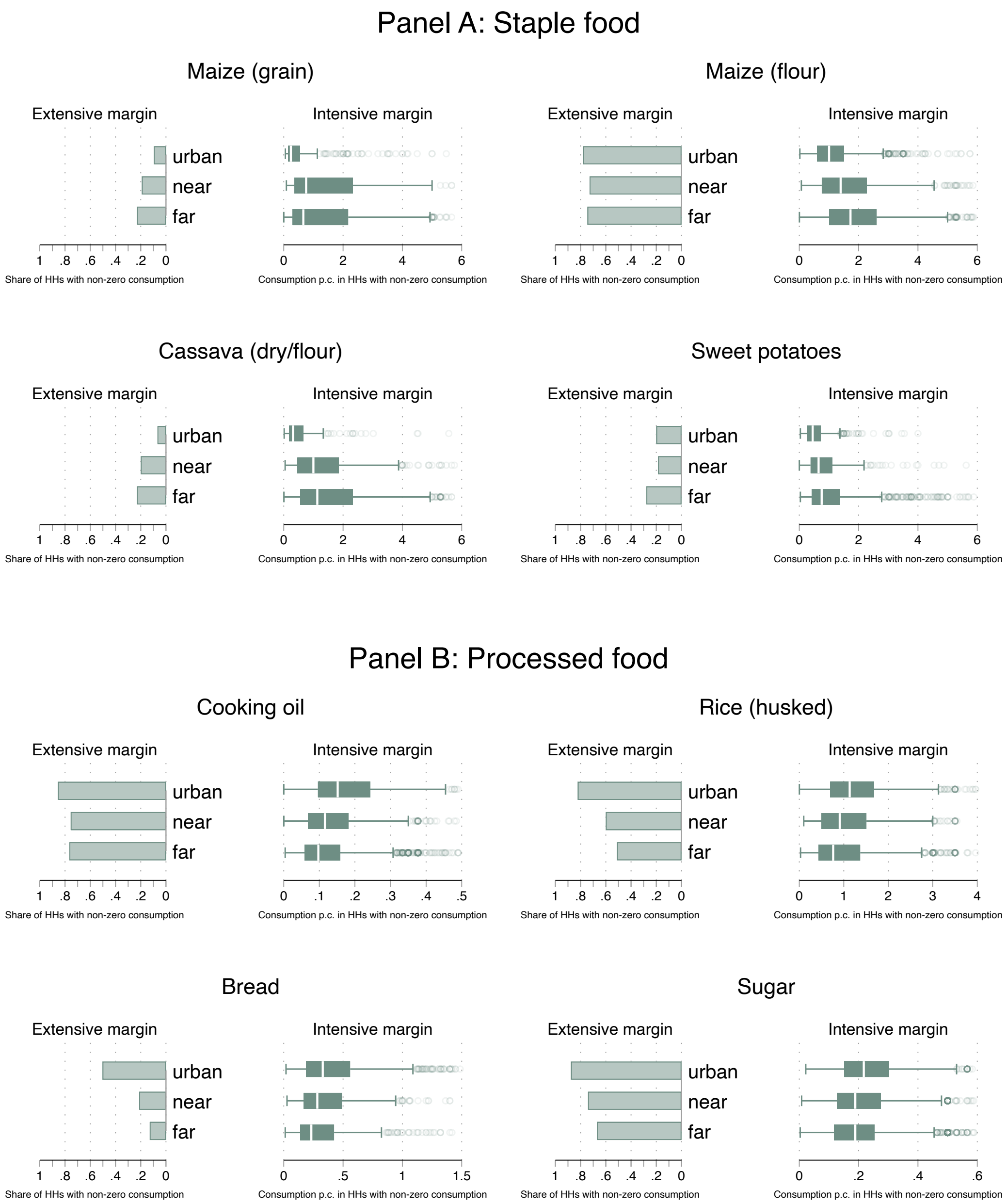


Note: * indicates that the category includes products of this food, e.g. 'Wheat*' refers to 'Wheat and products'. Net import dependence is calculated as 100*(imports-exports)/(production+imports-exports). Caloric contribution refers to FAOSTAT's 'Food supply (kcal/capita/day)' variable. Net import dependence and caloric contribution are mean values for the years 2010-2020. This figure only shows foods with a caloric contribution of at least 50 kcal/capita/day. Source: FAOSTAT Food Balance (Feb 20, 2023)

markets. For food, the situation in Tanzania is more complex: net import dependence is high for some types of food (mostly processed foods), while for other types of food Tanzania is self-sufficient or even a net exporter (mostly staple foods).

Spatial heterogeneity. There are strong spatial patterns in poverty, economic activity and consumption. This spatial heterogeneity matters critically for households' vulnerability to global price shocks. As Figure 2.4 shows, food consumption patterns vary significantly across space. Households in rural regions consume more domestically produced traditional staples like maize, cassava, and sweet potatoes. Urban households consume higher quantities of processed foods, such as cooking oil and sugar. These different consumption patterns imply a differential exposure to global price shocks, as a higher consumption share of processed foods in urban regions implies a higher dependence on imports.

Figure 2.4: Consumption of staple and processed food items by region



Note: This figure shows household consumption of selected staple food items (i.e. unprocessed and minimally processed foods) and processed food items (i.e. industrially processed food items) at the extensive and the intensive margin. The extensive margin is defined as the share of households with non-zero consumption of a given food item in the 7 days before the time of data collection. The intensive margin is defined as the quantity consumed per capita in kilograms (or litres) in the 7 days before the time of data collection. Per capita consumption is approximated as total household consumption adjusted for household size as measured in adult equivalents. Households are categorised as urban, near rural or far rural based on their geo-referenced location and the GHSL degrees of urbanisation. Source: LSMS, National Panel Survey (waves 1-4), food consumption questionnaires.

Policy response to recent shocks. In response to the recent global price spikes, Tanzania has implemented different measures to insulate the domestic economy from the effect of these shocks. These policy measures include "(i) providing targeted and temporary subsidies for the import of petroleum products via the waving of excise duties, (ii) reducing import duties on cooking oil by lowering important duty rates for refined edible oil and crude cooking oil, (iii) introducing fertiliser subsidies to support producers of agricultural products and fertilisers to boost local production and substitute part of the imports, and (iv) waiving of VAT for locally produced fertilisers and reduction of royalty charges on minerals used in the energy and fertiliser industries" (Rother et al., 2022, p. 30).

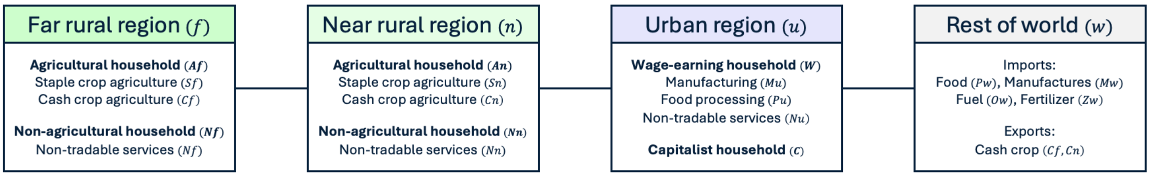
2.4 Model structure

We develop a multi-sector, multi-household, spatial general equilibrium model of a small open agricultural economy. The model defines four stylised locations: a remote ('far') rural economy involved in staple and cash-crop agriculture; an adjacent ('near') rural region, which trades with both the far rural region and the principal urban centre; an urban centre, which trades with the near rural region and with the rest of the world; and finally, the rest of the world, which constitutes the fourth region (see Figure 2.5). Internal trade (and hence market structure) is a function of transport linkages between these locations, which in turn, reflect recurrent transport costs and the degree of monopoly power.

The model captures the principal pass-through channels from world food, fuel and fertiliser price shocks to: (i) border prices for export cash crops and (ii) non-traded staple food crops (via fertiliser and domestic transport costs); and (iii) to final consumption of tradable processed food (through input costs). It is also able to track pass-through effects to the non-food economy, both through transport costs and, indirectly, through real exchange rate effects on the structure of domestic production and consumption.

There are nine spatially-differentiated sectors of production. The two rural regions produce a staple food crop and an export cash crop, along with a local non-tradable services. In the urban region, there is a manufacturing, food processing, and a local non-tradable services sector.

Figure 2.5: Schematic overview of the model geography



The model features six representative household types which differ by location and economic activities (Table 2.1). All households are hand-to-mouth, except for the capitalist household in the urban region, which saves and invests in fixed capital in the urban sector and in government bonds. All households except for the capitalist household receive direct net transfers from the government. Income is taxed at rates specific to household types. Households consume food, manufactures, and non-tradable services. There are subsistence requirements for food, which create non-homotheticities in consumption.

The economy is open to trade with the rest of the world, subject to import tariffs and export duties. The economy imports foreign varieties of both processed food and manufactures. In addition, it sources fertiliser and fuel, two non-competitive imports, from the rest of the world. The cash crop production from the two rural regions is exported in full, selling at the exogenous global price of the crop. All inter-regional domestic trade is subject to transport costs, which reflect fuel requirements and monopoly rents.

On the fiscal side, the model incorporates a range of tariff, tax and subsidy instruments, recurrent and government spending (all of which may be differentiated across sectors) and domestic and external sovereign borrowing, both concessional and non-concessional, allowing us to explore a range of public policy responses to price shocks under alternative financing arrangements. Given this structure, and in addition to examining heterogeneous pass-through from a positive perspective, the model lends itself to a granular normative welfare analysis, across locations and household types, of both price shocks themselves and of alternative policy responses.

2.4.1 Production

Agriculture

Agricultural production in the rural regions (far and near) is modeled using a nested CES production function, with labour (L) and fertiliser (Z) on the lower tier, and the resulting composite combined with land (A) on the upper tier[13]

$$Q_i = a_i \left[\lambda_i^{\frac{1}{\gamma_i}} A_i^{\frac{\gamma_i - 1}{\gamma_i}} + (1-\lambda_i)^{\frac{1}{\gamma_i}} \left(\delta_i^{\frac{1}{\theta_i}} L_i^{\frac{\theta_i - 1}{\theta_i}} + (1-\delta_i)^{\frac{1}{\theta_i}} Z_i^{\frac{\theta_i - 1}{\theta_i}} \right)^{\frac{\theta_i(\gamma_i - 1)}{(\theta_i - 1)\gamma_i}} \right]^{\frac{\gamma_i}{\gamma_i - 1}} \tag{2.1}$$

where $i = \{Sf, Sn, Cf, Cn\}$ denotes the type of crop (staple, cash) and production location (far, near). The parameters γ_i and θ_i are the usual elasticities; λ_i and δ_i are share parameters in the upper and lower level CES functions, and a_i is a crop- and region-specific productivity parameter. We assume that land allocations are fixed in the short run. The representative farmer in location l thus maximises profits by choosing quantities of labour and fertiliser:

$$\max_{L_i, Z_i} p_i^l \, Q_i - b_i \, A_i - w_l \, L_i - p_Z^l \, Z_i,$$

where $i = \{Sf, Sn, Cf, Cn\}$ denotes the type of crop (staple, cash) and $l = \{f, n\}$ denotes the production location (far, near). The rental cost of land is denoted by b_i, while w_i is the wage paid to labour, and p_Z^l is the price of fertiliser in location l.

Manufacturing

Production in the manufacturing sector in the urban region (Mu) uses labour (L) and capital (K) to produce output:

$$Q_{Mu} = a_{Mu} \, L_{Mu}^{\alpha_{Mu}} \, K_{Mu}^{(1-\alpha_{Mu})}, \tag{2.2}$$

where α_{Mu} is the output elasticity of labour and a_{Mu} is a sector-specific productivity parameter. The representative firm maximises profits by choosing the quantities of labour (L) and capital (K):

$$\max_{L_{Mu}, K_{Mu}} p_{Mu}^u \, Q_{Mu} - w_u \, L_{Mu} - r_u \, K_{Mu},$$

[13]We use this production structure to allow for limited (zero in the limit) substitutability of labour for fertiliser.

where w_u is the wage paid to labour in the urban region and r_u is the rental rate of capital in the urban region.

Food processing

Production in the food processing sector in the urban region (Pu) is modeled using a Leontief specification with composite staple crop ($\bar{S}$) and composite manufactures ($\bar{M}$) as intermediate inputs, and labour (L) and capital (K) as factors of production.

In turn, the composite staple food crop used in food processing ($\bar{S}_P$) is a CES aggregate of the staple crop produced in the far rural region (Sf) and the staple crop produced in the near rural region (Sn):

$$\bar{S}_P = \left[\iota_S^{1/\rho_S}\, Sf_P^{(\rho_S-1)/\rho_S} + (1-\iota_S)^{1/\rho_S}\, Sn_P^{(\rho_S-1)/\rho_S}\right]^{\rho_S/(\rho_S-1)} \tag{2.3}$$

Similarly, the composite manufactured good used in food processing ($\bar{M}_P$) is a CES aggregate of manufactures produced domestically in the urban region (Mu) and manufactures imported from the rest of the world (Mw):

$$\bar{M}_P = \left[\iota_M^{1/\rho_M}\, Mu_P^{(\rho_M-1)/\rho_M} + (1-\iota_M)^{1/\rho_M}\, Mw_P^{(\rho_M-1)/\rho_M}\right]^{\rho_M/(\rho_M-1)} \tag{2.4}$$

Total intermediate input use in the urban food processing sector (Pu) is thus:

$$IN_P = \phi_{\bar{S}P}\, Q_{Pu} + \phi_{\bar{M}P}\, Q_{Pu}, \tag{2.5}$$

with $\phi_{\bar{S}P}$ and $\phi_{\bar{M}P}$ denoting the input requirement of (composite) staple crop and (composite) manufactures per unit of output of processed food (Pu).[14]

Gross output of the food processing sector (Q_{Pu}) is produced by combining factors (value added) and intermediates. The value of output is given by:

$$p_{Pu}^u\, Q_{Pu} = p_{Pu}^{va}\, va_{Pu} + \left(p_{\bar{S}}^u\, \phi_{\bar{S}P} + p_{\bar{M}}^u\, \phi_{\bar{M}P}\right) Q_{Pu}, \tag{2.6}$$

where p_{Pu}^u is the factory-gate output price for processed food in the urban region and $p_{\bar{S}}^u$ and $p_{\bar{M}}^u$ are the market prices for composite staple crop and composite manufactures in the urban region. Real value added is defined as:

$$va_{Pu} = a_{Pu}\, K_{Pu}^{\alpha_{Pu}}\, L_{Pu}^{(1-\alpha_{Pu})} \tag{2.7}$$

[14]The input-output coefficients, ϕ_{iP}, are calibrated directly as $\phi_{iP} = n_{iP}^o/Q_P^o$, where the o superscript denotes initialisation values.

,

where α_{Pu} is the output elasticity of capital and a_{Pu} is a sector-specific productivity parameter. Note that the value-added price is defined implicitly from

$$\left(p^u_{Pu} - p^u_{\bar{S}}\,\phi_{\bar{S}P} - p^u_{\bar{M}}\,\phi_{\bar{M}P}\right) Q_{Pu} = p^{va}_{Pu}\,va_{Pu} \tag{2.8}$$

where $p^u_{\bar{S}}$ ($p^u_{\bar{M}}$) is the tariff, tax and transport cost inclusive price of composite staple crop (manufactures) in the urban region. The representative firm maximises profits by choosing the quantities of labour (L) and capital (K)

$$\max_{L_{Pu},K_{Pu}} p^u_{Pu}\,Q_{Pu} - w_u\,L_{Pu} - r_u\,K_{Pu} - \left(p^u_{\bar{S}}\,\phi_{\bar{S}P} + p^u_{\bar{M}}\,\phi_{\bar{M}P}\right) Q_{Pu}$$

Non-tradable services

In each of the three regions there exists a sector that produces local non-tradable services (N). The production of these non-tradable services is linear in labour, the single factor of production

$$Q_{Nl} = a_{Nl}\,L_{Nl} \tag{2.9}$$

where $l = \{f, n, u\}$ denotes the region (far rural, near rural and urban) and a_{Nl} is a sector- and region-specific productivity parameter. The representative firm in region l maximises profits by choosing the quantity of labour (L)

$$\max_{L_{Nl}} p^l_{Nl}\,Q_{Nl} - w_l\,L_{Nl}$$

2.4.2 Household consumption

There are six representative households in the economy, two in each of the three regions: an agricultural and a non-agricultural household in the far rural region, an agricultural and a non-agricultural household in the near rural region, and a wage-earning and a capitalist household in the urban region. Table 2.1 gives an overview of households' factor endowments and corresponding factor income. All households except for the capitalist supply labour. Agricultural land in a given region is owned by the local agricultural household. The capitalist household earns income from returns on capital and from transport rents.

All households consume food, a local non-tradable, and the manufactured good. Household consumption, c^h_{lt} is defined over seven commodities

Table 2.1: Household types

Location	Type	Factor	Sector	Factor income	
Far (f)	Agricultural (A)	Labour, land	Staple crop	w_{Sf}	wage
				b_{Sf}	land rents
			Cash crop	w_{Cf}	wage
				b_{Cf}	land rents
	Non-agricultural (N)	Labour	Non-tradables	w_{Nf}	wage
Near (n)	Agricultural (A)	Labour, land	Staple crop	w_{Sn}	wage
				b_{Sn}	land rents
			Cash crop	w_{Cn}	wage
				b_{Cn}	land rents
	Non-agricultural (N)	Labour	Non-tradables	w_{Nn}	wage
Urban (u)	Wage-earning (W)	Labour	Food processing	w_u	wage
			Manufacturing	w_u	wage
			Non-tradables	w_u	wage
	Capitalist (C)	Capital	Food processing	r_u	capital rents
			Manufacturing	r_u	capital rents
			All sectors	ψ_j	transport rents

$\{Sf, Sn, Pu, Pw, Mu, Mw, Nl\}$. All household types $h = \{Af, Nf, An, Nn, W, C\}$ across the three regions $l = \{f, n, u\}$ in the economy have the same preferences across goods. Consumption is modeled as a three-tier nested CES structure to capture all key margins of substitution, both between different goods and between different varieties of the same good (see Figure 2.6 for a schematic overview). The top level of the CES consumption structure combines composite food ($c_{\bar{F}}$), composite manufactures ($c_{\bar{M}}$), and region-specific non-tradable services (c_{Nl}) into aggregate household consumption, c^h

$$c^h = \left[\eta_{\bar{F}}^{1/\epsilon}\, c_{\bar{F}}^{h\;(\epsilon-1)/\epsilon} + \eta_{\bar{M}}^{1/\epsilon}\, c_{\bar{M}}^{h\;(\epsilon-1)/\epsilon} + (1-\eta_{\bar{F}}-\eta_{\bar{M}})^{1/\epsilon}\, c_{Nl}^{h\;(\epsilon-1)/\epsilon}\right]^{\epsilon/(\epsilon-1)} \tag{2.10}$$

where ϵ is the elasticity parameter and $\eta_{\bar{F}}$ and $\eta_{\bar{M}}$ are share parameters. On the middle tier, composite food ($c_{\bar{F}}$) consumption is constructed as a combination of the consumption of composite staple food ($c_{\bar{S}}$) and composite processed food ($c_{\bar{P}}$), to capture the way in which households substitute between these different food categories.

$$c_{\bar{F}}^h = \left[\eta_{\bar{S}}^{1/\xi}\,(c_{\bar{S}}^h - \bar{c}_{\bar{S}})^{(\xi-1)/\xi} + (1-\eta_{\bar{S}})^{1/\xi}\,(c_{\bar{P}}^h - \bar{c}_{\bar{P}})^{(\xi-1)/\xi}\right]^{\xi/(\xi-1)} \tag{2.11}$$

where ξ is the elasticity parameter, $\eta_{\bar{S}}$ is the share parameter, and $\bar{c}_{\bar{S}}$ and $\bar{c}_{\bar{P}}$ are subsistence parameters that generate non-homotheticities in consumption. Below that, we add another layer to reflect the substitution between staple food produced in far rural or near rural on the one hand, and imported and domestically processed food on the other hand. On this bottom tier of the nested CES structure, composite staple food ($c_{\bar{S}}$) is a combination of staple crop produced in the far rural region (c_{Sf}) and staple crop produced in the near rural region (c_{Sn})

$$c^h_{\bar{S}} = \left[\nu_S^{1/\zeta_S}\, {c^h_{Sf}}^{(\zeta_S-1)/\zeta_S} + (1-\nu_S)^{1/\zeta_S}\, {c^h_{Sn}}^{(\zeta_S-1)/\zeta_S}\right]^{\zeta_S/(\zeta_S-1)} \tag{2.12}$$

where ζ_S is the elasticity parameter and ν_S is the share parameter. Composite processed food ($c_{\bar{P}}$) is a combination of processed food produced domestically in the urban region (c_{Pu}) and imported processed food (c_{Pw})

$$c^h_{\bar{P}} = \left[\nu_P^{1/\zeta_P}\, {c^h_{Pu}}^{(\zeta_P-1)/\zeta_P} + (1-\nu_P)^{1/\zeta_P}\, {c^h_{Pw}}^{(\zeta_P-1)/\zeta_P}\right]^{\zeta_P/(\zeta_P-1)} \tag{2.13}$$

where ζ_P is the elasticity parameter and ν_P is the share parameter. We use a similar nested structure for the manufactured good as well to reflect how households substitute between domestic and imported manufactures. Composite manufactures ($c_{\bar{M}}$) are a combination of manufactures produced domestically in the urban region (c_{Mu}) and imported manufactures (c_{Mw})

$$c^h_{\bar{M}} = \left[\nu_M^{1/\zeta_M}\, {c^h_{Mu}}^{(\zeta_M-1)/\zeta_M} + (1-\nu_M)^{1/\zeta_M}\, {c^h_{Mw}}^{(\zeta_M-1)/\zeta_M}\right]^{\zeta_M/(\zeta_M-1)} \tag{2.14}$$

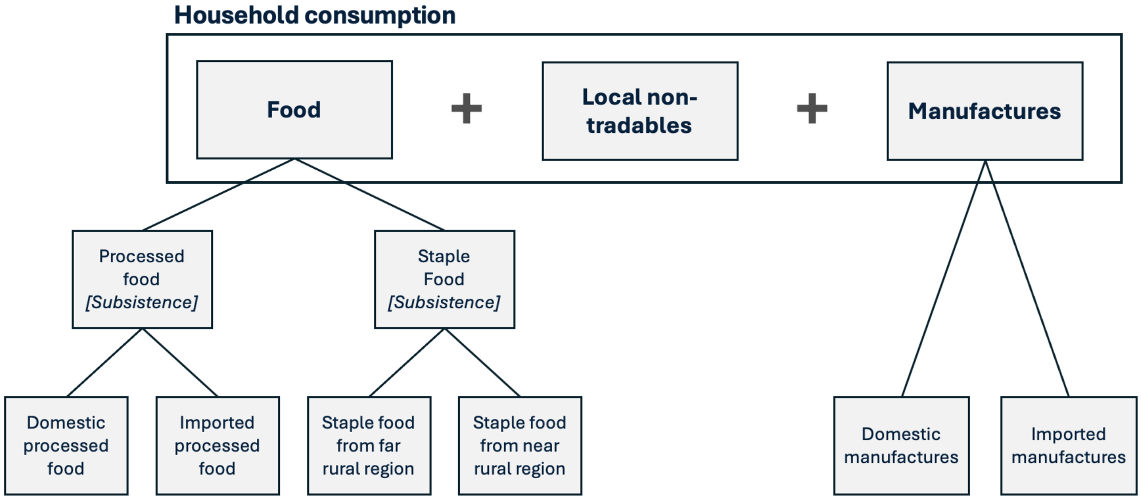
where ζ_M is the elasticity parameter and ν_M is the share parameter.

2.4.3 Transport costs

The movement of goods between regions is costly, which results in a transport cost wedge between the factory/farm gate price at the origin (p^x_j) and the consumption price at the destination (p^y_j). The transport cost wedge, i.e. the difference between the destination price and the origin price, consists of two components: (i) direct fuel costs and (ii) transport rents that accrue to the capitalist household in the urban region. The consumption price at destination is thus

$$p^y_j = p^x_j(1 + \underbrace{\psi^{rent}_{ij}}_{\text{rent}}) + \underbrace{\psi^{fuel}_{ij}}_{\text{fuel}} \tag{2.15}$$

Figure 2.6: Schematic overview of consumption



The rent component is a relative mark-up on the price at origin. The fuel cost component is the value of fuel required to move on unit of the commodity

$$p_j^y = p_j^x(1 + \underbrace{\psi_{ij}^{rent}}_{\text{rent}}) + \underbrace{(1 - t_O)Ep_O^w fuel_{ij}}_{\text{fuel}} \tag{2.16}$$

where t_O is the fuel subsidy, p_O^w is the international price of imported fuel, E is the exchange rate, $fuel_{ij}$ is the quantity of fuel required to move one unit of the commodity j from origin to destination. Both the fuel requirement ($fuel_{ij}$) and the rent parameter (ψ_{ij}^{rent}) are indexed by ij to indicate that they vary by commodity and trade dyad. Trade between the urban region and the far rural region moves via the near rural region. For simplicity we assume that transport costs are symmetric around the near rural location, such that goods transported between near and far, and goods transported between near and urban are subject to the same transport costs.

2.4.4 Prices

The small open economy is a price-taker in world markets. International prices of the four imported commodities (processed food, manufactures, fuel, and fertiliser) are denoted p_P^w, p_M^w, p_O^w, and p_Z^w. The economy's export good (the cash crop) sells at price p_C^w in international markets. The price of the imported manufactured good

(p^w_M) is held fixed at a value of one and functions as the numéraire. These international prices are 'at the quayside', i.e. before tariffs, duties or subsidies, and are denoted in foreign currency.

Table B1 in Appendix 2.B.7 shows an overview of region-specific consumption prices. Consumption prices at the destination are written as a function of the factory price at the origin. In addition to transport costs, tariffs, duties, subsidies and taxes introduce additional wedges between origin and destination prices for some goods. There is a single local price for each product, which applies to both final consumption by households and intermediate demand by firms.

2.4.5 Commodity market clearing

All commodity markets clear in each period, i.e. there is no storage of goods across periods. The cash crop production from both rural regions is fully exported to international markets. Processed food (both imported and domestically produced) is directly consumed by households. The staple crop (from far rural and near rural) and the manufactured good (imported and domestically produced) are consumed by households directly and are used as inputs in the production of processed food. In addition, imported manufactures are used in capital formation. The economy imports fertiliser and fuel to match input needs in agricultural production and transportation.

2.4.6 Factor market clearing

Labour markets clear within region, i.e. there is no migration between regions. Within a given region, however, labour is perfectly mobile across sectors. As our static model is focused on short-run effects, land allocations are held fixed; there can be no reallocation of land across crops or regions. Land rents adjust to satisfy first order conditions at these fixed allocations. Capital can be reallocated between the food processing and manufacturing sector to equalise returns in the two sectors. Investment is sufficient to maintain the overall capital stock at its steady state level.

2.4.7 Macroeconomic balances

To close the model, we assume closure rules for the three macroeconomic balances: the fiscal balance, the external balance and the savings-investment balance. While the choice of closure rules is arbitrary and does not alter the initial calibration of the model, by determining the form of overall macroeconomic adjustment, closure rules do affect the results in counterfactual model simulations.[15]

On the fiscal side, the government generates revenue from multiple sources: direct taxes on household income, import tariffs on processed food and manufactures, export duties on cash crops, and grant aid. Government expenditures cover subsidies on imported fertiliser, fuel and food, direct cash transfers to households, and debt servicing. The choice of closure rule for the fiscal balance varies by simulation to reflect differential fiscal financing strategies across the different policy counterfactuals. In simulation (1) (i.e. the 'no mitigation' case), all tax, tariff and subsidy rates are held constant, along with external fiscal financing so that government spending on household transfers adjusts to satisfy the fiscal budget constraint (see Equation 2.49 in Appendix 2.B.6). In simulations (2) and (4) (i.e. externally financed policies), transfers remain unaltered and external grant aid adjusts to finance the discretionary increase in government spending and to restore fiscal balance. In simulations (3) and (5) (i.e. domestically financed policies), it is the tax rate on the capitalist household's income that rises to close the fiscal financing gap that results from the increase in spending.

Foreign savings (both public and private) are exogenous and fixed. To ensure external balance, this then implies that adjustment is via the trade balance (i.e. exports and imports). This quantity adjustment to the trade balance is, in turn, facilitated by movements in the real exchange rate. Finally, we adopt an investment-driven closure where gross investment exactly offsets depreciation so that steady-state capital remains constant. With foreign savings held constant across steady states, private savings (and hence private consumption) by the capitalist house-

[15]For a comprehensive discussion of macro closure rules, see Lofgren et al. (2001). Macro closure rules are more readily handled in a dynamic model setting than in comparative statics. In a related forthcoming paper, we thus examine the effect of external price shocks under alternative fiscal mitigation measures and financing strategies in a dynamic model.

hold adjust to satisfy the 'savings-equal-investment' balance in the economy.[16]

2.5 Calibration and baseline equilibrium

We draw on a blend of aggregate macro data and high-quality geo-referenced micro survey data to calibrate our model. Following the approach adopted in Adam et al. (2018), we anchor the calibration of our model in the 2018 Social Accounting Matrix (SAM) for Tanzania (Thurlow, 2021). The SAM distinguishes between 42 activities and commodities and defines 10 different household types by income deciles. Translating the SAM into a structure that is consistent with our modelling framework requires a number of choices around classification and taxonomy in order to map activities and commodities reported in the SAM into the production and consumption structure of our model. For details, see Appendix 2.C.

One key challenge in using the SAM as a source of data to calibrate our model lies in its lack of spatial differentiation. We are thus required to make assumptions around the location of production and consumption to map the SAM's aggregate data into our spatial model structure. Our model features 9 spatially-differentiated production sectors, 13 commodities (8 domestically produced, 5 imported commodities), and 6 spatially-differentiated households. We draw on georeferenced microeconomic survey data from the World Bank's Living Standards Measurement Study (LSMS) to microfound the required assumptions around the spatial allocations of land and labour across locations in the economy.[17] The spatial distributions of land and labour that we derive from the LSMS data can be found in the calibration appendix 2.C.

We reshape the consumption vector given by the SAM, assigning total household consumption of each commodity type to the 6 household types in our model. As there exists no way to directly map from the income deciles in the SAM to our model' household structure, we again draw on LSMS data to inform the mapping. The calibration of transport cost wedges is based on Adam et al. (2018), subject to

[16]This 'Johansen closure' contrasts with a 'neo-classical closure' where savings are fixed and investment adjusts, and various Keynesian closures where the adjustment mechanism is via employment or unemployment.

[17]We draw on five waves of Tanzanian LSMS data from 2008-2019 (National Panel Survey data collection waves 1-4, plus the extended panel collected in 2019).

SAM control totals. Fiscal and macro aggregates are based on the SAM in conjunction with IMF Article IV documents and reports. The structure of the economy and factor intensity of production at baseline are summarised in Table 2.2. Table 2.3 reports the fiscal and external balance at baseline.

2.6 Comparative statics

In this section, we run a set of comparative static experiments, simulating the impact of import price shocks under different policy regimes. We first focus on the effects of a simultaneous 20% increase in the international prices of food, fuel and fertiliser (p_P^w, p_O^w, and p_Z^w) in the absence of any mitigating policy response, to explore the fundamental mechanisms of transmission for import price shocks in our general equilibrium framework. This 'no mitigation' experiment is the counterfactual against which we subsequently assess different mitigation policies.[18] Individual import price shocks can have complex (at times mutually reinforcing or offsetting) effects on the economy in general equilibrium. To guide our understanding of the fundamental mechanisms of shock transmission in general equilibrium, we also run simulations of isolated price shocks for food, fuel and fertiliser in the absence of a policy response. Comparing the results from these isolated price shocks with those of the joint shock experiment allows us to detect potential feedbacks and interactions from the joint shock in general equilibrium. Simulation results for isolated price shocks can be found in Appendix 2.D.

Building on the analysis of shock transmission in the 'no mitigation' experiment, we compare different fiscal mitigation strategies and financing arrangements. In a first set of experiments, we assume the government increases import price subsidies to mitigate the impact of the shock. We set subsidy rates on food, fuel and fertiliser such that the original price increase is fully offset by the higher subsidy. With these higher subsidy rates, domestic effective prices of imported food, fuel and fertiliser remain unchanged at the border. We consider two different financing arrangements for this policy. First, we assume the increase in subsidy

[18]This shock is smaller in magnitude than the global price shock observed in 2022 after Russia's invasion of Ukraine, particularly for fuel and fertiliser. Nonetheless, a simultaneous 20% increase in the prices of food, fuel and fertiliser is a useful benchmark case, as it constitutes a significant terms of trade shock which can still be accommodated comfortably by our modelling structure. Larger price shocks challenge the accuracy of the numerical solutions to the model.

Table 2.2: Structure of the domestic economy at baseline

	Cash crop	Staple crop	Food processing	Manufact.	Non-tradables
[1] Supply (producer prices as share of total output)					
Total	4.1 %	32.1 %	19.7 %	26.3 %	17.8 %
Far rural	2.4 %	22.9 %	-	-	3.0 %
Near rural	1.6 %	9.2 %	-	-	1.5 %
Urban	-	-	9.3 %	8.9 %	13.3 %
RoW	-	-	10.4 %	17.5 %	-
[2] Absorption (consumer prices as share of total absorption)					
Total	8.8 %	30.9 %	24.3 %	23.5 %	12.5 %
Final demand					
Far rural	-	11.4 %	7.3 %	7.9 %	2.1 %
Near rural	-	3.8 %	2.8 %	2.4 %	1.0 %
Urban	-	9.8 %	14.2 %	11.8 %	9.3 %
RoW	8.8 %	-	-	-	-
Intermediate demand					
Food processing	-	5.9 %	-	1.3 %	-
[3] Factor intensity (per unit of output)					
Labour					
Far rural	0.197	1.352	-	-	2.206
Near rural	0.233	1.284	-	-	1.500
Urban	-	-	0.437	0.364	0.728
Land					
Far rural	0.098	0.258	-	-	-
Near rural	0.058	0.201	-	-	-
Capital					
Urban	-	-	3.613	4.517	-
Fertiliser [*]					
Far rural	0.106	0.695	-	-	-
Near rural	3.035	3.106	-	-	-

Note: [1] Supply is reported as a share of total output at producer prices (total output is 135.5, i.e. 137% of GDP). [2] Final demand and intermediate demand are reported as a share of total absorption (total absorption is 193, i.e. 195% of GDP). [3] Factor intensity is reported per unit of output in quantity terms, so comparisons are only meaningful within commodity type. [*] Fertiliser intensity is multiplied by 100 to ease comparison of magnitudes across sectors.

Table 2.3: External and fiscal balance at baseline

External balance	% Value added	**Fiscal balance**	% Value added
Total Exports	17.2 %	**Total Revenue**	28.4 %
Cash crop – far	12.9 %	Tariffs	3.8 %
Cash crop – near	4.4 %	Export duties	1.7 %
Total imports		**Total Expenditure**	30.4 %
Processed food	13.0 %	Subsidies	2.4 %
Fuel	21.5 %	Transfers	25.9 %
Fertiliser	0.9 %	Debt service	2.2 %
Manufactures [a]	23.9 %	**Deficit**	2.0 %
Manufactures [b]	1.3 %		
Manufactures [c]	5.1 %	Aid	2.0 %

Note: Total value added at baseline is 98.8 (domestic production at producer prices plus fertiliser imports at farm-gate prices). [a] Manufactures imported for final consumption. [b] Manufactures imported for use as intermediates in the domestic production of processed food. [c] Manufactures imported for capital formation.

payments is fully externally financed by grant aid. In this simulation, the increased aid fully compensates for the increase in the cost of imports, while the perfectly targeted price subsidies fully insulate domestic production and consumption decisions. Second, to simulate domestic financing of the subsidy increase, we increase the direct tax rate on the capitalist household.[19] Increasing the rate results in a distortionary demand response from the capitalist and corresponding production shifts in the economy.

In a second set of experiments, we assume the government increases direct net transfers paid to households, while keeping subsidy rates at their baseline values. The total increase in the transfer budget is set equal to the increase in the subsidy budget in experiment (2), where subsidies are financed through an additional inflow of grant aid. We take the aid inflow required to finance the 'full offset' subsidies to be an estimate of the 'cost of the shock'. In the transfer experiments, we distribute these funds to households directly as uniform lump sum per capita payments. We again contrast the effects of the policy under two different financing arrangements. First, we again simulate external financing, increasing grant aid inflows to close the fiscal financing gap. Second, in a subsequent experiment, the increase in transfer payments is financed by an increase in the income tax rate

[19]Given that this household supplies no labour, household income consists of the gross return on capital plus transport rents.

on the capitalist household. In each of the experiments, we consider four broad sets of outcomes to analyse the economy-wide effects of the import price shock, summarised in Table 2.4: (i) fiscal aggregates, (ii) imports and exports, (iii) sectoral production responses, and (iv) household consumption.

2.6.1 Simultaneous price shock with no mitigation

In the 'no mitigation' experiment (1), we hold the inflow of grant aid, which is used to finance the fiscal deficit, fixed at the baseline value. Total government revenue drops by 9% as income from import tariffs and export duties falls due to declining trade flows. Revenue from direct taxes also declines by 9% as households' taxable incomes fall. Import price subsidy rates are kept fixed at the baseline values of 0% for food, 10% for fuel and 25% for fertiliser imports. As import prices rise, the total cost of these ad valorem subsidies thus increases. The resulting budget squeeze leads the government to cut direct transfers to households, the balancing item in this simulation, by 10.8% overall.

The simultaneous import price shock has a strong negative effect on the value of aggregate production (-6.8%) and final consumption (-8.2%). Farm- and factory-gate prices of all domestically produced goods fall as higher import prices, coupled with low elasticity of substitution between domestic and imported varieties of goods, reduce households' disposable income that can be spent on domestically produced goods. In addition, higher fuel prices drive up the cost of transporting goods between regions, exerting further downward pressure on farm- and factory-gate prices. Production quantities in agriculture and manufacturing fall, while production of domestic processed food increases and production of local non-tradable services edges up slightly.

The impact of the price shock on trade flows is quantitatively significant. With the global price of the export cash crop unchanged, the reduction in total value of exports by 6.3% can be attributed in full to a fall in production volume. As producers of the export cash crop are price takers in international markets, the higher fuel cost implies a larger transport cost wedge between local farm-gate prices and the price on world markets. This negative effect on farm-gate prices of the export crop

Table 2.4: Comparative statics results (% change over baseline)

	(1) No mitigation		(2) Subsidies Aid fin.		(3) Subsidies Tax fin.		(4) Transfers Aid fin.		(5) Transfers Tax fin.	
	Q %	V %	Q %	V %	Q %	V %	Q %	V %	Q %	V %
Fiscal aggregates										
Total Revenue	–	-9.0	–	-4.6	–	19.4	–	0.6	–	25.6
Tariffs	–	-9.4	–	-33.8	–	-43.6	–	6.9	–	-9.8
Export duties	–	-6.3	–	0.0	–	7.1	–	-7.5	–	-6.8
Direct Taxes	–	-9.2	–	0.0	–	30.8	–	0.1	–	33.9
Total Expenditure	–	-8.4	–	19.0	–	18.1	–	24.6	–	23.9
Subsidies	–	9.8	–	243.2	–	235.3	–	14.0	–	10.2
Transfers	–	-10.8	–	0.0	–	0.0	–	27.3	–	27.3
Debt Service	–	-0.4	–	0.0	–	-3.8	–	3.5	–	-1.6
Deficit	–	0.0	–	353.4	–	0.0	–	366.0	–	0.0
Aid	–	0.0	–	353.4	–	0.0	–	366.0	–	0.0
Trade										
Total Exports	–	-6.3	–	0.0	–	7.1	–	-7.5	–	-6.8
Cash crop	-6.3	-6.3	0.0	0.0	7.1	7.1	-7.5	-7.5	-6.8	-6.8
Total Imports	–	-1.7	–	10.8	–	1.9	–	9.2	–	-1.8
Processed food	-25.5	-10.5	0.0	20.0	-13.2	4.2	-15.6	1.3	-27.8	-13.4
Manufactures [a]	-9.3	-9.3	0.0	0.0	-15.1	-15.1	10.1	10.1	-8.2	-8.2
Manufactures [b]	-0.8	-0.8	0.0	0.0	-10.9	-10.9	5.9	5.9	-4.0	-4.0
Manufactures [c]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fertiliser	-29.5	-15.4	0.0	20.0	-3.4	15.9	-19.7	-3.6	-27.3	-12.7
Fuel	-6.3	12.4	0.0	20.0	0.0	20.0	-3.4	15.9	-6.2	12.6
Production										
Aggregate	–	-6.8	–	0.0	–	-8.6	–	5.1	–	-6.6
Cash crop	-6.3	-21.7	0.0	0.0	7.1	5.7	-7.5	-22.5	-6.8	-22.0
Staple crop	-1.6	-6.1	0.0	0.0	-4.0	-6.4	-0.5	7.8	-2.2	-3.5
Food processing	1.9	-4.8	0.0	0.0	-2.6	-14.9	1.7	3.0	1.1	-10.8
Manufacturing	-3.1	-7.2	0.0	0.0	6.3	-7.7	-0.1	6.7	5.4	-2.8
Non-tradables	0.1	-5.5	0.0	0.0	-1.6	-13.1	2.3	6.6	1.1	-8.2
Final consumption										
Aggregate	–	-8.2	–	0.0	–	-13.4	–	7.1	–	-8.2
Far-Agricultural	-10.5	-11.5	0.0	0.0	1.3	-2.3	9.0	15.7	11.8	11.5
Far-Non-agricult.	-12.6	-14.8	0.0	0.0	0.2	-3.4	12.3	17.6	11.8	11.0
Near-Agricultural	-12.2	-12.7	0.0	0.0	-0.8	-4.9	15.5	23.0	13.1	13.4
Near-Non-agricult.	-13.4	-12.7	0.0	0.0	-1.1	-4.7	16.1	23.3	13.0	14.3
Urban-Wage-earning	-4.6	-4.4	0.0	0.0	-1.6	-8.4	-0.5	5.4	-1.2	-3.6
Urban-Capitalist	-7.7	-8.4	0.0	0.0	-29.2	-33.8	-9.3	-3.7	-38.1	-39.8

Note: This table shows the impact of a simultaneous 20% increase in the import prices of food, fuel and fertiliser, given different mitigation policies. Effects are reported as changes in quantity (Q%) and changes in value (V%) over baseline in percent. (1) Joint price shock in the absence of any mitigating policy response; (2) Joint price shock with aid-financed 'full-offset' import price subsidies; (3) Joint price shock with tax-financed 'full-offset' import price subsidies; (4) Joint price shock with aid-financed net transfers to households; (5) Joint price shock with tax-financed net transfers to households. [a] Manufactures imported for final consumption. [b] Manufactures imported for use as intermediates in the domestic production of processed food. [c] Manufactures imported for capital formation. Consumption is adjusted for household size, i.e. refers to consumption per capita.

is more pronounced in the far than in the near rural region, as more fuel is required to transport goods from this more remote area to the urban port for exporting.

Due to substitution effects, import quantities respond strongly to the increase in prices.[20] For processed food and fertiliser, which can partly be substituted with domestically produced food and other factors of production, the relative reduction in import volumes exceeds the relative increase in price. The 20% price increase causes a 25.5% reduction in the imported quantity of processed food and 29.5% reduction in the imported quantity of fertiliser. Since there is no domestic substitute for fuel, the reduction in imported quantity of fuel is much smaller than for food and fertiliser. For fuel, the 20% price increase results in only a 6.3% reduction in the volume of imports, reflecting the lower level of domestic trade flows. Imported manufactured goods have three different uses in the domestic economy. Firstly, they are consumed by households, who can substitute imports with the domestically produced manufactured alternative in response to price shock. Secondly, imported manufactures are used as an intermediate input in the domestic production of processed food, where they can again be substituted with the domestic variety of manufactures. Thirdly, imported manufactures are the sole input into capital formation. With fixed investment, this requires the quantity imported for capital formation to remain unchanged even at the higher post-shock price.

While there is an overall reduction in domestic production, there is significant heterogeneity between sectors in the economy. Production quantities fall most strongly in the export cash crop sectors, as higher fuel costs depress farm-gate prices. Production quantities of the staple food crop also fall, as lower fertiliser use reduces production capacity. In the domestic manufacturing sector, the production volume falls by 1.7% due to weaker demand. In the local non-tradable services sectors, on the other hand, production quantities increase. Since local non-tradables do not incur transport costs, households shift some consumption from tradable

[20]The extent of import compression depends crucially on our chosen macro closure rules. In this first simulation, import demand responds very strongly to the price shock in order to restore external balance, since external financial inflows such as foreign savings, aid, and remittances are held fixed. In subsequent simulations (2) and (4), external grant aid flexibly adjusts to close the fiscal financing gap, which significantly reduces import compression vis-a-vis simulation (1). The different simulations reflect stylised closures at opposite ends of the spectrum; In reality, the price shock's effect on import demand is likely more muted than in simulation (1), but more pronounced than in simulation (2).

commodities to non-tradable services. While the effect on production *quantities* is very heterogeneous, with some sectors shrinking and others expanding, marked reductions in all farm- and factory-prices imply reductions in the *value* of production across all sectors in the economy.

Wages fall significantly across all regions in the economy.[21] Rural labour shifts from cash crop agriculture to staple crop agriculture and non-tradable services. This implies a reduction in the size of the agricultural household in both the far and near rural region, and a corresponding increase in the size of the non-agricultural households.[22] In the urban region, labour movements are smaller, with some labour reallocating from food processing and manufacturing to non-tradable services. Falling land rents across all agricultural sectors reflect less intensive farming due to lower fertiliser use.[23] Fertiliser use plummets in response to the price shock, which is amplified by the increase in fuel costs, further driving up the farm-gate price of fertiliser. The fertiliser reduction is stronger in cash than in staple crop production, and stronger in the far than in the near rural region. Some capital, which is employed in urban production only, is reallocated from manufacturing to food processing.

The import price shock has strong negative effects on households' real income and final consumption. In line with falling prices paid to all factors of production, real income from factors declines for all household types. Non-factor incomes, which constitute direct net transfers for all households except the capitalist, fall even more strongly than income from factors. This results in disposable income (post-tax factor income plus non-factor income) falling by 8.6% on average for all wage-earning households (i.e. excluding the capitalist).

The price shock and resulting production shifts cause strong movements in relative prices in the economy. In response, households adjust their consumption baskets in favor of locally produced commodities that incur lower fuel costs, and domestic substitutes to replace expensive imports. All households significantly re-

[21] Labour market clearing within region implies that wages equalise between sectors in a given region.

[22] There is no migration between regions in the current version of this model, so labour markets must clear within each region.

[23] As we examine only the short-run effects of the shock, land allocations are held fixed such that land cannot be moved between cash crop and staple crop agriculture.

duce their consumption of imported processed food in response to the 20% price increase. The strength of their response varies depending on the level of per capita consumption at baseline and, to some extent, on the household-specific share parameters that influence the degree of substitutability between imported and domestic processed food.[24] The further a household's per capita consumption at baseline is from the subsistence level on processed food, the stronger the response to the price increase. Households that at baseline consume quantities relatively close to the subsistence threshold have less scope to reduce their consumption in response to the price shock. All households increase their consumption of local non-tradable services, which do not incur transport costs. By contrast, consumption of commodities that require a relatively large expenditure on transport, such as imported manufactures consumed in the far rural region, is reduced significantly.

2.6.2 Externally financed subsidy increase

The first policy analysis we conduct is a straightforward experiment to illustrate the mechanisms that operate in the model. Subsidy rates on imported food, fuel and fertiliser are set to fully offset world price increases, with the cost of the increased subsidies met by an additional inflow of grant aid. The policy is therefore budget-neutral. This implies raising the subsidy on fuel (s_O) from 10% to 25% and the subsidy on fertiliser (s_Z) from 25% to 37.5%.[25] At baseline, there is a 10% tariff on food imports, but no subsidy. To keep the price of imported food at its baseline value, we introduce a subsidy of 8.3%.[26]

As these higher subsidy rates keep effective consumer prices for imported food, fuel and fertiliser at their baseline values, the price shock does not cause a demand response and import quantities remain unchanged (see column (2) in Table 2.4).

[24]We model the consumption of imported and domestic processed food in a CES aggregator. The elasticity of substitution in this CES structure (ζ_P) is the same for all household types. The share parameters (ν_P) are calibrated based on the common elasticity parameter and households' consumption quantities at baseline.

[25]The 'full-offset' subsidy rates are calculated as $s^{new} = 1 - ((1 - s^{base})p_w^{base}/p_w^{shock})$.

[26]At baseline, the international price of processed food is $p_{Pw}^{base} = 1$ and the effective price of food imports in the urban region, inclusive of the 10% tariff, is $p_{Pw}^{u} = 1.1$. The 20% price shock increases the international price of processed food to $p_{Pw}^{shock} = 1.2$. To bring the effective price in the urban region back to its baseline value, we introduce a subsidy (s_P) such that $s_P = (p_{Pw}^{u}/p_{Pw}^{shock}) - 1$, which yields $s_P = 8.3\%$.

Since international prices for food, fuel and fertiliser increase in global markets, the value of total imports rises by 11.7% overall. Expenditure on subsidies more than triples (+243%) under the higher 'full-offset' subsidy rates. This increase in the import bill is accounted for by an equivalent increase in grant aid. As a result, the country's aid budget more than quadruples (+353%). At the same time, tariff revenues fall by 33.8% as the government forgoes the tariff on food imports. The magnitudes of the increase in subsidies and aid required to perfectly offset the shock are large and would likely render this externally financed 'full offsetting' of the shock an infeasible policy. In any case, if the price shock is assumed to be permanent, the additional expenditure on subsidies would be recurring each period, and the government would have to secure sufficient grant aid financing in each period. However, this experiment yields useful insights on the 'cost of the shock', i.e. the funds that would be required to fully shield the economy from the effects of the shock in a given period. In addition, this experiment can be seen as a proof-of-concept to illustrate how the model works.

2.6.3 Domestically financed subsidy increase

In this experiment, we again increase subsidies on import prices such that the global price shock is fully offset and effective consumer prices remain at baseline levels. The additional expenditure on subsidies, however, is financed domestically through an increase in the income tax rate applied to the capitalist household's income[27] from 30% to 63.8%. In contrast to the previous experiment, where the policy is financed externally through grant aid, domestic financing implies a distortionary demand response from the capitalist household which has noticeable ripple effects throughout the domestic economy. At the higher tax rate, the capitalist's real disposable income declines by 24.1% vis-a-vis the baseline value. As a result, the capitalist household significantly reduces consumption of all commodities by between 31% to 36% in value terms. Since the capitalist's consumption expenditure accounts for approximately 27% of aggregate final consumption at baseline, this reduction has significant effects on all sectors in the economy in general equilibrium.

[27]The capitalist household pays income tax on income from returns on capital and transport rents. Returns on bonds and remittances are not taxed.

Total government revenue increases by 19.4%, owing to higher income from direct taxes and export duties, which more than offset the reduction in tariff income (see column (3) Table 2.4). Total expenditure increases by 18.1% due to the surge in spending on subsidies (+235%). Import quantities fall by 12.4% for manufactures, 13.2% for processed food, and 3.4% for fertiliser, while the import quantity of fuel is flat at +0.3%. Due to the offsetting price effects (international prices for food, fuel and fertiliser increase by 20%), the value of total imports still increases by 2%, despite the reduction in import quantities. While farm- and factory-gate prices of all other domestically produced goods fall in response to the capitalist's demand reduction, the higher subsidies on fuel and fertiliser cause the farm-gate price of the export cash crop to remain at its baseline value. The corresponding change in the relative price between staple and cash crops, and the balance of payments pressures exerted by the higher value of imports, lead to an increase in cash crop production and exports (+7.1%).

In the far rural region, labour moves from staple crop agriculture (-13.8%) into cash crop agriculture (+113%), which generates a 9.4% increase in cash crop production in the far rural region. In the near rural region, labour movements are smaller, with some labour moving out of staple crop agriculture and non-tradable services and into cash crop agriculture. In the urban region, labour and capital are pulled into the manufacturing sector to increase production of the domestic substitute for imported manufactures.

Nominal pre-tax factor incomes fall for all household types, with the reduction strongest for the urban wage-earning household (-14.4%) and weakest for the non-agricultural household in the far rural region (-2.1%). Disposable real incomes per capita, which are net of tax and transfers and deflated by households' consumer price indices, also fall for households in the urban region (-1.6% for the wage-earning household and -24.1% for the capitalist household). In the near rural region, the fall in disposable real incomes is smaller (-1.1% for the non-agricultural household and -0.8% for the agricultural household). In the far rural region, however, disposable real incomes increase slightly, as the reduction in consumer prices is stronger than the reduction in nominal incomes.

2.6.4 Externally financed transfer increase

We next consider direct household transfers to buffer the domestic population from the import price shock. We again start with an experiment in which the policy is financed externally through an increase in grant aid (see column (4) in Table 2.4). Instead of raising import price subsidies, in this experiment we distribute uniform per capita transfer payments to households. To ease comparison with the previous policy experiments, we set the increase in the transfer budget to equal the increase in the subsidy budget in experiment (2), i.e. the total increase in transfers is equivalent to the cost of implementing the 'full offset' import price subsidies in column (2).

As in the 'no mitigation' experiment (1), the price shock is fully passed through to the domestic economy in this experiment. However, households' disposable incomes are significantly higher than in the experiment with no mitigation due to the additional transfers payments, which props up domestic demand. There are no additional subsidies to offset the import price shock, such that domestic prices of imported goods increase. Unlike in experiments (2) and (3), where price shocks were offset, changes in relative prices and corresponding substitution effects can fully unfold in this experiment. Producers and consumers are exposed to the new prices, which appropriately signal opportunity costs, and consequently substitute imports with domestically produced goods to some extent where possible.

Due to the relatively high cost of the policy, and the limited increase in government revenues, the fiscal deficit expands significantly, with aid inflows increasing to close the fiscal financing gap. Total government expenditures rise more strongly than in experiments (2) and (3), because in addition to the discretionary increase in transfers, other subsidy payments, which are defined ad valorem, increase automatically as import prices rise. Debt servicing costs also increase, since the price index that is used to deflate interest payments on the capitalist's holdings of government bonds rises.

Cash crop exports decline, as farm-gate prices of the export cash crop fall due to increase in fuel costs, and production costs rise due to the increase in fertiliser cost. Imports of processed food, fertiliser, and fuel fall in response to the increase in price. Imports of the manufactured good, however, increase as the price remains unchanged and households' disposable incomes rise thanks to the transfer payments.

As price shocks are allowed to pass through to the domestic economy, production responses are broadly similar to the 'no mitigation' experiment (1). Since the shock is buffered by the transfer policy, however, shifts in production volumes are more muted than in the experiment with no mitigation. The shift of rural labour out of the cash crop sectors is stronger, and the shift into the staple crop sectors is weaker, as the aid-financed transfers reduce the pressure on the external balance and domestic food production. Transfer payments allow households to continue to cover a higher share of subsistence needs with imported food despite the price increase. The non-tradable services sectors in both rural regions expand, as households spend some of the additional disposable income on these goods which are not subject to transport costs and are thus not directly affected by the fuel price shock.

As a direct effect of the policy, households' disposable nominal incomes rise strongly. Real incomes rise less strongly, as higher prices erode part of the additional transfer payments. Transfers are paid uniformly per capita, such that they are distributed equally across household types in nominal terms. However, the increase in price levels is heterogeneous across households, depending on their specific consumption bundles, and baseline per capita income levels vary widely between households. As a result, the transfer payments, while uniformly distributed in nominal terms, have differential effects on relative changes in households' real income per capita. Households in rural regions benefit more strongly from the policy in relative terms, with consumption per capita increasing by between 9.0% and 16.1%. For the wage-earning household in the urban region, the income gains from transfers are not sufficient to compensate for the increase in prices, such that consumption falls slightly in quantity terms. The capitalist household does not receive a transfer payment, and their income is eroded by higher prices and lower income from transport rents, as households' consumption becomes more local in response to the increase in fuel prices.

2.6.5 Domestically financed transfer increase

Finally, we contrast the previous policy experiments with a domestically financed increase in transfer payments to households. In this experiment, the overall increase in the transfer budget is again identical to the aid-financed transfer increase in experiment (4), but the fiscal financing gap is closed domestically by raising the capitalist's income tax rate from 30% to 69.4%. Domestic financing again implies a distortionary demand response from the capitalist, as the capitalist's real income declines by 32.4% vis-a-vis the baseline value. This leads the capitalist household to significantly reduce consumption of all commodities by between 37% to 45% in value terms. The total fiscal financing gap in the transfer experiment with domestic financing is larger than in the subsidy experiment with domestic financing, because in addition to the increase in transfers, ad valorem subsidies automatically add to the fiscal financing gap when import prices rise.

All trade flows, including both exports and imports, decrease strongly. Imports of processed food, fuel and fertiliser decline directly in response to the import price shock. Imports of the manufactured good decline as a result of weaker demand from the capitalist household, who at baseline accounts for 24% of total consumption of imported manufactures. Lower trade flows result in lower revenue from export duties and import tariffs. However, as the tax rate on the capitalist's income is raised to finance the additional transfer payments, increased revenue from direct taxes more than offsets the reduction in trade taxes, leading to an expansion of total government revenue by 25.6%. Government expenditure rises discretionary for the increase in transfer payments, and automatically for the ad valorem subsidies on imports, with subsidy rates kept at baseline values. As in the domestic financing experiment (3), debt servicing costs fall slightly due to the capitalist's holdings of government bonds being deflated at a higher price index.

The total value of domestic production declines by 6.6%. This reduction is significantly smaller than in the tax-financed subsidy experiment (-8.6%), but only slightly smaller than in the 'no mitigation' experiment (-6.8%). Due to increased fuel and fertiliser costs, agricultural production volumes of both staple and cash crop decline most strongly. Manufacturing production expands in quantity terms to replace imports, but declines in value terms due to price deterioration. Food processing and non-tradable services expand slightly in quantity terms, but production declines strongly in value terms, which is again due to a reduction in producer prices.

Real disposable incomes per capita increase in both rural regions, decline slightly for the urban wage-earning household, and fall strongly for the capitalist due to the higher tax. The reduction in aggregate consumption is quantitatively equivalent to the reduction in the 'no mitigation' experiment (both -8.2%). However, the distribution across households is very different from the 'no mitigation' experiment. In the transfer experiment, the brunt of the reduction in consumption is born by the capitalist household, and to a lesser extent by the urban wage-earning household. For rural households, the two experiments differ substantially, with income changes having a similar absolute magnitude, but with opposite signs. While the value of consumption for rural households declines by between 11.5% to 14.8% in the 'no mitigation' experiment, it rises by between 11.0% to 14.3% in the experiment with tax-financed transfers.

2.7 Conclusion

World prices of food, fuel and fertiliser increased sharply in 2021-2022. Such global price shocks have heterogeneous effects across households and sectors within low-income countries. This paper documents that the spatial dimension is key to understanding this heterogeneity, as spatial modelling makes it possible to capture households' differential exposure to international markets, account for knock-on effects of fuel price shocks in transport-intensive remote regions, and realistically simulate the internal supply and demand shifts caused by the shock in general equilibrium.

We develop a spatial general equilibrium model of a small open agricultural economy to study the heterogeneous spatial transmission of shocks to households and to contrast different policy measures that can be used to mitigate these effects. This framework is particularly valuable since there are strong spatial patterns of poverty in Tanzania (and similar countries). When the outcomes of interest – e.g., poverty or food insecurity – show strong spatial patterns, spatial models are uniquely well suited to policy analysis. When poverty is concentrated in remote rural areas, or in poor urban clusters, a spatial model is essential for appropriately capturing the poverty impacts of different policies.

The model we develop is able to capture the spatial consumption patterns observed in the data in response to global price shocks. Model simulations show that, consistent with their higher net food consumption, urban wage-earning households are more adversely affected than agricultural households by global food price shocks. In the case of a simultaneous shock to global food, fuel and fertiliser prices, rural households are affected more strongly than urban wage-earning households. With no mitigation, the negative impact on real income per capita is approximately three times stronger in relative terms for the non-agricultural household in near, the worst-hit household, than for the wage-earning household in urban (-13.4% vs. -4.6%). While urban households' consumption baskets are more exposed to global food price shocks, remote rural households' consumption and production are strongly dependent on fuel and fertiliser imports.

Counterfactual policy experiments show that the negative effects of a combined food, fuel and fertiliser shock can be partially mitigated through import price subsidies and net household transfers. We contrast different financing arrangements for these policies and conclude that even policy measures that are fully domestically financed through higher income taxes can partially mitigate the negative effect of global price shocks on households' real incomes. In our current analysis we assume import price subsidies are set to fully offset the global price shocks. In reality, neither donors nor domestic authorities are likely to provide the financing required to fund such 'full mitigation' subsidies. In the absence of full mitigation, we cannot therefore fully neutralise the changes in relative prices through subsidy offsets, and so the distinction between subsidies and transfers will diminish in this case.

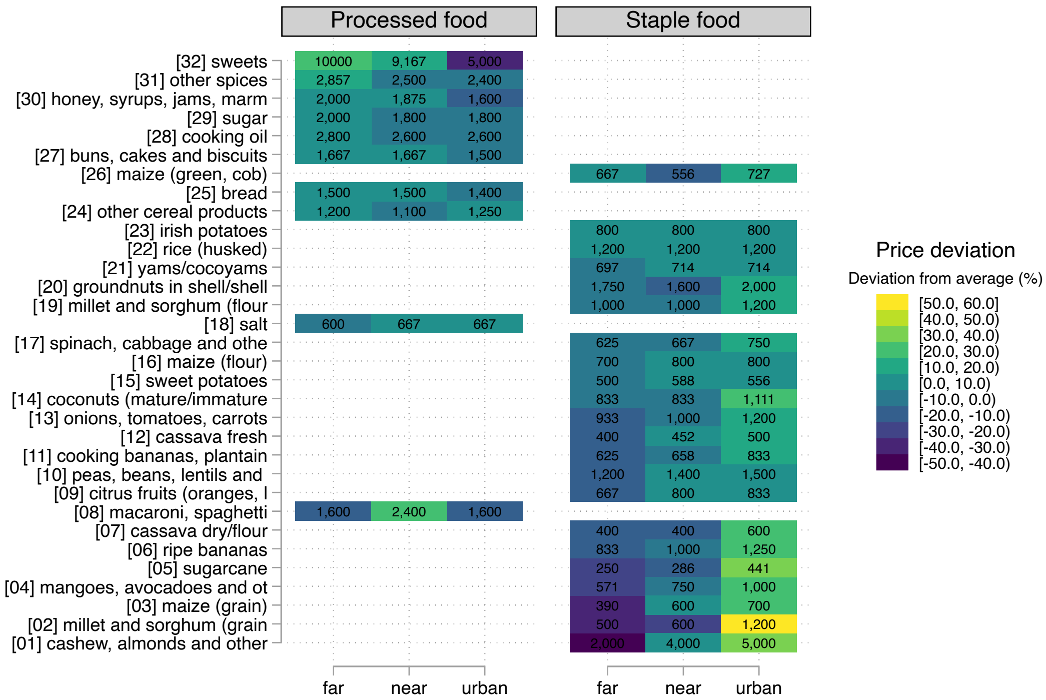
In ongoing work, we extend the policy experiments to include additional policy measures (such as VAT reductions) and more complex financing arrangements (e.g. partial-aid-partial-domestic adjustment using a mix of domestic instruments and external debt and grant aid). In addition, we solve a dynamic version of the model, which will be important for understanding smoothing and adjustment paths. The dynamic model will allow for transitory price shocks, richer public and private adjustment, partial inter-temporal smoothing with adjustment costs in the private sector, and rule-based fiscal responses. As such, the dynamic model will be crucial for understanding smoothing and adjustment paths, a key consideration in the design of future policy responses to global price shocks.

Appendices of Chapter 2

Appendix 2.A Additional figures

2.A.1 Spatial heterogeneity in food prices

Figure A1: Regional differences in consumer prices of food items (LSMS 2010-11)



Note: This figure shows median prices of different food items in far rural, near rural, and urban regions. Prices are calculated from households' expenditure and consumption quantities. Values reported in this figure are median prices in each of the regions, shown as Tanzanian shilling per kilogram/litre. The color scale indicates whether a given regional price is high or low compared to the other regions. The color shading codes a given price's deviation from the mean across all three regions; e.g. for [01] cashew, almonds and other nuts, the price in far rural was 2,000 tsh/kg, which was 45% lower than the average across the three regions (2,000+4,000+5,000)/3=3,666. Source: LSMS, National Panel Survey (wave 2, 2010-11), food consumption questionnaires.

Appendix 2.B Model appendix

2.B.1 First order conditions

Agricultural production

$$\frac{\partial}{\partial A_i} = p_i^l\, Q_i^{\frac{1}{\gamma_i}} \left(\frac{\lambda_i}{A_i}\right)^{\frac{1}{\gamma_i}} - b_i \tag{2.17}$$

$$\begin{aligned}\frac{\partial}{\partial L_i} =& p_i^l\, Q_i^{\frac{1}{\gamma_i}} (1-\lambda_i)^{\frac{1}{\gamma_i}} \left(\delta_i^{\frac{1}{\theta_i}}\, L_i^{\frac{\theta_i-1}{\theta_i}} + (1-\delta_i)^{\frac{1}{\theta_i}}\, Z_i^{\frac{\theta_i-1}{\theta_i}}\right)^{\frac{\theta_i(\gamma_i-1)-\gamma_i(\theta_i-1)}{\gamma_i(\theta_i-1)}} \\ & \left(\frac{\delta_i}{L_i}\right)^{\frac{1}{\theta_i}} - w_i\end{aligned} \tag{2.18}$$

$$\begin{aligned}\frac{\partial}{\partial Z_i} =& p_i^l\, Q_i^{\frac{1}{\gamma_i}} (1-\lambda_i)^{\frac{1}{\gamma_i}} \left(\delta_i^{\frac{1}{\theta_i}}\, L_i^{\frac{\theta_i-1}{\theta_i}} + (1-\delta_i)^{\frac{1}{\theta_i}}\, Z_i^{\frac{\theta_i-1}{\theta_i}}\right)^{\frac{\theta_i(\gamma_i-1)-\gamma_i(\theta_i-1)}{\gamma_i(\theta_i-1)}} \\ & \left(\frac{\delta_i}{Z_i}\right)^{\frac{1}{\theta_i}} - p_Z^l\end{aligned} \tag{2.19}$$

Manufacturing

$$\frac{\partial}{\partial L_M} = \omega\, p_M^u\, a_M \left(\frac{L_M}{K_M}\right)^{\omega-1} - w_M \tag{2.20}$$

$$\frac{\partial}{\partial K_M} = (1-\omega)\, p_M^u\, a_M \left(\frac{L_M}{K_M}\right)^{\omega} - r_M \tag{2.21}$$

Food processing

Since the optimal volume of intermediates is defined from the Leontief relation, the firm's first order conditions are defined over labour (L) and capital (K) as

$$\frac{\partial}{\partial L_P} = (1-\alpha)\, p_P^{va}\, va_P - w_P\, L_P \tag{2.22}$$

$$\frac{\partial}{\partial K_P} = \alpha\, p_P^{va}\, va_P - r_P\, K_P \tag{2.23}$$

where w_F is the wage and r_F the rental rate of capital in the food processing sector.

Non-tradable services

The firm's first order condition with respect to labour (L) is

$$\frac{\partial}{\partial L_{Nl}} = p^l_{Nl} a_{Nl} - w_l \tag{2.24}$$

Budget constraints

The agricultural household in the far rural region (Af) spends on the consumption goods $\{Sf, Sn, Pu, Pw, Mu, Mw, Nf\}$, generates income from wages ($w_{jf}\, L_{jf}$) and returns on land ($b_{jf}\, A_{jf}$) in the two agricultural sectors $\{S, C\}$ which is taxed at rate f_A, and receives a share of total transfers ($\mu_T^{Af}\, T$). The household thus faces the following budget constraint, where $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nf\}$ denotes consumption goods and $j = \{S, C\}$ denotes sectors of production

$$\sum_i \left[p^f_i\, c^{Af}_i\right] = (1 - f_A) \sum_j \left[w_{jf}\, L_{jf} + b_{jf}\, A_{jf}\right] + \mu_T^{Af}\, T \tag{2.25}$$

The non-agricultural household in the far rural region (Nf) spends on the consumption goods $\{Sf, Sn, Pu, Pw, Mu, Mw, Nf\}$, generates income from wage labour in the non-tradables sector ($w_{Nf}\, L_{Nf}$) which is taxed at rate f_N, and receives a share of total transfers ($\mu_T^{Nf}\, T$). The household thus faces the following budget constraint, where $i = \{Sf, Sn, Pu, Pw, Mu, Mw,\, Nf\}$ denotes consumption goods

$$\sum_i \left[p^f_i c^{Nf}_i\right] = (1 - f_N)\left[w_{Nf}\, L_{Nf}\right] + \mu_T^{Nf}\, T \tag{2.26}$$

The agricultural household in the near rural region (An) spends on the consumption goods $\{Sf, Sn, Pu, Pw, Mu, Mw, Nn\}$, generates income from wage labour ($w_{jn}\, L_{jn}$) and returns on land ($b_{jn}\, A_{jn}$) in the two agricultural sectors $\{S, C\}$ which is taxed at rate f_A, and receives a share of total transfers ($\mu_T^{An}\, T$). The household thus faces the following budget constraint, where $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nn\}$ denotes consumption goods and $j = \{S, C\}$ denotes sectors of production

$$\sum_i \left[p^n_i\, c^{An}_i\right] = (1 - f_A) \sum_j \left[w_{jn}\, L_{jn} + b_{jn}\, A_{jn}\right] + \mu_T^{An}\, T \tag{2.27}$$

The non-agricultural household in the near rural region (Nn) spends on the consumption goods $\{Sf, Sn, Pu, Pw, Mu, Mw, Nn\}$, generates income from wage labour in the non-tradables sector ($w_{Nn}\, L_{Nn}$) which is taxed at rate f_N, and receives a share of total transfers ($\mu_T^{Nn}\, T$). The household thus faces the following budget constraint, where $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nn\}$ denotes consumption goods

$$\sum_i \Big[p_i^n c_i^{Nn}\Big] = (1 - f_N)\Big[w_{Nn}\, L_{Nn}\Big] + \mu_T^{Nn}\, T \tag{2.28}$$

The wage-earning household in the urban region (W) spends on the consumption goods $\{Sf, Sn, Pu, Pw, Mu, Mw, Nu\}$, generates income from wage labour ($w_{ju}\, L_{ju}$) in the three sectors $\{F, M, N\}$ which is taxed at rate f_W, and receives a share of total transfers ($\mu_T^W\, T$). The household thus faces the following budget constraint, where $j = \{F, M, N\}$ denotes sectors of production and $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nu\}$ denotes consumption goods

$$\sum_i \Big[p_i^u c_i^W\Big] = \Big(1 - f_W\Big) \sum_j \Big[w_{ju}\, L_{ju}\Big] + \mu_T^W\, T \tag{2.29}$$

The capitalist household in the urban region (C) spends on the consumption goods $\{Sf, Sn,\ Pu, Pw, Mu, Mw, Nu\}$. The capitalist household generates income three sources: returns on capital used in the sectors $\{P, M\}$ and rents from the transport sector (both of which are taxed at rate f_C), and interest from matured bonds. In addition, the capitalist household receives (or pays) remittances (R). The household thus faces the following budget constraint in each period t, where $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nu\}$ denotes consumption goods, $k = \{Sf, Sn, Cf, Cn, Pu, Pw, Mu, Mw, Z\}$ denotes commodities that are transported between regions, and $j = \{P, M\}$ denotes sectors of production

$$\sum_i \Big[p_i^u\, c_i^C\Big] = \Big(1 - f_C\Big)\Big[r_u\,\big(K_{Pu} + K_{Mu}\big) + rents\Big] + r\, P_C\, B + R \tag{2.30}$$

where transport rents for a given trade flow are calculated as the value of the trade flow at origin, multiplied by the commodity-specific wedge parameter.

2.B.2 Non-tradables market clearing

The markets for non-tradable services clear in each region $l = \{f, n, u\}$

$$Q_{Nl} = \sum_h c_{Nl}^h \tag{2.31}$$

2.B.3 Commodity market clearing

Staple crop

Staple crop is produced in both rural regions and is not traded internationally. It is used as an intermediate input in food processing in the urban region and for final consumption in all of the three domestic regions.

$$Q_{Sf} = \sum_h c^h_{Sf} + Sf_P \tag{2.32}$$

$$Q_{Sn} = \sum_h c^h_{Sn} + Sn_P \tag{2.33}$$

where $h = \{Af, Nf, An, Nn, W, C\}$ denotes households.

Cash crop

Cash crop is produced in both rural regions and the domestic production is fully exported.

$$Q_{Cf} = X_{Cf} \tag{2.34}$$

$$Q_{Cn} = X_{Cn} \tag{2.35}$$

Processed food

Processed food is produced domestically in the urban region and is imported from the rest of the world. Processed food is consumed in each of the three domestic regions.

$$Q_{Pu} = \sum_h c^h_{Pu} \tag{2.36}$$

$$D_{Pw} = \sum_h c^h_{Pw} \tag{2.37}$$

where $h = \{Af, Nf, An, Nn, W, C\}$ denotes households.

Manufactured good

The manufactured good is produced domestically in the urban region and is imported from the rest of the world. Manufactures are consumed in each of the three domestic regions, used as an intermediate input in production of processed food, and used in the production of capital.

$$Q_{Mu} = \sum_h c^h_{Mu} + Mu_P \tag{2.38}$$

$$D_{Mw} = \sum_h c^h_{Mw} + Mw_P + I_M + I_P \tag{2.39}$$

where $h = \{Af, Nf, An, Nn, W, C\}$ denotes households.

Fertiliser

There is no domestic production of fertiliser, such all fertiliser that is used is fully imported from the rest of the world. It is used in the production of both agricultural crops (staple, cash) in both rural regions (far, near)

$$D_Z = \sum_j Z_j \tag{2.40}$$

where $j = \{Sf, Cf, Sn, Cn\}$ denotes agricultural sectors.

Fuel

$$\begin{aligned} D_O =& fuel_1 \Big[\sum_{Nn,An} \left(c^h_{Pu} + c^h_{Pw} + c^h_{Mu} + c^h_{Mw} \right) + c^W_{Sn} + c^C_{Sn} + Z_{Sn} + Z_{Cn} + Q_{Cn} \Big] \\ &+ fuel_2 \Big[c^{Nf}_{Sn} + c^{Af}_{Sn} + c^{Nn}_{Sf} + c^{An}_{Sf} \Big] \\ &+ \left(fuel_1 + fuel_2 \right) \Big[\sum_{Nf,Af} \left(c^h_{Pu} + c^h_{Pw} + c^h_{Mu} + c^h_{Mw} \right) + c^W_{Sf} + c^C_{Sf} \\ &+ Z_{Sf} + Z_{Cf} + Q_{Cf} \Big] \end{aligned} \tag{2.41}$$

2.B.4 Factor market clearing

Labour

Labour markets are integrated within region, resulting in three separate labour markets in the economy. Within a given labour market, labour moves frictionlessly between sectors. Labour in the far rural region (f) can be allocated to staple crop farming (S), cash crop farming (C) or the non-agricultural sector (N).

$$\bar{L}_f = L_{Sf} + L_{Cf} + L_{Nf} \tag{2.42}$$

Similarly, labour in the near rural region (n) can be allocated to staple crop farming (S), cash crop farming (C) or the non-tradable services sector (N).

$$\bar{L}_n = L_{Sn} + L_{Cn} + L_{Nn} \tag{2.43}$$

In the urban region (u), labour can be allocated to food processing (P), manufacturing (M), or the non-tradable services sector (N)

$$\bar{L}_u = L_{Pu} + L_{Mu} + L_{Nu} \tag{2.44}$$

Land

Agricultural land in the far rural region (f) can be allocated to staple crop farming (S) or cash crop farming (C)

$$A_f = A_{Sf} + A_{Cf} \tag{2.45}$$

Similarly, agricultural land in the near rural region (n) can be allocated to staple crop farming (S) or cash crop farming (C)

$$A_n = A_{Sn} + A_{Cn} \tag{2.46}$$

Capital

Capital is only used in production processes in the urban region and can be allocated to the food processing or the manufacturing sector

$$\bar{K} = K_P + K_M \tag{2.47}$$

2.B.5 Current account

$$\begin{aligned} p_{Cw}\left(Q_{Cf} + Q_{Cn}\right) + Aid = & \, p_{Mw}\left(\sum_h c^h_{Mw} + Mw_P\right) + p_{Pw}\left(\sum_h c^h_{Pw}\right) \\ & + p_{Zw}\left(\sum_i Z_i\right) + p_{Ow}\, fuel \\ & + p_K\left(I_{Pu} + I_{Mu}\right) + r^d D + R + \epsilon_{BOP} \end{aligned} \tag{2.48}$$

where p^y_{it} denotes the consumption price of good i in destination y in period t, p^x_{it} denotes the factory/farm-gate price of good i at origin x in period t, $h = \{Af, Nf, An, Nn, W, C\}$ denotes households, and $i = \{Sf, Sn, Pu, Pw, Mu, Mw, Nf, Nn, Nu\}$ denotes consumption goods.

2.B.6 Fiscal adjustment

$$
\begin{aligned}
r^d D + r^g\, P_C\, B + s_Z\, E\, p_{Zw}\, D_Z + s_O\, E\, p_{Ow}\, fuel + T = Aid &+ f_A \sum_i \Big(w_i\, L_i + b_i\, A_i\Big) \\
&+ f_N \sum_j \Big(w_j\, L_j\Big) + f_W \sum_k \Big(w_k\, L_k\Big) \\
&+ f_C \Big(p_K\, r_u\, (K_{Pu} + K_{Mu}) + rents\Big) \\
&+ t_M\, E\, p_{Mw}\, D_M + t_P\, E\, p_{Pw}\, D_P \\
&+ t_C\, E\, p_{Cw} \Big(Q_{Cf} + Q_{Cn}\Big)
\end{aligned} \tag{2.49}
$$

where $i = \{Sf, Cf, Sn, Cn\}$, $j = \{Nf, Nn\}$, $k = \{Pu, Mu, Nu\}$, $l = \{Pu, Mu\}$. Transport rents for a given trade flow are calculated as the value of the trade flow at origin, multiplied by the wedge parameter (ψ^{rent}).

2.B.7 Domestic prices

Table B1: Local prices

Product	**Origin**		**Local prices at destination**			
			Far rural	**Near rural**	**Urban**	**RoW**
Non-tradables	Far	Nf	p^f_{Nf}	-	-	-
	Near	Nn	-	p^n_{Nn}	-	-
	Urban	Nu	-	-	p^u_{Nu}	-
Staple crop	Far	Sf	p^f_{Sf}	$p^n_{Sf} = p^f_{Sf}(1+\psi^{rent}_2) + \psi^{fuel}_2$	$p^u_{Sf} = p^f_{Sf}(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	-
	Near	Sn	$p^f_{Sn} = p^n_{Sn}(1+\psi^{rent}_2) + \psi^{fuel}_2$	p^n_{Sn}	$p^u_{Sn} = p^n_{Sn}(1+\psi^{rent}_1) + \psi^{fuel}_1$	-
Cash crop	Far	Cf	$p^f_{Cf} = \frac{Ep^w_C - \psi^{fuel}_1 - \psi^{fuel}_2}{(1+\psi^{rent}_1+\psi^{rent}_2)(1+t_C)}$	-	-	Ep^w_C
	Near	Cn	-	$p^n_{Cn} = \frac{Ep^w_C - \psi^{fuel}_1}{(1+\psi^{rent}_1)(1+t_C)}$	-	Ep^w_C
Manufactures	Urban	Mu	$p^f_{Mu} = p^u_{Mu}(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	$p^n_{Mu} = p^u_{Mu}(1+\psi^{rent}_1) + \psi^{fuel}_1$	p^u_{Mu}	-
	RoW	Mw	$p^f_{Mw} = (1+t_M)Ep^w_M(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	$p^n_{Mw} = (1+t_M)Ep^w_M(1+\psi^{rent}_1) + \psi^{fuel}_1$	$p^u_{Mw} = (1+t_M)Ep^w_M$	Ep^w_M
Processed food	Urban	Pu	$p^f_{Pu} = p^u_{Pu}(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	$p^n_{Pu} = p^u_{Pu}(1+\psi^{rent}_1) + \psi^{fuel}_1$	p^u_{Pu}	-
	RoW	Pw	$p^f_{Pw} = (1+t_P)Ep^w_P(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	$p^n_{Pw} = (1+t_P)Ep^w_P(1+\psi^{rent}_1) + \psi^{fuel}_1$	$p^u_{Pw} = (1+t_P)Ep^w_P$	Ep^w_M
Fertiliser	RoW	Z	$p^f_Z = (1+t_Z)Ep^w_Z(\psi^{rent}_1 + \psi^{rent}_2) + \psi^{fuel}_1 + \psi^{fuel}_2$	$p^n_Z = (1+t_Z)Ep^w_Z(1+\psi^{rent}_1) + \psi^{fuel}_1$	-	Ep^w_Z
Fuel	RoW	O	-	-	-	Ep^w_O

Appendix 2.C Calibration appendix

2.C.1 Introduction

The basis for calibration of the model is the 2018 Social Accounting Matrix (SAM) for Tanzania compiled by IFPRI, the International Food Policy Research Institute (Thurlow, 2021). This is a *national* SAM that distinguishes 42 activities and commodities and 10 households, consisting of five rural households, ordered by income quintile, and five similarly ordered urban households. Factors of production consist of three classes of labour, differentiated by level of educational achievement; land, which is used only in agricultural production; and capital, which in employed in all sectors. The SAM is defined in value terms and traces flows of spending across the economy. Household spending on commodities accrues to producers, net of taxes, who in turn purchase intermediate inputs from other firms and the rest of the world and employ factors of production owned by households. Some share of domestic production is exported and some proportion of final expenditures on consumption goods, intermediate inputs and capital goods, is imported. A government account tracks revenue mobilisation through direct, indirect and trade taxes, and spending on goods and services, debt service and on the accumulation and maintenance of infrastructure capital. Net flows of saving by households, government and the rest of the world constitute the residual vector of flows that balance the SAM for each activity and commodity, for each household, for the government account and the aggregate external balance of payments.

Transforming this SAM into a form that is consistent with the specific structure of production and consumption used in the model, including their spatial distribution, requires a set of aggregation and allocation choices. These are implemented on the basis of data from other sources and are ultimately disciplined by the constraints of our model structure and by our own judgements, subject to the objective that the overall control totals in the SAM, including for the overall level and structure of GDP, are preserved as best as possible.

2.C.2 Initial aggregation over commodities and households

The first step involves aggregating from the 42 SAM commodities to the six model commodities: staple food; cash crops; processed food; manufactured goods; non-tradable goods and services; and natural resources and tourism. The allocation between staples, cash crops and processed foods follows directly from 1.B.2. Manufactured goods consist of all remaining tangible commodities in the SAM,[28] while the non-tradable sector is a simple aggregation of all the remaining activities in the SAM except for mining and hotels which are aggregated into the natural resources and tourism sector. This final sector, which has a high net export content, is treated in the model as an enclave sector employing no domestic factors of production and exporting its entire output: it plays no substantive role in our simulation analysis but its existence helps preserve the coherence of the model with the underlying data.

This first aggregation over commodities and activities is done at the national level. Likewise, on the household side, we initially define a single rural household as the aggregate of all five rural quintiles and likewise a single urban household. Flows of labour income are aggregated accordingly by sector and household type. All land is deemed to be owned by the rural household who receive total rental incomes. Gross returns to capital accrue to a 'capitalist' household. This is not separately defined in the IFPRI SAM: we describe the construction of this household account below. Finally, at this initial stage, we make a number of further alterations and re-allocations to the SAM to map the data to our model structure. These are as follows:

- All elements in the input-output structure other than intermediate inputs of staple food and manufactured goods into processed food production are set to zero. This necessarily changes the value of gross output and gross expenditure in each sector; the row-column balance in the SAM is restored by absorbing the net expenditure by sector in the final consumption vector.

[28] These are textiles, wood, chemicals, metal and non-metal goods, machinery and other manufactures.

- All investment spending by sector of origin, including via change in stocks, is reallocated to the manufactured goods sector (this being the only input into investment spending in our model).
- Household-specific savings and investment spending are allocated to the capitalist household, the only non hand-to-mouth household in the model (see below).
- All exports of staples, cash crops and processed foods are allocated to cash crops (and will later be disaggregated by location). All other exports are treated as elements of the non-modeled 'natural resource and tourism' export-only sector.
- All imports of staples, cash crops and processed foods are reallocated to processed food imports. Mining and tourism imports are netted off against exports from this sector and all other imports are treated as imports of manufactures. Residual imports of labour, labour services and final direct imports by households are all set to zero.

Taken as a whole, this set of adjustments to the SAM, which are necessary to provide for a model-consistent calibration, would appear to be radical. In practice, however, with the exception of our suppression of the majority of the input-output structure, most of these changes are marginal and do not radically alter the overall structure of production and consumption. For example, it is the case that the majority of food exports are of cash crops and the majority of non-resource, non-agricultural exports are of manufactured goods. Likewise, most imports are of processed foods and manufactured goods. Similarly, we regard it as a reasonable approximation that most households in Tanzania are hand-to-mouth and do not save and investment (other than, of course, enjoying title to rural land).

2.C.3 Spatial distribution of activity

The preceding adjustments serve to re-shape the aggregate *national* SAM. The next step involves using geo-referenced microeconomic survey data from the World

Bank's Living Standards Measurement Study (LSMS) to inform the spatial allocations of land and labour and hence of employment and consumption across locations in the economy. As noted in 1.B.3 we assume three regions: far, near and urban. Staple and cash crops are produced using land and labour in both the far and near regions while processed food and other manufactured goods are produced using capital and labour in the urban region. We assume region-specific non-tradable goods are produced using only local labour in each of the three regions. Each region is populated by two households each: in far and near these are an agricultural household and a non-agricultural household while in the urban region there is a wage-earning and a capitalist household. To effect the spatial allocation of production and consumption across these three locations, the following adjustments are made to the SAM:

- We draw on five waves of Tanzanian LSMS data from 2008-2019 (National Panel Survey data collection waves 1-4, plus the extended panel collected in 2019), to estimate the spatial distributions of land and labour across sectors. These are shown in Table C2. The same LSMS data also generates data on the average intensity of fertiliser use between staple and cash crops across regions. Land rental rates are not fully equalised across crops and locations: we assume for convenience that rental in far are similar but that rental rates are higher in near and higher for cash crops in near than staple crops.

- The model assumes that labour markets clear by location hence equalising wages across locations. By dint of allocation of labour and production of staples, cash crops and non-tradable goods and services between far and near locations, this implies a baseline wage premium of approximately 47% accruing to those employed in the near region relative to the far region.

- The foregoing aggregations and sectoral allocation of production, trade and factors, combined with the model assumption that all the non-capitalist households consume their entire income period-by-period, the assumptions on public finance, and the treatment of transport costs, imply a residual matrix of final consumption. Relative to the SAM, our constructed capitalist household, discussed below, gives us an additional degree of freedom which allows for small re-allocations of final consumption between households to ensure the final consumption vector is consistent with evidence on household expenditures in low-income countries while respecting the balance requirements of the SAM.[29]

- Transport costs consist of two elements, the first being fuel costs and the second rents accruing to the capitalist household as the owner of the transport and distribution system. The IFPRI SAM includes a separate transport cost wedge. Using earlier work by two of the authors (Adam et al., 2018) and data on total fuel imports from (IMF, 2022) this transport cost wedge is partitioned into a fuel-cost component (which forms part of imports) and a rental component (accruing to the capitalist household) and allocated across final consumption and intermediate goods in terms of their location of consumption.

- The final household consumption vectors at producer and consumer prices are shown in Table C1.

2.C.4 The capitalist household

The capitalist household does not appear in the IFPRI SAM but rather is created to absorb items of income and expenditure in the national economy that cannot be booked elsewhere, given the restrictions of our model structure. Hence this household is the sole owner of capital and is thus the sole recipient of gross profits. Likewise, it owns the transport and distribution sector and therefore receives the rents accruing from the movement of goods from producers to consumers in different locations. Finally, as the only saving household in the model, all private savings are booked though the capitalist household which then allocates these to the accumulation of indexed domestic government debt and/or investment in tangible capital in the urban food processing and manufactured goods sectors. The capitalist household account is balanced by a residual 'net remittances' flow to the rest of the world.

[29]The final income-expenditure balance for the non-capitalist households is guaranteed by deriving net household transfers from government.

Table C1: Final consumption vector

	Far Agric.	Far Non-ag.	Near Agric.	Near Non-ag.	Urban Worker	Urban Capital
Real consumption (per capita)						
Staple food – Far	0.35	0.24	0.13	0.06	0.10	1.22
Staple food – Near	0.04	0.05	0.32	0.16	0.13	0.36
Processed food – Urban	0.06	0.04	0.08	0.07	0.26	0.62
Processed food – Imports	0.05	0.04	0.09	0.07	0.28	0.66
Manufactured good – Urban	0.06	0.05	0.08	0.04	0.18	0.71
Manufactured good – Imports	0.11	0.12	0.16	0.12	0.37	2.04
Non-tradables – Far	0.06	0.13	–	–	–	–
Non-tradables – Near	–	–	0.14	0.07	–	–
Non-tradables – Urban	–	–	–	–	0.39	2.95
Consumer Prices						
Staple food – Far	1.00	1.00	1.53	1.53	2.05	2.05
Staple food – Near	1.53	1.53	1.00	1.00	1.53	1.53
Processed food – Urban	3.21	3.21	2.66	2.66	2.10	2.10
Processed food – Imports	1.81	1.81	1.46	1.46	1.10	1.10
Manufactured good – Urban	1.48	1.48	1.24	1.24	1.00	1.00
Manufactured good – Imports	1.61	1.61	1.36	1.36	1.10	1.10
Non-tradables – Far	1.00	1.00	–	–	–	–
Non-tradables – Near	–	–	1.00	1.00	–	–
Non-tradables – Urban	–	–	–	–	1.00	1.00
Consumption expenditure shares						
Staple food – Far	35%	26.5%	14.7%	11.8%	9.3%	22.7%
Staple food – Near	5.5%	8.3%	24.1%	19.7%	9.1%	5.1%
Processed food – Urban	17.7%	13.9%	16.9%	22.4%	24.8%	11.9%
Processed food – Imports	8.4%	8.1%	10%	12.3%	13.7%	6.6%
Manufactured good – Urban	9.2%	8.3%	7.4%	5.2%	7.9%	6.4%
Manufactured good – Imports	18.1%	20.4%	16.3%	20%	18%	20.4%
Non-tradables – Far	6.1%	14.5%	–	–	–	–
Non-tradables – Near	–	–	10.6%	8.5%	–	–
Non-tradables – Urban	–	–	–	–	17.3%	26.9%

Note: This table shows the final consumption vector at baseline. Columns refer to households (Agricultural – Far, Non-agricultural – Far, Agricultural – Near, Non-agricultural – Near, Wage-earning – Urban, Capitalist – Urban) and rows refer to goods (type of good and source).

2.C.5 The final spatial SAM

The final step consists in transforming the original IFPRI SAM fiscal and external balance accounts to a model-consistent form. Most of the required adjustments follow directly from the transformations to the household production and consumption accounts described above, but some final adjustments to tax revenues, expenditures and external balancing items are implemented to ensure a broad alignment with the fiscal and external balance accounts reported for 2018 by the IMF (see IMF, 2022). The resulting spatial structure of the model economy and factor intensity of production at baseline are summarised in Table 2.2, and Table 2.3 reports the associated fiscal and external balance at baseline.

Table C2: Factor allocation at baseline

Region	Sector	Labour	Land	Capital
Far rural	Staple crop	42%	59%	-
	Cash crop	5%	19%	-
	Non-tradables	9%	-	-
Near rural	Staple crop	16%	19%	-
	Cash crop	2%	4%	-
	Non-tradables	3%	-	-
Urban	Processed food	6%	-	46%
	Manufacturing	4%	-	54%
	Non-tradables	13%	-	-

Note: This table shows the relative allocation of labour, land, and capital across the nine spatially differentiated sectors of production at baseline.

Appendix 2.D Comparative statics appendix

Table D1: Joint and isolated price shocks (% change over baseline)

	(1) No mitigation Joint shock		**(2) No mitigation Fuel only**		**(3) No mitigation Fertiliser only**		**(4) No mitigation Food only**	
	Q %	V %	Q %	V %	Q %	V %	Q %	V %
Fiscal aggregates								
Total Revenue	–	-9.0	–	-8.6	–	-0.7	–	-0.1
Tariffs	–	-9.4	–	-9.2	–	-0.3	–	-0.3
Export duties	–	-6.3	–	-3.4	–	-2.6	–	-1.9
Direct Taxes	–	-9.2	–	-8.9	–	-0.6	–	0.0
Total Expenditure	–	-8.4	–	-8.0	–	-0.6	–	-0.1
Subsidies	–	9.8	–	11.5	–	-1.4	–	-0.9
Transfers	–	-10.8	–	-10.3	–	-0.6	–	-0.2
Debt Service	–	-0.4	–	-1.7	–	0.1	–	1.0
Deficit	–	0.0	–	0.0	–	0.0	–	0.0
Aid	–	0.0	–	0.0	–	0.0	–	0.0
Trade								
Total Exports	–	-6.3	–	-3.4	–	-2.6	–	-1.9
Cash crop	-6.3	-6.3	-3.4	-3.4	-2.6	-2.6	-1.9	-1.9
Total Imports	–	-1.7	–	-0.9	–	-0.7	–	-0.5
Processed food	-25.5	-10.5	-7.0	-7.0	-0.3	-0.3	-19.9	-3.9
Manufactures [a]	-9.3	-9.3	-10.7	-10.7	-0.2	-0.2	1.3	1.3
Manufactures [b]	-0.8	-0.8	-4.8	-4.8	-0.6	-0.6	5.1	5.1
Manufactures [c]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fertiliser	-29.5	-15.4	-16.8	-16.8	-16.7	-0.1	-0.1	-0.1
Fuel	-6.3	12.4	-4.6	14.5	-1.6	-1.6	-1.0	-1.0
Production								
Aggregate	–	-6.8	–	-8.1	–	0.1	–	0.9
Cash crop	-6.3	-21.7	-3.4	-19.5	-2.6	-2.5	-1.9	-1.6
Staple crop	-1.6	-6.1	-1.1	-8.4	-0.8	1.3	0.6	0.4
Food processing	1.9	-4.8	-0.6	-7.1	-0.4	-1.8	3.4	4.5
Manufacturing	-3.1	-7.2	-1.3	-7.8	0.2	-0.1	-2.3	0.4
Non-tradables	0.1	-5.5	0.9	-5.9	-0.1	-0.4	-0.6	0.6
Final consumption								
Aggregate	–	-8.2	–	-8.2	–	-0.4	–	0.1
Far-Agricultural	-10.5	-11.5	-9.0	-11.8	-0.5	0.1	-1.2	-0.3
Far-Non-agricult.	-12.6	-14.8	-12.3	-15.5	0.0	0.4	-0.8	-0.4
Near-Agricultural	-12.2	-12.7	-12.1	-14.7	0.6	1.1	-1.6	0.0
Near-Non-agricult.	-13.4	-12.7	-13.0	-14.6	0.8	1.0	-2.3	0.1
Urban-Wage-earning	-4.6	-4.4	-1.6	-4.8	-0.4	-0.5	-2.6	0.8
Urban-Capitalist	-7.7	-8.4	-4.0	-6.9	-1.4	-1.2	-2.2	-0.5

Note: This table shows quantity (Q%) and value (V%) effects of joint and isolated price shocks without policy mitigation. (1) Joint 20% price shock for fuel, fertiliser, and food; (2) Isolated 20% fuel price shock; (3) Isolated 20% fertiliser price shock; (4) Isolated 20% food price shock. [a] Final consumption. [b] Intermediates in food processing. [c] Capital formation. Consumption is adjusted for household size (consumption per capita).

CHAPTER 3

Predicting Household Food Insecurity From Readily Available Data

Abstract Accurately identifying food insecure households is essential for effective targeting of policy interventions and humanitarian aid. However, food insecurity is a multi-dimensional phenomenon that is difficult to measure. Numerous different indicators exist, all of which require expensive and time-consuming data collection through household surveys. This paper addresses the challenge of measuring food insecurity in two ways. First, I use detailed survey data that allows to compare different food security indicators for the same set of households. This analysis confirms previous findings in the literature that different food security indicators are only weakly correlated and appear to capture largely independent phenomena. Second, I draw on the same detailed household survey data and additional publicly available data on weather, prices, and geography to train and test machine learning models for the prediction of households' food security status. The models achieve overall accuracy of 54-65% at a true positive rate of 70%. This approach provides a way to assess household food security without the need for costly and time-consuming primary data collection.

3.1 Introduction

Food insecurity remains an important issue in many regions of the world. The number of people affected by hunger has been increasing again since 2014, particularly in Sub-Saharan Africa. This deterioration sped up and intensified in recent years as a result of climate change effects, the COVID-19 pandemic, and Russia's invasion of Ukraine. These simultaneous adverse developments have shown that, in addition to local armed conflict, other factors such as extreme weather events, supply chain fragility, and excessive import dependence pose a threat to food security in many low- and middle-income countries (Brown & Funk, 2008; Hakobyan et al., 2022; Swinnen & McDermott, 2020; Verschuur et al., 2021).

Accurately identifying currently and future food insecure households is crucial for effective targeting of humanitarian aid during acute local crises and long-term policy interventions aimed at reducing chronic or recurring food insecurity. However, food security cannot be observed directly, and the various experience-based, consumption-based, monetary, and biological indicators all capture different dimensions. Discordance between these indicators raises questions about their suitability and reliability for both current assessments as well as for forecasting future food security states. Further, the collection of primary data through household surveys, necessary for constructing these indicators, can be expensive and time-consuming.

This paper addresses the challenge of measuring food insecurity in two ways. First, I use detailed survey data that allows to compute and compare different food security indicators for the same set of households. The analysis confirms previous findings in the literature that different food security indicators are only weakly correlated and appear to capture largely independent phenomena (Vaitla et al., 2017). Second, I draw on the same detailed household survey data and additional publicly available data on weather, prices, and geography to predict households' current food security status. This methodology provides a way to assess household food security without the need for costly and time-consuming primary data collection.

Food security prediction can refer to two distinct tasks: (i) inferring current food security using contemporaneous data (to improve targeting while obviating the need for primary data collection), or (ii) forecasting future food security states (e.g. for famine early warning). Forecasting introduces additional complexities, often requiring prediction of future driver states, such as future agricultural production conditions and future market prices. This paper focuses on methods for contemporaneous prediction to that can aid policymakers and humanitarian actors in rapid situation assessments when direct food security measurements are unavailable. In that, the approach is similar in spirit to other studies that use machine learning methods to bridge data gaps in low-income countries for improved policy targeting (Aiken et al., 2022; Blumenstock et al., 2015; Jean et al., 2016). This paper relates closely to recent studies that employ machine learning models and secondary data to predict the prevalence of food insecurity at various geographical scales (Lentz et al., 2019; Martini et al., 2022; Zhou et al., 2021).

To facilitate prediction, I draw on two waves of data from the Living Standards Measurement Study (LSMS) for Tanzania, and train machine learning models to detect patterns in the data that imply a high probability of food insecurity. I focus on two different food security indicators: the Food Consumption Score (FCS) is a consumption-based indicator that reflects dietary diversity, while the reduced Coping Strategies Index (rCSI) is an experience-based measure of food security that captures the cost of constrained access to food (Vaitla et al., 2017). I predict households' FCS and rCSI status from data that is publicly available or easily accessible, such as domestic and international food prices, climatic conditions, geographical variables, and basic indicators of household assets. Motivated by my previous work on the role of domestic trade frictions for food consumption patterns (Chapter 1) and spatially differentiated pass-through of price shocks (Chapter 2), I also include predictors that reflect the remoteness of households' locations. These remoteness indicators, such as distance to the nearest major road and distance to the nearest agricultural market, capture households' market access and may mediate the effect of market price shocks or local production shocks on food security outcomes. In addition to weather, prices, and geographic variables, previous research has highlighted the importance of including some household-level information in order to achieve sufficient accuracy of predictions (Lentz et al., 2019; Zhou et al., 2021). Following these previous studies, I include roof type and cellphone ownership as straightforward asset indicators that can potentially be accessed without primary data collection, through remote sensing and through cellphone companies' records. I also include basic household demographics (household size and household head characteristics) to check how much these can improve prediction accuracy. The main models are specified without household demographics, as analysis shows that such characteristics add little predictive power and I prioritise ease of data access for predictors.

Since errors of exclusion (failing to detect food insecure households) can be thought to weigh heavier than errors of inclusion, I set the recall rate to 70% across all models and compare resulting accuracy levels.[1] This implies that the models correctly identify 7 out of 10 food insecure households, but such high recall comes at the cost lower overall accuracy due to associated inclusion errors. At this recall level, models predicting FCS achieve higher overall accuracy than models predicting rCSI (65% vs. 54%). I compare two approaches to predicting whether households are food secure or insecure. First, I map FCS and rCSI values to food security status using threshold values before prediction, and apply binary classification algorithms. Second, I predict raw FSC and rCSI values using regression algorithms first and only subsequently infer binary food security status using threshold values on the predicted values. For FCS, accuracy is slightly higher using the second, more granular, approach (65% vs. 63%) while for rCSI both methods yield equivalent accuracy (54%). Comparison of models with varying predictor sets reveals the importance of household assets for predictive performance. Household demographics add very little to model performance for FCS, but appear slightly more important for the prediction of rCSI. The remoteness indicators display only limited predictive power, but given their ease of accessibility (compared to e.g. household demographics) I still include them as predictors in the final model.

The remainder of this paper is structured as follows: Section 3.2 conceptualises food security, explains the challenges involved with measurement and prediction, and summarises recent advances in the literature. Section 3.3 presents different

[1]The performance of binary prediction models can be evaluated using out-of-sample accuracy, precision, and recall. Accuracy refers to the overall proportion of correctly classified households. Precision measures the proportion of predicted insecure households that are truly insecure. Recall, conversely, measures the proportion of truly insecure households correctly identified as such by the model.

types of food security indicators and details on the construction of the indicators used in this study. Section 3.4 characterises food insecurity in Tanzania, comparing different indicators, mapping food security over time and space, and assessing correlations with potential predictors. Section 3.5 describes the machine learning methods and presents prediction results. Section 3.6 concludes.

3.2 The challenge of predicting food insecurity

This paper links to a large body of literature on food security and famine (Barrett, 2002; Ravallion, 1997; Sen, 1980, 1983). Previous work has highlighted the complex multidimensionality of food security and the challenges this implies for measurement. Food insecurity encompasses not only realised macro- or micronutrient deficiency but also risk and instability over time. As such, food security is inherently unobservable and cannot be measured directly. Observable nutritional phenomena such as hunger and malnutrition are sufficient but not necessary conditions for food insecurity. The most widely adopted definition, produced at the 1996 World Food Summit in Rome, describes food security as a situation in which "all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life" (United Nations, 1996).

There has been considerable debate about the causes of food insecurity. In two seminal works, Sen (1980, 1983) rejects the previously dominant food availability approach[2] and instead advocates for what he terms the *entitlement approach*. Sen stresses that food availability decline (FAD) is neither necessary nor sufficient for famine to occur. Instead, he views starvation as a result of *entitlement failure* – namely, people losing command over sufficient amounts of food because of an adverse shock to either their endowments or terms of trade. Sen's entitlement framework does not contradict, but rather nests, the FAD case in which famine is

[2]The food availability approach was initially a corollary of the Malthusian view that famines are nature's way of keeping population growth in check. A number of studies have since found this view to be inadequate (Caldwell and Caldwell (1992), Fogel (1992), Watkins and Menken (1985) as cited in Ravallion (1997)).

caused by a negative shock to food supply in response to e.g. crop failure.[3]

In line with Sen's argument that food security cannot be reduced to food availability, policy institutions nowadays assess food security along four dimensions which, in addition to *availability*, include *access*, *utilisation* and *stability* (FAO, 2006). The *access* dimension of food security corresponds closely to Sen's concept of entitlements and encompasses both physical and economic access to food, as highlighted by the UN's definition. *Utilisation* refers to all factors which jointly determine an individual's nutritional status, such as clean water and sanitation, without which an individual can still be left food insecure, even if food is available and accessible. The fourth dimension, *stability*, was added to stress that food security implies stable access to food over time, not just in a given period.

A large body of literature discusses the challenge of reliably measuring food security (Barrett, 2010; Vaitla et al., 2015). Proxy measures of food security span a range of different dimensions, including biological indicators such as stunting and wasting, consumption-based indicators such as measures of dietary diversity and average caloric intake, experience-based indicators that reflect coping strategies and psychological stress, and monetary indicators such as food price indices and food expenditure shares. Due to the difficulties in measurement and the fact that each of these indicators captures a slightly different dimension of food security, there are usually significant discrepancies between food security figures generated by different institutions. This is true even when food security is measured for the same population, for the same time period, and at the same level of aggregation (individual, household, community, country).

A few studies systematically examine the relationship between different food security indicators (Manikas et al., 2023; Vaitla et al., 2015, 2017). Vaitla et al. (2017) draw on 21 survey datasets from ten countries to conduct factor analysis for four different food security indicators. They find that different indicators seem to capture two quite separate food security phenomena. Of the four widely used indicators included in their analysis, the two experience-based indicators (HHS and rCSI) reflect a different dimension than the two consumption-based indicators

[3] Ravallion (1997) identifies the reliance on a functioning legal system as one of the main shortcomings of Sen's entitlement approach to famine. In fact, a number of scholars attribute the majority of present-day famines to violent conflict, which oftentimes entails a collapse of the legal system (De Waal, 2018; Maxwell, 2019).

(HDDS and FCS). The authors conclude that these two separate dimensions may reflect either food quantity vs. food quality (as hypothesised in earlier work by the authors, Vaitla et al. (2015)), or a difference in the cost of access vs. realised access. The latter interpretation implies that experienced-based indicators are sensitive to stressed food access, such as reducing portion size or borrowing food, which may not be reflected in consumption-based indicators that measure dietary diversity. Vaitla et al. (2017) find strong discordance between the severity categories of different food security indicators. While one indicator may classify a household as food secure, another indicator may classify the same household as severely food insecure. Such discordance is even larger for forecasts of future food security states than it is for figures that are based on past observational data.

While highly policy-relevant, forecasting of future food insecurity crises is more challenging still than the measurement of current food security states. There are a number of institutional forecasting initiatives, one of the most prominent being the USAID-funded Famine Early Warning Systems Network (FEWS NET).[4] Willenbockel and Robinson (2014) provide an overview of different CGE and PE models that are used to forecast food security, often with a particular focus on agricultural production and food availability rather than issues around distribution and access. There are a number of models and analyses that focus on very long-run food insecurity and climate change, forecasting food insecurity at the country level in 2050 or even 2080. For example, building on the methodology for constructing a food insecurity vulnerability index developed by Krishnamurthy et al. (2014), a joint initiative by the Met Office and the World Food Program projects country-level food insecurity under different climate change scenarios (MetOffice, 2015).

A relatively new but fast growing strand of academic literature has focused on using machine learning techniques to forecast food security at a more granular spatial scale and for a shorter time horizon than CGE and PE models (Andrée et al., 2020; Balashankar et al., 2023; Gholami et al., 2022). Other studies focus on nowcasting rather than forecasting, i.e. predicting contemporaneous food security from secondary data (Deléglise et al., 2022; Lentz et al., 2019; Martini et al.,

[4]Other institutional entities engaged in food security forecasting include the WFP's Vulnerability Assessment and Mapping Unit (VAM), the World Bank's Famine Early Action Mechanism (FAM), the Food Security Cluster (FSC) and Global Nutrition Cluster (GNC), and the Modelling Early Risk Indicators to Anticipate Malnutrition (MERIAM) project. See Maxwell (2019) for a brief overview.

2022; Villacis et al., 2023; Zhou et al., 2021). While most of these studies use a combination of weather, geographic, price, and (remotely sensed) household asset data for prediction, Balashankar et al. (2023) instead use news streams to forecast food insecurity. Employing relatively simple linear prediction techniques, Lentz et al. (2019) show that statistical methods can significantly improve on existing, often scenario-based, institutional forecasting approaches which could be used to inform humanitarian action and policy making. In a more recent study, the authors show that random forest techniques, which are better suited to capture non-linear relationships and complex interaction effects, can significantly improve on linear models (Zhou et al., 2021).

In addition to contributing to the literature on measuring and predicting food security, this paper also relates to the literature on the relationship between food security and market integration (Burgess & Donaldson, 2012; Dithmer & Abdulai, 2017; Gao & Lei, 2021; Janssens et al., 2020). The relationship between market integration and food security in a given location is theoretically ambiguous. Consider a hypothetical location which is nearly autarkic and in which agriculture is primarily rain-fed. In such a setting, local *real* incomes, which are a crucial determinant of food security, are highly sensitive to local productivity shocks such as extreme weather events. As trade costs fall and a location's openness to trade increases, local prices become less responsive to local productivity shocks. At the same time, local *nominal* incomes from agriculture become more responsive to local productivity shocks, since trade openness reduces the scope for offsetting price movements. The effect of local productivity shocks on *real* incomes, which determine food consumption, is thus theoretically ambiguous. There exists little quantitative empirical research on the role of market integration for the incidence of acute food insecurity. A notable exception is a study by Burgess and Donaldson (2012), who empirically investigate whether improvements in transportation infrastructure have an effect on the responsiveness of real incomes to local productivity shocks. They find that in colonial India, where agriculture was primarily rain-fed and famines were frequent and widespread, railroads reduced real income volatility and thus the incidence of famine. To account for this relationship, I include different variables on households' remoteness in the predictor set to proxy for market integration.

3.3 Measuring food insecurity

Food security is a multidimensional phenomenon that can be difficult to measure accurately and comprehensively. There exists no single indicator that is suitable for all contexts and purposes. No single indicator has sufficient validity, reliability, and comparability across time and space to reflect all four dimensions of food insecurity including availability, access, utilisation, and stability (Maxwell et al., 2013). A multitude of different indicators are used to summarise food security at the level of individuals, households, communities, and countries (Barrett, 2010). The most widely used indicators are experience-based, consumption-based, or monetary indicators which capture economic access to food. In addition, biological measures (anthropometric and medical) are used to measure longer term under- and malnutrition. All available indicators have different strengths and shortcomings. For instance, caloric intake, a widely used measure of current access to food, does not reflect nutritional quality, coping strategies employed to realise the observed level of consumption, or stability of food intake over time. Experience-based measures, on the other hand, can be context-specific and pick up a more subjective aspect of food security. Monetary indicators, such as food prices and expenditure shares, are only indirect measures of food security and do not constitute direct goals for policy-makers and international humanitarian organisations. In addition to measuring somewhat different aspects of food security, different indicators also vary widely in terms of the cost of data collection. Experience-based indicators, which can be constructed from a questionnaire that can be administered remotely, are more cost-effective than detailed nutritional indicators or anthropometric measures. The remainder of this section first discusses different types of indicators which measure food security at the level of the household, and subsequently presents distributions and correlations for selected indicators using the Tanzanian LSMS dataset.

3.3.1 Experience-based measures

Experience-based measures reflect households' self-reported experience of food security. Examples of this type of measure are the (reduced) Coping Strategy Index (rCSI), the Household Food Insecurity Access Scale (HFIAS), the Food Insecurity Experience Scale (FIES), and the Household Hunger Scale (HHS), which are widely used by international organisations such as USAID, FAO, and WFP. The constituent items of these indicators include psychological questions, such as feeling worried about having enough food, self-reported consumption behavior, and coping strategies that are indicative of stresses around food security which may not be picked up by consumption-based measures. The resulting overall indicators summarise the frequency of occurrence of each strategy employed and each situation experienced, weighted by their respective severity.

The food security questionnaire in the World Bank's LSMS data includes a set of experiential and behavioral questions that allow to compute experience-based measures of food security. The reduced Coping Strategies Index (rCSI) can be calculated straightforwardly from the LSMS data, as all rCSI items are covered in the food security questionnaire, and the LSMS recall period of seven days is identical to the rCSI's standard recall period.[5]

Reduced Coping Strategies Index (rCSI)

The reduced CSI draws on a subset of five questions about coping strategies that the household employed in the last seven days. The questions record the frequency of occurrence of each of the coping strategies, which are subsequently weighted to account for differential severity, and then summed to yield the overall score. The score ranges from zero (no coping strategies used in the last seven days) to 56 (all five coping strategies used every day in the last seven days, adjusted for severity). A higher value indicates more reliance on coping strategies and thus lower food security. The five coping strategies and associated severity weights used to construct the rCSI are presented below.

[5]Note that, while the LSMS food security questionnaire does include additional questions beyond those required to derive the rCSI, the other widely used experience-based indicators cannot properly be calculated, since (a) the LSMS does not include the full set of constituent question and/or (b) the recall period is too short.

In the previous seven days, if there have been times when you did not have enough food or money to buy food, how often has your household had to ...?

Coping Strategy		Weight
1.	Rely on less preferred and less expensive foods	[1]
2.	Borrow food or rely on help from friends or relatives	[2]
3.	Limit portion size at mealtime	[1]
4.	Restrict consumption by adults in order for small children to eat	[3]
5.	Reduce the number of meals eaten in a day	[1]

3.3.2 Consumption-based measures

Consumption-based measures of food security draw on consumption questionnaires, which record the quantity and frequency of consumption. There are two broad sub-categories. One set of indicators focuses on dietary diversity, reflecting the number of food groups (sometimes weighted by nutritional density) consumed over a specified number of days. The widely used Food Consumption Score (FCS) and Household Dietary Diversity Score (HDDS) fall in this sub-category. Another set of indicators focuses on the quantity consumed, measuring either total energy consumption or consumption of specific macro- or micro-nutrients. In this study, I calculate two consumption-based measures, one from each sub-category, using the LSMS food consumption questionnaire.

Food Consumption Score (FCS)

The Food Consumption Score (FCS), which was developed by the World Food Programme and continues to be used in their Vulnerability Assessment and Mapping (VAM) unit, is a measure of dietary diversity. In addition to capturing diversity, the FCS also reflects the quantity dimension of food security and has been shown to correlate closely with measures of energy sufficiency (Maxwell et al., 2013). To construct the score, food items are first classified into eight specified food groups and the frequency of consumption in the last seven days is recorded for each group. Raw frequencies of consumption are weighted to reflect differential nutritional density of different food groups, and the weighted frequencies are summed to yield the overall FCS. The score ranges from 0 (no food group was consumed in the last seven days) to 112 (every food group was consumed every day in the last seven days). Higher FCS values indicate greater food security. An overview of the food groups and associated weights is presented below.

Food group	Weight	Food group	Weight
Main staples	2	Meat/Fish	4
Pulses	3	Milk	4
Fruit	1	Oil	0.5
Vegetables	1	Sugar	0.5

Average daily caloric intake per capita

Again drawing on the LSMS food consumption questionnaire, I calculate household average dietary energy consumption, i.e. caloric intake per capita per day. To arrive at this indicator, I use information on the average energy density of different food items (using a food composition table provided by the Food and Agriculture Organization), and adjust households' average daily caloric intake for household size measured in adult equivalents.

3.3.3 Monetary measures

Monetary measures of food security, which include expenditure shares and price indices, can signal stressed food availability and access. In addition, food price volatility can be used as an indicator for the stability dimension of food security, flagging seasonal patterns or recurring issues. In this study, I focus on households' food expenditure share as a monetary measure of food security.

Food expenditure share

The share of total expenditure that is allocated to food is a useful indicator of economic access to food. Since food consumption is highly inelastic, households tend to reduce other expenditure to sustain a given level of food consumption when food prices rise. As Engel (1857) noted, food expenditure shares tend to fall as incomes rise. Food expenditure shares can be indicative of both quantity and quality of the food consumed. When food prices rise, poor households who are already spending a large share of income on food and are unable to reduce other expenditure may be unable to sustain food consumption and instead would substitute to lower quantity or quality of food. If income is unchanged, then an unresponsive food expenditure share in the face of rising prices signals reduced quantity or quality of food. For net producers, the relationship is different, as rising food prices may increase income from sales.

From the LSMS consumption and expenditure questionnaires, I calculate the food expenditure share by: (1) calculating weekly average expenditure on durables (recall for durables expenditure is 12 months), (2) calculating weekly average expenditure on consumables (recall for consumables expenditure is 30 days), (3) calculating weekly average expenditure on food, using median cluster prices to infer the value of food that was not purchased, such as food from own production or from gifts and donations.

3.4 Characterising food insecurity in Tanzania

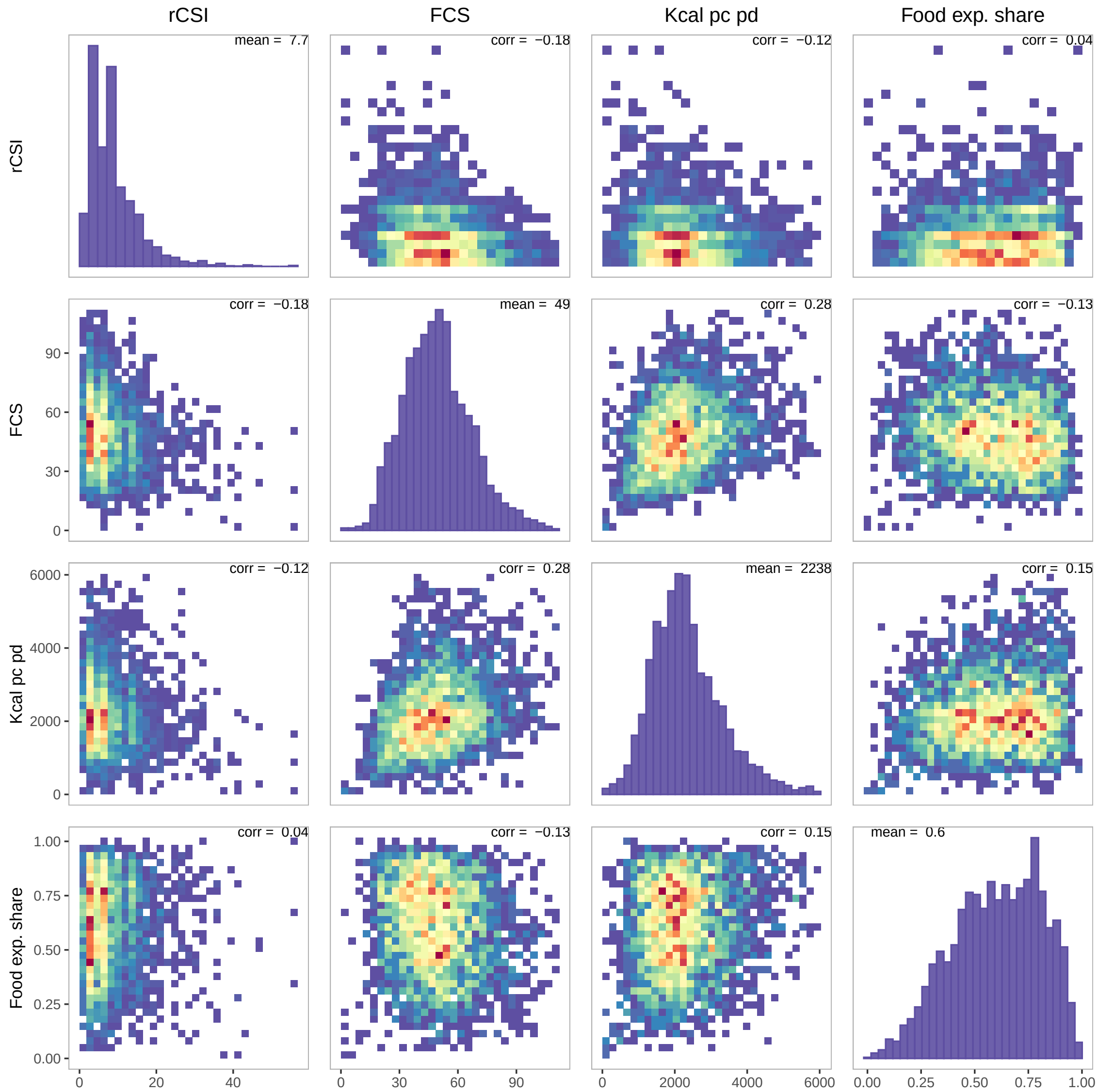
To lay the groundwork for the machine learning prediction in Section 3.5, this section characterises food security in Tanzania using two waves of LSMS data: the 2010/11 wave, which functions as the training dataset in prediction, and the 2012/13 wave, which is used for out-of-sample testing. I first assess correlations between the four different food security indicators described in detail in Section 3.3: the reduced Coping Strategies Index (rCSI), the Food Consumption Score (FCS), caloric intake (kcal pc pd), and the household's food expenditure share. I subsequently focus on two of the four indicators (rCSI and FCS) for the remainder of the paper, as they allow to classify households' food security status according to pre-defined categorical bounds, which facilitates the use of classification models in addition to regression models. In order to document the variation that exists in the prediction target, I map rCSI and FCS over time and across space. The final part of this section focuses on assessing the relationship between food security and potential predictors, before presenting prediction results in Section 3.5.

3.4.1 Comparing different food security indicators

Correlations between various different food security indicators are weak. Figure 3.1 shows a correlation matrix of the four food security indicators discussed in more detail above, namely rCSI (experience-based), FCS and caloric intake (consumption-based), and food expenditure share (monetary). The correlation matrix shows pairwise relationships between these indicators for a subsample of households that experienced at least *some* food insecurity as measured by the rCSI, i.e. households that have an rCSI score greater than zero. Pairwise correlation coefficients are generally low, ranging from 0.04 (rCSI and food expenditure share) to 0.28 (FCS and caloric intake). The distribution of rCSI scores is highly skewed and the relationship with other indicators appears extremely noisy at lower levels of coping (lower rCSI scores). The correlation is strongest between FCS and caloric intake as both of these indicators are consumption-based, and so essentially measure a very similar dimension of food security. Previous research has shown the FCS to be closely linked with overall energy supply (Leroy et al., 2015). Negative correlation coefficients are to be expected, as for some indicators higher values imply higher food security (FCS, caloric intake), while for other indicators lower values imply higher food security (rCSI, food expenditure share).

These weak correlations between food security indicators, particularly between indicators from different categories (e.g. experience-based vs. consumption-based), confirms previous findings in the literature (Maxwell et al., 2013; Vaitla et al., 2015, 2017). Vaitla et al. (2017) study the relationship between different indicators using data from 21 survey datasets and 10 countries. They find that concordance, the degree to which indicators agree on a household's food security status and assign it the same food security status, is extremely low. Factor analysis reveals that the different types of indicators capture two different dimensions of food security, which appear to be only weakly correlated. While factor analysis does not directly suggest a theoretical interpretation of the different dimensions, the authors hypothesise that rCSI and HHS, which are experience-based, measure the cost of access to food, while FCS and HDDS, which are consumption-based, measure realised food consumption. The argument is that experience-based indicators capture behavioural differences in coping strategies that are not necessarily picked up by consumption-based measures. For example, if a household reduces portion size or borrows food from neighbors, this will be picked up by the experience-based measures, while consumption-based measures of realised dietary diversity will not reflect this household's stresses around food access. In the

Figure 3.1: Correlation between different measures of food security

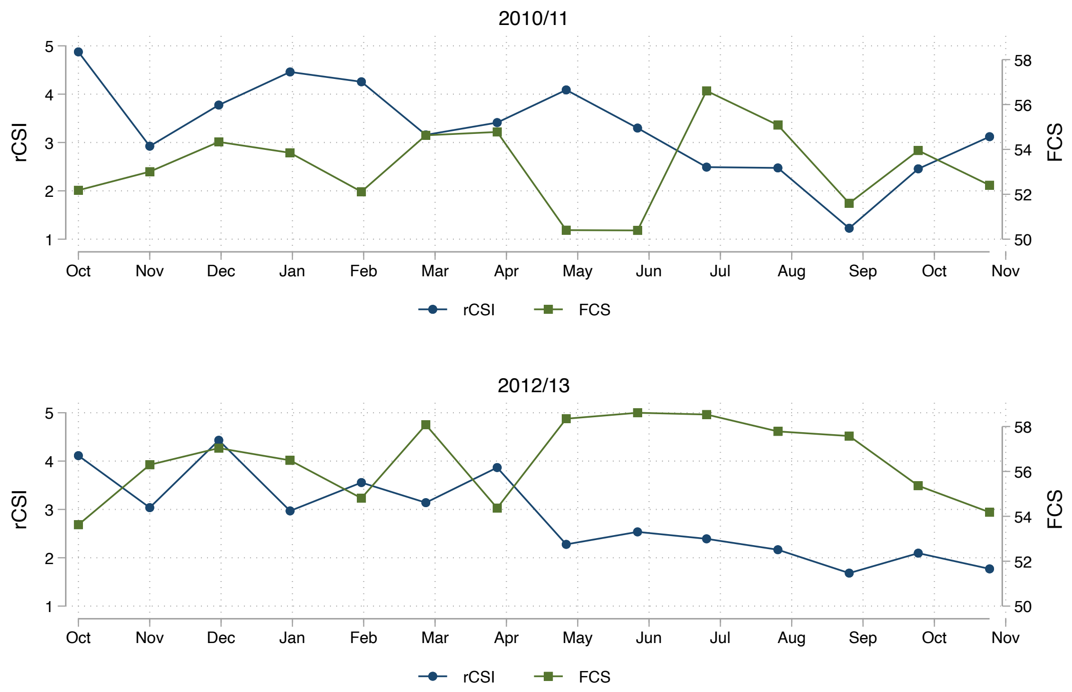


Note: This figure shows the pairwise relationships between different food security indicators. Scatterplots are 2d bin scatters, where red indicates the highest density of data. Subplots on the diagonal are histograms of the four indicators. The data shown is a pooled subsample of the 2010/11 and 2012/13 waves of LSMS data for Tanzania. This subsample of 2,935 households (1,429 from 2010/11 and 1,506 from 2012/13) includes only households with an rCSI 0. In addition, I remove households that report kcal pc pd = 0, food expenditure share = 0, and the top 2% of kcal pc pd (p(98) = 5,889 kcal), to exclude extremely low and high values of food consumption which likely reflect measurement error and which would distort the figure.

prediction exercise in this paper, I take rCSI and FCS as representatives of the two different dimensions identified by Vaitla et al. (2017), and compare how well these two indicators can be predicted from observable features such as location, weather, prices, and household assets and demographics.

3.4.2 Mapping food security over time and across space

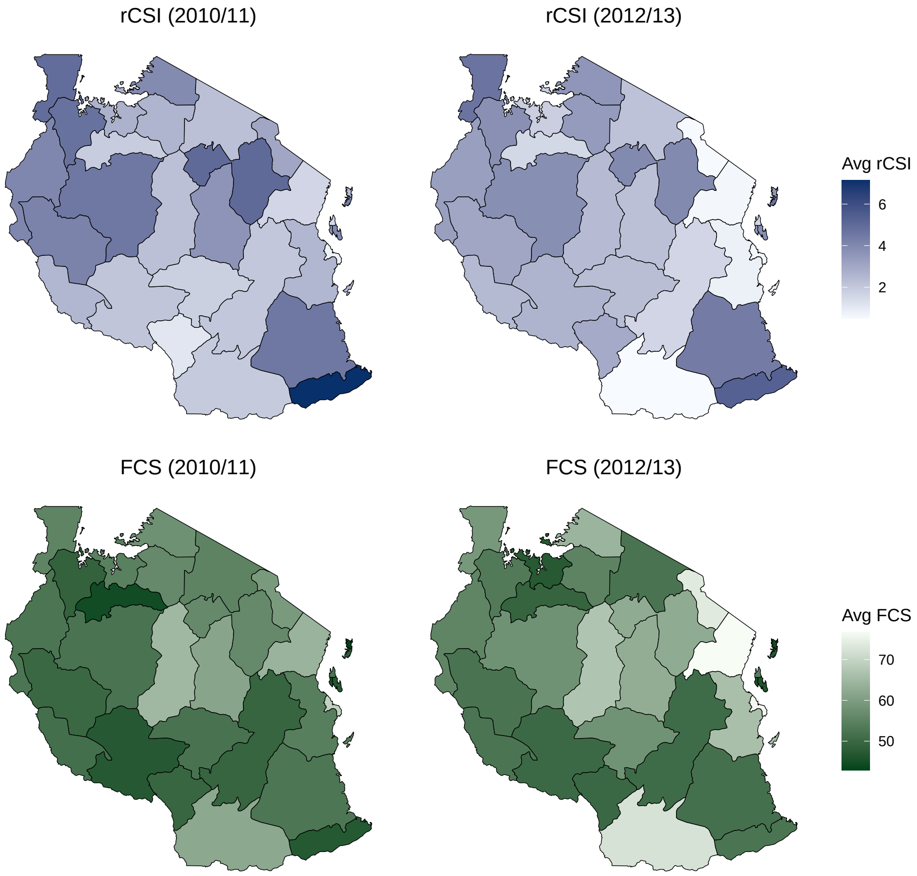
Figure 3.2: Monthly average of rCSI and FCS scores



Note: This figure shows monthly average household rCSI and FCS scores. Higher rCSI implies lower food security; higher FCS implies higher food security. The data are a subsample of households from the 2010/11 (wave 2) and 2012/13 (wave 3) LSMS data for Tanzania where observations are excluded due to missing values or extreme outliers in caloric intake; subsample shown here includes 3,365 households in wave 2 and 4,377 households in wave 3.

Figures 3.2 and 3.3 in this section illustrate the substantial variation in food security that exists across time and space. Average monthly food security values vary over the course of a year, in line with harvest and lean seasons (see the seasonal calendar in Figure 3.14 in Appendix 3.D for details). Changes in monthly averages of rCSI and FCS move largely in opposite directions, since an improvement in food security is reflected by a deline in rCSI but an increase in FCS. The second half of the year, roughly from May or June onwards, registers lower rCSI and higher FCS, reflective of improved food security during harvest season. This effect is particularly visible in the 2012/13 wave of LSMS data. These seasonal patterns suggest that the inclusion of month fixed effects may improve the prediction of food security substantially.

Figure 3.3: Regional average of rCSI and FCS scores

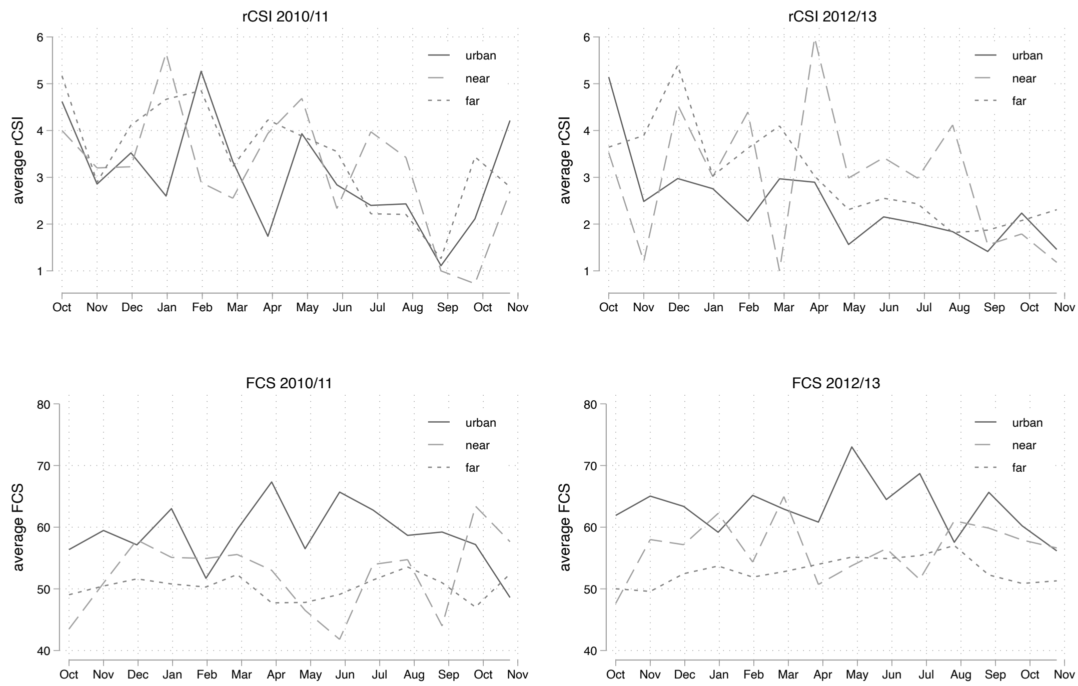


Note: This figure shows average household rCSI and FCS scores by region. Darker colors indicate lower food security; higher rCSI implies lower food security; higher FCS implies higher food security. The data are a subsample of households from the 2010/11 (wave 2) and 2012/13 (wave 3) LSMS data for Tanzania where observations are excluded due to missing values or extreme outliers in caloric intake; subsample shown here includes 3,365 households in wave 2 and 4,377 households in wave 3.

While there is significant within-location variation in food security, average food security also varies strongly across regions. Figure 3.3 maps average rCSI and FCS values by region for both of the two waves of LSMS data used in this study. Across maps, darker colors indicate lower food security. There is some limited agreement between the two indicators about relative differences between regions. Mtwara in the southeast has the highest average rCSI value in both the 2010/11 as well as the 2012/13 wave. Mtwara's average FCS scores are also among the lowest ones, but the difference to other regions is not as striking as in the rCSI map. While the two indicators agree that Singida in the center, Kilimanjaro and

Tanga on the northeastern coast, and Ruvuma in the south are relatively food secure on average, there are some striking discrepancies, too. An example of such a discrepancy is Shinyanga in the north, which appears food secure in the rCSI map, but relatively food insecure in the FCS map. Another striking feature of these maps is the limited spatial correlation between bordering regions' average food security values. The difference is particularly stark between Mtwara, which has low average food security, and bordering Ruvuma, which has substantially higher food security, in the south of Tanzania.

Figure 3.4: Average rCSI and FCS by type of region



Note: This figure shows average household rCSI and FCS scores by type of region and LSMS wave. Higher rCSI implies lower food security; higher FCS implies higher food security. The data are a subsample of households from the 2010/11 (wave 2) and 2012/13 (wave 3) LSMS data for Tanzania where observations are excluded due to missing values or extreme outliers in caloric intake; subsample shown here includes 3,365 households in wave 2 and 4,377 households in wave 3.

In addition to distinguishing locations by administrative region, I classify household locations by degree of remoteness. Previous literature has shown that trade openness can have a strong effect on food security by changing the degree to which local crop production determines food availability (Burgess & Donaldson, 2012; Janssens et al., 2020). To account for this effect in the prediction of food security, I incorporate remoteness as a measure of households' market access and trade openness. The LSMS data includes household coordinates at the cluster level[6], which I combine with the Global Human Settlement Layer's Degree of Urbanisation (DoU) raster data to arrive at a three-class categorisation of locations. I map the DoU's eleven categories into three region types; an urban region, an adjacent rural region which I call 'near', and a remote rural region which I call 'far'. For details on the construction of region types see Appendix 1.B.3. Urban households tend to have higher average food security than households in either of the rural region types. Average FCS values are consistently lower and less volatile in the remote 'far' rural region than in the urban region.

3.4.3 Assessing correlations with potential predictors

Figure 3.5 shows pairwise correlations between rCSI and FCS and potential predictors. At first glance, these potential predictors appear to be more strongly correlated with FCS than with rCSI. FCS and rCSI tend to have opposite signs on these correlation coefficients, as higher rCSI indicates lower food security, while higher FCS indicates higher food security. The columns refer to different subsamples of the data, pooled across region types, or differentiated by the degree of remoteness of the households' location. In the pooled sample, household location correlates with both rCSI and FCS. For FCS in particular, households in far (remote) rural locations have lower food security scores than households in urban locations, with households in near (adjacent) rural locations being an intermediate case.

Somewhat surprisingly, the rainfall measure does not correlate strongly with the food security indicators. In the near rural region, the direction of the correlation is as expected, with higher rainfall improving households' food security sta-

[6]Coordinates are randomly perturbed for respondents' privacy. The data documentation states that "the coordinate modification strategy relies on random offset of cluster centerpoint coordinates (or average of household GPS locations by EA in TZNPS2) within a specified range determined by an urban/rural classification. For urban areas a range of 0-2 km is used. In rural areas, where communities are more dispersed and risk of disclosure may be higher, a range of 0-5 km offset is used. An additional 0-10 km offset for 1% of rural clusters effectively increases the known range for all rural points to 10 km while introducing only a small amount of noise. Offset points are constrained at the district level, so that they still fall within the correct district for spatial joins, or point-in-polygon overlays" (Tanzania - National Panel Survey 2010-2011, Wave 2 documentation, p.9).

Figure 3.5: Correlations between food security indicators and potential predictors

		rCSI				FCS			
		All	Far	Near	Urban	All	Far	Near	Urban
Location	Far rural region	0.01				-0.18			
	Near rural region	0.03				-0.01			
	Urban region	-0.03				0.21			
Rainfall	CHIRPS (12 months)	0.01	0.02	-0.06	0.05	0.00	0.02	0.06	-0.06
Domestic prices	Maize price	0.01	0.01	0.04	-0.02	0.11	0.07	0.10	0.13
	Rice price	0.02	0.03	0.02	0.00	0.05	0.01	0.05	0.07
Global prices	FAO Food Price Index	0.04	0.05	0.02	0.04	-0.07	-0.08	-0.08	-0.05
	FAO Cereals Price Index	0.07	0.10	0.01	0.06	-0.05	-0.08	-0.04	0.00
	FAO Dairy Price Index	-0.07	-0.10	0.02	-0.09	0.04	0.07	-0.04	0.03
	FAO Meat Price Index	-0.03	-0.04	0.01	-0.04	0.03	0.01	-0.01	0.07
	FAO Oils Price Index	0.05	0.06	0.01	0.05	-0.08	-0.08	-0.07	-0.09
	FAO Sugar Price Index	0.04	0.04	0.00	0.04	-0.07	-0.07	-0.03	-0.09
Assets	Natural roof	0.10	0.09	0.15	0.09	-0.26	-0.23	-0.24	-0.10
	Number of cellphones	-0.14	-0.14	-0.19	-0.12	0.38	0.36	0.40	0.29
Demographics	Education of HH head	-0.08	-0.07	-0.05	-0.09	0.10	0.07	0.05	0.04
	Age of HH head	0.11	0.10	0.11	0.11	-0.05	-0.04	-0.03	0.02
	Female HH head	0.11	0.10	0.14	0.10	-0.10	-0.07	-0.15	-0.14
	Size of HH (adult equiv.)	0.02	-0.01	-0.04	0.10	0.14	0.14	0.23	0.17
	Number of adults	-0.01	-0.03	-0.07	0.06	0.16	0.14	0.22	0.16
	Number of children	0.06	0.03	0.02	0.14	0.06	0.08	0.14	0.10

Note: This table shows pairwise correlation coefficients for two food security indicators (rCSI and FCS) and potential predictors at the household level in a subsample of households from the 2010/11 (wave 2) and 2012/13 (wave 3) LSMS data for Tanzania. Some observations are excluded due to missing values (or extreme outliers in caloric intake); the subsample shown here includes 3,365 households in wave 2 and 4,377 households in wave 3. Higher rCSI indicates lower food security. Higher FCS indicates higher food security. Columns refer to different subsamples of the data, pooled across location types, or differentiated by the degree of remoteness of the households' location.

tus. In the urban region, higher rainfall is associated with a deterioration in food security, and in the far rural region, the correlation is near zero. Pairwise correlations with different measures of domestic and international food prices are broadly as expected, such that higher food prices are associated with lower food security outcomes. Exceptions to this pattern are the correlations between domestic prices of maize and rice and the FCS indicator, where positive correlations imply that higher prices are associated with higher food security, and the FAO dairy and meat price indices, where higher price levels are associated with higher food security as measured by both rCSI and FCS.

Household assets and demographics exhibit the strongest relationships with both food security indicators. This is intuitive, as these are the most granular measures and can differentiate adjacent households that have different food security levels but are exposed to the same weather, prices, and market access. Households whose dwelling's roof is constructed of natural materials such as bamboo, leaves or mud have lower food security than households whose roof is constructed of non-natural materials such as metal sheets. The number of cellphones owned by the household is particularly strongly associated with both rCSI and FCS, as this appears to be an informative measure of household assets. Household head characteristics also correlate relatively strongly with both food security indicators. The household head's education level is positively associated with food security, while age and female gender are associated with lower food security outcomes. Household size is positively correlated with FCS across all subsamples, whereas correlations with rCSI are more mixed. Larger household size is associated with higher food security in rural locations, but lower food security in urban locations, as measured by the rCSI score.

3.5 Predicting food insecurity

This section focuses on predicting the Food Consumption Score (FCS) and the reduced Coping Strategies Index (rCSI) as indicators of household food security. These indicators offer advantages over alternatives like caloric intake or food expenditure share because rCSI and FCS have pre-defined categorical thresholds for classifying households as either food secure or insecure. These thresholds allow training not only regression models for predicting raw scores but also classification models to predict binary food security status.

I employ a tree-based ensemble method (eXtreme Gradient Boosting) for both classification and regression. LSMS wave 2 (2010/11) data is used for training and LSMS wave 3 (2012/13) data is used for out-of-sample testing. While I do not aim to forecast future food security states, the temporal separation between training and testing data allows to assess whether model coefficients are sufficiently stable over time, such that the method could reliably be applied to predict contemporaneous food insecurity in future periods. To prevent overfitting, 10-fold cross-validation is applied to the training set. Grid search hyperparameter tuning optimises model performance by setting an appropriate minimum node size and the number of trees. Since class sizes are unbalanced and rCSI and FCS distributions are skewed, fold construction for cross-validation is stratified by rCSI (FCS) deciles to ensure each fold covers the full distribution.

The analysis reveals two key findings: First, FCS prediction yields higher accuracy than rCSI prediction. From a policy-maker's perspective, it may be particularly desirable to minimise errors of exclusion such that most food insecure households are correctly detected as such. I fix the proportion of correctly identified food insecure households to 70%, and compare overall prediction accuracy for the two scores. Detecting 70% of food insecure households based on FCS results in fewer inclusion errors compared to rCSI.

Second, predictor relevance differs between the two indicators. Time, region, market prices, and climatic conditions are used as predictors, alongside variables reflecting remoteness and market access, selected household assets (cellphones owned, roof type), and basic household demographics. Household assets hold significant predictive value for both FCS and rCSI. Demographics, on the other hand, offer no benefit for FCS prediction but appear slightly more important for rCSI. Remoteness indicators have low predictive power but are cheaper to collect than demographics and assets. Hence, they are included in the main predictor set despite their relatively low relevance.

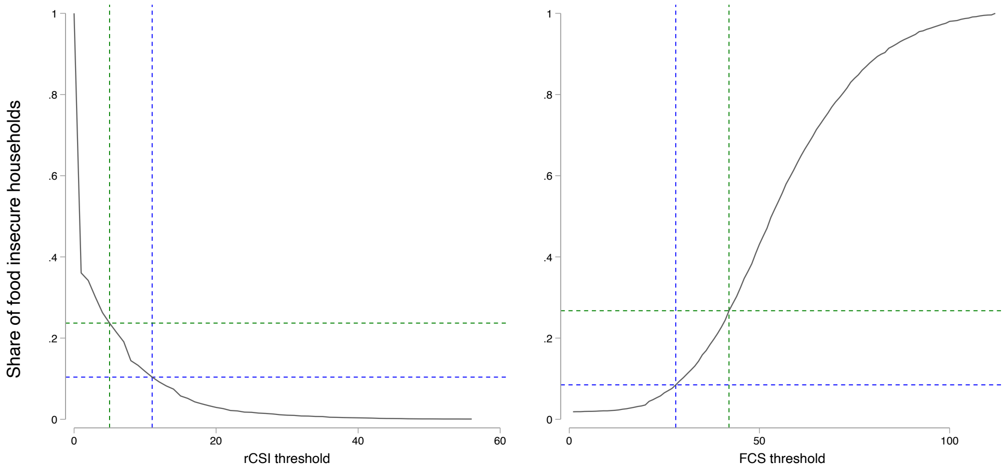
3.5.1 Defining the prediction target

While pre-defined thresholds for rCSI and FCS exist, some discretion has to be used in fixing exact values. I follow Maxwell et al. (2014) in using rCSI categorical bounds of 0 to 4 to indicate food secure status, 5 to 10 to indicate moderately food insecure status, and greater than 11 to indicate severely food insecure status. For the FCS, I follow WFP's guidance on setting thresholds: Scores greater than or equal to 42 indicate acceptable food security status, scores between 28 and 42 indicate borderline food security status, and scores less than or equal to 28 indicate poor food security status.[7] The choice of threshold matters crucially for the extent of class imbalance present in the sample. Since the thresholds cut right through

[7] As noted in Vaitla et al. (2017), WFP uses lower FCS thresholds of 21 and 35 in settings where oil and sugar are not consumed daily. In the Tanzanian sample, however, 83% of households consume oil daily and 78% consume sugar daily, which suggests that using the higher cut-off values is appropriate in this setting.

the 'thick' part of the distribution, even marginal shifts in the thresholds yield substantial shifts in class sizes. Figure 3.6 illustrates the relationship between the choice of categorical bounds for rCSI and FCS and the share of the population that is classified as food insecure as a result.

Figure 3.6: Categorical bounds for rCSI and FCS and the prevalence of food insecurity



Note: This figure shows the effect of the categorical bounds (cutoff values) on the share of households defined as food insecure in a pooled dataset of Tanzanian LSMS data for wave 2 (2010/11) and wave 3 (2012/13). For rCSI, a higher value indicates lower food security; choosing a higher value for the threshold thus decreases the share of households defined as food insecure. For FCS, a higher value indicates higher food security; choosing a higher threshold thus increases the share of households defined as food insecure. The vertical green line (at rCSI = 5 and FCS = 42) indicates moderate food insecurity. The vertical blue line (at rCSI = 11 and FCS = 28) indicates severe food insecurity. Horizontal lines mark the corresponding share of households in each category.

Using the thresholds described above yields a three-step categorisation for both indicators, i.e. 'food secure', 'moderately food insecure', and 'severely food insecure' for the rCSI, and 'acceptable', 'borderline', and 'poor' for the FCS. While the number of observations in each of the three classes is similar for rCSI and FCS, the cross-tabulation presented in Figure 3.7 shows substantial discordance between the two indicators. This is not surprising, given the low correlation in raw scores shown in Figure 3.1 above and previous findings in the literature on low concordance between rCSI and FCS (Vaitla et al., 2017). However, low concordance does raise the question whether both indicators can be predicted from the same set of predictors, and whether prediction accuracy is higher for one indicator than the other. The cross-tabulation also highlights that class sizes are imbalanced. Since (severe) food insecurity is luckily relatively rare in Tanzania, the '(severe) food insecure' class is substantially smaller than the 'food secure' class. Class imbalance makes it more difficult for the model to learn to identify the minority class, i.e. food insecure households.[8] I group moderately and severely food insecure households into one class, which I label 'food insecure'. The two resulting classes ('secure'/'insecure') serve as the prediction targets for the binary classification models, and as a way to assess model performance in the prediction of raw FCS and rCSI values.

Figure 3.7: Cross-tabulation of rCSI and FCS categories

		FCS (%)			
		Acceptable	Borderline	Poor	Total
rCSI (%)	Food secure	60.0	11.0	5.3	76.3
	Moderately food insecure	8.4	3.3	1.6	13.3
	Severely food insecure	6.4	2.4	1.6	10.4
	Total	74.9	16.6	8.5	100.0

Note: This figure shows a cross-tabulation of households' rCSI and FCS categories for a pooled dataset of Tanzanian LSMS data for wave 2 (2010/11) and wave 3 (2012/13). Categorical bounds on the rCSI and FCS scores follow Vaitla et al. (2017); rCSI cut-offs are [0-4] food secure; [5-10] moderately food insecure; >11 severely food insecure; FCS cut-offs are ≥42 acceptable; (28-42) borderline; ≤28 poor.

3.5.2 Choosing the set of predictors

The set of variables that are chosen as predictors can have strong effects on prediction accuracy. More granular data, for example on household assets, tends to improve prediction accuracy but typically implies higher data collection costs (Lentz et al., 2019). While seasonal patterns, food prices, market access, and agricultural production conditions have some predictive power, there remains substantial within-location variation in food security at any given point in time. This remaining within-location variation in food security outcomes can only be explained by household-level characteristics, such as demographics and assets. Household-level variables typically require costly and time-consuming data collection through

[8] Possible ways to address this could be oversampling techniques, which can alleviate the issues of class imbalance and skewness, or cost-sensitive learning techniques, which introduce a penalty for mis-classifying the minority class.

surveys. The analysis in this section shows that the inclusion of household asset indicators is crucial for model performance, whereas household demographics add little predictive power. Consequently, the main models include two basic indicators of household assets, which could conceivably be estimated without the need for primary data collection, but no other household characteristics.

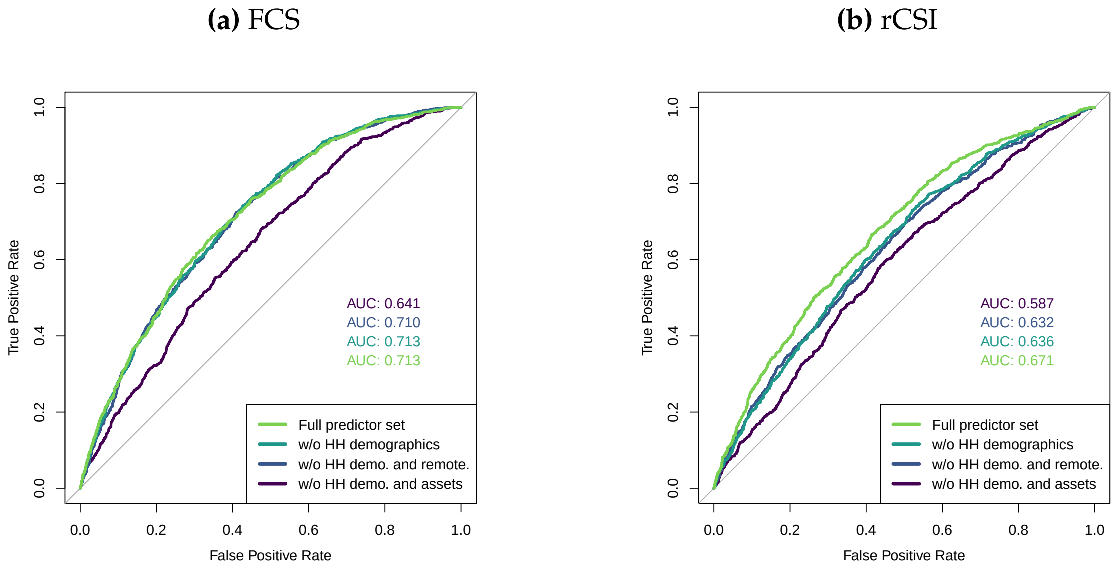
I evaluate the relevance of different predictors by comparing model performance with and without them as part of the predictor set (Figure 3.8). For simplicity, I focus here on the binary classification of food security status ('secure' or 'insecure') rather than the prediction of raw FCS and rCSI values. In binary classification, the Area Under the ROC Curve (AUC) serves as a key indicator of model performance. The ROC curve plots the True Positive Rate (TPR) and False Positive Rate (FPR) at different class probability cutoffs.[9] Curves that lie further to the top left and have a higher AUC value represent better model performance.[10]

Predictive performance of the binary classification models for both FCS and rCSI food security status is highest when the full set of predictors is used. The full set includes (i) time and region indicators, (ii) publicly available data on food prices and climatic conditions, (iii) remoteness indicators that can be derived from households' geo-coordinates, (iv) basic indicators of household assets, and (v) household demographics. Time and region fixed effects include dummy variables for each month of the year and each of the 26 administrative regions in Tanzania. The second category, publicly available data on food prices and climatic conditions, includes maize and rice prices in the nearest of the 20 major markets monitored by WFP, FAO global food price indices (both overall and for the five constituent food groups), and total rainfall at a household's location in the past 12

[9]Based on the patterns observed in the training data, the binary classifier attaches class probabilities to each household in the test dataset. By choosing a probability cutoff, these probabilities can be mapped to one of the two possible outcomes, classifying each household as either 'food secure' or 'food insecure'. Different cutoffs result in varying levels of error, yielding the different combinations of True Positive Rate and False Positive Rate that constitute the ROC curve.

[10]If the classification worked perfectly, all truly food insecure households would be classified as such, while no food secure households would mistakenly be classified as being food insecure, yielding a True Positive Rate of one and a False Positive Rate of zero (the top left corner in ROC plots). The 45-degree line in ROC plots indicates the performance of a hypothetical classifier that assigns outcomes at random. An ROC curve that lies above this line has higher performance than random assignment. Since different ROC curves might cross, which would complicate selecting the 'highest' one, the Area Under the Curve (AUC) is commonly used as an indicator that summarises overall model performance.

Figure 3.8: Performance of binary classification model using different predictor sets



Note: This figure shows ROC curves and corresponding AUC values for binary classification models that use different sets of predictors. Panel (a) shows performance of models predicting binary food security status according to FCS. Panel (b) shows performance of models predicting binary food security status according to rCSI. All models use the same algorithm (Gradient Boosted Decision Trees (GBDT)) and training data (LSMS wave 2 2010/11), and performance is always evaluated on the same test dataset (LSMS wave 3 2012/13).

months (as reported by CHIRPS). In addition, I include different variables that indicate the degree of remoteness of a household's location. These remoteness variables include a household's Degree of Urbanisation category (ranging from 'very low density rural' to 'urban centre') and distance to the nearest agricultural market and nearest major road. The fourth category of predictors includes two variables that reflect household assets: the number of cellphones owned and a dummy variable indicating the type of roof of the household's dwelling (whether it is made of natural materials such as leaves and bamboo or non-natural materials such as metal sheets). In this study, both asset variables are taken from the LSMS household survey, but they could conceivably be remotely sensed or provided by cellular companies, rather than collected through a survey (Lentz et al., 2019). The final category of predictors subsumes basic demographic variables: household size (measured in adult equivalents) and composition (the number of adults and number of children), and household head characteristics including age, gender, and a dummy variable to indicate completion of primary education.

The relevance of different predictors varies depending on the specific prediction target, i.e. the food insecurity indicator that is predicted. Variables that reflect household assets are strongly predictive of both FCS and rCSI status (see 'w/o HH demo. and assets' ROC curves in panels (a) and (b) of Figure 3.8). Excluding household demographics and remoteness indicators, on the other hand, has little to no effect on the predictive performance for FCS status, and only a small effect on the predictive performance for rCSI status. I choose the predictor set that excludes household demographics, but retains household assets and remoteness variables (the 'w/o HH demographics' ROC curve) for the models in the remainder of this section, as this predictor set strikes a balance between predictive performance and data collection costs.

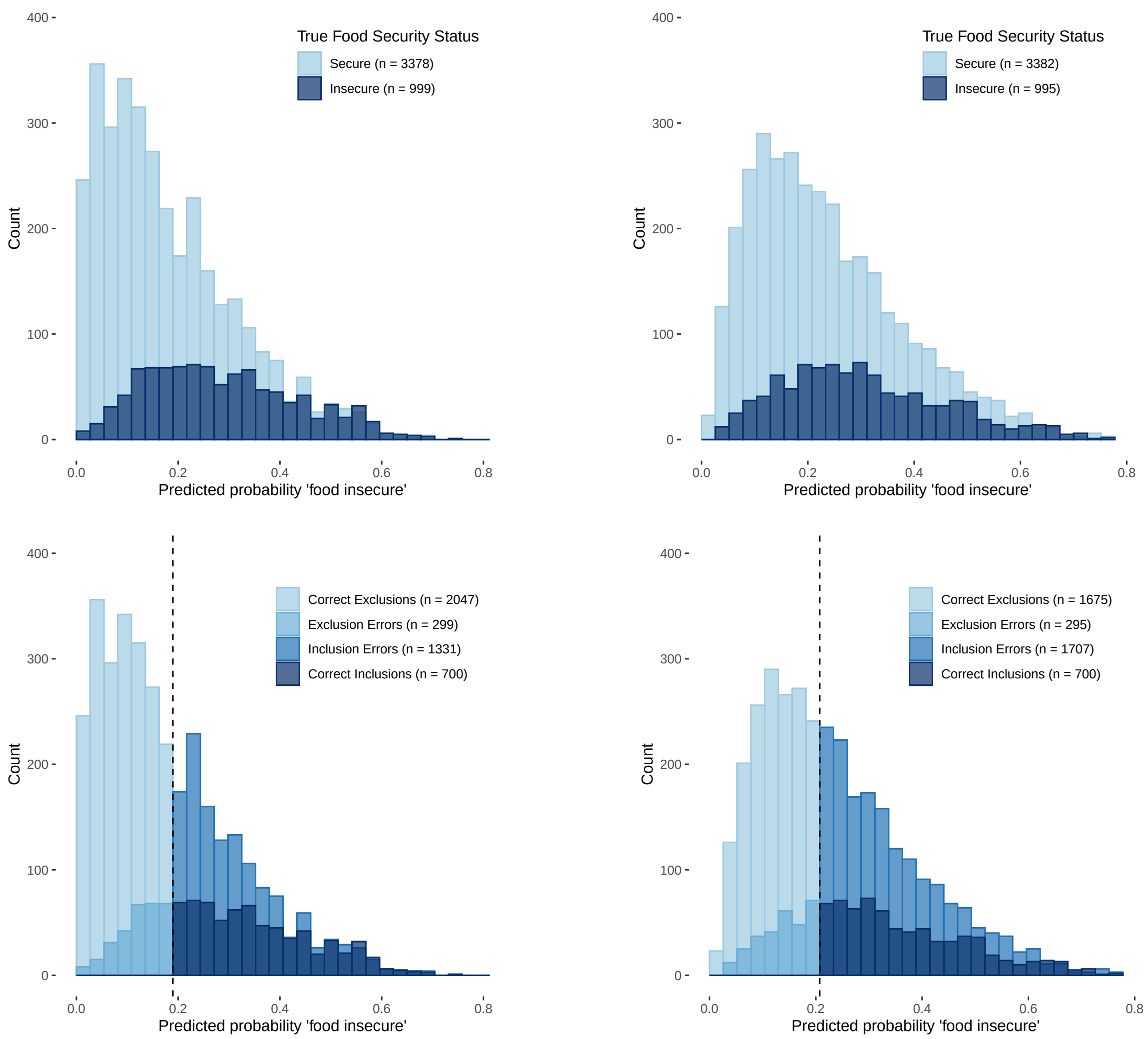
3.5.3 Prediction of binary FCS and rCSI status

This section presents results of the prediction of binary household food security status. From observed FCS and rCSI values, households' true food security status can be determined in both waves of LSMS data, such that the 'ground truth' is known both in-sample (in the 2010/11 training dataset) and out-of-sample (in the 2012/13 test dataset). I use the cutoff values described in Section 3.5.1 to categorise households' status as secure or insecure based on the values of FCS and rCSI. Since concordance between FCS and rCSI is relatively low (see cross-tabulation in Figure 3.7), I predict FCS status and rCSI status separately.

I use a tree-based ensemble method – eXtreme Gradient Boosting (XGBoost) – for prediction, as this typically achieves better results than linear models such as logistical regression (Martini et al., 2022; Zhou et al., 2021). I employ the full predictor set, excluding household demographics, as described in Section 3.5.2. To prevent overfitting, I implement 10-fold cross-validation, with stratification on deciles of FCS and rCSI, respectively, to ensure all folds encompass the entire distribution of the prediction target. The model assigns a probability score to each household, reflecting the predicted likelihood of food insecurity. By setting a specific threshold for the predicted likelihood of food insecurity, households can be categorised as either 'secure' or 'insecure'. Since households' true food security

Figure 3.9: Prediction of binary FCS and rCSI status

(a) FCS **(b)** rCSI



Note: This figure plots predicted probability of food insecurity by true food security status for households in the test dataset (LSMS wave 3, 2012/13). Panels on the left show results for FCS; Panels on the right show results for rCSI. The x-axis shows the predicted probability of a household's status being 'food insecure'; the y-axis shows the absolute number of households whose probability that lies in the respective bin. The top histograms show the distribution of predicted probabilities by true food security status. The bottom two histograms show how setting the probability threshold (dashed line) to 0.19 (for FCS) and 0.21 (for rCSI) to achieve 70% recall.

status is known from the LSMS data, we can assess the model's performance by comparing predictions to ground truth.

Figure 3.9 visually demonstrates how varying the probability threshold impacts classification errors in the prediction of FCS status (panel (a) on the left) and rCSI status (panel (b) on the right). In addition to ROC-AUC as discussed in Section 3.5.2, the performance of binary prediction models can be evaluated using out-of-sample accuracy, precision, and recall. Accuracy refers to the overall proportion of correctly classified observations. When class sizes are equal and different prediction errors are equally 'costly', accuracy is a sufficient performance metric. However, when classes are imbalanced, or when errors of exclusion weigh heavier than errors of inclusion (or vice versa), then accuracy can be misleading. In the setting of this study, the 'food insecure' class is significantly smaller than the 'food secure' class (23% vs. 77%), and errors of exclusion (failing to detect a food insecure household) can reasonably be considered 'more costly' for policy-makers than errors of inclusion (mistakenly classifying a secure household as insecure). Precision and recall are additional metrics which help to assess model performance in such a setting. Precision measures the proportion of predicted insecure households that are truly insecure. Recall, conversely, measures the proportion of truly insecure households correctly identified by the model.

Given the class imbalance and the importance of minimising errors of exclusion, I prioritise recall over other performance metrics and set the probability threshold such that it yields a recall of 70%. This translates to the model correctly detecting 7 out of every 10 truly food insecure households. Due to the fact that the probability distributions for the 'secure' and the 'insecure' class overlap substantially, achieving such a high recall rate necessitates a relatively low probability threshold of around 0.2 (i.e. 20% predicted probability of class 'food insecure'). This threshold results in overall accuracy of 63% and precision of 34% for FCS prediction, and accuracy of 54% and precision of 29% for rCSI prediction. These levels of accuracy and precision might appear relatively low at first glance, but they align with previous findings in the literature. Studies by Zhou et al. (2021) and Lentz et al. (2019) report broadly similar levels of performance when predicting food security status at the cluster level. Models that achieve substantially higher performance include lagged food security values or additional household asset indicators as predictors (Lentz et al., 2019; Martini et al., 2022; Zhou et al., 2021).[11]

3.5.4 Prediction of FCS and rCSI values

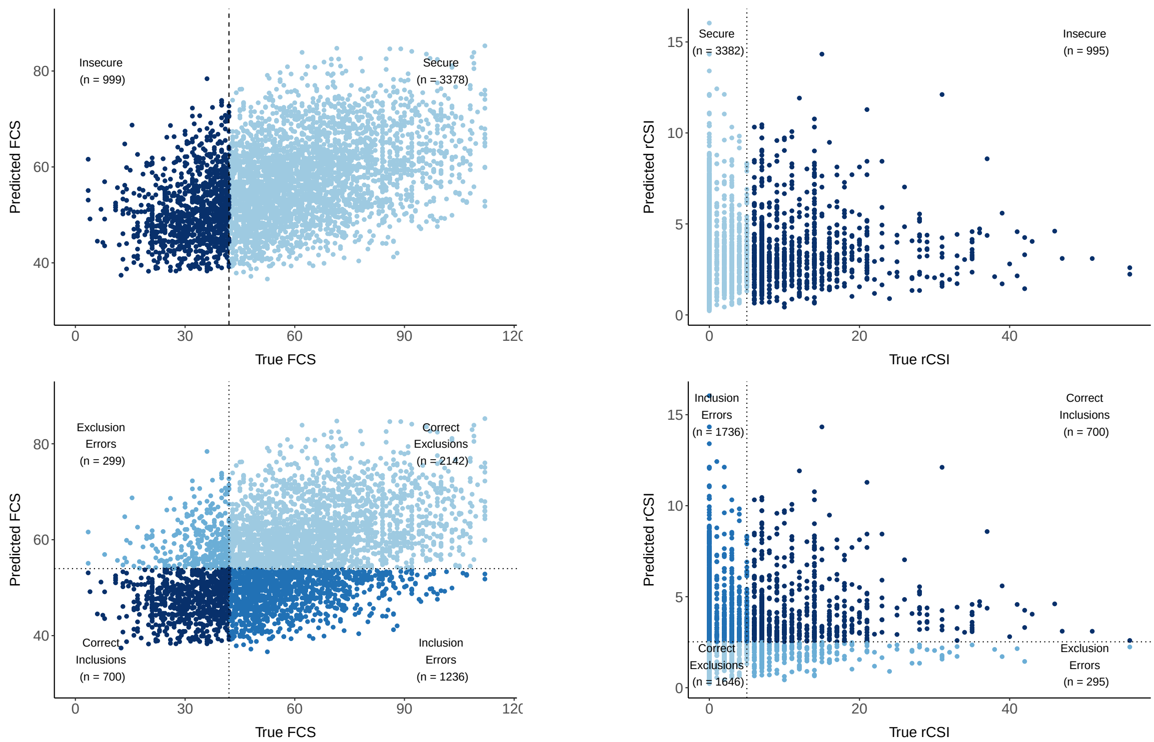
In this section, I use a more granular 'bottom-up' approach to predict households' food security status. Instead of directly predicting whether a household is food secure or insecure as in the binary classification model in Section 3.5.3, I now predict raw FCS and rCSI values using a regression model. Subsequently, I classify households as 'secure' or 'insecure' based on the predicted values. The rationale behind this approach is that, while the binary classification model has no way of differentiating by the level of severity of food insecurity, the regression model uses the full variation in underlying FCS and rCSI values, which may result in more accurate predictions.

I use the same eXtreme Gradient Boosting (XGBoost) model – though this time in regression rather than classification mode – and the same datasets for training and testing. Figure 3.10 plots observed and predicted FCS and rCSI values for households in the test dataset. For the observed 'true' FCS and rCSI values, I maintain the same cutoffs describes in Section 3.5.1: households with a true FCS value below 42 or a true rCSI value above 4 are classified as food insecure. The two top scatterplots in Figure 3.10 show how these cutoffs separate the sample into 'secure' and 'insecure' households. The scatterplots show under-prediction of FCS and

[11] Lentz et al. (2019) use LSMS household survey data for Malawi to predict average FCS, rCSI and HDDS at the cluster (village) level and subsequently analyse errors of inclusion and exclusion by applying cutoffs to construct food security categories. For FCS (rCSI), their main model achieves 76% (65%) accuracy at 83% (86%) recall. These higher levels of predictive performance are most likely due to the inclusion of additional predictors (e.g. lagged IPC food security assessments, additional household assets, and household demographics). The authors do not report the performance of models that use a more parsimonious set of predictors, but they do concede that the inclusion of lagged IPC assessments and household assets as predictors is crucial for model performance. In a subsequent study (Zhou et al., 2021), the authors predict binary food insecurity status at the cluster level, defined as at least 20% of households in the village experiencing food insecurity as measured by FCS or rCSI. For Tanzania, Zhou et al. (2021) report accuracy of 81-84% at 70% recall for FCS, which is higher than the performance achieved in my main models. For rCSI, however, they report accuracy of 52-53% at 70% recall, which is slightly lower than the performance achieved in my main models. Again, the authors include additional predictors on household assets and demographics, among others. While the prediction target in both of these previous studies is defined at the cluster rather than the household level, the performance metrics still provide useful comparisons for the models in this paper. This is particularly true for the results reported in Zhou et al. (2021), as their predictions for Tanzania draw on the same waves of LSMS data for training and testing the models in this paper.

Figure 3.10: Prediction of FCS and rCSI values

(a) FCS **(b)** rCSI



Note: This figure plots true and predicted FCS and rCSI values by true food security status for households in the test dataset (LSMS wave 3, 2012/13). Panels on the left show results for FCS; Panels on the right show results for rCSI. The x-axis shows true values; the y-axis shows predicted values. The top scatterplots color codes households as food secure (FCS above 42, rCSI below 5) or insecure based on observed FCS and rCSI values. The bottom scatterplots show how setting the food insecurity cutoff to 54 for FCS and 2.5 for rCSI to achieve 70% recall affects errors of inclusion.

rCSI values, especially for rCSI. Since the distribution of rCSI is strongly skewed to the right, with many households reporting zero use of coping strategies, I use Poisson regression. In addition, Appendix 3.E shows that the inclusion of household demographics as additional predictors can partially reduce the downward bias in the prediction of rCSI scores.

To facilitate comparison with the prediction results in Section 3.5.3, I again enforce a recall rate of 70%. Due to the under-prediction of raw FCS and rCSI values, correct identification of 70% of food insecure households requires adjusting the classification cutoffs that are applied to predicted values. The bottom two scatterplots in Figure 3.10 illustrate the relationship between cutoff values and prediction errors. For FCS, the threshold on predicted values is set to 54 (30% higher than the true FCS cutoff). This threshold results in overall accuracy of 65% and precision of 36%, implying a small improvement over the binary classification results. For rCSI, setting the threshold to 2.5 (50% lower than the true rCSI cutoff) correctly identifies 70% of food insecure households. The resulting accuracy of 54% and precision of 29% are approximately the same as the values achieved in binary classification (although the number of inclusion errors is marginally higher in the prediction of raw rCSI values; 1736 vs. 1707 in the binary classification).

3.6 Conclusion

Accurate measurement of household food security, while essential for effective policy targeting, is complex, costly and time-consuming. The analysis in this chapter confirms previous findings in the literature that the correlation between different food security indicators is weak, suggesting that the various experience-based, consumption-based, monetary, and biological measures capture independent dimensions of food security. For the two indicators analysed in detail in this study, the Food Consumption Score (FCS) as an example of a consumption-based indicator and the reduced Coping Strategies Index (rCSI) as an example of an experience-based indicator, machine learning models yield promising results for the prediction of households' food security status from readily available secondary data when primary data collection is infeasible.

Predictive performance of these machine learning models is higher for FCS than rCSI. This may be due to the fact that FCS is derived directly from realised consumption – rather than psychological questions and use of coping strategies – which appears to have a stronger association with observable drivers such as market prices and weather. The inclusion of remoteness indicators in the predictor set does not add much predictive performance in the current models. This may be due to the relatively strong correlations between remoteness indicators and more granular predictors such as household assets, such that remoteness indicators add little information beyond these other predictors. Future research could investigate whether training separate models for different types of region (remote rural, adjacent rural, urban) using a larger dataset might improve prediction accuracy compared to a model trained on pooled data. I find suggestive evidence that 'bottom-up' prediction of food security status (prediction of continuous scores and subsequent classification according to cutoff values) can improve predictive performance relative to 'top-down' binary classification. Future research could explore whether this also holds for geographic aggregation, i.e. whether granular bottom-up prediction of a region's food security yields higher accuracy than top-down classification of a region's food security status. In addition, future research could explore the use of alternative model types, such as neural networks, to improve predictive performance.

There remain some limitations for the real-world application of the prediction methodology proposed in this paper: First, asset indicators (roof type and cellphone ownership), which are key for predictive performance, are taken from LSMS survey data in this study. In a context in which policy-makers seek to infer food security from secondary data, such household survey data is likely unavailable. While roof type and cellphone ownership may be accessible through remote sensing or data provided by private cellphone companies, the quality of predictions would have to be assessed separately using these alternative data sources. Second, the approach does not currently take into account potentially realised policy interventions or distributions of food aid. The presence of such programs may impact the relationship between predictors and food security outcomes. Zhou et al. (2021) provide some evidence showing that predictive performance is weaker in settings where food aid disrupts the relationship between predictors and food security outcomes. Third, the approach developed in this paper may not be applicable in settings of extreme food insecurity or famine, especially when this occurs as a result of armed conflict. Some secondary data (such as price data) may not be available during episodes of conflict. Hence, the proposed prediction approach is intended to complement, but not replace, food security monitoring via convergence-of-evidence and scenario development, as conducted by e.g. the Famine Early Warning Systems Network (FEWS NET) and the Integrated Food Security Phase Classification (IPC).

Despite these limitations, this study underscores the considerable potential of machine learning methods and secondary data for improving the measurement of food security in low-income countries. Rapid advances in the academic literature in recent years, and institutions' nascent adoption of innovative approaches – e.g. the use of machine learning models to complement data from phone surveys in WFP's 'HungerMap' for real-time monitoring of food security – are promising developments for improving the targeting of food policy.

Appendices of Chapter 3

Appendix 3.A Food security measures

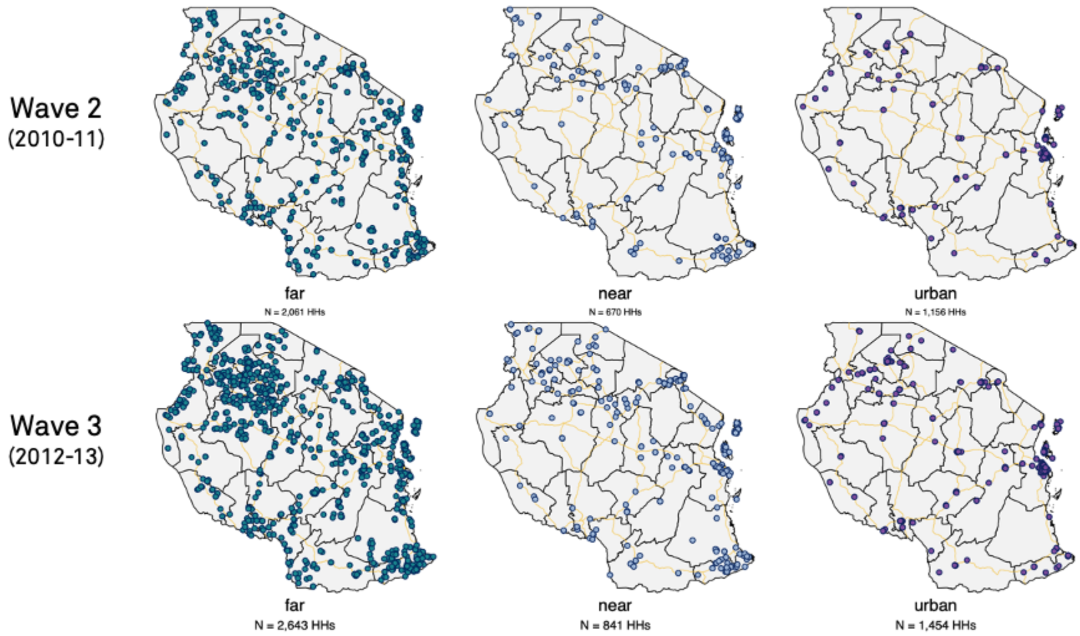
Table 3.1: Constituent items of rCSI and FCS

Indicator	Item name	Values	Item description
rCSI	rcsi_less_preferred	0-3	In the past 7 days, did your household rely on less preferred foods?
	rcsi_limit_variety	0-3	In the past 7 days, did your household limit the variety of foods eaten?
	rcsi_limit_portion	0-3	In the past 7 days, did your household limit portion size at meal-times?
	rcsi_fewer_meals	0-3	In the past 7 days, did your household reduce the number of meals eaten in a day?
	rcsi_restrict_adult	0-3	In the past 7 days, did your household restrict consumption by adults for small children to eat?
	rcsi_borrow_food	0-3	In the past 7 days, did your household borrow food, or rely on help from a friend or relative?
FCS	fcs_staples	0-7	Frequency of staples consumption in last 7 days
	fcs_pulses	0-7	Frequency of pulses consumption in last 7 days
	fcs_vegetables	0-7	Frequency of vegetables consumption in last 7 days
	fcs_fruits	0-7	Frequency of fruits consumption in last 7 days
	fcs_meat	0-7	Frequency of meat consumption in last 7 days
	fcs_dairy	0-7	Frequency of dairy consumption in last 7 days
	fcs_oil	0-7	Frequency of oil consumption in last 7 days
	fcs_sugar	0-7	Frequency of sugar consumption in last 7 days

Note: rCSI items take values 0-7, with the value indicating the number of days that a given coping strategy was used in the last 7 days. Raw scores are weighted using severity weights to derive overall rCSI score. FCS items take values 0-7, with the value indicating the number of days that a given food group was consumed in the last 7 days. Raw scores are weighted to account to nutritional density before aggregating to overall score.

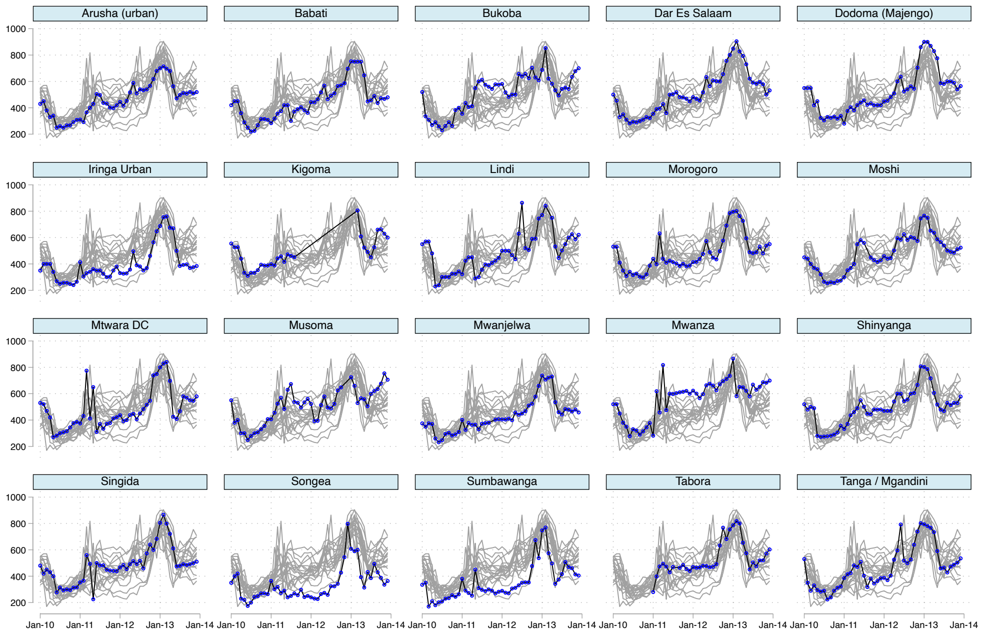
Appendix 3.B Household location by type of region

Figure 3.11: LSMS household locations by type of region



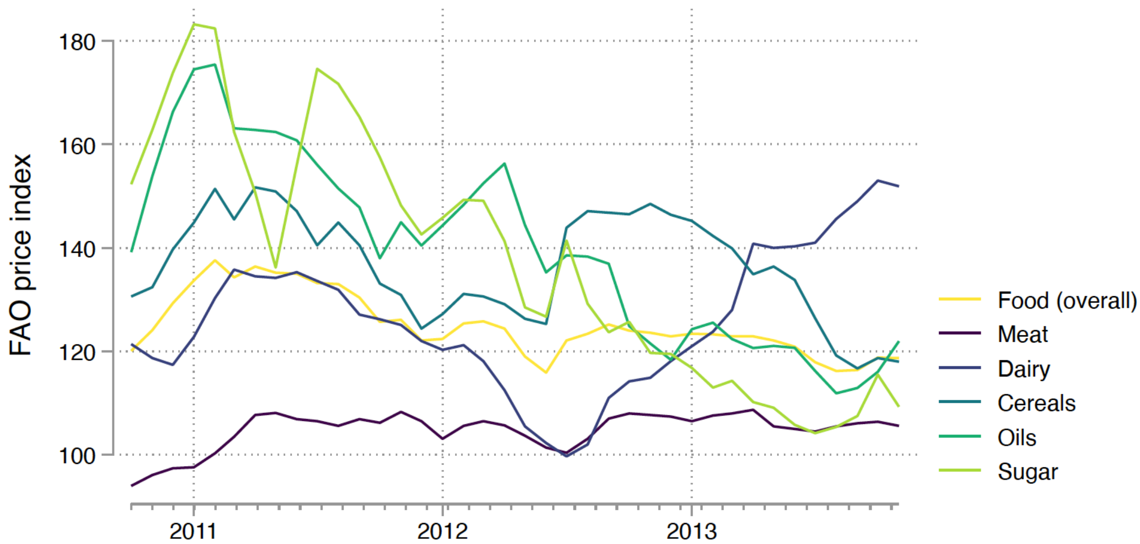
Appendix 3.C Food prices

Figure 3.12: Local market price of Maize, 2010-2013 (tsh/kg)



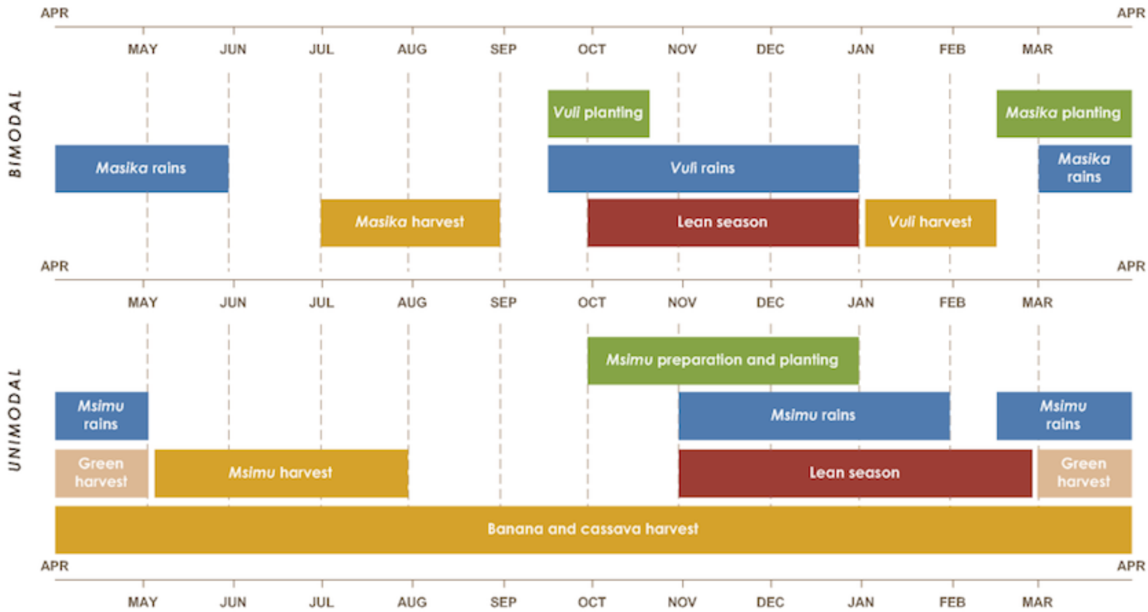
Source: WFP VAM's sub-national food price data

Figure 3.13: FAO Global food price indices



Appendix 3.D Seasonal calendar

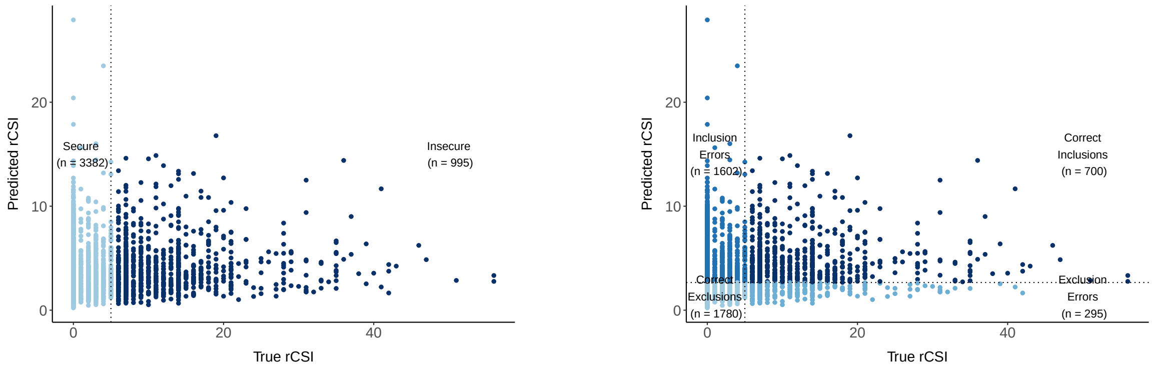
Figure 3.14: Tanzania Seasonal Calendar (FEWS NET)



Source: https://fews.net/east-africa/tanzania

Appendix 3.E Additional prediction results

Figure 3.15: Prediction of rCSI values with HH demographics in predictor set



Note: This figure plots true and predicted rCSI values by true food security status for households in the test dataset (LSMS wave 3, 2012/13). The x-axis shows true values; the y-axis shows predicted values. The left scatterplot color-codes households as food secure (rCSI below 5) or insecure based on observed rCSI values. The right scatterplot shows how setting the food insecurity cutoff to 2.5 for rCSI to achieve 70% recall affects errors of inclusion. In contrast to the main model presented in Section 3.5, the model used here includes household demographics as additional predictors.

Conclusion

Food production and consumption play a key role in economic development and consequently receive particular attention from policy makers, with food security ranking highly among the 17 Sustainable Development Goals (Goal 2: Zero Hunger by 2030). Domestic transport costs remain particularly high in low-income countries (Moszoro & Soto, 2022). This thesis speaks to the intersection of the centrality of food and the salience of domestic trade frictions in low-income countries. In this concluding section, I discuss potential future extensions for each of the three chapters.

In Chapter 1, I develop a stylised spatial general equilibrium model of a small open agricultural economy to study processes of structural change in open economies with high domestic trade frictions. I find that in such an imperfectly open economy, the food problem can continue to bind, causing a high share of the population to remain in agriculture despite low agricultural productivity; this is true even though the economy has the ability to import food from the rest of the world. Results of the comparative static analysis show that productivity improvements in the transport sector, agricultural sectors and non-agricultural sectors can all make the food problem less binding, but these differ with respect to their effects on trade patterns, the spatial distribution of economic activity, and food consumption.

To ensure analytical tractability of the model, I make assumptions about the tradability of different goods. While in the current version of the model the only export is an agricultural good, an interesting future extension would be to allow for the urban manufacturing good to be exported. Other useful modifications could include the addition of a second far rural region, or to allow for food processing to take place in the near rural region. I presume that the parameterisation would again be critical for model predictions with such extensions. For example, depending on global prices and the degree of substitutability between domestic and imported food, and assuming that urbanisation is not restricted (e.g. by congestion externalities or migration frictions), the economy's ability to export the urban manufactured good could considerably ease the pressures exerted by the food problem.

In Chapter 2, we analyse the spatially differentiated pass-through of global food, fuel and fertiliser price shocks to heterogeneous households in the domestic economy. We find that urban households, whose food consumption is more heavily dependent on imports, are more adversely affected by isolated food price shocks. When food price shocks occur simultaneously with fuel and fertiliser price shocks, remote rural households are equally negatively affected, due to the transport-cost intensity of their consumption and the reliance on fertiliser in agricultural production.

In our current policy analysis we assume import price subsidies are raised to a level at which they fully offset the global price shocks. In reality, neither donors nor domestic authorities are likely to provide the financing required to fund such 'full mitigation' subsidies. In the absence of full mitigation, we cannot therefore fully neutralise the changes in relative prices through subsidy offsets, and so the distinction between subsidies and transfers will diminish in this case. In ongoing work, we solve a dynamic version of the model that incorporates realistic levels of adjustment costs and intertemporal smoothing. The dynamic model will facilitate analysis of temporary price shocks and more realistic policy counterfactuals.

Chapter 3 is concerned with accurate targeting of food policy. I show that household food security can be predicted at reasonable levels of accuracy using machine learning methods and secondary data. A remaining limitation of the proposed methodology is the reliance on household asset indicators which – while they could conceivable be accessed through remote sensing or provided by e.g. cellular companies – are currently still taken from household survey data. In addition, the prediction models do not currently account for existing food aid which may affect the relationship between the predictors and food security outcomes. Lastly, the approach is likely not well suited to predicting acute food insecurity in conflict settings, as some of the secondary data (such as price data) may not be accessible. During such episodes of extreme food insecurity, there may be breaks in the relationship between predictors and foot security outcomes.

Despite these limitations, food security prediction using machine learning holds considerable potential for improving policy targeting in developing countries. When used as a complement to food security monitoring via rapid phone surveys, convergence-of-evidence and scenario development, as conducted by e.g. the Famine Early Warning Systems Network (FEWS NET) and the Integrated Food Security Phase Classification (IPC), machine learning methods can be a powerful way to fill gaps in the data.

References

Adam, C., Bevan, D., & Gollin, D. (2018). Rural–urban linkages, public investment and transport costs: The case of Tanzania. *World Development*, *109*, 497–510.

Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, *603*, 864–870.

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

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, *129*(3), 1085–1139.

Amaglobeli, D., Hanedar, E., Hong, G. H., & Thévenot, C. (2022). *Fiscal policy for mitigating the social impact of high energy and food prices* (IMF Note No. 2022/001). International Monetary Fund. Washington, DC.

Andrée, B. P. J., Chamorro, A., Kraay, A., Spencer, P., & Wang, D. (2020). Predicting food crises. *World Bank Policy Research Working Paper*, *9412*.

Arndt, C., Benfica, R., Maximiano, N., Nucifora, A. M. D., & Thurlow, J. T. (2008). Higher fuel and food prices: Impacts and responses for Mozambique. *Agricultural Economics*, *39*, 497–511.

Arndt, C., Diao, X., Dorosh, P., Pauw, K., & Thurlow, J. (2023). The Ukraine war and rising commodity prices: Implications for developing countries. *Global Food Security*, *36*. https://doi.org/10.1016/j.gfs.2023.100680

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

Balashankar, A., Subramanian, L., & Fraiberger, S. P. (2023). Predicting food crises using news streams. *Science Advances*, *9*(9). https://doi.org/10.1126/sciadv.abm3449

Baptista, D., Farid, M., Fayad, D., Kemoe, L., Lanci, L., Mitra, P., Muehlschlegel, T., Okou, C., Spray, J., Tuitoek, K., & Unsal, F. (2022). *Climate change and chronic food insecurity in sub-Saharan Africa* (Departmental Paper Series No. DP/2022/016). International Monetary Fund. Washington, Dc.

Baptista, D., Spray, J., & Unsal, D. F. (2023). Coping with climate shocks: Food security in a spatial framework. *Working Paper*.

Barrett, C. (2002). Food security and food assistance programs. *Handbook of Agricultural Economics*, *2*.

Barrett, C. (2010). Measuring food insecurity. *Science*, *327*, 825–828.

Bekkers, E., Francois, M. B. J., & Yang, F. (2017). Local food prices and international price transmission. *World Development*, *96*, 216–230.
Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*, 1073–1076.
Brown, M. E., & Funk, C. C. (2008). Food security under climate change. *Science*, *319*, 580–581.
Burgess, R., & Donaldson, D. (2012). Railroads and the demise of famine in colonial India. *Working Paper*.
Burke, M., Bergquist, L. F., & Miguel, E. (2019). Sell low and buy high: Arbitrage and local price effects in Kenyan markets. *Quarterly Journal of Economics*, *134*, 785–842.
Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*(6), 1320–1365.
Caldwell, J. C., & Caldwell, P. (1992). Mortality and society in Sub-Saharan Africa. In E. Van De Walle, G. Pison, & M. Sala-Diakanda (Eds.). Clarendon Press.
Cudjoe, G., Breisinger, C., & Diao, X. (2010). Local impacts of a global crisis: Food price transmission, consumer welfare and poverty in Ghana. *Food Policy*, *35*, 294–302.
De Waal, A. (2018). *Mass starvation: The history and future of famine*. Cambridge.
Deléglise, H., Interdonato, R., Bégué, A., Maître d'Hôtel, E., Teisseire, M., & Roche, M. (2022). Food security prediction from heterogeneous data combining machine and deep learning methods. *Expert Systems with Applications*, *190*. https://doi.org/10.1016/j.eswa.2021.116189
Dithmer, J., & Abdulai, A. (2017). Does trade openness contribute to food security? A dynamic panel analysis. *Food Policy*, (69), 218–230.
Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899–934.
Emerick, K. (2018). Agricultural productivity and the sectoral reallocation of labor in rural India. *Journal of Development Economics*, *135*, 488–503.
Engel, E. (1857). Die vorherrschenden Gewerbszweige in den Gerichtsämtern mit Beziehung auf die Productions und Consumtionsverhältnisse des Königreichs Sachsen. *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Ministeriums des Innern*.
Fajgelbaum, P., & Redding, S. (2022). Trade, structural transformation, and development: Evidence from Argentina 1869–1914. *Journal of Political Economy*, *130*(5), 1249–1318.
FAO. (2006). *Food security* (Policy Brief No. 2). Food and Agriculture Organization (FAO) of the United Nations.
Farrokhi, F., & Pellegrina, H. S. (2023). Trade, technology, and agricultural productivity. *Journal of Political Economy*, *131*(9), 2289–2595.
Fogel, R. W. (1992). Nutrition and poverty. In S. R. Osmani (Ed.). Oxford University Press.

Gao, P., & Lei, Y.-H. (2021). Communication infrastructure and stabilizing food prices: Evidence from the telegraph network in China. *American Economic Journal: Applied Economics*, *13*(3), 65–101.

Gholami, S., Knippenberg, E., Campbell, J., Andriantsimba, D., Kamle, A., Parthasarathy, P., Sankar, R., Birge, C., & Ferres, J. L. (2022). Food security analysis and forecasting: A machine learning case study in southern Malawi. *Data & Policy*, *4*. https://doi.org/10.1017/dap.2022.25

Gollin, D., Parente, S., & Rogerson, R. (2002). The role of agriculture in development. *American Economic Review (Papers and Proceedings)*, *92*(2), 160–164.

Gollin, D., Parente, S., & Rogerson, R. (2007). The food problem and the evolution of international income levels. *Journal of Monetary Economics*, *54*, 1230–1255.

Gollin, D., & Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, *107*, 38–48.

Gollin, D., & Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. *Journal of Political Economy*, *129*(1).

Hakobyan, S., Fuje, H., Rawlings, H., Reyes, I., & Zhang, Q. (2022). *Building a more food-secure sub-Saharan Africa* (In Regional Economic Outlook: Sub-Saharan Africa—Living on the Edge). International Monetary Fund. Washington, DC.

IMF. (2022). *Request for a 40-month arrangement under the extended credit facility* (Country Report No. 22/269). International Monetary Fund. Washington, DC.

Janssens, C., Havlík, P., Krisztin, T., Baker, J., Frank, S., Hasegawa, T., Leclère, D., Ohrel, S., Ragnauth, S., Schmid, E., Valin, H., Lipzig, N. V., & Maertens, M. (2020). Global hunger and climate change adaptation through international trade. *Nature Climate Change*, *10*, 829–835.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794.

Krishnamurthy, P., Lewis, K., & Choularton, R. (2014). A methodological framework for rapidly assessing the impacts of climate risk on national-level food security through a vulnerability index. *Global Environmental Change*, *25*, 121–132.

Lentz, E., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, *122*, 399–409.

Leroy, J. L., Ruel, M., Frongillo, E. A., Harris, J., & Ballard, T. J. (2015). Measuring the food access dimension of food security: A critical review and mapping of indicators. *Food and Nutrition Bulletin*, *36*(2), 167–195. https://doi.org/10.1177/0379572115587274

Lofgren, H., Harris, R. L., & Robinson, S. (2001). *A standard computable general equilibrium (cge) model in gams* (TMD discussion papers No. 75). International Food Policy Research Institute (IFPRI).

Manikas, I., Ali, B., & Sundarakani, B. (2023). A systematic literature review of indicators measuring food security. *Agriculture & Food Security*, *12*(10). https://doi.org/10.1186/s40066-023-00415-7

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., & Omodei, E. (2022). Machine learning can guide food security efforts when primary data are not available. *Nature Food*, 716–728. https://doi.org/10.1038/s43016-022-00587-8

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of Economic Theory*, *58*, 317–334.

Maxwell, D. (2019). *Famine early warning and information systems in conflict settings: Challenges for humanitarian metrics and response* (tech. rep.). Conflict Research Programme.

Maxwell, D., Coates, J., & Vaitla, B. (2013). *How do indicators of household food insecurity measure up? an empirical comparison from ethiopia* (tech. rep.). Feinstein International Center, Tufts University: Medford, USA.

Maxwell, D., Coates, J., & Vaitla, B. (2014). Machine learning can guide food security efforts when primary data are not available. *Food Policy*, 107–116. https://doi.org/10.1016/j.foodpol.2014.04.003

McCaig, B., & Pavcnik, N. (2013). Moving out of agriculture: Structural change in Vietnam. *NBER Working Paper*, *19616*.

MetOffice. (2015). *Food insecurity and climate change technical report* (tech. rep.). Met Office Hadley Centre for Climate Science and Services.

Meyimdjui, C., & Combes, J.-L. (2021). *Food price shocks and household consumption in developing countries: The role of fiscal policy* (IMF Working Paper, Strategy, Policy, & Review No. WP/21/12). International Monetary Fund. Washington, DC.

Minot, N., & Dewina, R. (2015). Are we overestimating the negative impact of higher food prices? Evidence from Ghana. *Agricultural Economics*, *46*, 579–593.

Moser, C., Barrett, C., & Minten, B. (2009). Spatial integration at multiple scales: Rice markets in Madagascar. *Agricultural Economics*, *40*(3), 281–294.

Moszoro, M., & Soto, M. (2022). *Road quality and mean speed score* (IMF Working Papers No. 2022/095). International Monetary Fund. Washington, DC.

Nath, I. B. (2020). The food problem and the aggregate productivity consequences of climate change. *NBER Working Paper*, *27297*.

Ngai, R., & Pissarides, C. (2007). Structural change in a multisector model of growth. *American Economic Review*, *97*(1).

Okou, C., Spray, J., & Unsal, D. F. (2022). *Staple food prices in sub-Saharan Africa: An empirical assessment* (IMF Working Papers, Research Department No. WP/22/135). International Monetary Fund. Washington, DC.

Ravallion, M. (1997). Famines and economics. *Journal of Economic Literature*, *35*(3), 1205–1242.

Regmi, A., Gehlhar, M., Wainio, J., Vollrath, T., Johnston, P., & Kathuria, N. (2005). *Market access for high-value foods* (Agricultural Economic Report No. AER-840). U.S. Department of Agriculture, Economic Research Service.

Rother, B., Sosa, S., Kohler, L., Pierre, G., et al. (2022). *Tackling the global food crisis: Impact, policy response, and the role of the IMF?* (IMF Note No. 2022/004). International Monetary Fund. Washington, DC.

Schultz, T. (1953). *The economic organization of agriculture*. McGraw Hill.
Sen, A. (1980). Famines. *World Development*, *8*, 613–621.
Sen, A. (1983). *Poverty and famines: An essay on entitlement and deprivation*. Oxford University Press.
Simola, A., Boysen, O., Ferrari, E., Nechifor, V., & Boulanger, P. (2022). Economic integration and food security – The case of the AfCFTA. *Global Food Security*, *35*, 100651. https://doi.org/10.1016/j.gfs.2022.100651
Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, *128*(7), 2690–2738.
Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, *79*, 159–209.
Swinnen, J., & McDermott, J. (2020). COVID-19 and global food security [Online; accessed 27 January 2022].
Thurlow, J. (2021). *2018 Social Accounting Matrix for Tanzania: A Nexus Project SAM* (Data Paper). International Food Policy Research Institute (IFPRI). Washington, DC. https://doi.org/10.2499/p15738coll2.134808
Tombe, T. (2015). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, *7*(3), 226–258.
United Nations. (1996). World Food Summit concludes in Rome | Meetings coverage and press releases [Accessed: 2022-02-02].
Vaitla, B., Coates, J., Glaeser, L., Hillbruner, C., Biswal, P., & Maxwell, D. (2017). The measurement of household food security: Correlation and latent variable analysis of alternative indicators in a large multi-country dataset. *Food Policy*, *68*, 193–205. https://doi.org/10.1016/j.foodpol.2017.02.006
Vaitla, B., Coates, J., & Maxwell, D. (2015). *Comparing household food consumption indicators to inform acute food insecurity phase classification* (FHI 360). Food and Nutrition Technical Assistance III Project (FANTA), Washington, DC.
Van Campenhout, B. (2007). Modelling trends in food market integration: Method and an application to Tanzanian maize markets. *Food Policy*, *32*, 112–127.
Verschuur, J., Li, S., Wolski, P., & Otto, F. E. L. (2021). Climate change as a driver of food insecurity in the 2007 Lesotho-South Africa drought. *Sci Rep*, *11*, 3852.
Villacis, A. H., Badruddoza, S., Mishra, A. K., & Mayorga, J. (2023). The role of recall periods when predicting food insecurity: A machine learning application in Nigeria. *Global Food Security*, *36*. https://doi.org/10.1016/j.gfs.2023.100671
Watkins, S. C., & Menken, J. (1985). Famines in historical perspective. *Population and Development Review*, *11*(4), 647–675.
WFP. (2022). *Tanzania country brief* (December 2022). World Food Program.
Willenbockel, D., & Robinson, S. (2014). Global food security modelling: A review of approaches and results. *Report for the Department for Environment, Food and Rural Affairs*.
Wollburg, P., Bentze, T., Lu, Y., Udry, C., & Gollin, D. (2024). Crop yields fail to rise in smallholder farming systems in sub-Saharan Africa. *Proceedings of the National Academy of Sciences*, forthcoming.

Zhou, Y., Lentz, E., Michelson, H., Kim, C., & Baylis, K. (2021). Machine learning for food security: Principles for transparency and usability. *Applied Economic Perspectives and Policy*. https://doi.org/10.1002/aepp.13214