Inequality, Redistribution and Optimal Trade Policy: A Public Finance Approach*

Roozbeh Hosseini
University of Georgia
roozbeh@uga.edu

Ali Shourideh
Carnegie Mellon University
ashourid@andrew.cmu.edu

February 15, 2024

Abstract

In this paper, we explore the relationship between optimal trade and redistributive policies when the gains from trade are unequally distributed. We use a competitive trade model with input-output linkages where trade affects relative wages and the reallocation of workers across various sectors is frictional. We study how income taxes and trade policies should be designed in order to balance the efficiency gains from trade with the costs associated with the resulting increased inequality. We show that for a large class of global production structures, the global trade of goods and services must be undistorted even when personal taxes are incomplete. In other words, barriers to trade such as tariffs are never optimal. In contrast, producer taxes in the form of value-added taxes (VAT) that are differentially levied on different sectors play a crucial role in redistributing the gains from trade. We provide formulas that highlight the main determinants of optimal VAT and non-linear income taxes. Finally, in a quantitative version of our model, we study the optimal response to the rise of China in international trade. Our quantitative analysis establishes that differential VAT taxes play the main role of redistributing the gains from trade, while income taxes do not.

*We would like to thank Laurence Ales, Andy Atkeson, Mostafa Beshkar, Ariel Burstein, V. V. Chari, Pablo Fajgelbaum, Hugo Hopenhayn, Brian Kovak, Ahmad Lashkaripour, Chris Sleet, Florian Scheuer, Aleh Tsyvinski, Jonathan Vogel, and various conference and seminar participants. We specially thank Ivan Werning and Arnaud Costinot for their extensive comments on an earlier version of our paper. We are also thankful to Simon Galle, Andres Rodriguez-Clare and Moises Yi for sharing their data and codes with us. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

Recent evidence suggests that international trade and global reallocation of production has significantly changed the allocation of labor and inequality of income in the United States.[1] According to Autor et al. (2013), the rising import competition from China has led to a significant decline in manufacturing employment in the United States. Under this narrative, the import competition stemming from China's relatively high productivity in manufacturing together with the inability of workers to move to sectors or locations with comparative advantage (due to the specialization of their skills) is at the root of this problem. Given such distributional effects of free trade and modern governments' desire for redistribution, an important question arises: How should a redistributive policy be designed to distribute these unequal gains from trade? In particular, as governments have a variety of policy instruments at their disposal (i.e., tariffs, income taxes, production taxes, etc.), what policy instruments should be used to achieve this goal? In this paper, we set out to answer these questions from a theoretical and quantitative perspective.

We study a multi-country competitive model of trade where global production occurs through input-output linkages and workers exhibit skill specialization and imperfect mobility. In this environment, we show that for a large class of production functions, tariffs and other distortions to free trade are not optimal. Instead, sector-specific VAT taxes should be used to redistribute the gains from the global reallocation of production. Finally, we provide a quantitative assessment of the government response in a version of the model calibrated to the recent rise of China in international trade. Using this quantitative framework, we show that, unlike VAT, income taxes do not play an essential role in redistributing the gains from trade.

In our model, production is done competitively across countries where goods produced in one country can be used for production in others. Thus, production occurs through a network of input-output linkages across locations. We assume that workers are heterogeneous in terms of their ability to move across production units, à la Roy (1951). In particular, we assume that the population of workers is divided into groups (e.g., differing by their skills), each of which has a pattern of sector-specific productivities. As a result, their sectoral choice is determined by this pattern together with wages and an idiosyncratic productivity shock. We assume that that these idiosyncratic productivity shocks have a type-2 extreme value distribution, which allows us to characterize the elasticity matrix of workers sectoral choice in response to changes in sectoral wages in a tractable fashion.

We use our model to study the joint determination of the optimal trade policy and redistributive income taxes. The key restriction is that personal income taxes can depend only on income

[1]There is a large literature trying to understand the effect of trade on inequality in developed and developing countries. Papers include Goldberg and Pavcnik (2007), Verhoogen (2008), Helpman et al. (2010), Lagakos and Waugh (2013), Helpman et al. (2017) , Artuç et al. (2010), and Caliendo et al. (2017), among many others.

and not on other characteristics of workers. In order for countries to be able to fully realize the global gains from trade, we consider the determination of trade under cooperation among countries. That is, we characterize the world Pareto frontier when the objective function within each country exhibits redistributive motives. One can interpret these policies as being determined under a trade agreement where the outcome of the negotiations is efficient.

In this environment, we first investigate the optimality of free trade or efficiency of the allocation of goods across countries. We show that in our benchmark model, the allocation of intermediate inputs across countries and sectors is efficient. We refer to this efficiency as *pseudo-efficiency of production* since we do not allow for reallocation of labor across sectors (in contrast with the standard notion of production efficiency). This result resembles the seminal production efficiency result of Diamond and Mirrlees (1971), despite that at the personal level, taxes are incomplete and thus the assumptions of Diamond and Mirrlees (1971) do not hold in this framework. Intuitively, we obtain this result because in our baseline model, the intermediate inputs do not directly affect the distribution of labor productivity in the economy. Since the distribution of labor productivity determines the deadweight loss of income taxation, it is efficient not to distort production. Our analysis further identifies the conditions under which the undistorted allocation of goods across sectors fails to be optimal. Such allocation occurs when the degree of substitution between labor and other factors (intermediate inputs) varies for workers with different skills.

Having established the pseudo-efficiency of production, we study the best mix of instruments that can redistribute the gains from trade. We show that, in general, using producer taxes is a powerful tool to redistribute the gains from trade. These taxes should take the form of a sector-specific value-added tax under which firms are allowed to deduct the cost of intermediate inputs from their tax bill. We show that when the labor force exhibits skill specialization (i.e., when groups of workers differ in terms of their comparative advantage across sectors), VAT taxes must be different across sectors. We provide formulas that describe the key components of optimal VAT taxes. These components include the social marginal value of increasing the income of workers in a sector, the fiscal externality of raising VAT taxes on income tax revenue, and the sectoral relocation effect of raising VAT taxes. In addition to VAT taxes, income taxes remain a viable tool to redistribute the gains from trade.

While our theoretical analysis identifies VAT taxes as a potential policy instrument to redistribute the gains from trade, its importance compared to that of income taxes is a quantitative question. To answer it, we use the recent rise of China in global production (*the China shock*) as an example to test the importance of each policy instrument. In particular, we consider a variant of the quantitative framework developed by Galle et al. (2017)[2] to measure the distribution of

[2]Our main difference is that we allow for an endogenous intensive margin of labor supply and government policies. Adjusting for these differences, we obtain welfare calculations that broadly line up with theirs.

changes in welfare caused by the China shock across different groups of workers in the United States which are distinguished by their Commuting Zone (see Autor et al. (2013)) and education. In our baseline quantitative exercise, we look for optimal policies that compensate the losers from the China shock while keeping the winners as well off as they were prior to the shock. [3]

Our quantitative exercise provides two main insights: First, VAT taxes are an integral part of optimal policies to compensate groups of workers that lose from the China shock; second, optimal income taxes do not play a role in redistributing the gains from trade. Regarding optimal VAT taxes, our results suggest significant variation across sectors, with 'Textile & Leather' and 'Electrical Equipment' receiving large subsidies, while sectors such as 'Coke & Petroleum' and 'Transport Equipment' paying VAT taxes. In particular, the VAT taxes and subsidies range from -18% (subsidy to 'Textile & Leather') to 5% (tax on 'Chemicals').

The sign and the magnitude of the VAT taxes are fairly correlated with the employment change in that sector as well as the distribution of the welfare changes of the groups of workers concentrated in each sector. This is mainly because the data suggest there is significant specialization of skill among a select group of workers.

As for income taxes, our main result is that they do not play a role in redistributing gains from trade. In particular, we show that marginal income taxes are zero for all levels of income above 5% of the average income. This result is mainly because the distribution of welfare changes resulting from the China shock is almost uncorrelated with group incomes. In other words, the gains and losses from the rise of China are mainly determined by the patterns of skill specialization rather than income. Therefore, income taxes are not beneficial to redistributing the gains from the China shock.

Our results have very stark implications for the design and structure of trade agreements and domestic policy response to such agreements. VAT taxes are already prevalent in many countries, with the United States being an important exception. The European Union as a large free trade area already has a centralized VAT tax system which determines how much VAT different sectors in different countries must pay. The standard reasoning for using them is that they keep production efficient while they raise significant revenue. In our paper, we show that in addition to such benefits, when imposed differentially on different sectors, VAT taxes can also be used very effectively to redistribute the gains from trade across different groups of workers. Our analysis suggests that in the presence of distributional concerns in trade agreements, VAT taxes should be utilized to ensure that the gains from trade are distributed more equally across the population.

Additionally, our analysis identifies the conditions required for optimality of VAT taxes and free trade. As we show, optimality of free trade requires labor provided by different types of

[3] Additionally, we solve for optimal policies that ensure a least level of gain from the China shock for all groups of workers. Our results remain fairly unchanged.

workers to have the same degree of substitution with intermediate inputs. This implies that measuring the heterogeneity in substitution of intermediate inputs is a crucial element of trade policies and should be subject of further empirical studies.

1.1 Related Literature

Our paper builds on and makes a contribution to several strands of literature in public finance and international trade.

The public finance literature has classically studied the problem of the optimal design of direct and indirect taxes. This dates back to the work by Ramsey (1927) as well as later work by Diamond and Mirrlees (1971), Atkinson and Stiglitz (1976), and Deaton (1981), among many others. The seminal work by Diamond and Mirrlees (1971) has established the optimality of production efficiency under constant returns to scale and a rich tax structure. While later studies such as that of Naito (1999)[4] have shown that departures from production efficiency could be necessary when the tax structure is incomplete, the precise theoretical and quantitative determinants of optimal distortions to production are still not fully explored.[5] To our knowledge, our paper is one of the few that study the determinants of optimal producer taxes in an economy with incomplete direct taxes, specifically, an economy where income taxes do not depend on workers' sectoral choice. Our main contribution to this literature is the identification of the main force that leads to the failure of production efficiency; see Proposition 4 below. Our analysis provides a condition that can be measured and tested empirically. Moreover, we provide a framework that encapsulates quantitative models of international trade and thus allows us to quantitatively assess the importance of various policies.

As our model includes a discrete choice by workers (i.e., the workers' sectoral choice), our paper builds on the literature on optimal income taxation in presence of sectoral choice. These papers include Saez (2002), Rothschild and Scheuer (2013), and Gomes et al. (2017). Our main contribution to this line of research is providing a fairly tractable framework using discrete choice with extreme value distribution.

A recent paper by Costinot and Werning (2018) investigates a similar question to ours. They develop a model in which new technologies that do not employ labor are introduced into an economy that is otherwise similar to our model – a special case of this can be thought of as the case of a small open economy. They show that in such an economy production inefficiency arises similar to Naito (1999) and provide novel formulas that govern optimal taxes and distortions to

[4]For other papers along this line of research, see Guesnerie (1998), Spector (2001), and Jacobs (2015).

[5]Saez (2004) argues that in the long-run, workers switch to sectors with higher wages and thus there is no need for distortions to production. This is in contrast with Naito (1999)'s result on lack of production efficiency. In a world where the technology frontier keeps expanding, it is hard to imagine that such differentials diminish even in the long-run.

production. They map these formulas into sufficient statistics that are measurable in the data and calculate optimal values of distrotions to production efficiency. In their environment, new technologies can have differential effect on workers wages and thus production distortions can be used to reduce the deadweight loss of income taxation. In contrast, our analysis highlights that unequal gains from trade (or new technologies) do not necessarily lead to inefficiency of production, since they could be caused by specialization of the labor force. As a result VAT taxes can be an integral part of redistributing the gains from trade and keeping production efficient. Furthermore, in our quantitative exercise, the network structure of production is used to design redstributive VAT taxes.

Moreover, our paper is related to a more recent literature on optimal taxation in trade and spatial models (e.g., Costinot and Werning (2018), Lyon and Waugh (2017), Antràs et al. (2017), Fajgelbaum and Gaubert (2018), and Ales and Sleet (2017)). In particular, Lyon and Waugh (2017) study an optimal income taxation problem in a small open economy where prices change exogenously and workers move across sectors. They show that progressive income taxes can be used to redistribute the gains from trade. In Antràs et al. (2017), trade increases inequality and redistributive income taxes adjust to mitigate this rise in inequality. Consequently, a more progressive income taxes leads to distortions to trade and reduces welfare gains from trade. Our focus instead is on a more general class of policy instruments, namely we allow for policies that directly affect producers. In particular, we show that the gains from trade can be redistributed using VAT taxes on producers. In Fajgelbaum and Gaubert (2018) and Ales and Sleet (2017) the role of taxes is to correct externalities.

While our paper is somewhat related to the literature on optimal trade policy (e.g., Bagwell and Staiger (1999), Opp (2010), Costinot et al. (2015), and Beshkar and Lashkaripour (2017) among others), our paper is different in that this literature often assumes strategic motives of the government as the main determinant of trade policies. As a result, these policies help governments manipulate their terms of trade in their favor. In our setup, this motive is shut down intentionally to focus on optimal policies under cooperation. Thus, our paper is mainly related to the design of trade agreements and discussion of the type of industrial policies allowed under such agreements. In the trade literature, our paper is also related to the work by Dixit and Norman (1980) and Dixit and Norman (1986), who show that it is always possible to redistribute gains from trade with lump-sum and proportional taxes. While their analysis is insightful, it does not identify the main policy instruments that are required to redistribute gains from trade. In contrast, we provide a theoretical and quantitative analysis of optimal policy instruments which are beneficial in redistributing gains from trade.

Our quantitative exercise builds upon recent models in the international trade literature that try to measure and account for differential changes in welfare and employment caused by international trade (with a focus on the recent rise of China), as documented by Autor et al. (2013). These include Caliendo et al. (2015) and Galle et al. (2017). More specifically, we use a variant of the model in Galle et al. (2017) to perform our quantitative exercise. To do so, we allow for the intensive margin of labor supply in their environment as well as redistributive taxes.

The rest of the paper is organized as follows: In section 2 we describe a general model of trade with imperfect mobility. In section 3 we describe and characterize the optimal policy problem in such a model, and in section 4, we discuss the determinants of optimal policies. In section 5, we present our quantitative results. We conclude in section 6.

2 A Model of Trade with Imperfect Mobility

In this section, we provide the basic framework of analysis in this paper. This framework is based on an economy where workers' movement across sectors is determined by their sector specific productivity, as in the model proposed by Roy (1951). This framework allows us to provide our basic theoretical result on optimal trade policy and optimal income taxes.

Geography. Production and consumption occur in C countries, and each country is represented by $c \in \{1, \cdots, C\}$. In each country there is a unit continuum of workers who are potentially heterogeneous with respect to their working opportunities, as we clarify below.

Production. There are a total of N goods produced globally. In particular, suppose that each good, $i \in \{1, \cdots, N\}$, can be produced in country c according to the production function

$$Y_i^c = G_i^c \left(L_i^c, \{Q_{ij}^c\}_{j=1}^N \right),$$

where L_i^c is the total effective units of labor in this sector, and Q_{ij}^c is the total amount of good j used in production of good i in country c. We assume that the above production function exhibits constant returns to scale with respect to all factors, the marginal product of each intermediate good converges to infinity as its quantity goes to zero, and G_i^c is strictly concave with respect to all of its inputs. Moreover, the producers of each good are price takers; they take as given the prices of their products, their inputs, and the wages of their workers.

Workers' and Sectoral Choice. In each country c there is a unit continuum of workers.[6] They are assumed to have preferences over a vector of consumption goods $\mathbf{x} = (x_1, \cdots, x_N)$, where x_i is the consumption of good i. Aside from choosing a consumption bundle $\mathbf{x}$, workers can work

[6]We can also work with a version of our model where countries are heterogeneous in their population.

in firms that produce a good j and choose how much effort to put in their work. This choice of working status depends on the workers' types and their idiosyncratic productivity shock. In particular, a worker draws a type $\theta \in \Theta$, which is distributed according to probability measure μ^c. A worker of type θ can choose to work in sector j with labor productivity given by $z = a_j^c(\theta)\,\epsilon_j$, where ϵ_j is a random variable that is distributed i.i.d. and according to a Fréchet distribution given by

$$H_j^c(\epsilon) = \Pr(\epsilon_j \le \epsilon) = e^{-\epsilon^{-\sigma^c}}$$

for $\sigma^c > 0$. In other words, a worker is identified by a group θ and idiosyncratic productivity type $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_N)$. The group θ can be associated with education, location, and other characteristics of a worker. We let the c.d.f. of $\boldsymbol{\epsilon}$ be given by $H^c(\boldsymbol{\epsilon}) = \prod_{j=1}^N H_j^c(\epsilon_j)$.

Finally, the utility of a worker that consumes a bundle $\mathbf{x}$ and supplies ℓ units of labor and works in sector j is given by

$$U^c(\mathbf{x}) - v^c(\ell)\,. \tag{1}$$

We assume that workers' preferences satisfy the following properties:

Assumption 1. *The prefences of workers in country c satisfy the following:*

1. *The utility from consumption, $U^c(\mathbf{x})$ is strictly concave, homothetic, and increasing. Moreover, its implied indirect value function is linear in wealth.*
2. *The disutility of effort $v^c(\cdot)$ satisfies $v^c(\ell) = \frac{\ell^{1+1/\varepsilon^c}}{1+1/\varepsilon^c}$.*
3. *The parameter governing the distribution of ϵ satisfies $\sigma^c > 1 + \varepsilon^c$.*

The above assumptions on preferences are mainly made for tractability of the analysis. The homotheticity assumption implies that Atkinson and Stiglitz (1976) and Deaton (1981)'s result on uniform commodity taxation applies and consumers in each country must pay the same taxes on consumption of different goods. Moreover, the fact that the indirect utility function associated with U^c is linear in wealth implies that there is no income effect which simplifies the analysis of optimal taxes.[7] Finally, the last part of the assumption ensures that the mean incomes are well-defined.[8]

The above assumptions about the structure of worker mobility are made to allow for a somewhat general structure yet maintain tractability. In particular, our specification of productivity is flexible enough to allow for various patterns of mobility:

[7] It is fairly straightforward to extend the results to non-homothetic preferences but we skip this for the ease of exposition.

[8] Without taxes, income within a group has a Fréchet distribution with shape parameter $\frac{\sigma^c}{1+\varepsilon^c}$ which has a well-defined mean if $\sigma^c > 1 + \varepsilon^c$.

1. Absolute Advantage: Suppose that $a_j^c(\theta) = \alpha_j^c \beta^c(\theta)$. That is, a worker of type θ, whose $\beta^c(\theta)$ is high, is on average more productive in all sectors. In this sense, workers with high $\beta^c(\theta)$ have an absolute advantage over those with low $\beta^c(\theta)$.

2. Specialization: Suppose that

$$a_j^c(\theta) = \alpha^c(\theta) - \psi^c (j - l^*(\theta))^2$$

   In this setting, a worker of type θ is on average most productive in sector $l^*(\theta) \in \{1, \cdots, N\}$. For sectors whose index is different from a worker's most productive sector, the worker experiences an average loss in productivity given by $\psi^c (j - l^*(\theta))^2$. The parameter ψ^c controls a worker's extent of specialization.

Additionally, the assumption of Fréchet distribution implies that the labor supply in each sector can be characterized in a fairly tractable way. We will discuss this tractability when we analyze the properties of equilibrium.

Markets. We assume that for each good $i \in \{1, \cdots, N\}$, there is a competitive market where the producers of this good in country c sell their product, while the producers of other goods and consumers purchase their demand from this market. The market price of good i is given by $\hat{p}_i$.

Seemingly, the assumption above suggests that all goods are traded in all countries and trade is free and, therefore, our specification cannot handle trade costs. As we will clarify later, this literal interpretation of the model is at best incomplete. In particular, we argue that our specification is general enough and can encompass a model with iceberg trade costs à la Samuelson, non-tradable goods, autarky, etc.

In addition to markets for goods and services, we also assume that in each country there are competitive labor markets. This implies that there is a wage associated with working in sector j in country c, which is given by $w_j^c, \forall j \in \{1, \cdots, N\}, c \in \{1, \cdots, C\}$.

Governments and Policies. We assume that in each country c, there is a government that has access to income taxes, producer taxes, and consumer taxes. In particular, we assume that the government in country c imposes a general non-linear tax on earnings given by $T^c(y)$ where y is a worker's labor income.

Moreover, upon consumption of good i by a consumer in country c, the government imposes an ad-valorem tax rate given by $t_i^{x,c}$. Finally, a producer of good i in country c faces a tax rate of $t_i^{p,c}$ on its revenue and a tax rate $t_{i,j}^{p,c}$ on its purchases of intermediate inputs of good j. For now, we assume that governments do not have any expenditure and therefore must have a balanced budget.

Given the above structure for the global economy, a competitive equilibrium given governments' policies can be defined as: (i) consumption, and leisure allocations together with sectoral choice by each worker in group θ and idiosyncratic type $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_N)$, $\left\{\hat{\mathbf{x}}^c(\theta,\boldsymbol{\epsilon}), \hat{\ell}^c(\theta,\boldsymbol{\epsilon}), j^c(\theta,\boldsymbol{\epsilon})\right\}_{\theta\in\Theta,\boldsymbol{\epsilon}\in\mathbb{R}_+^N}$; (ii) production in each sector, $L_i^c, \left\{Q_{ij}^c\right\}_{j=1}^N$; (iii) vector of prices, $\hat{\mathbf{p}} = (\hat{p}_1, \cdots, \hat{p}_N)$, and wages, $\mathbf{w}^c = \{w_i^c\}_{i=1}^N$, such that:

1. Workers in each country c maximize utility, taking as given prices and government policies:

$$\hat{\mathbf{x}}^c(\theta,\boldsymbol{\epsilon}), \hat{\ell}^c(\theta,\boldsymbol{\epsilon}), j^c(\theta,\boldsymbol{\epsilon}) \in \arg\max_{\mathbf{x},\ell,j} U^c(\mathbf{x}) - v^c(\ell)$$

   subject to

$$\sum_{i=1}^N \hat{p}_i\left(1+t_i^{x,c}\right)x_i \le w_j^c a_j^c(\theta)\epsilon_j\ell - T^c\left(w_j^c a_j^c(\theta)\epsilon_j\ell\right)$$

2. Firms maximize their profits, taking as given prices and government policies:

$$L_i^c, \left\{Q_{ij}^c\right\} \in \arg\max_{L,Q_{ij}} \hat{p}_i\left(1-t_i^{p,c}\right)G_i^c\left(L, \{Q_{ij}\}_{j=1}^N\right) - w_i^c L - \sum_{j=1}^N\left(1+t_{ij}^{p,c}\right)\hat{p}_j Q_{ij}$$

3. Government budget constraint holds:

$$\sum_{c=1}^C\sum_i \hat{p}_i t_i^{x,c} Y_i^c + \sum_{c=1}^C\sum_{i=1}^N\sum_{j=1}^N \hat{p}_j t_{ij}^{p,c} Q_{ij}^c + \sum_{c=1}^C\sum_{i=1}^N \hat{p}_i t_i^{x,c}\int_\Theta\int_{\boldsymbol{\epsilon}} x_i^c(\theta,\boldsymbol{\epsilon})\,dH^c(\boldsymbol{\epsilon})\,d\mu^c$$
$$+\sum_{c=1}^C\int_\Theta\int_{\boldsymbol{\epsilon}} T^c\left(w_{j^c(\theta,\boldsymbol{\epsilon})}^c a_{j^c(\theta,\boldsymbol{\epsilon})}^c(\theta)\epsilon_j\ell^c(\theta,\boldsymbol{\epsilon})\right)dH^c(\boldsymbol{\epsilon})\,d\mu^c = 0$$

4. Markets clear:

$$\sum_{c=1}^C\int_\Theta\int_{\boldsymbol{\epsilon}} x_i^c(\theta,\boldsymbol{\epsilon})\,dH^c(\boldsymbol{\epsilon})\,d\mu^c + \sum_{c=1}^C\sum_{k=1}^N Q_{ki}^c = \sum_{c=1}^C G_i^c\left(L_i^c, \left\{Q_{ij}^c\right\}\right), \forall i, c \tag{2}$$

$$L_i^c = \int_\Theta\int_{\mathbb{R}_+^N} a_{j^c(\theta,\boldsymbol{\epsilon})}^c(\theta)\,\epsilon_{j^c(\theta,\boldsymbol{\epsilon})}\ell^c(\theta,\boldsymbol{\epsilon})\,\mathbf{1}\left[j^c(\theta,\boldsymbol{\epsilon}) = i\right]dH^c(\boldsymbol{\epsilon})\,d\mu^c, \forall i, c \tag{3}$$

Note that in our definition of equilibrium, we have a single budget constraint for the governments. In other words, our equilibrium concept allows for transfers across countries. As a result and in general trade can be unbalanced. In our environment, imposing a budget balance introduces additional distortions, which we want to abstract from. In our quantitative exercise, we study the role of these inter-government transfers.

Generality of the Model. As we have claimed before, our model encompasses various versions of the models that are popular in the international trade literature. Here, we provide some remarks to illustrate this:

1. Iceberg trade costs: While we have assumed that markets are competitive, it is possible to map a trade model with iceberg costs into the setup above. In particular, consider a model with iceberg trade costs. In order to map such a model into our setup, we first extend the set of goods so that each good is produced in a different country. That is, we define an extended set of goods $\{1, \cdots, N\}$ and partition it into goods produced in each country given by $\mathcal{I}^c$ where

$$\bigcup_{c=1}^{C} \mathcal{I}^c = \{1, \cdots, N\}, \mathcal{I}^c \cap \mathcal{I}^{c'} = \emptyset.$$

Let $c^*(i)$ be the country in which good i is produced. Suppose that the iceberg cost of shipping good $i \in \mathcal{I}^c$ from country c to c' is given by $d_i^{c,c'}$ with $d_i^{c,c} = 1$ for all c and $i \in \mathcal{I}^c$. We then refer to $x_i^{c'}$, consumption of good i in country c', as this consumption including the iceberg trade cost. Similarly, $Q_{ji}^{c'}$ is defined by the intermediate input demand of sector j in country c' of good i which includes the iceberg cost $d_i^{c,c'}$. Given this relabeling, the utility functions and production functions in any country c' are given by

$$U^{c'}(\mathbf{x}) = U^{c'}\left(\frac{x_1}{d_1^{c^*(1),c'}}, \frac{x_2}{d_2^{c^*(2),c'}}, \cdots, \frac{x_N}{d_N^{c^*(N),c'}}\right)$$
$$Y_i^{c'} = G_i^{c'}\left(L_i^{c'}, \left\{\frac{Q_{ij}^{c'}}{d_j^{c^*(j),c'}}\right\}\right), \forall i \in \mathcal{I}^{c'}$$

Note that in the above we relied heavily on the extension of the set of goods so that each good is produced in a unique country. This is often possible in neoclassical models of trade. For example, in an Armington model, each country produces a set of differentiated products that are imperfect substitutes and hence, no extension is necessary. In Ricardian models, such as that of Dornbusch et al. (1977) or Eaton and Kortum (2002), the goods produced in each country are perfect substitutes and thus the set of goods can be easily relabeled so that certain classes of goods are perfectly substitutes in utility and production functions. For example, if the set of goods are cloth and wine and the set of countries are England and Portugal, we can extend the set of goods to be English Wine, Portuguese Wine, English Cloth and Portuguese Cloth. In this case, the wine and cloth produced in each country are perfect substitutes for all consumers and producers.

Note that the types of trade costs are technological as opposed to government imposed. In particular, the goal of the paper is to find the optimal government-imposed trade costs given the technological ones.

2. Non-tradable goods: It is fairly straightforward to see that the non-tradable goods are captured in our model by requiring certain goods to be produced and used (by consumers or producers) in one country. That is, the non-tradable goods do not enter the utility of workers and production function of firms in other countries.

3. Two layers of mobility: It is possible to extend the above model to allow for costless mobility across some sectors, similar to the models in Caliendo et al. (2017) and Galle et al. (2017). In particular, one can partition the set of goods in country c into sectors $\mathcal{I}_i^c, i = 1, \cdots, N_J^c$ where $\cup_{i=1}^{N_J^c} \mathcal{I}_i^c = \{1, \cdots, N\}$. We then assume that idiosyncratic productivity shocks ϵ_j's are the same for all goods produced in the same sector, $j \in \mathcal{I}_i^c$. This is the approach taken when we use a quantitative version of this model in section 5.

Given the above discussion, it is easy to see that various neoclassical models of trade such as the Armington model, and Ricardian models of Dornbusch et al. (1977) and Eaton and Kortum (2002) are special cases of our general model.

Generality of the Policies Despite the focus of our paper on trade policy, we have not explicitly introduced tariffs in the above environment. This is in part because tariffs are a special case of consumer and producer taxes. To see this, suppose that the government in country c imposes an ad valorem tariff, τ_i^c, on good i which is on net being imported into country c. Therefore, if the (international) market price of good i is $\hat{p}_i$, the price faced by producers and consumers in country c is $\hat{p}_i\left(1+\tau_i^c\right)$. In other words, a tariff imposes a tax on the use of a good, by consumer and producers, while at the same time, it imposes a subsidy on the production/making of good i in country c. That is, under this tariff, we have:

$$t_i^{x,c} = \tau_i^c, t_i^{p,c} = -\tau_i^c, t_{ji}^{p,c} = \tau_i^c.$$

The above example illustrates that the indirect (commodity) tax policies considered in our model include the possibility of tariff. Another way to see this is to realize that in the presence of tariffs,τ_i^c, the consumer price vector in country c is given by[9]

$$\hat{\mathbf{p}}^c = \left(\left(1+\tau_1^c\right)\left(1+t_1^{x,c}\right)\hat{p}_1, \cdots, \left(1+\tau_N^c\right)\left(1+t_N^{x,c}\right)\hat{p}_N\right)$$

while the set of after tax/tariff prices faced by producers in sector i is given by

$$\left(1-t_i^{p,c}\right)\left(1+\tau_i^c\right)\hat{p}_i, \left\{\left(1+t_{ij}^{p,c}\right)\left(1+\tau_j^c\right)\hat{p}_j\right\}_{j=1}^N.$$

[9]This particular way of defining tariffs works as a tax on imports and subsidy on exports.

Since consumers and producers care only about after tax/tariff prices, it is possible to redefine consumer and producer taxes so that the resulting equilibrium allocations and prices will be the same as the ones in the economy with tariffs. Because of this equivalence of tax systems, we focus our attention to indirect tax policies that involve only consumer and producer taxes.

2.1 Characteriation of Equilibria

In this section, we provide a characterization of allocations that can arise in competitive equilibria, which are distorted by government policies. This characterization helps us in solving the optimal taxation problem in Section 3.

We first observe that the workers' sectoral choice is independent of taxes and depends only on wages. In other words, a worker identified by $(\theta, \boldsymbol{\epsilon})$ choses the sector that gives the worker the highest labor productivity. That is,

$$j^c(\theta, \boldsymbol{\epsilon}) \in \arg\max_j w_j^c a_j^c(\theta) \epsilon_j$$

This is because personal income taxes depend only on total income and not on wages or the sector in which a worker works.[10] We can then define labor productivity for a worker given by $z = \max_j w_j^c a_j^c(\theta) \epsilon_j$. Given this level of labor productivity which is determined by the sectoral choice of the worker, a worker's choice of earning is simply given by

$$\max_\ell V^c(z\ell - T^c(z\ell); \mathbf{q}^c) - v^c(\ell)$$

where $\mathbf{q}^c = \left(\hat{p}_1\left(1+t_1^{x,c}\right), \cdots, \hat{p}_N\left(1+t_N^{x,c}\right)\right)$ is the after-tax vector of prices faced by consumers in country c, and $V^c(I; \mathbf{q})$ is the indirect utility of a worker in country c with after-tax income I who faces after-tax vector of prices $\mathbf{q}$. This illustrates a key property that significantly simplifies the analysis of optimal taxes in this paper: The earning decision depends only on the labor productivity of an individual together with income tax T^c, while the sectoral choice decision of a worker only depends on wages.

This separation of decisions allows us to express the distribution of types in each sector and income level in a tractable fashion. Particularly, let $\lambda_j^c\left(z, \theta; \left\{w_j^c\right\}\right)$ be the density of workers of type θ in sector j that have productivity z. In the Appendix, we show that the extreme value assumption about the distribution of $\boldsymbol{\epsilon}$ implies that

$$\lambda_j^c\left(z, \theta; \left\{w_j^c\right\}\right) = \sigma^c z^{-1-\sigma^c}\left(w_j^c a_j^c(\theta)\right)^{\sigma^c} e^{-z^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}$$

[10] A similar separation result emerges in Rothschild and Scheuer (2013).

Similarly, if we define $\lambda^c\left(z,\theta;\left\{w_j^c\right\}\right)$ as the distribution of labor productivity among workers of type θ, then

$$\lambda^c\left(z,\theta;\left\{w_j^c\right\}\right)=\sigma^c z^{-1-\sigma^c}\sum_i\left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}e^{-z^{-\sigma^c}\sum_i\left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}$$

In other words, the distribution of labor productivity among workers of type θ is Fréchet with a mean value of $\Gamma\left(1-\frac{1}{\sigma}\right)\left[\sum_i\left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}\right]^{\frac{1}{\sigma^c}}$ and shape parameter σ^c. Moreover, conditional on sectoral choice j, the distribution of labor productivity is Fréchet with a shape parameter of σ^c and the same mean value. The extreme value assumption also implies the fraction of workers of type θ that work in sector j is given by

$$\Lambda_j^c\left(\theta\right)=\frac{\left(w_j^c a_j^c\left(\theta\right)\right)^{\sigma^c}}{\sum_i\left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}}.$$

The parameter σ^c is related to the elasticity of mobility of workers across sectors with respect to wages. When $\sigma^c=1+\varepsilon^c$, labor supply in sector j is only a function of wages in that sector and thus any reallocation of workers across sectors is irrelevant. In other words, it is as if workers are attached to sectors and cannot move across them. At the other extreme where $\sigma^c=\infty$, only sectors j in which $a_j^c\left(\theta\right)w_j^c=\max_k a_k^c\left(\theta\right)w_k^c$ holds will have a non-zero measure of workers of type θ.

In addition to these distributions, recall that by Assumption 1, the indirect utility function associated with $U^c\left(\mathbf{x}\right)$, which we denote by $V^c\left(I;\mathbf{q}\right)$, is linear in I. As a result a price index $\nu^c\left(\mathbf{q}\right)$ exists so that $V^c\left(I;\mathbf{q}\right)=\frac{I}{\nu^c(\mathbf{q})}+\vartheta^c\left(\mathbf{q}\right)$. Therefore, the choice of work effort by a worker of type $\left(\theta,\boldsymbol{\epsilon}\right)$, $\hat{\ell}^c\left(\theta,\boldsymbol{\epsilon}\right)$, must satisfy

$$\hat{\ell}^c\left(\theta,\boldsymbol{\epsilon}\right)\in\arg\max_{\ell}\frac{z\left(\theta,\boldsymbol{\epsilon}\right)\ell-T^c\left(z\left(\theta,\boldsymbol{\epsilon}\right)\ell\right)}{\nu^c\left(\mathbf{q}^c\right)}-v^c\left(\ell\right)$$

This implies that effort depends only on labor productivity z; we refer to this as $\ell^c\left(z\right)$. Finally, we can calculate the welfare of workers of type θ as follows:

$$\hat{u}^c\left(\theta\right)=\int_0^\infty u^c\left(z\right)\lambda^c\left(z,\theta;\left\{w_j^c\right\}\right)dz$$

where $u^c\left(z\right)=\frac{z\ell^c(z)-T^c(z\ell^c(z))}{\nu^c(\mathbf{q}^c)}-v\left(\ell^c\left(z\right)\right)$. We can thus summarize the necessary and sufficient conditions that allocations in an equilibrium must satisfy in the following proposition:

Proposition 1. *In any equilibrium, the allocation of consumption $\hat{\mathbf{x}}^c\left(\theta,\boldsymbol{\epsilon}\right)$, effort $\hat{\ell}^c\left(\theta,\boldsymbol{\epsilon}\right)$, and sectoral choice $j^c\left(\theta,\boldsymbol{\epsilon}\right)$ satisfy the following:*

i. *Sectoral choice is independent of income taxes, i.e.,*

$$j^c(\theta,\boldsymbol{\epsilon}) = \arg\max_j \epsilon_j a_j^c(\theta)\, w_j^c,$$

ii. *If we let* $z(\theta,\boldsymbol{\epsilon}) = \max_j \epsilon_j a_j^c(\theta)\, w_j^c$ *be the labor productivity of a worker of type* $(\theta,\boldsymbol{\epsilon})$*, then effort choice of a worker of type* $(\theta,\boldsymbol{\epsilon})$ *depends only on the worker's labor productivity* z*. That is, a function* $\ell^c(z)$ *exists such that* $\hat{\ell}^c(\theta,\boldsymbol{\epsilon}) = \ell^c(z(\theta,\boldsymbol{\epsilon}))$*.*

iii. *A function* $I^c(z)$ *must exist such that*

$$z \in \arg\max_{\hat{z}} \frac{I^c(\hat{z})}{\nu^c(\mathbf{q}^c)} - v\left(\frac{\hat{z}\ell^c(\hat{z})}{z}\right)$$

and $\hat{\mathbf{x}}^c(\theta,\boldsymbol{\epsilon}) = \mathbf{x}^c(I^c(z);\mathbf{q}^c)$*, where* $\mathbf{x}^c(I;\mathbf{q}) = (x_1^c(I;\mathbf{q}),\cdots,x_N^c(I;\mathbf{q}))$ *is the demand function associated with the utility function* $U^c(\mathbf{x})$*. The function* $I^c(z)$ *represents the disposable income of a worker with labor productivity* z*.*

The above proposition is useful in solving the optimal taxation problem, since it simplifies the constraint set of the problem. In what follows, we use this characterization of equilibria in order to discuss the solution of the optimal taxation problem.

2.2 Winners and Losers of Trade

As Galle et al. (2017) have shown, the above model can be used to shed light on the distribution of welfare gains of trade. In particular, depending on the pattern of specialization, a reorganization of production in the global economy can generate winners and losers. This reorganization of production could come from a productivity shock to certain sectors outside of a particular country; the literature often models the rise of China (what we refer to as the China shock) in the production of manufacturing goods as an increase in manufacturing productivity (see Caliendo et al. (2017) and Galle et al. (2017), among others).

In order to provide a better understanding of the effect of such a change, we discuss how a change in the structure of wages (something that can be caused by a trade shock) affects the welfare of workers. The following lemma describes the main result:

Lemma 1. *Consider a marginal change in wages in country* c*, given by* δw_j^c*. Then, the marginal change in the welfare of workers of type* θ *is given by*

$$\delta\hat{u}^c(\theta) = \delta u^c(0) + \sigma\left(\sum_j \Lambda_j^c(\theta)\frac{\delta w_j^c}{w_j^c}\right)\int z\,(u^c)'(z)\,\lambda\left(z,\theta;\left\{w_j^c\right\}\right)dz \tag{4}$$

where $\Lambda_j^c(\theta)$ is the fraction of workers of type θ in sector j in country c.

The proof can be found in the Appendix.

The above formula illustrates that whether a group of workers of type θ becomes a winner or loser from a trade shock depends on the distribution of workers across sectors and on relative changes of wages. For example, if the trade shock causes wages in manufacturing to decline, the groups that lose the most (or win the least) are those that have a high fraction of workers in manufacturing. To the extent that $\Lambda_j^c(\theta)$ is determined by the function $a_j^c(\theta)$, the pattern of specialization determines the gains and losses from trade. To see this, suppose that $a_j^c(\theta)$ exhibits absolute advantage, i.e., $a_j^c(\theta) = \alpha_j^c \beta^c(\theta)$. Then, in this case,

$$\Lambda_j^c(\theta) = \frac{\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{\sum_k \left(a_k^c(\theta)\, w_k^c\right)^\sigma} = \frac{\left(\alpha_j^c w_j^c\right)^\sigma}{\sum_k \left(\alpha_k^c w_k^c\right)^\sigma}$$

In other words, the distribution of workers across sectors is independent of their types, θ. This implies that when $\delta u^c(0) = 0$, either all workers gain from trade or all workers lose from it, albeit to a varying degree given equation (4).[11] Note that the response of welfare depends on how the utility of workers with zero labor productivity changes, in other words, on the response of transfers to the trade shock. As we will show in the next section, the optimal policy response to a trade shock also depends on the degree of specialization across different groups of workers.

3 The Optimal Policy Problem

The model developed in the previous section clarifies the type of policies that can be used by the governments in each country. Furthermore, it highlights how various groups of workers can be differentially affected from international trade and global relocation of production. In this section, we describe the optimal policy problem faced by these governments under cooperation and derive formulas that govern the behavior of the optimal trade policy.

Starting with the government in each country c, we assume that the government evaluates the allocation of resources across workers of types θ according to a social welfare function given by

$$\int_\Theta g^c(\theta)\, \hat{u}^c(\theta)\, d\mu^c$$

where $g^c(\theta) > 0$ is the welfare weight on workers of group θ. For now, we assume that $g^c(\theta)$ is an exogenous welfare weight. In section 5 where we study the problem of compensating the losers

[11]When $u^c(z) = Az^\alpha$ for some A and α, then the percent change increase in welfare for all the groups will be the same. An example of this is the laissez-faire economy where $u^c(z) \propto z^{1+\varepsilon}$. In this case, (4) becomes the basis of the formulas provided by Galle et al. (2017).

from a trade shock, these welfare weights arise from constraints that ensure that each group of worker's utility is at least as high as its pre-shock level.

Our main assumption about the determination of policy is that it is determined under cooperation. In particular, we assume that the optimal trade policy problem is given by

$$\max_{\{\mathbf{t}^{p,c},\mathbf{t}^{x,c},T^c\}_{c=1,\cdots,C}} \sum_{c=1}^{C} \psi^c \int_\Theta g^c(\theta)\,\hat{u}^c(\theta)\,d\mu^c \tag{5}$$

where $\hat{u}^c(\theta)$ is the utility profile arising from a competitive equilibrium of the economy described above given the policy choices $\{\mathbf{t}^{p,c},\mathbf{t}^{x,c},T^c\}_{c=1,\cdots,C}$. One way to rationalize this assumption is to consider trade agreements whereby governments negotiate with each other on coordination of their tax policies. We assume that the outcome of this negotiation process is efficient in that it is equivalent to maximizing a weighted average of aggregate welfare in each country. Equation (5) represents such an objective, where $\psi^c > 0$ is the welfare weight of country c implied by the negotiation process.

Given this objective for the government, the optimal taxation problem for the government is to maximize social welfare given that the utility profile $\hat{u}^c(\theta)$ arises from a competitive equilibrium, as defined above. As in Atkinson and Stiglitz (1976) or Diamond and Mirrlees (1971), it is often convenient to write this problem in terms of (after-tax) consumer and producer prices defined by

$$\begin{aligned} q_i^c &= \hat{p}_i\left(1+t_i^{x,c}\right) \\ p_i^c &= \hat{p}_i\left(1-t_i^{p,c}\right), p_{ij}^c = \hat{p}_j\left(1+t_{ij}^{p,c}\right) \end{aligned}$$

where q_i^c is the after-tax price paid by a worker/consumer for good i in country c while p_i^c and $\left\{p_{ij}^c\right\}_{j=1}^N$ are the after-tax prices faced by producers of good i. Note that producer prices affect the production decision of firms. However, since the set of taxes on producers is general enough, we can show that any demand vector for intermediate inputs by a firm i in country c, $\left\{Q_{ij}^c\right\}_{j=1}^N$, can be implemented by carefully choosing the vector of producer taxes, $\left\{t_{ij}^{p,c}\right\}_{j=1}^N$.

Following the characterization of equilibria in section 2.1 and the discussion above, the optimal policy problem can be stated as in the following proposition:

Proposition 2. *Any solution of the optimal taxation problem – and its associated consumer and producer prices and wages – must solve the following optimization problem*

$$\max_{\mathbf{q}^c,\{I^c(z),\ell^c(z),u^c(z)\},\mathbf{w}^c,\left\{Q_{ij}^c\right\}} \sum_{c=1}^{C} \psi^c \int_\Theta g^c(\theta) \int_0^\infty u^c(z)\,\lambda^c(z,\theta;\mathbf{w}^c)\,dz d\mu^c \tag{P}$$

subject to

$$\int_{\Theta}\int_{0}^{\infty} x_i^c\left(\mathbf{q}^c; I^c(z)\right)\lambda^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c + \sum_{c=1}^{C}\sum_{k=1}^{N} Q_{ki}^c = \sum_{c=1}^{C} G_i^c\left(L_i^c, \left\{Q_{ij}^c\right\}_{j=1}^N\right), \forall i \tag{6}$$

$$V^c\left(\mathbf{q}^c; I^c(z)\right) - v^c\left(\ell^c(z)\right) = u^c(z)$$

$$\frac{\ell^c(z)\, v'\left(\ell^c(z)\right)}{z} = \left(u^c\right)'(z) \tag{7}$$

$$\frac{1}{w_j^c}\int_{\Theta}\Lambda_j^c(\theta)\int_0^{\infty} z\ell^c(z)\,\lambda^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c = L_j^c$$

where $\Lambda_j^c(\theta) = \left(a_j^c(\theta)\, w_j^c\right)^{\sigma} / \left[\sum_i \left(a_i^c(\theta)\, w_i^c\right)^{\sigma}\right]$.

Conversely, any solution of the above optimization problem that satisfies $d\left(z\ell^c(z)\right)/dz \geq 0$ *can be used to construct a solution of the optimal taxation problem.*

The proof of this proposition is relegated to the Appendix.

As the above proposition establishes, the only restrictions that competitive equilibrium imposes on allocations is feasibility and optimal choice of labor supply. The reasoning for the above proposition is straightforward. Given that consumption is calculated using the demand function of households, the households' budget constraints must be satisfied. Furthermore, since feasibility is satisfied, these conditions imply that the government budget constraint must hold. Finally given that consumption comes from the demand function and optimality of choice of labor supply, the consumption and labor supply allocation must satisfy optimality for workers.

Note that in the above, we are not choosing producer prices, because of two reasons: First, as we mentioned above, by choosing intermediate quantities $\left\{Q_{ij}^c\right\}$, one effectively chooses a vector of producer taxes $\left\{t_{ij}^p\right\}$ that implement such intermediate quantities. Second, the labor demand by firms is determined by

$$p_i^c \frac{\partial G_i^c}{\partial L} = w_i^c$$

Therefore, given intermediate quantities and aggregate labor supply, a change in producer prices (implemented by a change in producer tax t_i^c) translates into a change in wage w_i^c. Thus, choosing the vector wages $\mathbf{w}^c$ is equivalent to choosing after-tax prices $\{p_i^c\}$. This discussion also clarifies the role of commodity taxes in this model. Specifically, producer taxes redistribute resources to workers in a certain sector. At the same time, they change the labor supply of workers, by affecting the intensive and extensive margin of labor supply, and thus create distortions. This is also true for earnings taxes, although earnings taxes redistribute resources across different productivity levels z. The above planning problem finds a balance between the redistribution and distortions via producer taxes and earnings taxes. Finally, since consumer and producer prices are sufficient to characterize allocations, the actual producer and consumer taxes are indeterminate.

This result is standard, as it is typically irrelevant whether to tax consumer or producers. Optimal allocations, however, determine the ratio $\frac{1-t_i^p}{1+t_i^x} = \frac{p_i}{q_i}$.

Remark. While in the above, we have assumed that the optimal policies are determined under co-operation, there are alternative formulations of the problem that can result in the same outcome. An example of this is a model of one small open economy where the entire vector of relative prices is exogenously determined. In this case, since the country cannot affect interntional prices, it will act very similarly to a country that is cooperating with the rest of the world, i.e., it has no motives for terms of trade manipulations. [12]

4 Properties of Optimal Policies

In this section, we discuss the qualitative features of optimal redistributive and trade policies. We start our discussion by analyzing production efficiency and then we discuss the determinants of optimal policies.

4.1 Production Efficiency

Since the seminal result of Diamond and Mirrlees (1971), it is known that when the government has access to a rich set of tax instruments, one for every good that households care about (including leisure and consumption goods), then production must be efficient, i.e., it is optimal to not distort firms' production decision. In our context, this result would translate into optimality of free-trade as it implies that firms' decision to use intermediate inputs is undistorted. Note that despite the generality of their result, it does not apply in our setting since the government does not have access to a rich set of taxes. In particular, the application of the result of Diamond and Mirrlees (1971) to our environment requires the government to be able to tax workers in different sectors differently. A question that arises then is whether productive efficiency holds in our environment. In what follows, we show that a version of production efficiency holds, which in turn leads to optimality of VAT taxes.

Consider an allocation of production across countries given by $\left\{\left\{\ell^c\left(z\right), \lambda_j^c\left(z,\theta;\mathbf{w}^c\right)\right\}_{z\in\mathbb{R}_+,\theta\in\Theta}, \left\{Q_{ij}^c, Y_i^c\right\}_{1\le i,j\le N}\right\}_{c\in\{1,\cdots,C\}}$ with

$$Y_i^c = G_i^c\left(\int\int z\ell^c\left(z\right)\lambda_j^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c, \left\{Q_{ij}^c\right\}\right).$$

We call this allocation *psuedo-production efficient* if there does not exist an alternative

[12] See Bagwell and Staiger (1999) for an interpretation of this assumption using the reciprocity clause often used in trade agreements.

$\left\{ \hat{Q}_{ij}^c, \hat{Y}_i^c \right\}_{i,j,c}$ such that

$$\hat{Y}_i^c = G_i^c \left(\int \int z\ell^c(z) \lambda_j^c(z, \theta; \mathbf{w}^c) \, dz d\mu^c, \left\{ \hat{Q}_{ij}^c \right\} \right)$$

$$\sum_{c=1}^{C} \hat{Y}_i^c - \sum_{c=1}^{C} \sum_{j=1}^{N} \hat{Q}_{ji}^c \geq \sum_{c=1}^{C} Y_i^c - \sum_{c=1}^{C} \sum_{j=1}^{N} Q_{ji}^c, \forall i \in \{1, \cdots, N\}$$

with the inequality holding strictly for at least one sector and one country. In words, an allocation is pseudo-production efficient if any reshuffling of intermediate inputs across sectors and countries cannot improve total consumption of all goods in the world. This notion of production efficiency is a weaker notion that its standard definition, since it does not allow for reshuffling of labor across sectors. The following proposition establishes a key property that allocations that with pseudo-production efficiency must exhibit.

Lemma 2. *A production allocation* $\left\{ \left\{ \ell^c(z), \lambda_j^c(z, \theta; \mathbf{w}^c) \right\}_{z \in \mathbb{R}_+, \theta \in \Theta}, \left\{ Q_{ij}^c, Y_i^c \right\}_{1 \leq i,j \leq N} \right\}_{c \in \{1, \cdots, C\}}$ *is efficient if and only if there exists a vector of positive numbers* $\{\rho_i\}_{i \in \{1, \cdots, N\}}$ *such that*

$$\left\{ Q_{ij}^c \right\}_{i,j,c} \in \arg \max_{\left\{ \hat{Q}_{ij}^c \right\}} \sum_{i=1}^{N} \rho_i \left[\sum_{c=1}^{C} G_i^c \left(L_i^c, \left\{ \hat{Q}_{ij}^c \right\} \right) - \sum_{c=1}^{C} \sum_{j=1}^{N} \hat{Q}_{ji}^c \right] \tag{8}$$

where $L_i^c = \int \int z\ell^c(z) \lambda_i^c(z, \theta; \mathbf{w}^c) \, dz d\mu^c$.

The proof is relegated to the Appendix.

The proof of the above lemma uses a standard separating hyper-plane argument to show the claim. We can think about the vector ρ_i's as the vector of prices. The above claim then states that for production to be pseudo-efficient, the marginal rate of transformation between any two inputs must be equated across all sectors and countries. Given the above lemma, our production efficiency result is fairly straightforward:

Corollary 1. *Consider the solution of the optimization problem (P). Then the production allocation* $\left\{ \left\{ \ell^c(z), \lambda_j^c(z, \theta; \mathbf{w}^c) \right\}_{z \in \mathbb{R}_+, \theta \in \Theta}, \left\{ Q_{ij}^c, Y_i^c \right\}_{1 \leq i,j \leq N} \right\}_{c \in \{1, \cdots, C\}}$ *is pseudo-efficient.*

Proof. We know that at the optimum, Lagrange multipliers $\{\rho_i\}_{i \in \{1, \cdots, N\}}$ should exist for resource constraints (6). Then, at the optimum, the optimality condition for each Q_{ij}^c satisfies

$$\rho_i \frac{\partial G_i^c}{\partial Q_{ij}} = \rho_j$$

since G_i^c satisfies the Inada condition with respect to each intermediate input.

Note further that each $\rho_i > 0$. This is because a uniform increase in income of all workers in a country that likes good i increases welfare. Since all G_i^c's are strictly concave, the above condition is equivalent to the condition stated in lemma 2. This concludes the proof. $\square$

The key feature of our model which leads to this result despite the incompleteness of income taxes, is that different kinds of labor are perfect substitutes in the production function. This implies that the wage of a worker of type θ whose productivity is z and works in sector j is the marginal product of labor in sector j multiplied by z. As a result, relative wages are independent of the quantitities of intermediate inputs. Consequently, distortions to allocation of intermediate goods cannot affect the deadweight loss of income taxation and, as a result, it must not be distorted.

The efficiency properties of production imply that taxes on firms must be designed so that all sectors face the same after tax-prices. This means that any sales taxes paid by firms in sector j on their sales must be partially offset by a subsidy on intermediate inputs. One implementation of efficiency would be to have the sales tax rate be equal to the subsidy rate on intermediate inputs. In other words, optimal production taxes take the form of a VAT tax. We, thus, have the following proposition:

Proposition 3. *Optimal allocations in (P) can be implemented by VAT taxes where*

$$t_i^c = -t_{ij}^c, \forall i, j \in \{1, \cdots, N\}.$$

Proof. In any competitive equilibrium, we have

$$\hat{p}_i (1 - t_i^c) \frac{\partial G_i^c}{\partial Q_{ij}} = \hat{p}_j \left(1 + t_{ij}^c\right)$$

Since prices in a competitive equilibrium can be freely picked, we set $\hat{p}_i = \rho_i$ where ρ_i is the Lagrange multiplier on the resource constraint (6). Then, the pseudo-efficiency of production implies that

$$\frac{1 + t_{ij}^c}{1 - t_i^c} = 1$$

which in turn implies that $t_{ij}^c = -t_i^c, \forall i, j$. $\square$

As we illustrate below and in our numerical results, sector-specific VAT taxes play an important role in redistributing gains from trade across different groups in the population.

Remark. It is important to compare our results to other results in the literature on production efficiency, for example, those of Naito (1999) and Costinot and Werning (2018) as they have shown that distortions to production efficiency can be optimal. This is because, in their work, a change in the allocation of intermediate goods affects relative wages and thus can reduce the deadweight loss from taxation of income. Our discussion above on the interplay between allocation of intermediate inputs and relative wages clarifies the main forces behind the difference between our results and those of Naito (1999) and Costinot and Werning (2018). In particular, one can consider a more general production function of the form $G_i^c\left(\{L_i^c(\theta)\},\{Q_{ij}^c\}\right)$ where

$$L_i^c(\theta) = \int_{\mathbb{R}_+^N} a_i^c(\theta)\,\epsilon_i \ell^c(\theta,\epsilon)\,\mathbf{1}\left[i = j^c(\theta,\boldsymbol{\epsilon})\right] dH^c(\boldsymbol{\epsilon})\,d\mu^c \tag{9}$$

where in the above $w_i^c(\theta)$ is the wages of workers in sector i country c. In this case, it is possible to find conditions so that production remains pseudo-efficient. The following proposition states this result:

Proposition 4. *Suppose that good i in country c is produced using the production function $Y_i^c = G_i^c\left(\{L_i^c(\theta)\},\{Q_{ij}^c\}\right)$ where $L_i^c(\theta)$ is the total effective labor supply of workers of type θ in sector i defined by (9). Then under optimal policies, production is pseudo-efficient if the following condition holds:*

$$\frac{\partial}{\partial Q_{ij}^c}\frac{\frac{\partial G_i^c}{\partial L^c(\theta)}}{\frac{\partial G_i^c}{\partial L^c(\theta')}} = 0, \forall i,j \in \{1,\cdots,N\},\theta,\theta' \in \Theta.$$

The proof is relegated to the appendix.

As the above proposition states, for distortions to production to be optimal, it must be that intermediate inputs have different degrees of substitution with different types of labor, e.g., the intermediate inputs must be complements to some types of skills and substitutes for others. Let us clarify this with an example. Consider a producer in country c who has a choice of building a machine using intermediate inputs or hiring low-skilled workers to produce. In a closed economy and when the price of intermediate inputs are high, the producer is incentivizied to rely more on low-skilled labor. Alternatively, in an open economy, lower prices of intermediate inputs encourage the producer to build the machine and reduce its demands for low-skilled labor. In this environment, a tax on such intermediate-inputs increases the relative wages of low-skilled workers and reduces the deadweight-loss of income taxation and is efficient.

Our analysis here implies that in order to understand the best policy response to unequal gains from trade, it is imperative to measure the substitution of intermediate inputs with different types of labor. In other words, there are two channels through which some worker groups might lose from trade: First, a change in the structure of wages combined with the specialization of their skills can reduce their welfare; Second, cheaper intermediate inputs can lead to firms substituting away from these groups' labor. The literature has often discussed the unequal effect of international trade; see for example, Autor et al. (2013), Galle et al. (2017), among many others. However, it is still unknown whether this is due to the specialization or the substitution effect. Our analysis highlights the importance of the underlying mechanism for optimal policy response: with specialization there is no need for distortion of trade while with substitution it might be optimal to distort trade. In our quantitative exercise, we focus on the specialization story, based on the work of Galle et al. (2017), and show how to design VAT taxes in order to redistribute the gains from trade.

4.2 Optimal VAT Taxes

As we have shown, production taxes must take the form of VAT taxes. Here, we describe the behavior of optimal VAT taxes and their role in redistributing the gains from trade across workers of different groups.

The following proposition describes the determinants of optimal VAT taxes:

Proposition 5. *Optimal VAT taxes, t_j^c, satisfy*

$$\begin{aligned}(\sigma^c - 1)\frac{t_j^c}{1-t_j^c}Y_j^c = &\left[1-\overline{\mathcal{W}}_j^c\right]Y_j^c \\ &+\int \Lambda_j^c \int y^c\left[(1-\tau_\ell^c)(1+\varepsilon^c) - \sigma^c - \varepsilon^c z\left(\tau_\ell^c\right)'\right]\lambda^c dz d\mu^c \\ &+\sum_i \int \frac{\Lambda_i^c \Lambda_j^c}{1-t_i^c}\int y^c\left[\sigma^c - 1 - \varepsilon^c + \varepsilon^c \frac{z\left(\tau_\ell^c\right)'}{1-\tau_\ell^c}\right]\lambda^c dz d\mu^c\end{aligned} \tag{10}$$

where $\overline{\mathcal{W}}_j^c$ is the social marginal value of increasing the income of workers in sector j by 1% and is given by

$$\overline{\mathcal{W}}_j^c = \frac{\int g^c(\theta)\Lambda_j^c(\theta)\int\left(1-\tau_\ell^c(z)\right)y^c(z)\lambda^c(z,\theta)dzd\mu^c}{\int g^c(\theta)d\mu^c\int\Lambda_j^c(\theta)\int y^c(z)\lambda^c(z,\theta)dzd\mu^c};$$

$y^c(z)$ is the before-tax income of a worker with productivity z; $\tau_\ell^c(z)$ is the marginal income tax rate for a worker with labor productivity z; and Y_j^c is total labor income earned in sector j, i.e., $Y_j^c = \int \Lambda_j^c \int y^c \lambda dz d\mu^c$.

The above formula characterizes optimal VAT taxes. Its intuition is best understood by considering a 1% decrease in VAT tax on sector i in country c and considering its impact on government budget and welfare. The left hand side of the equation is the change in government revenue due to the behavioral response of workers to a 1% decrease in taxes. The term $\sigma^c - 1$ can then be thought of as the elasticity of labor supply to this tax change; this term captures three effects: When VAT tax in sector j declines and as a result its wages increase, workers respond by choosing sector j more frequently - with elasticity σ^c; by providing more hours – with elasticity ε^c;

and by reducing the productivity in that sector – with elasticity $1+\varepsilon^c$. The superimposition of these effects leads to an elasticity of $\sigma^c - 1$.

The right hand side of the equation captures the costs of this change. The first term is the mechanical decrease in government revenue net of its *welfare effect.*[13] The second term is related to the fiscal externality of changing VAT taxes on the revenue from income taxes. Finally, the last term captures the fact that a reduction in VAT taxes in sector j causes an increase in productivity of other sectors – since less productive workers in j move out of other sectors and into sector j. We refer to the former as *the relocation effect.*

The above formula has the following important implications:

Optimal taxes depend on the specialization of workforce.

So far we have emphasized the role of specialization in generating winners and losers from trade shocks. Here, we show how specialization affects optimal VAT taxes. To do so, we consider the case without specialization, i.e., when there is absolute advantage. In this case, as discussed in section 2.2, the fraction of workers in sector j of type θ, $\Lambda_j^c(\theta)$ is independent of θ. As a result, we can write the formula (10) as

$$\begin{aligned}(\sigma^c - 1)\frac{t_j^c}{1-t_j^c}\Lambda_j^c Y^c = & \left[1-\overline{\mathcal{W}}_j^c\right]\Lambda_j^c Y^c \\ & + \Lambda_j^c \int\int y^c\left[\left(1-\tau_\ell^c\right)\left(1+\varepsilon^c\right)-\sigma^c-\varepsilon^c z\left(\tau_\ell^c\right)'\right]\lambda^c dz d\mu^c \\ & + \sum_i \frac{\Lambda_i^c\Lambda_j^c}{1-t_i^c}\int\int y^c\left[\sigma^c - 1 - \varepsilon^c + \varepsilon^c\frac{z\left(\tau_\ell^c\right)'}{1-\tau_\ell^c}\right]\lambda^c dz d\mu^c\end{aligned}$$

Note that in the above, we have used the fact that Λ_j^c is independent of θ. In the Appendix, we show that $\overline{\mathcal{W}}_j^c$ is the same for all sectors. Roughly speaking, this is because the distribution of workers across different sectors is identical. We can thus write it as

$$\begin{aligned}(\sigma^c - 1)\frac{t_j^c}{1-t_j^c} = & 1-\overline{\mathcal{W}}^c \\ & + \frac{1}{Y^c}\int\int y^c\left[\left(1-\tau_\ell^c\right)\left(1+\varepsilon^c\right)-\sigma^c-\varepsilon^c z\left(\tau_\ell^c\right)'\right]\lambda^c dz d\mu^c \\ & + \frac{1}{Y^c}\sum_i \frac{\Lambda_i^c}{1-t_i^c}\int\int y^c\left[\sigma^c - 1 - \varepsilon^c + \varepsilon^c\frac{z\left(\tau_\ell^c\right)'}{1-\tau_\ell^c}\right]\lambda^c dz d\mu^c\end{aligned}$$

[13] We have $\overline{\mathcal{W}}_j^c = \int g^c\Lambda_j^c\int\left(1-\tau_\ell^c\right)y^c\lambda^c dz d\mu^c$. Notice that there is no compositional effect on welfare because of the extreme value assumption. This assumption implies that the welfare of all the workers of the same type conditional on their sectoral choice is independent of their sectoral choice.

Thus, t_j^c is independent of j. In other words, VAT taxes will not be used to redistribute resources across different groups, and we have the following proposition:

Proposition 6. *Suppose $a_j^c(\theta) = \alpha_j^c \beta^c(\theta)$. Then optimal VAT taxes are equal.*

The proof is relegated to the Appendix.

The above proposition highlights the main forces behind using VAT taxes to redistribute the gains from trade. When the distribution of workers across sectors is the same, VAT taxes cannot prove a useful tool to redistribute the gains of trade. This is because VAT taxes in this model work similar to other tags as discussed by Akerlof (1978), wherein the sectoral choice of a worker can be used as a signal of that worker's group θ. With absolute advantage, sectors and θ are uncorrelated. Thus, VAT taxes cannot be used as a tag. However, in the presence of specialization, losers from trade are often concentrated in sectors that do not experience gains from trade. As a result, VAT taxes can be used to redistribute the gains from trade to the groups of workers that are concentrated in sectors whose wages decline.

Optimal taxes are independent of the pattern of trade.

As the above formula shows, trade or, more generally, the parameters of technology across countries, only affect optimal taxes through their effect on the distribution of income. In other words, in this model, the source of inequality is irrelevant for optimal taxes. Moreover, trade elasticities and the network structure of trade do not affect optimal taxes directly. A stark illustration of this can be done by considering the model in which the distribution of workers across sectors is irrelevant for their productivity, i.e., $\sigma^c = 1 + \varepsilon^c$ and when income taxes are linear. Under this restriction, the formula for optimal producer taxes becomes

$$\frac{t_j^c}{1 - t_j^c}\varepsilon^c = 1 - \tau_\ell^c (1 + \varepsilon^c) - \mathcal{W}_j^c$$

Note that this formula also holds when we ignore general equilibrium effects and assume that producer taxes do not affect the structure of wages. In fact, when we set $\mathcal{W}_j^c = 0$, i.e., when the government does not care about workers in a sector, the above formula is simply the tax rate at the peak of Laffer curve,

$$\frac{1}{\left(1 - t_j^c\right)(1 - \tau^c)} = \frac{1 + \varepsilon^c}{\varepsilon^c}.$$

In other words, general equilibrium effects (and, as a result, the global process of production) are irrelevant for optimal VAT taxes. The main reason behind this result is that even though we have assumed away sector-specific income taxes, our tax structure is quite rich. A rich tax schedule implies that the value of each good to the government in each country is simply the price of that good. In other words, the government in each country c can perturb its tax schedule (consumer taxes and producer taxes) and increase only the supply of one particular good (that is, either used or produced in each country) by a small amount. Since all taxes are at the optimum, this perturbation of taxes has no first-order effect and, thus, the valuation of the government is proportional to the price of this good. That each government marginal valuation of each good is simply its price implies that the government can simply ignore the general equilibrium effect and use current prices to evaluate the change in demand and supply coming from perturbation of $t_j^{p,c}$.

Finally, we should mention how our results on production taxes compare to other forms of taxation of firms and, most importantly, the so-called border tax adjustments. Since our VAT taxes are non-uniform across sectors, an implication of this result is that border tax adjustments – wherein firms deduct their exports from their tax bill but cannot do so for imports – are suboptimal. In fact, border tax adjustments create a wedge between the price paid for domestic inputs and the price paid for foreign inputs. [14]

4.3 Optimal Income Taxes

As we have described above, the tools that are optimally chosen by the government to redistribute the gains from trade are VAT taxes and income taxes. To the extent that a group of workers' average income is correlated with whether that group loses or wins from a trade shock, non-linear income taxes can be used to redistribute the gains from trade. Here, we describe the determinants of optimal income taxes.

Before describing optimal taxes, it is useful to define the unconditional distribution of labor productivity $f^c(z)$ given by

$$f^c(z) = \int_\Theta \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c, F^c(z) = \int_0^z f^c(\hat{z})\, d\hat{z}$$
$$f_j^c(z) = \int_\Theta \lambda_j^c(z,\theta;\mathbf{w}^c)\, d\mu^c = \int_\Theta \Lambda_j^c \lambda^c d\mu^c$$

In the above, $f^c(z)$ is the countrywide density of labor productivity, while $f_j^c(z)$ is the same for each sector. The following proposition describes optimal income taxes:

Proposition 7. *Optimal income taxes must satisfy*

$$\frac{1}{1-\tau_\ell^c(z)}\frac{1}{1-\bar{t}^c(z)} - 1 = \frac{1-F^c(z)}{zf^c(z)}\left(1+\frac{1}{\varepsilon^c}\right)\int_z^\infty \left[1-\mathcal{W}^c(\hat{z})\right]\frac{dF^c(\hat{z})}{1-F^c(z)} \tag{11}$$

[14]This point is also shown originally by Grossman (1980); also see Dixit (1985). For a related discussion, see Costinot and Werning (2017).

where

$$\frac{1}{1-\bar{t}^{c}(z)} = \sum_{j} \frac{f_j(z)}{f(z)} \frac{1}{1-t_j^c}$$

$$\mathcal{W}^c(z) = \frac{\int_{\Theta} g^c(\theta)\,\lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c}{f^c(z)\int g^c(\theta)\,d\mu^c}.$$

The proof is relegated to the Appendix.

The above formula is reminiscent of the classic formula that characterizes optimal non-linear taxes (see Mirrlees (1971), Diamond (1998), and Saez (2001)). It is, however, modified to allow for a difference from the canonical model. Since the distortions to the decision to earn income are created by both VAT taxes and income taxes, the right hand side of the equation contains the labor wedge, τ_ℓ^c, and a weighted average of VAT taxes $\bar{t}^{c}(z)$. This weighted average depends on the level of labor productivity, since the sectoral composition of the labor force at each productivity level changes and the VAT differs across sectors.

Additionally, the social welfare weights, $\mathcal{W}^c$, depend on the composition of types at each labor productivity level, their welfare weights, and their correlation with income. To see this, consider the hypothetical case where $g^c(\theta)$ and $\lambda^c(z,\theta;\mathbf{w}^c)$ are uncorrelated, i.e.,

$$\int g^c(\theta)\,d\mu^c \int \lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c = \int g^c(\theta)\,\lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c.$$

Then

$$\mathcal{W}^c(z) = \frac{\int g^c(\theta)\,d\mu^c \int \lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c}{f^c(z)\int g^c(\theta)\,d\mu^c} = \frac{\int \lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c}{f^c(z)} = 1.$$

In other words, the right hand side of (11) is zero. That is, the only role of income taxes is to cancel out the effect of VAT taxes on the labor supply. This case identifies the importance of the correlation between θ and labor productivity in the ability of income taxes to redistribute the gains from trade. As we argue in our quantitative exercise, the fact that the gains from trade are not perfectly correlated with income reduces the importance of redistributive income taxes and emphasizes the role of VAT taxes as a powerful mechanism to redistribute the gains from trade.

Another extreme case is the one with absolute advantage. In this case, VAT taxes across all sectors are equal. As a result, we can simply set them to zero and then the above formula becomes identical to the classic Mirrlees-Diamond-Saez formula. In essence, the best case for using non-linear income taxes to redistribute the gains from trade can be made in an economy where the distribution of workers across sectors is the same.

Our theoretical analysis has highlighted a few important qualitative features:

1. Production taxes should resemble VAT taxes.

2. In the presence of specialization, differential VAT taxes can be useful in redistributing the gains from trade.

3. Non-linear income taxes are useful to redistribute the gains from trade to the extent that income is correlated with the gains from trade.

These theoretical insights point toward a need for a quantitative evaluation of the forces involved. In particular, it is important to understand which policy tools are the right ones to redistribute the gains from trade. In what follows, we pursue a quantitative analysis to shed light on this issue.

5 A Quantitative Exercise

In this section, we provide a quantitative evaluation of the optimal policy response to a trade shock, namely, the rise of China in the global goods market (China shock henceforth). To do so, we use a quantitative framework, based on the work of Galle et al. (2017), which can be used to measure the differential impact of the China shock on the welfare of different types of workers. We use this framework to quantitatively assess the optimal policies that are Pareto improving, i.e., keep workers at least as well off as the economy without the China shock.

5.1 A Model of Trade with Intermediate Goods

Our applied framework closely follows Galle et al. (2017). The world economy consists of C countries in each of which production occurs in N_J sectors. There are two types of goods in each sector j of country c. Firms in each sector j in country c can produce many varieties of tradable intermediate goods and one variety of non-tradable composite good. In particular, in each sector there is a continuum of intermediate goods represented by $\omega_j \in [0,1]$. The output of a variety ω_j in sector j in country c is given by the following constant returns to scale production function:

$$b_j^c\left(\omega_j\right)\left(L_j^c\right)^{\chi_j^c}\prod_{k=1}^{N_J}\left(Q_{j,k}^c\right)^{\gamma_{j,k}^c},$$

which uses labor and the composite good that is produced domestically, as in Caliendo and Parro (2015). Following Eaton and Kortum (2002), we assume b_j^c has a Fréchet distribution with CDF $e^{-A_j^c z^{-\nu}}$. Parameter A_j^c is the average efficiency of production in sector j, country c. Parameter ν is the dispersion of productivity across varieties and will be the elasticity of trade flow to changes in trade costs.

The composite good in sector j, country c is produced by combining intermediate goods ω_j according to the production function:

$$\left[\int r_j^c\left(\omega_j\right)^{1-\frac{1}{\rho}}\, d\omega_j\right]^{\frac{\rho}{\rho-1}},$$

where $r_j^c\left(\omega_j\right)$ is the amount of intermediate goods of variety ω_j used in the production of the composite goods. The composite good, while non-tradable, can be produced using intermediate goods' varieties produced in any countries in order to minimize the cost of production. Additionally, the composite good is used in the domestic production of intermediate goods and for domestic consumption.

We assume trade across countries is costly, and we explicitly model this as an iceberg cost à la Samuelson (1954). More specifically, we denote the cost of shipping an intermediate good produced in sector j from country c' to country c as $d_j^{c,c'} \geq 1$ with $d_j^{c,c} = 1$. This means for one unit of good j to arrive at country c, $d_j^{c,c'}$ unit must be shipped from country c'.

Finally, we assume that the governments impose VAT taxes t_j^c on intermediate goods produced in sector j and country c. This, together with wages in country c and prices of composite goods given by P_j^c, determines the unit cost of producing an intermediate good ω_j in country c. If we use the production function above, the unit cost of producing variety ω_j in country c is given by $\psi_j^c/b_j^c\left(\omega_j\right)$, where

$$\psi_j^c = \left(\frac{w_j^c}{\chi_j^c\left(1-t_j^c\right)}\right)^{\chi_j^c} \prod_k \left(\frac{P_k^c}{\gamma_{jk}^c}\right)^{\gamma_{jk}^c}. \tag{12}$$

Finally, following Eaton and Kortum (2002), the optimal decision to import intermediate goods from the lowest cost origin implies that the share of good j expenditure in country c that is imported from country c' is given by

$$\pi_j^{c,c'} = \frac{A_j^{c'}\left(d_j^{c,c'}\psi_j^{c'}\right)^{-\nu}}{\sum_{c''} A_j^{c''}\left(d_j^{c,c''}\psi_j^{c''}\right)^{-\nu}}. \tag{13}$$

Moreover, the price of the composite good j in country c is given by

$$P_j^c = \Gamma\left(1+\frac{1-\rho}{\nu}\right)^{\frac{1}{1-\rho}} \left[\sum_{c'} A_j^{c'}\left(d_j^{c,c'}\psi_j^{c'}\right)^{-\nu}\right]^{\frac{-1}{\nu}}. \tag{14}$$

Finally, we assume that workers are perfectly mobile within sectors (across the continuum of varieties), while their sectoral choice is described by the Roy model in section 2, where the productivity of a worker of type θ in sector j is given by $a_j^c\left(\theta\right)\epsilon_j$, where ϵ_j is drawn from a

Frechét distribution with shape parameter σ^c, as described in section 2. Additionally, workers' utility function is given by $U^c(\mathbf{x}) = \prod_j x_j^{\alpha_j^c}$.

5.2 Calibration and the China Shock

In this section, we describe the calibration of the baseline model and the China shock. We follow Caliendo et al. (2017) and Galle et al. (2017) and model the China shock as an increase in sector specific TFP, i.e., by changing A_j^{China}. Here, we briefly describe the calibration of the model ingredients and relegate the calibration details to the Appendix.

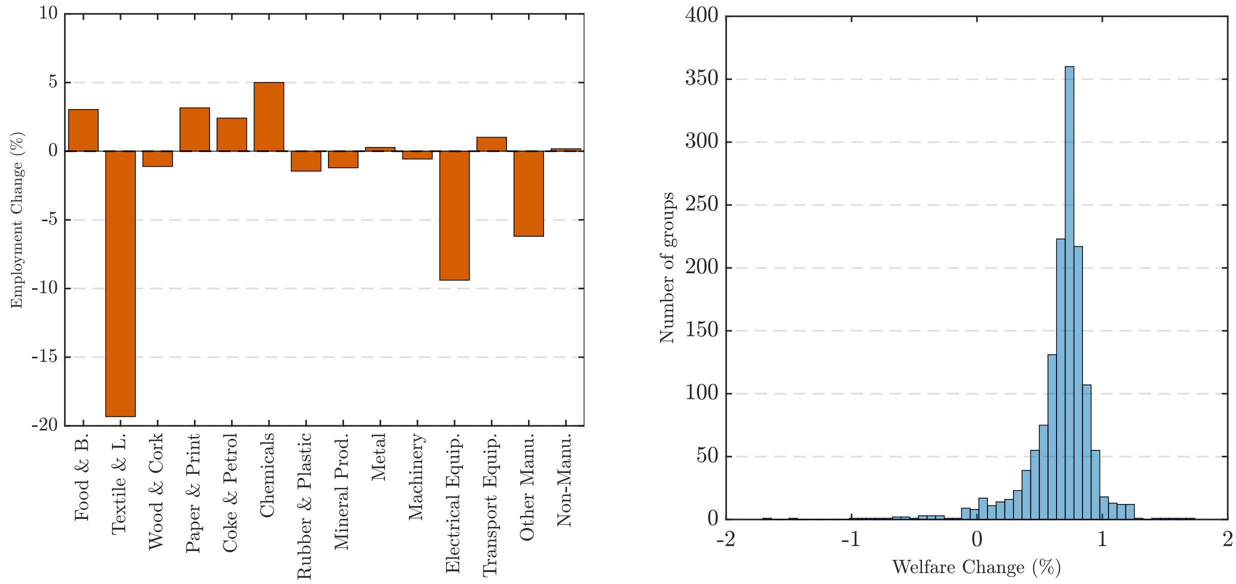
Data and Benchmark Calibration. We calibrate the model in Section 5 by taking a laissez-faire version of this economy – one without government policies – to the data. As a robustness check, in Section 5 we allow for the presence of redistributive taxes in our benchmark calibration. We choose some of the parameters of the model similar to those chosen in the literature. In line with many estimates of the intensive elasticity of labor supply, we choose $\varepsilon^c = 0.5$; see Chetty et al. (2011) for an extensive discussion. We choose the dispersion of productivity to be $\nu = 5$. This is based on estimated trade elasticities in Head and Mayer (2014). The elasticity of substitution in the final goods production function is $\rho = 4$. Moreover, we choose $\sigma^c = 1.6$ which is close to the benchmark used in Galle et al. (2017). [15] Finally, we assume that the elasticity of substitution across different consumption goods is 1. The parameters of technology, productivities, and production functions are chosen to match the following data.

We use the data on bilateral trade and employment by sector from the World Input-Output Database (WIOD) in year 2000 and 2011. [16] We restrict our analysis to the U.S. and the following nine countries: Australia, Canada, China, Denmark, Finland, France, Germany, Japan, and Spain. Our sectors are listed in Table 1 and consist of 13 manufacturing sectors and one non-manufacturing sector. Our manufacturing sectors roughly correspond to two-digit ISIC Rev. 3 codes. We use these data to calculate bilateral trade shares and calibrate the share of labor and intermediate goods in the production function in (12).

We follow Autor et al. (2013) and Galle et al. (2017) and define workers' types or groups – associated with θ in the model – based on their geographic location and education in the United States. More specifically, we use commuting zones (CZs) to define local labor markets. We split each commuting zone into two education groups based on whether workers hold at least an Associate degree or not. This implies that there are a total of 1,444 types (722 CZs $\times$ 2 skills). All

[15] Galle et al. (2017) estimate a range of values for this parameter. They choose 1.5 as their preferred values. Since we have elastic labor supply within each sector we have to choose a value that is larger than $1 + \varepsilon^c$. Notably, our results are not sensitive to changes in the value of σ^c in the range of 1.6–3.

[16] See Timmer et al. (2015) for a detailed description of the data.



(a) Percentage changes in employment by sector

(b) Distribution of welfare gains

Figure 1: Employment and welfare effect of the China shock. Panel (a) shows the percentage changes in employment by sector after the China shock. Panel (b) shows the distribution of welfare gains across different types. The rise of China leads to aggregate (income weighted) welfare gain of 0.59%. Mean welfare gains across types is 0.67% with coefficient of variation of 0.40. The minimum gain is -1.70% and maximum gain is 1.74%. Overall, 2.5 percent of all types experience a welfare loss.

other countries have only a single type. [17] In each commuting zone we use the data on wages and sectoral employment shares by education from the 2000 Census. [18]

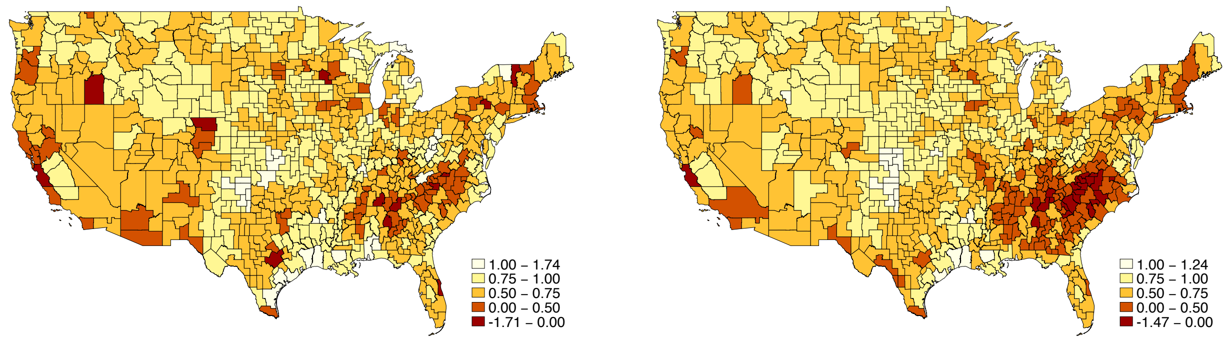
Finally, we use the data on relative prices in each sector and country from the Groningen Growth and Development Centre (GGDC) Productivity Level Database.[19] We use these data to calibrate the sectoral TFP parameters in the benchmark. For more details on the calibration procedure, refer to the Appendix B.

Calibration of the China Shock. We calibrate the China shock as the rise in the sector-specific TFP parameter A_j^{China} in year 2011 relative to the calibrated values in year 2000. Our calibration procedure closely follows that of Autor et al. (2013) and Galle et al. (2017). In order to calibrate the change in TFP in China, we use the change in trade shares between the U.S. and China as a target. In particular, in order to determine how much of the change in expenditure

[17]Note that our definition of groups based on location and education implies no mobility across local labor markets and education categories. As Galle et al. (2017) argue, there is little evidence of trade exposure causing population shifts across local labor markets. See Autor et al. (2013) and Dix-Carneiro and Kovak (2017).

[18]We use the crosswalk provided by David Dorn to map Census Public Use Micro-data Areas (PUMAs) into commuting zones. See `https://www.ddorn.net/data.htm`.

[19]See Inklaar and Timmer (2014) for a detailed description of these data.



(a) Low skilled workers

(b) High skilled workers

Figure 2: Changes in welfare across commuting zones in the United States in response to China shock. Panel (a) depicts the percentage change in welfare for the less-educated workers – those with education less than an Associate degree. Panel (b) depicted the percentage change in welfare for educated workers. For the less-educated workers, 13 commuting zones experience a welfare loss. For the educated workers, 23 commuting zones experience a welfare loss.

shares on Chinese goods in the U.S. is driven by the China shock, we first run the following regression

$$\hat{\pi}_j^{US,China} = \beta_0 + \beta_1 \hat{\pi}_j^{Others,China} + \epsilon_j$$

where $\hat{\pi}_j^{US,China} = \frac{\pi_j^{US,China} \text{in year 2011}}{\pi_j^{US,China} \text{in year 2000}}$ and $\hat{\pi}_j^{Others,China} = \frac{\sum_{c \in \{Other\}} \pi_j^{c,China} \text{in year 2011}}{\sum_{c \in \{Other\}} \pi_j^{c,China} \text{in year 2000}}$. This regression, inspired by Autor et al. (2013), is aimed at isolating the portion of the changes in U.S. expenditure share on Chinese goods that is only due to the rise of China. The $\hat{\pi}_j^{Others,China}$ is calculated using the following countries: Australia, Denmark, Finland, Germany, Spain, and Japan. The regression coefficient is 1.32, with R^2 of 0.975. These estimates are close to those of Autor et al. (2013) and Caliendo et al. (2017). We then use the predicted values from this regression as the target for calibration of the change in TFP. These predicted values are reported in the third column of Table 1.

To better understand the results of our quantitative exercise, it is useful to discuss the welfare and employment changes implied by the model. The left panel in Figure 1 shows the percentage change in employment within each sector. The sectors that are most negatively affected are 'Textile and Leather Products', 'Electrical Equipments', and 'Other Manufacturing', as they suffer the largest loss from the China shock. In terms of employment share, 'Textile and Leather Products' and 'Other Manufacturing' are very small, with only about 0.5% percent of total employment. However, 'Electrical Equipment' employs more than 2.2% of workers in our sample.

The right panel of Figure 1 shows the distribution of welfare gains across types. The gains range from -1.70% to 1.74%, with about 2.5% of all types experiencing a welfare loss (about 4.1% of

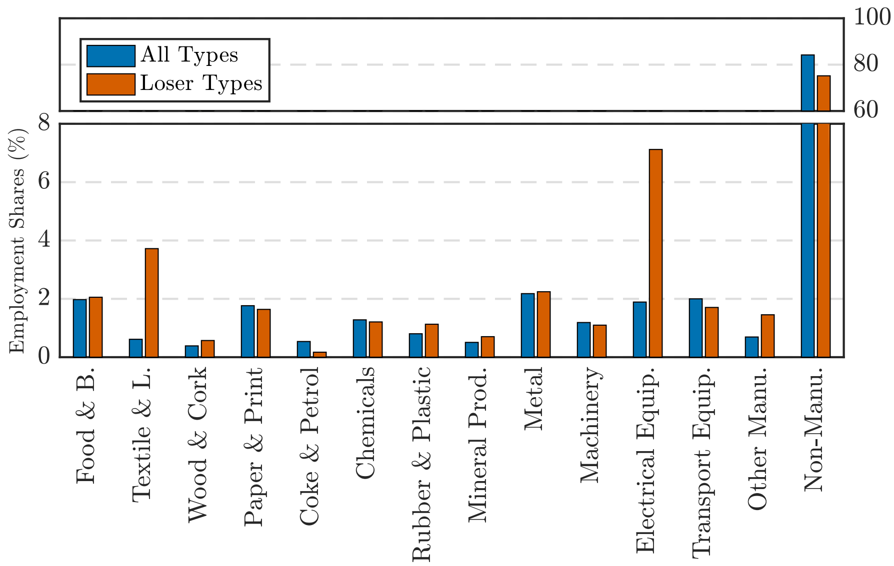


Figure 3: Employment share across sectors. Comparing types who lose from the China shock to the rest. Notice that the types who lose from the rise of China are disproportionately employed in “Textile & Leather”, “Electrical Equipments” and “Other Manufacturing” .

population in our sample). The aggregate (income weighted) welfare gain is 0.59% and the mean welfare gain is 0.67%. These values are higher than those reported in Galle et al. (2017), mainly because we allow for an elastic intensive margin of labor supply. This alone scales all welfare changes by a factor of $1 + \varepsilon^c$ (1.5 in our calibration).

Finally, Figure 2 depicts the geographic distribution of welfare gains across the U.S. In line with Galle et al. (2017), the main area that experiences welfare losses is the southeastern United States. As Figure 3 shows, this is mainly because this area has employment concentrated in sectors that are hit the most: “Textile and Leather”, “Electrical Equipment”, and “Other Manufacturing”. Note that as equation (4) states, the sign of the change in welfare for a group is determined by its employment distributions across sectors.

5.3 Pareto Optimal Policies

In this section, we solve for optimal redistributive policies in the environment described above. Our main assumption on the objective function is that policies must be Pareto improving with respect to the welfare of workers before the China shock. More specifically, we assume that the world economy faces the rise of TFP in China, yet all groups of workers in the United States must be as well off as they were prior to the China shock. To do so, we solve a planning problem that maximizes a weighted average of utility of all countries other than the United States, while

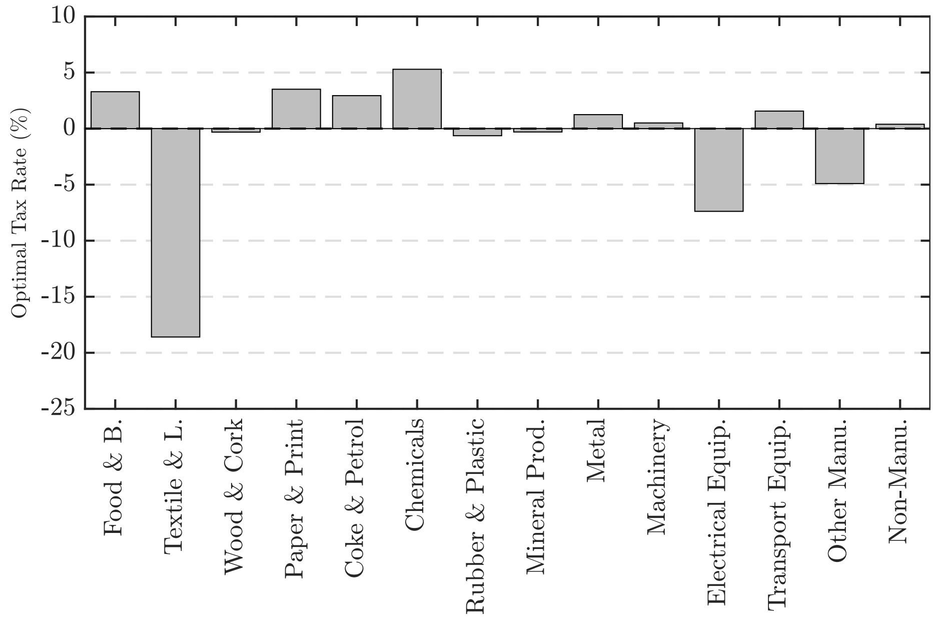


Figure 4: Optimal VAT taxes in the United States

we impose that worker groups in the United States must receive a utility at least as high as that of the laissez-faire economy prior to the China shock. Note that the question of how to redistribute the gains from trade – which depends on the institutional and political details in each country – is beyond the scope of this paper. Our criterion is in line with that used by Werning (2007) and Hosseini and Shourideh (2018). In Section 5.4.2, we discuss another criterion for the determination of the optimal policy that ensures that all groups achieve at least a fraction of the total gains from the trade shock.

As discussed above, our optimal policy objective is to maximize a weighted average of welfare in all countries other than the United States. In particular, we choose a welfare function of the form

$$\sum_{c' \neq U.S.} \psi^{c'} \hat{u}^{c'},$$

where $\hat{u}^c$ is the utility of the representative consumer in country c. We assume that ψ^c is proportional to the population of country c in 2000. This, assumption implies that the distribution of welfare across other countries is not pinned down and hence transfers are indeterminate. In other words, we intentionally choose the weights so that transfers are not pinned down and cross-country redistributive motives are absent (among countries other than the U.S.). This implies that our policies do not arise because we care differentially about workers in the countries other than the United States.

As for workers in the U.S., the goal is to make sure that their welfare does not fall below its level before the China shock. Therefore, we maximize this objective subject to equilibrium

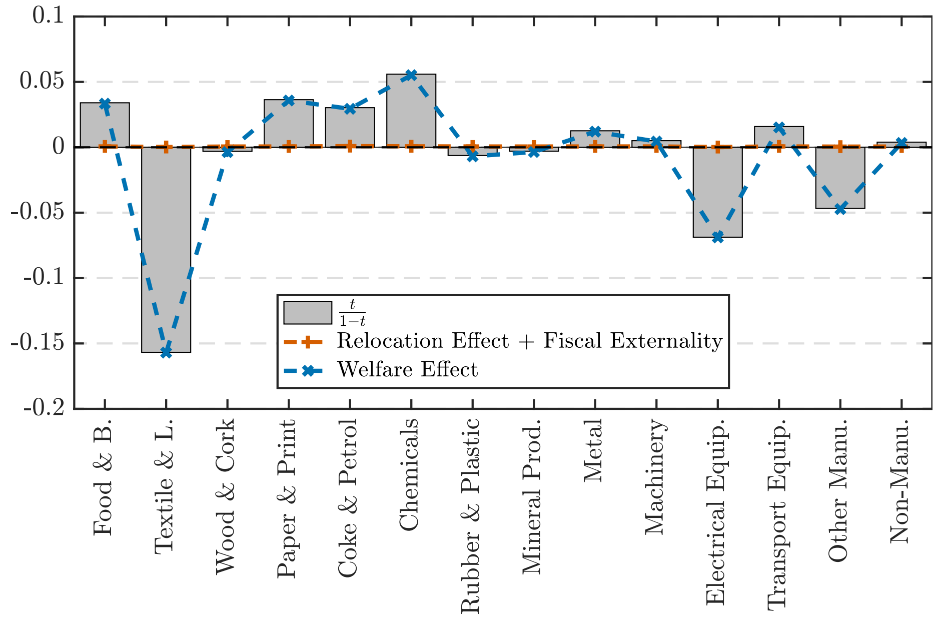


Figure 5: Breakdown of optimal VAT taxes into welfare effect and the rest: fiscal externality on income taxes and the relocation effect.

constraints, as described in Proposition 2, together with a constraint of the form $\hat{u}^{US}(\theta) \geq \hat{u}^{US,pre}(\theta)$, where $\hat{u}^{US}(\theta)$ is the utility of workers in group θ in the U.S and $\hat{u}^{US,pre}(\theta)$ is its value in the laissez-faire economy above before the China shock. Note that solving this problem is equivalent to solving the problem studied in sections 2–4 where the welfare weights $\psi^c g^c(\theta)$ for U.S. workers are determined by the Lagrange multipliers on constraint $\hat{u}^{US}(\theta) \geq \hat{u}^{US,pre}(\theta)$. Therefore, the same formulas govern the behavior of optimal VAT taxes and those of income taxes.

Optimal VAT Taxes Figure 4 depicts optimal VAT taxes in the United States. In line with significant employment drops in 'Textile & Leather', 'Electrical Equipment', and 'Other Manufacturing', these sectors receive significant subsidies. In fact, 'Textile & Leather' has the largest drop in employment (around 18%; see Figure 1), and it receives approximately 18.6% in VAT subsidies. Note that despite this relatively large subsidy, the employment share of the workers in this sector is fairly small and around 0.6% of employment.

An important observation is that the magnitude of the welfare changes (ranging from -1.70% to 1.74%) is much smaller than that of the VAT tax and subsidies (ranging from -18.6% to 5.3%). The key to this observation can be understood from the term capturing the *welfare effect* in equation 10, the term $\overline{\mathcal{W}}_j^c$. The main determinant of this effect is the welfare weights, $g^c(\theta)$, which are the Lagrange multipliers on the constraint $\hat{u}^{US}(\theta) \geq \hat{u}^{US,pre}(\theta)$. Therefore, they are negatively correlated with welfare losses from the China shock.This effect, together with differences between

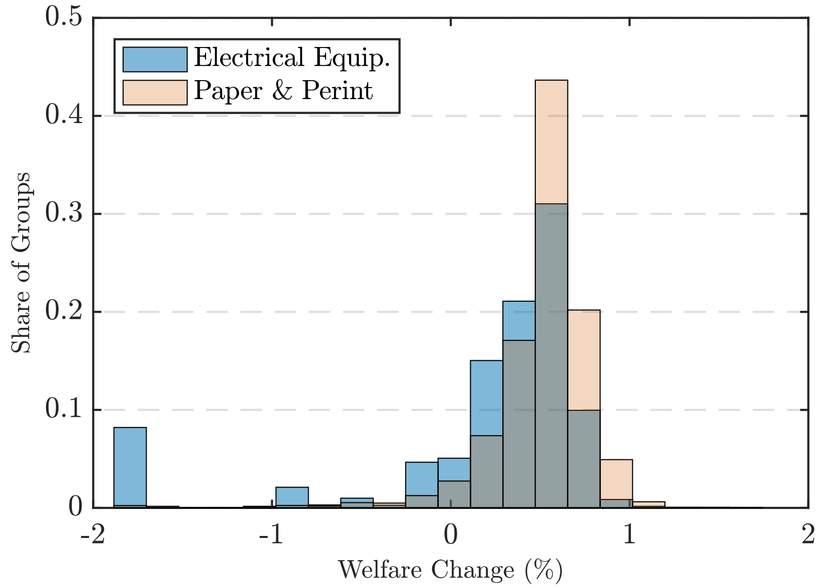


Figure 6: Distribution of welfare gains with 'Electrical Equipment' and 'Paper and Print' Sector.

the sectoral employment of different groups of workers, implies a higher degree of variation in VAT taxes. Note that, as Figure 5 depicts, the main force that shapes optimal VAT taxes is the welfare effect, while the relocation effects and externality effects are constant and insignificant across sectors.

To gain insight on what determines the differences between these welfare effects across different sectors, we plot the distribution of welfare gains/loses from China shock for two different sectors in Figure 6. The blue bars show the distribution of welfare gains/loses in 'Electrical Equipment' sector. A notable observation is the concentration of the mass at the bottom of the distribution. Approximately, 10 percent of workers in this sector experience very large losses. These large losses resulting from the China shock in this sector imply that to compensate these workers large subsidies are required. Indeed this sector received second highest subsidy rate (of 7.4%). The red bars in 6 show the distribution of welfare gains/loses in 'Paper and Print' sector. In this sector, very few workers experience big loses (or any loses). Moreover, the distribution is almost symmetric around its mean. Therefore, there is very little motive to compensate loses for these workers (if any). In fact, this sector pays value added taxes (of 3.5%) that are used to compensate losers in many other sectors (such as Electrical equipment').

At the heart of these results is the pattern of specialization of the labor force. As Figure 3 shows, groups that experience losses are the groups who have higher employment ratio in sectors that experience highest employment losses. This is particularly apprent when considering the employment shares of the losing groups in 'Electrical Equipments' and 'Textile & Leather'. Respectively, these groups of workers have employment shares of 4% and 7.3% in these two sec-

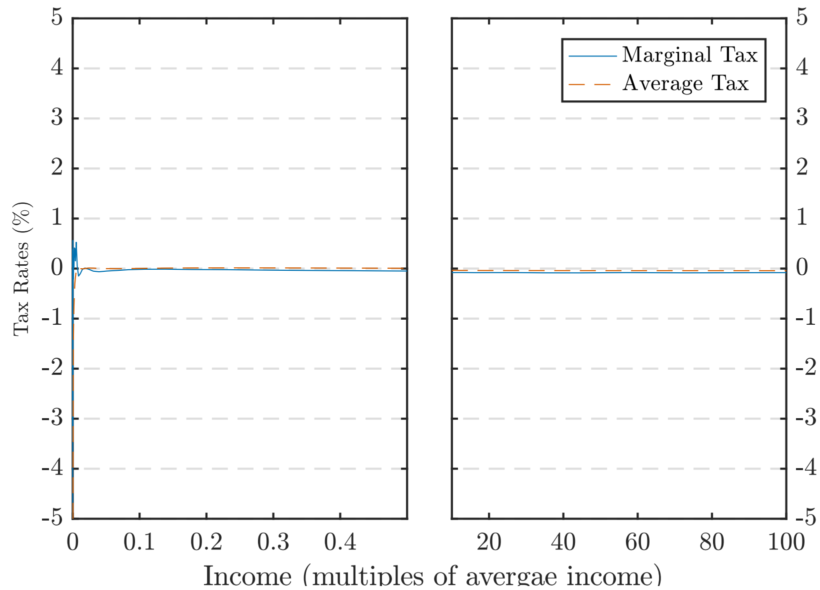


Figure 7: Optimal income taxes

tors, compared to the national employment shares of 0.6% and 2%. In other words, these groups have higher productivity in these sectors and are specialized. Therefore, cannot respond to the China shock by switching to other sectors . Providing the VAT subsidy to these sectors transfers resources to these groups of workers.

Optimal Income Taxes Our simulations imply that optimal income taxes are largely zero, except at the very bottom of the income distribution. The optimal (nonlinear) income tax rates are shown in Figure 7, where we plot marginal tax rates and average tax rates as a function of individual income (normalized by income per capita). To highlight the pattern of income tax rates at the very bottom of the income distribution, we have split the figure into two. The left panel shows the tax rates for those at the bottom 1 or 2% of the distribution. The right panel shows the tax rate for the rest of the income distribution. As we see, aside from extremely low income levels, optimal income taxes are zero.

This feature of the optimal policy is surprising, as income taxes can be potentially useful in redistributing the gains from the China shock. However, the main force that explains this result is the fact that the average income and welfare gains or losses are not highly correlated among different groups. We show this in Figure (8) by plotting each group's welfare gains or losses from the China shock against the average income (relative to per capital income). The correlation between the two is positive but very small.

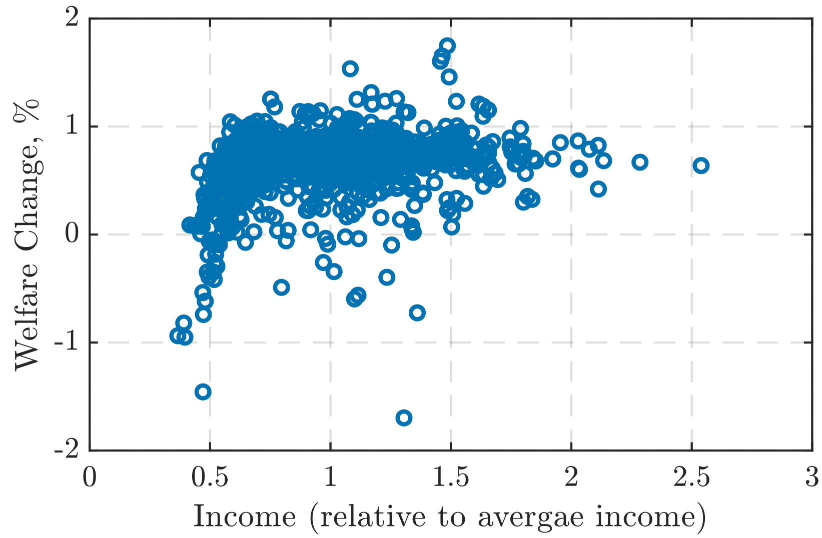


Figure 8: Welfare gains and average income across groups in the United States

5.4 Robustness

In this section we study two extensions of our main quantitative exercise. First, we consider an economy where the status quo includes a redistributive income tax. Second, we consider an alternative welfare criterion that guarantees a minimum level of welfare gain for all groups of workers in response to the China shock.

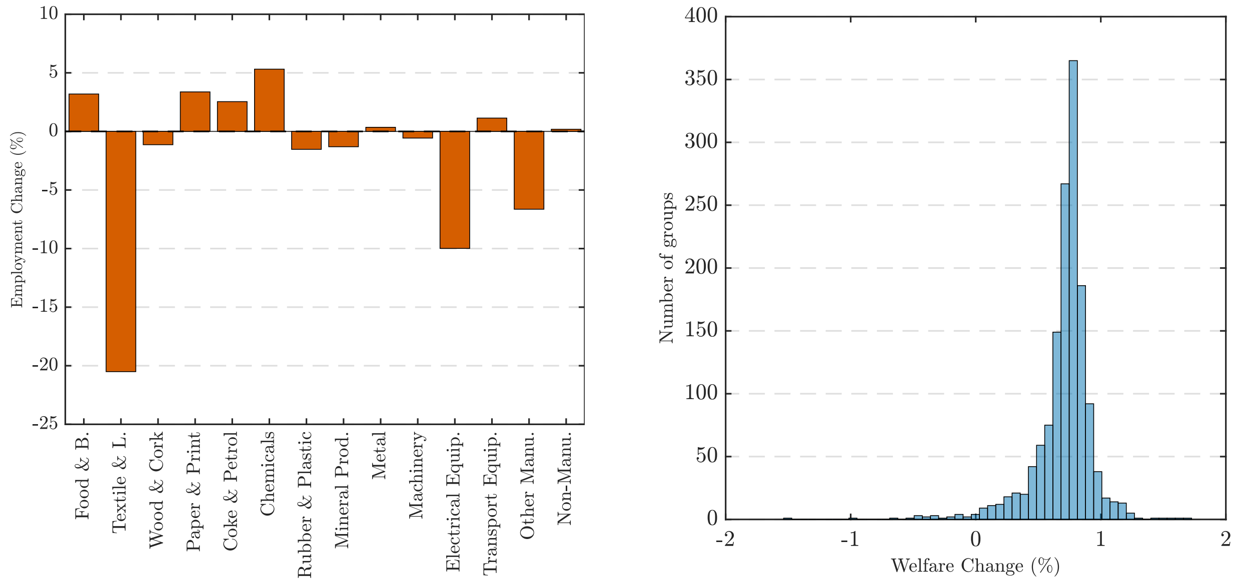
5.4.1 Redistributive Tax in Status Quo

Recall that our benchmark calibration was performed under the assumption of no government policy. In this section, we relax this assumption and assume there is an affine income tax and transfer is in place in the U.S. (for the other countries we continue with the assumption of no government policy). More precisely, let y be the before-tax labor income. Then, each worker pays the following taxes on their labor income

$$T^{US}(y) = \tau^{US} y - \overline{T}^{US}.$$

For this exercise we assume a tax rate $\tau^{US} = 0.1$ and the transfer $\overline{T}^{US}$ is such that all tax revenues are rebated to workers. These parameters are chosen so that we roughly match the level of transfers in the United States.

The rest of the parameters (and the China shock) are calibrated exactly as described in Section 5.2. Using this model, we first repeat the China shock exercise. Figure 9 shows the distribution



(a) Percentage changes in employment by sector

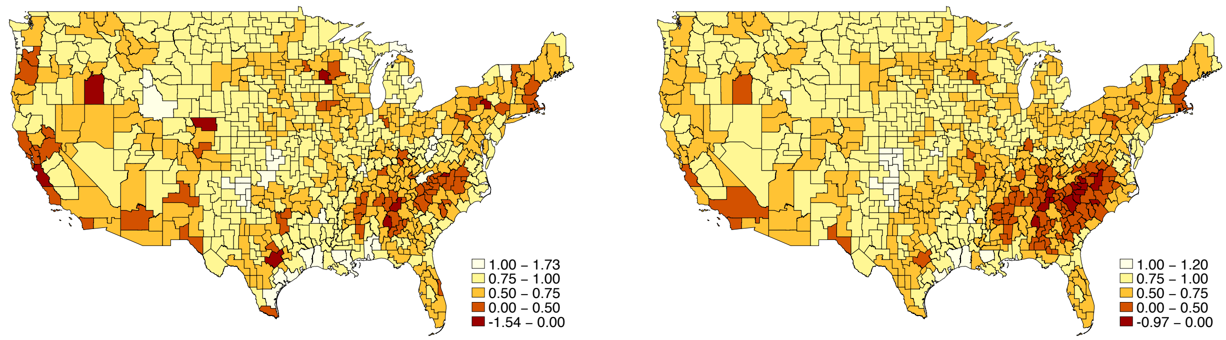
(b) Distribution of welfare gain

Figure 9: Employment and Welfare Effect of the Rise of China in economy with redistributed taxes. Panel (a) shows the percentage changes in employment by sector after the China shock. Panel (b) shows the distribution of welfare gains across different types. The rise of China leads to aggregate (income weighted) welfare gain of 0.63%. Mean welfare gains across types is 0.70% with coefficient of variation of 0.33. The minimum gain is -1.54% and maximum gain is 1.73%. Overall, 1.6 percent of all types experience a welfare loss.

of welfare gains (panel (a)) and employment changes (panel (b)) after the rise of China. As the figures demonstrate, inclusion of redistributive income taxes has a minimal effect on the outcome of the China shock in our baseline economy. Indeed, the distribution of welfare gains is very similar to the one reported for the benchmark exercise. The gains range from -1.54% to 1.73%. The aggregate (income weighted) welfare gain is 0.63%, and the mean welfare gain is 0.70%. The slightly higher minimum and high average welfare are due to the fact that the redistributive tax and transfer partially mitigate the negative employment effects. Indeed, the fraction of groups with negative gain also falls to 1.5% (from 2.5%).

Figure 10 shows the changes in welfare across commuting zones. These graphs also demonstrate the same patterns of gains and loses across commuting zones as the ones reported for the benchmark model in Section 5.2.

Using the baseline economy with redistributive income taxes, we can calculate the optimal policy response to the China shock. As Figure 11 demonstrates, the presence of the redistributive income tax has very little effect on the pattern and magnitude of taxes and subsidies across sectors. As discussed in Section 5.3, the main force behind these taxes and subsidies is welfare effects. This can be clearly seen in Figure 12, which shows the breakdown of these optimal taxes into welfare



(a) Low skilled workers

(b) High skilled workers

Figure 10: Change in welfare across commuting zones in the United States in response to China shock in economy with redistributive tax and transfer. Panel (a) depicts the percentage change in welfare for the less-educated workers – those with education less than an Associate degree. Panel (b) depicted the percentage change in welfare for educated workers. For the less-educated workers, 9 commuting zones experience a welfare loss. For the educated workers, 13 commuting zones experience a welfare loss.

effects and relocation effects.

Note that the presence of income tax generates a fiscal externality. In other words, the changes in value-added taxes and employment in a sector affect the income tax collected from workers in that sector. For this reason, the relocation effect and welfare effects have been shifted (in the opposite direction) in Figure 12 relative to Figure 5. The main insight remains the same. The pattern and magnitude of taxes and subsidies are driven by the welfare effects. The relocation terms show very little variability across sectors, and they are not an important determinant of optimal value-added taxes.

Finally, Figure 13 shows the optimal income taxes. This figure conforms to our earlier finding that a Pareto improving policy response aimed at redistributing gains from the China shock does not involve a reform or response of the income tax. As we see in Figure 13, the optimal marginal income tax is 10%, which is identical to the one assumed in the calibration. This result resembles that in Werning (2007) and Hosseini and Shourideh (2018), namely, that many income tax functions can be Pareto efficient. Therefore, a Pareto improving policy reform does not necessarily require a reform of the income tax function.

5.4.2 Alternative Welfare Criterion

So far, our policy experiments have focused on Pareto improvement relative to the status quo before the rise of China. By construction, this policy response transfers all the potential (aggregate) gains from the rise of China to the rest of the world. In this section we report the results of an optimal policy experiment that guarantees no one is worse off relative to the status quo

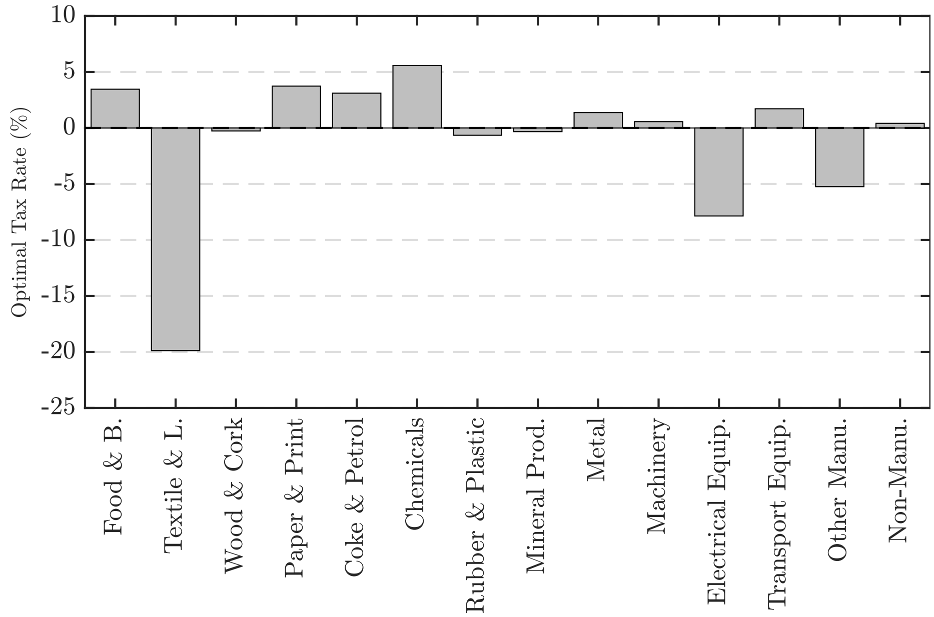


Figure 11: Optimal VAT taxes in the United States

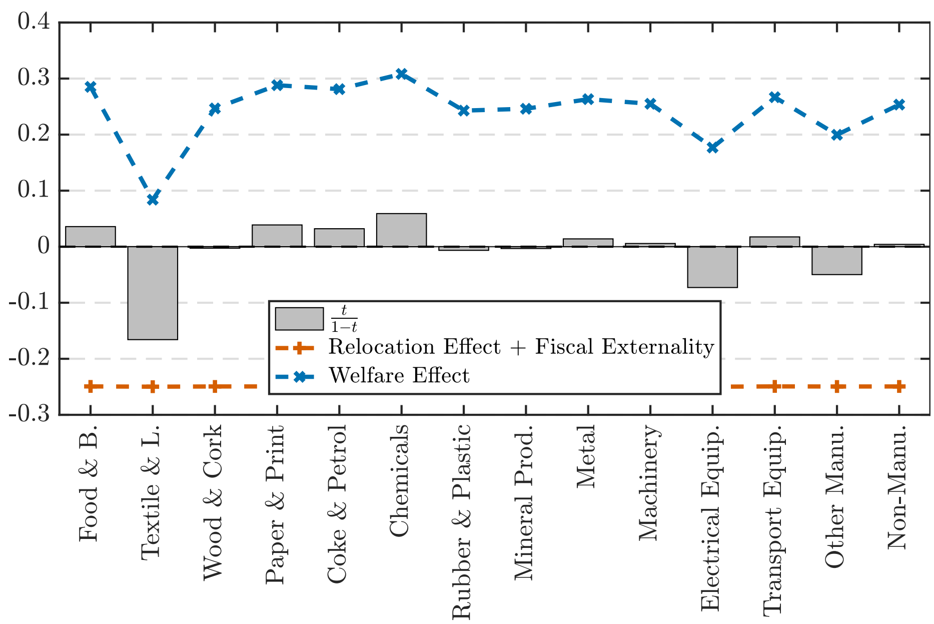


Figure 12: Breakdown of optimal VAT taxes into welfare effect and the rest: fiscal externality on income taxes and the relocation effect.

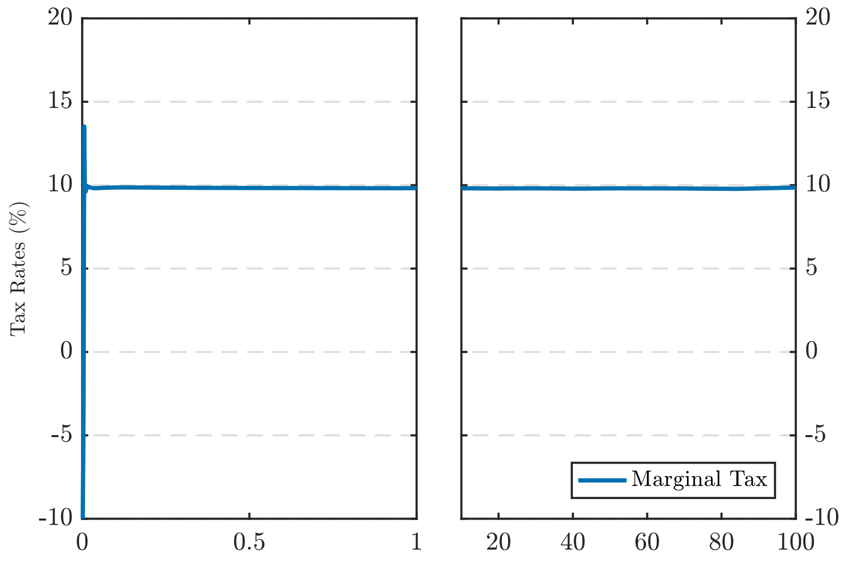


Figure 13: Optimal income tax

prior to the China shock and yet the aggregate welfare rises to the level of the post-China shock. To this end, we repeat our main exercise with a small modification. We scale up the status quo welfare (prior to the China shock) so that the average (scaled up) welfare matches the average welfare after the China shock. Recall that in Section 5.2 we reported that the rise of China results in a 0.67% rise in the average welfare. This exercise guarantees that under optimal policies the average welfare from the rise of China is realized and at the same time no group is harmed.

The resulting optimal taxes are shown in Figure 14. The optimal value-added taxes fall (and subsidies rise) relative to those in the main exercise (by about 1 or 2 percentage points). The aggregate gains, however, are larger because under these policies there are larger transfers from the rest of the world to the U.S. (or smaller transfers from the U.S. to the rest of the world).

Finally, Figure 15 compares the distribution of welfare gains across groups for three scenarios: 1) rise of China and no policy response, 2) Pareto improving policy response that guarantees the status quo welfare, and 3) Pareto improving policy response that guarantees the status quo welfare and delivers the aggregate welfare from the rise of China. The policies that deliver only the status welfare (green bar) have lower aggregate welfare gains. On the other hand, the policies that deliver the status quo and the aggregate welfare from the rise of China (red bar) have higher average welfare gains that match the average welfare gains from the rise of China in the absence of a policy response.

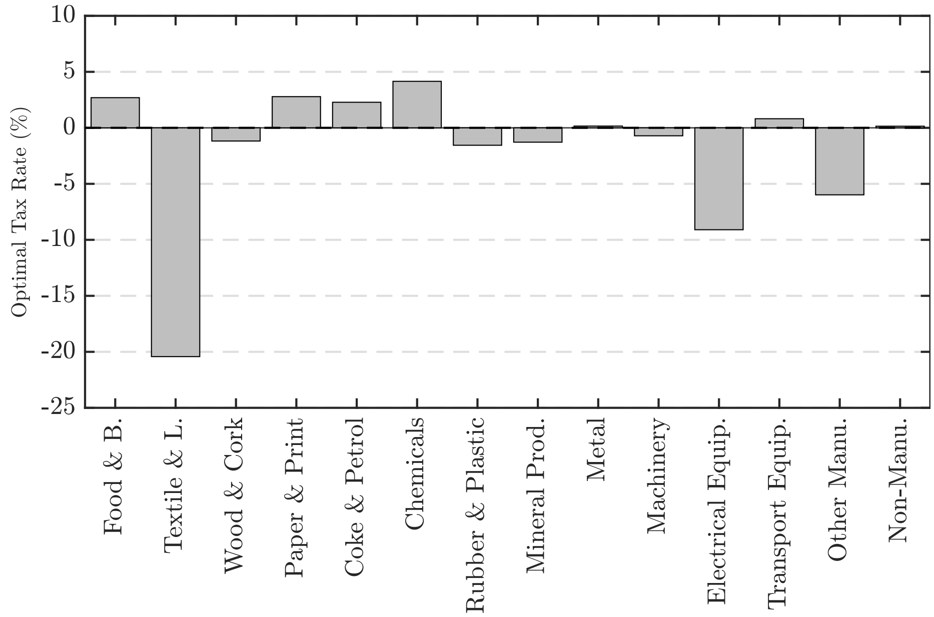


Figure 14: Optimal VAT taxes in the United States for Pareto improving policies that maintain the aggregate gain from the rise of China

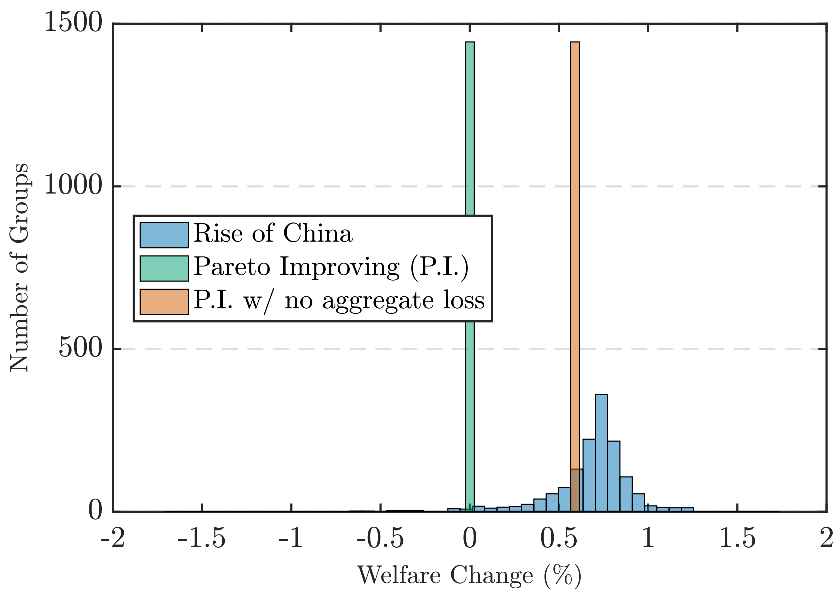


Figure 15: Distribution of welfare for three scenarios. The blue bar is the histogram of the welfare gains from the rise of China in the absence of a policy response. The green bar is the histogram of the welfare gains from the rise of China under policies that guarantee no loss relative to the status quo (prior to the rise of China). The red bar is the histogram of the welfare gains from the rise of China under policies that guarantee no loss relative to the status quo (prior to the rise of China) and deliver average welfare gains from the rise of China.

6 Conclusion

In this paper, we have provided a theoretical and quantitative analysis of optimal policies when international trade creates winners and losers and when income taxes cannot target winners and losers (i.e., lump-sum tax and transfers are unavailable). We show that for a large class of production functions, production must be efficient and, as a result, trade policy should be designed to respect that. Consequently, tariffs and other distortionary trade policies are not optimal. We show theoretically and quantitatively that producer taxes in the form of value-added taxes play an important role in redistributing the gains from trade. In contrast, income taxes are not an integral part of redistributing the gains from trade.

While our analysis provides a foundation for optimal government response to trade shocks (and potentially other technological changes), some issues remain unexplored. For example, in our model we have treated the skill decomposition of each country as given. While such an assumption might be suitable in the short-run, in the long-run workers' skills respond to changes in international trade and technology. This implies that dynamic and transitional considerations can potentially be an important determinant of optimal policies. Moreover, we have focused on tax and transfer policies, while a potentially important aspect of policy response to technological changes is training policies. These are important extensions of our framework that we leave for future work.

In our analysis, we have assumed that policies are determined under cooperation in order to focus on the design of trade agreements. Much of trade between countries is organized using a form of trade agreements: WTO, European Union, NAFTA, etc. are all examples of situations where countries negotiate trade policies. Evidently, a country has incentives to unilaterally deviate and choose different set of production and income taxes. This is an important aspect of optimal policy determination which is an important direction of future research.

References

Akerlof, G. A. (1978): "The economics of" tagging" as applied to the optimal income tax, welfare programs, and manpower planning," *The American economic review*, 68, 8–19.

Ales, L. and C. Sleet (2017): "Spatial Attraction: Optimal Geographic Social Insurance," *Working Paper, Carnegie Mellon University.*

Antràs, P., A. De Gortari, and O. Itskhoki (2017): "Globalization, inequality and welfare," *Journal of International Economics*, 108, 387–412.

Artuç, E., S. Chaudhuri, and J. McLaren (2010): "Trade shocks and labor adjustment: A structural empirical approach," *American economic review*, 100, 1008–45.

Atkinson, A. B. and J. E. Stiglitz (1976): "The design of tax structure: direct versus indirect taxation," *Journal of public Economics*, 6, 55–75.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *The American Economic Review*, 103, 2121–2168.

Bagwell, K. and R. W. Staiger (1999): "An economic theory of GATT," *American Economic Review*, 89, 215–248.

Beshkar, M. and A. Lashkaripour (2017): "Interdependence of Trade Policies in General Equilibrium," .

Caliendo, L., M. Dvorkin, and F. Parro (2015): "The impact of trade on labor market dynamics," Tech. rep., National Bureau of Economic Research.

——— (2017): "Trade and Labor Market Dynamics," *Working Paper, Yale University.*

Caliendo, L. and F. Parro (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82, 1–44.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *American Economic Review*, 101, 471–75.

Costinot, A., D. Donaldson, J. Vogel, and I. Werning (2015): "Comparative advantage and optimal trade policy," *The Quarterly Journal of Economics*, 130, 659–702.

Costinot, A. and I. Werning (2017): "The Lerner Symmetry Theorem: Generalizations and Qualifications," Tech. rep., National Bureau of Economic Research.

——— (2018): "Robots, Trade, and Luddism," *Working Paper, MIT.*

Deaton, A. (1981): "Optimal taxes and the structure of preferences," *Econometrica: Journal of the Econometric Society*, 1245–1260.

Diamond, P. A. (1998): "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, 83–95.

Diamond, P. A. and J. A. Mirrlees (1971): “Optimal taxation and public production I: Production efficiency,” *The American Economic Review*, 61, 8–27.

Dix-Carneiro, R. and B. K. Kovak (2017): “Trade Liberalization and Regional Dynamics,” *American Economic Review*, 107, 2908–46.

Dixit, A. (1985): “Tax policy in open economies,” *Handbook of public economics*, 1, 313–374.

Dixit, A. and V. Norman (1980): *Theory of international trade*, Cambridge University Press.

——— (1986): “Gains from trade without lump-sum compensation,” *Journal of International Economics*, 21, 111–122.

Dornbusch, R., S. Fischer, and P. A. Samuelson (1977): “Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods,” *The American Economic Review*, 67, 823–839.

Eaton, J. and S. Kortum (2002): “Technology, geography, and trade,” *Econometrica*, 70, 1741–1779.

Fajgelbaum, P. D. and C. Gaubert (2018): “Optimal Spatial Policies, Geography and Sorting,” .

Galle, S., A. Rodriguez-Clare, and M. Yi (2017): “Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade,” *NBER Working Paper*.

Goldberg, P. K. and N. Pavcnik (2007): “Distributional effects of globalization in developing countries,” *Journal of economic Literature*, 45, 39–82.

Gomes, R., J.-M. Lozachmeur, and A. Pavan (2017): “Differential taxation and occupational choice,” *The Review of Economic Studies*, 85, 511–557.

Grossman, G. M. (1980): “Border tax adjustments: Do they distort trade?” *Journal of international economics*, 10, 117–128.

Guesnerie, R. (1998): “Peut-on toujours redistribuer les gains à la spécialisation et à l’échange? Un retour en pointillé sur Ricardo et Heckscher-Ohlin,” *Revue économique*, 555–579.

Head, K. and T. Mayer (2014): “Gravity equations: Workhorse, toolkit, and cookbook,” in *Handbook of international economics*, Elsevier, vol. 4, 131–195.

Helpman, E., O. Itskhoki, M.-A. Muendler, and S. J. Redding (2017): “Trade and inequality: From theory to estimation,” *The Review of Economic Studies*, 84, 357–405.

Helpman, E., O. Itskhoki, and S. Redding (2010): "Inequality and unemployment in a global economy," *Econometrica*, 78, 1239–1283.

Hosseini, R. and A. Shourideh (2018): "Retirement Financing: An Optimal Reform Approach," *Working Paper, Carnegie Mellon University.*

Inklaar, R. and M. P. Timmer (2014): "The relative price of services," *Review of Income and Wealth*, 60, 727–746.

Jacobs, B. (2015): "Optimal Inefficient Production," *Working Paper, Erasmus University, Rotterdam.*

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

Lewis, L. T., R. Monarch, M. Sposi, and J. Zhang (2018): "Structural Change and Global Trade," *Working Paper, Federal Reserve Back of Dallas.*

Lyon, S. and M. E. Waugh (2017): "Redistributing the Gains From Trade through Progressive Taxation," *Working Paper, New York University.*

Mirrlees, J. A. (1971): "An exploration in the theory of optimum income taxation," *The review of economic studies*, 38, 175–208.

Naito, H. (1999): "Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency," *Journal of Public Economics*, 71, 165–188.

Opp, M. M. (2010): "Tariff wars in the Ricardian Model with a continuum of goods," *Journal of International Economics*, 80, 212–225.

Ramsey, F. P. (1927): "A Contribution to the Theory of Taxation," *The Economic Journal*, 37, 47–61.

Rothschild, C. and F. Scheuer (2013): "Redistributive taxation in the Roy model," *The Quarterly Journal of Economics*, 128, 623–668.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3, 135–146.

Saez, E. (2001): "Using elasticities to derive optimal income tax rates," *The review of economic studies*, 68, 205–229.

——— (2002): "Optimal income transfer programs: intensive versus extensive labor supply responses," *The Quarterly Journal of Economics*, 117, 1039–1073.

——— (2004): "Direct or indirect tax instruments for redistribution: short-run versus long-run," *Journal of Public Economics*, 88, 503–518.

Samuelson, P. A. (1954): "The transfer problem and transport costs, II: Analysis of effects of trade impediments," *The Economic Journal*, 64, 264–289.

Spector, D. (2001): "Is it possible to redistribute the gains from trade using income taxation?" *Journal of International Economics*, 55, 441–460.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. De Vries (2015): "An illustrated user guide to the world input–output database: the case of global automotive production," *Review of International Economics*, 23, 575–605.

Verhoogen, E. A. (2008): "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 123, 489–530.

Waugh, M. (2010): "International trade and income differences," *The American Economic Review*, 100, 2093–2124.

Werning, I. (2007): "Pareto Efficient Income Taxation," *Working Paper, MIT*.

Appendix

A Proofs

A.1 Derivation of the Distribution of Labor Productivity

Suppose that wages are given by $\mathbf{w}^c = \left\{w_j^c\right\}$. Then the distribution of labor productivity for workers of type θ can be drived as follows

$$\begin{aligned}
\Pr\left(z \leq \hat{z}\right) &= \sum_{j=1}^{N} \int_{\mathbb{R}_+^N} \mathbf{1}\left[w_j^c a_j^c\left(\theta\right)\epsilon_j \geq w_i^c a_i^c\left(\theta\right)\epsilon_i, \forall i\right] \mathbf{1}\left[\hat{z} \geq w_j^c a_j^c\left(\theta\right)\epsilon_j\right] dH^c\left(\boldsymbol{\epsilon}\right) \\
&= \sum_{j=1}^{N} \int_0^{\frac{\hat{z}}{w_j^c a_j^c(\theta)}} \prod_{i\neq j} e^{-\left(\frac{w_j^c a_j^c(\theta)\epsilon_j}{w_i^c a_i^c(\theta)}\right)^{-\sigma^c}} \sigma^c \epsilon_j^{-\sigma^c-1} e^{-\epsilon_j^{-\sigma^c}} d\epsilon_j \\
&= \sum_{j=1}^{N} \int_0^{\frac{\hat{z}}{w_j^c a_j^c(\theta)}} e^{-\left(w_j^c a_j^c(\theta)\right)^{-\sigma^c} \epsilon_j^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}} \sigma \epsilon_j^{-\sigma^c-1} d\epsilon_j \\
&= \sum_{j=1}^{N} \int_0^{\frac{\hat{z}}{w_j^c a_j^c(\theta)}} \frac{\left(w_j^c a_j^c\left(\theta\right)\right)^{\sigma^c}}{\sum_i \left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}} d\left(e^{-\left(w_j^c a_j^c(\theta)\right)^{-\sigma^c} \epsilon_j^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}\right) \\
&= \sum_{j=1}^{N} \frac{\left(w_j^c a_j^c\left(\theta\right)\right)^{\sigma^c}}{\sum_i \left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}} d\left(e^{-\left(w_j^c a_j^c(\theta)\right)^{-\sigma^c} \left(\frac{\hat{z}}{w_j^c a_j^c(\theta)}\right)^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}\right) \\
&= \sum_{j=1}^{N} \frac{\left(w_j^c a_j^c\left(\theta\right)\right)^{\sigma^c}}{\sum_i \left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}} e^{-\hat{z}^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}} \\
&= e^{-\hat{z}^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}
\end{aligned}$$

This implies that labor productivity of workers of type θ has a Frechét distribution with shape parameter σ^c and mean of $\Gamma\left(1 - \frac{1}{\sigma^c}\right)\left[\sum_i \left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c}\right]^{\frac{1}{\sigma^c}}$. This shows that

$$\begin{aligned}
\lambda\left(z, \theta; \mathbf{w}^c\right) &= \frac{\partial}{\partial z} e^{-z^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}} \\
&= \sigma^c z^{-\sigma^c-1} \sum_i \left(w_i^c a_i^c\left(\theta\right)\right)^{\sigma^c} e^{-z^{-\sigma^c} \sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}
\end{aligned}$$

Additionally, the fraction of type θ workers that work in sector j is given by

$$\begin{aligned}
\Lambda_j^c(\theta) &= \int_{\mathbb{R}_+^N} \mathbf{1}\left[w_j^c a_j^c(\theta)\epsilon_j \geq w_i^c a_i^c(\theta)\epsilon_i, \forall i\right] dH^c(\boldsymbol{\epsilon}) \\
&= \int_0^\infty \frac{\left(w_j^c a_j^c(\theta)\right)^{\sigma^c}}{\sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}} d\left(e^{-\left(w_j^c a_j^c(\theta)\right)^{-\sigma^c}\epsilon_j^{-\sigma^c}\sum_i\left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}\right) \\
&= \frac{\left(w_j^c a_j^c(\theta)\right)^{\sigma^c}}{\sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma^c}}
\end{aligned}$$

A similar derivation as above establishes that $\lambda_j^c(z,\theta;\mathbf{w}^c) = \Lambda_j^c(\theta)\lambda^c(z,\theta;\mathbf{w}^c)$.

A.2 Proof of Lemma 1

Proof. The welfare of a worker of type θ is given by

$$\begin{aligned}
\hat{u}^c(\theta) &= \int_0^\infty u^c(z)\lambda^c(z,\theta;\mathbf{w}^c)\,dz \\
&= \int_0^\infty u^c(z)\,d\left(e^{-\left[\sum_j\left(a_j^c(\theta)w_j^c\right)^{\sigma^c}\right]z^{-\sigma^c}}\right) \\
&= -\int_0^\infty u^c(z)\,d\left(1-e^{-\left[\sum_j\left(a_j^c(\theta)w_j^c\right)^{\sigma^c}\right]z^{-\sigma^c}}\right) \\
&= -\,u^c(z)\left(1-e^{-\left[\sum_j\left(a_j^c(\theta)w_j^c\right)^{\sigma^c}\right]z^{-\sigma^c}}\right)\Bigg|_0^\infty \\
&\quad + \int_0^\infty\left(1-e^{-\left[\sum_j\left(a_j^c(\theta)w_j^c\right)^{\sigma^c}\right]z^{-\sigma^c}}\right)\left(u^c\right)'(z)\,dz \\
&= u^c(0) \\
&\quad + \int_0^\infty\left(1-e^{-\left[\sum_j\left(a_j^c(\theta)w_j^c\right)^{\sigma^c}\right]z^{-\sigma^c}}\right)\left(u^c\right)'(z)\,dz
\end{aligned}$$

where in the above we have used integration by parts. Differentiating both sides with respect to wages implies that

$$
\begin{aligned}
\delta \hat{u}^c(\theta) &= \delta u^c(0) + \sigma^c \sum_{j=1}^{N} \int_0^\infty z^{-\sigma^c} \left(a_j^c(\theta) w_j^c\right)^{\sigma^c} \frac{\delta w_j^c}{w_j^c} e^{-\left[\sum_j \left(a_j^c(\theta) w_j^c\right)^{\sigma^c}\right] z^{-\sigma^c}} \left(u^c\right)'(z)\, dz \\
&= \delta u^c(0) + \sigma^c \sum_{j=1}^{N} \left(a_j^c(\theta) w_j^c\right)^{\sigma^c} \frac{\delta w_j^c}{w_j^c} \int_0^\infty z^{-\sigma^c} e^{-\left[\sum_j \left(a_j^c(\theta) w_j^c\right)^{\sigma^c}\right] z^{-\sigma^c}} \left(u^c\right)'(z)\, dz \\
&= \delta u^c(0) + \sigma^c \sum_{j=1}^{N} \frac{\left(a_j^c(\theta) w_j^c\right)^{\sigma^c}}{\sum_i \left(a_i^c(\theta) w_i^c\right)^{\sigma^c}} \frac{\delta w_j^c}{w_j^c} \int_0^\infty z \left(u^c\right)'(z)\, d\left(e^{-\left[\sum_j \left(a_j^c(\theta) w_j^c\right)^{\sigma^c}\right] z^{-\sigma^c}}\right) \\
&= \delta u^c(0) + \sigma^c \sum_{j=1}^{N} \Lambda_j^c(\theta) \frac{\delta w_j^c}{w_j^c} \int_0^\infty z \left(u^c\right)'(z)\, \lambda^c(z, \theta; \mathbf{w}^c)\, dz
\end{aligned}
$$

which proves the claim. $\square$

A.3 Proof of Proposition 2

Proof. That any allocations and consumer and producer price vectors must satisfy the constraints in problem (P) is straightforward from the definition of indirect utility function, demand function, and optimality condtion by firms for their choice of intermediate goods.

Now suppose that allocations $\left\{I^c(z), \ell^c(z), \lambda^c(z, \theta; \mathbf{w}^c), \Lambda_j^c(\theta), u^c(z)\right\}_{z \in \mathbb{R}_+, c \in \{1, \cdots, C\}}$, and $\left\{\left\{Q_{ij}^c\right\}, L_i^c\right\}_{c \in \{1, \cdots, C\}, i \in \{1, \cdots, N\}}$ together with consumer prices and wage $\left\{\mathbf{q}^c, \mathbf{w}^c\right\}_{c \in \{1, \cdots, C\}}$ satisfy the constraints in (P). We construct a set of commodity taxes and show that under these policies the resulting allocations and prices are a competive equilibrium.

Using the allocation, we can simply construct producer prices according to

$$
p_i^c = \frac{w_i^c}{\frac{\partial}{\partial L} G_i^c\left(L_i^c, \left\{Q_{ij}^c\right\}\right)}
$$

Consider an arbitrary vector of international prices $\hat{\mathbf{p}} = (\hat{p}_1, \cdots, \hat{p}_N)$ where $\hat{p}_i > 0$. For each good i and country c, we can define

$$
t_i^{x,c} = \frac{q_i^c}{\hat{p}_i} - 1
$$

and

$$
\begin{aligned}
t_i^{p,c} &= \frac{p_i^c}{\hat{p}_i} - 1 \\
t_{ij}^{p,c} &= \frac{p_i^c}{\hat{p}_j} \frac{\partial G_i^c}{\partial Q_{ij}} \left(\left\{Q_{ij}^c\right\}, L_i^c\right) - 1
\end{aligned}
$$

Obviously, at prices given by $\hat{\mathbf{p}}$ and taxes as defined above the firms' first order conditions are satisfied. Since the production functions are assumed to be concave, then we must have that firms' optimality conditions are satisfied. Moreover, by definition of the demand function and indirect value function, the workers' optimality conditions are also satisfied. This implies that we only need to check that government budget constraints are satisfied. Since all workers budget constraints hold - given the definition of the demand function, all feasibility constraints are satisfied - the first constraint in problem (P), and transfers between governments are allowed, Walras law implies that government budget constraints are satisfied. This completes the proof. □

A.4 Proof of Lemma 2

Proof. If $\left\{Q_{ij}^c\right\}_{i,j,c}$ solves the optimization in the statement for some positive vector ρ_i, then obviously there cannot exist an alternative $\left\{\hat{Q}_{ij}^c\right\}$ which increases final consumption of all goods – since this such an allocation would improve the objective of the optimization (8).

To show the reverse consider an allocation $\left\{Q_{ij}^c\right\}$ that is pseduo-efficient. Define the following two sets:

$$\mathcal{A} = \left\{(x_1, \cdots, x_N); \exists \left\{\hat{Q}_{ij}^c\right\}, x_i \leq \sum_{c=1}^{C} G_i^c\left(L_i^c, \left\{\hat{Q}_{ij}^c\right\}\right) - \sum_{c=1}^{C}\sum_{j=1}^{N} \hat{Q}_{ji}^c\right\}$$
$$\mathcal{B} = \left\{(x_1, \cdots, x_N); \sum_{c=1}^{C} G_i^c\left(L_i^c, \left\{Q_{ij}^c\right\}\right) - \sum_{c=1}^{C}\sum_{j=1}^{N} Q_{ji}^c \geq x_i\right\}$$

By definition of psuedo-efficiency, $\mathcal{A} \cap \mathcal{B} = \{(X_1, \cdots, X_N)\}$ where X_i is the total consumption of good i under the allocation $\left\{Q_{ij}^c\right\}$. Moreover, since G_i^c is concave, the sets $\mathcal{A}$ and $\mathcal{B}$ are convex. Therefore, by the separating hyperplane theorem, there must exists a non-zero vector $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_N)$ such that

$$\forall x \in \mathcal{A}, y \in \mathcal{B}, \sum_{i=1}^{N} \rho_i x_i \leq \sum_{i=1}^{N} \rho_i y_i$$

Note that for all i, $(X_1, \cdots, X_{i-1}, X_i + \varepsilon, X_{i+1}, \cdots, X_N) \in \mathcal{B}$, since $(X_1, \cdots, X_N) \in \mathcal{A}$, the above inequality implies that $\rho_i \geq 0$. Moreover, since $(X_1, \cdots, X_N) \in \mathcal{B}$, for all $(x_1, \cdots, x_N) \in \mathcal{A}$, we must have

$$\sum_{i=1}^{N} \rho_i x_i \leq \sum_{i=1}^{N} \rho_i X_i$$

This concludes the proof. □

A.5 Proof of Proposition 5

Proof. Consider the optimization (P) and let ρ_i be the multiplier on the resource constraint for good i, ζ_i be the lagrange multiplier on labor market clearing for sector i country c and let $\gamma^c(z)$ be the lagrange multiplier on the second constraint in program (P).

We, first, show that there must exist a constant ρ such that $q_i^c = \rho^{-1}\rho_i$. This proof closely follows Deaton (1981).

Moreover, let $V^c(I;\mathbf{q}) = \frac{I}{\nu^c(\mathbf{q})} + \vartheta^c(\mathbf{q})$. Note that under this assumption, Roy's identity implies that $x_i^c(I;\mathbf{q}) = s_i^c(\mathbf{q}) I + b_i^c(\mathbf{q})$ where $s_i^c(\mathbf{q}) = \frac{\frac{\partial}{\partial q_i}\nu^c(\mathbf{q})}{\nu^c(\mathbf{q})}$ and $b_i^c(\mathbf{q}) = -\nu^c(\mathbf{q})\frac{\partial}{\partial q_i}\vartheta^c(\mathbf{q})$. Note that by budget constraint, we must have that $\sum_i q_i s_i^c(\mathbf{q}) = 1$ and $\sum_i q_i b_i^c(\mathbf{q}) = 0$.

Now, if we take first order condition with respect to q_i^c, we have

$$-\sum_j \rho_j \left[\frac{\partial s_j^c}{\partial q_i}\bar{I}^c + \frac{\partial b_j^c}{\partial q_i}\right] + \int \gamma^c(z)\left[-\frac{\frac{\partial}{\partial q_i}\nu^c(\mathbf{q})}{\nu^c(\mathbf{q})^2} I^c(z) + \frac{\partial}{\partial q_i}\vartheta^c(\mathbf{q})\right] dz = 0$$

where $\bar{I}^c = \int_\Theta \int I^c(z)\lambda^c(z,\theta;\mathbf{w}^c)\, dz d\mu^c$. Moreover, the first order condition with respect to $I^c(z)$ gives us

$$-\sum_j \rho_j s_j^c \int_\Theta \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c + \frac{\gamma^c(z)}{\nu^c(\mathbf{q})} = 0$$

This implies that

$$\int \gamma^c(z)\, dz = \nu^c(\mathbf{q}) \sum_j \rho_j s_j^c(\mathbf{q}) \tag{15}$$

Now, if we set $q_i^c = \frac{1}{\rho}\rho_i$ for some ρ, we have

$$\begin{aligned}
\sum_j \rho_j \frac{\partial s_j^c}{\partial q_i} &= \rho \sum_j q_j^c \frac{\partial s_j^c}{\partial q_i} = -\rho s_i^c = -\rho\frac{\frac{\partial}{\partial q_i}\nu^c(\mathbf{q})}{\nu^c(\mathbf{q})} \\
\sum_j \rho_j s_j^c &= \rho \sum_j q_j^c s_j^c = \rho \\
\sum_j \rho_j \frac{\partial b_j^c}{\partial q_i} &= \rho \sum_j q_j^c \frac{\partial b_j^c}{\partial q_i} = -\rho b_i^c = \rho\nu^c(\mathbf{q})\frac{\partial}{\partial q_i}\vartheta^c(\mathbf{q})
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
-\sum_j \rho_j \frac{\partial s_j^c}{\partial q_i}\bar{I}^c - \frac{\frac{\partial}{\partial q_i}\nu^c(\mathbf{q})}{\nu^c(\mathbf{q})^2}\int \gamma^c(z) I^c(z)\, dz = \\
\rho s_i^c \bar{I}^c - \frac{\frac{\partial}{\partial q_i}\nu^c(\mathbf{q})}{\nu^c(\mathbf{q})}\rho\bar{I}^c = 0
\end{aligned}$$

and

$$
\begin{aligned}
-\sum_j \rho_j \frac{\partial b_j^c}{q_i} + \frac{\partial}{\partial q_i}\vartheta^c \int \gamma^c(z)\, dz = \\
\rho b_i^c + \frac{\partial}{\partial q_i}\vartheta^c \nu^c(\mathbf{q})\, \rho = 0
\end{aligned}
$$

In other words, the FOC is satisfied when $q_i^c = \frac{1}{\rho}\rho_i$. We refer the reader to Deaton (1981) for the proof of why this must be satisfies at every solution of the FOC.

Now, consider the FOC with respect to L_j^c. It is given by

$$
-\zeta_i^c + \rho_i \frac{\partial}{\partial L} G_i^c = 0
$$

Finally, the FOC with respect to w_j^c is given by

$$
\begin{aligned}
\psi^c \int g^c(\theta) \int u^c(z) \frac{\partial}{\partial w_j^c} \lambda^c(z, \theta; \mathbf{w}^c)\, dz d\mu^c \\
-\sum_i \rho_i s_i^c \int \int I^c(z) \frac{\partial}{\partial w_j^c} \lambda^c(z, \theta; \mathbf{w}^c)\, dz d\mu^c \\
-\zeta_j^c \frac{1}{\left(w_j^c\right)^2} \int \Lambda_j^c \int y^c \lambda^c dz \mu^c \\
+\sum_i \zeta_i^c \int \Lambda_i^c \int z\ell^c \frac{\partial}{\partial w_j^c} \lambda^c(z, \theta; \mathbf{w}^c)\, dz d\mu^c \\
+\sum_i \zeta_i^c \int \frac{\partial \Lambda_i^c}{\partial w_j^c} \int z\ell^c \lambda^c dz d\mu^c = 0
\end{aligned}
$$

Note that since $\lambda^c(z, \theta; \mathbf{w}^c) = \frac{d}{dz}\left(e^{-z^{-\sigma}\sum_j \left(a_j^c(\theta) w_j^c\right)^\sigma}\right)$, therefore

$$
\begin{aligned}
\frac{\partial}{\partial w_j^c} \lambda^c(z, \theta; \mathbf{w}^c) &= \frac{\partial^2}{\partial z \partial w_j^c}\left(e^{-z^{-\sigma}\sum_j \left(a_j^c(\theta) w_j^c\right)^\sigma}\right) \\
&= \frac{\sigma\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{w_j^c} \frac{\partial}{\partial z}\left(-z^{-\sigma} e^{-z^{-\sigma}\sum_j \left(a_j^c(\theta) w_j^c\right)^\sigma}\right)
\end{aligned}
$$

Therefore, for any differentiable function $\varphi(z)$, we have

$$\begin{aligned}
\int \varphi(z) \frac{\partial}{\partial w_j^c} \lambda^c(z,\theta;\mathbf{w}^c)\, dz &= \frac{\sigma\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{w_j^c} \int \varphi(z)\, d\left(z^{-\sigma} e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma}\right) \\
&= \frac{\sigma\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{w_j^c} \int \varphi(z)\, d\left(-z^{-\sigma} e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma}\right) \\
&= \frac{\sigma\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{w_j^c} \left[-\varphi(z)\, z^{-\sigma} e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma}\Big|_0^\infty\right] \\
&\quad + \frac{\sigma\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{w_j^c} \int \varphi'(z)\, z^{-\sigma} e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma} dz \\
&= \frac{\sigma}{w_j^c} \frac{\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{\sum_i\left(a_i^c(\theta)\, w_i^c(\theta)\right)^\sigma} \int \varphi'(z)\, z^{-\sigma} \sum_i \left(a_i^c(\theta)\, w_i^c(\theta)\right)^\sigma e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma} dz \\
&= \frac{1}{w_j^c} \frac{\left(a_j^c(\theta)\, w_j^c\right)^\sigma}{\sum_i\left(a_i^c(\theta)\, w_i^c(\theta)\right)} \int z\varphi'(z)\, \sigma z^{1-\sigma} \sum_i \left(a_i^c(\theta)\, w_i^c(\theta)\right)^\sigma e^{-z^{-\sigma}\sum_j\left(a_j^c(\theta) w_j^c\right)^\sigma} dz \\
&= \frac{1}{w_j^c} \Lambda_j^c(\theta) \int z\varphi'(z)\, \lambda^c(z,\theta;\mathbf{w}^c)\, dz
\end{aligned}$$

We can therefore write the FOC for w_j^c as

$$\begin{aligned}
\psi^c \int g^c(\theta)\, \Lambda_j^c(\theta) \int z\,(u^c)'(z)\, \lambda^c(z,\theta;\mathbf{w}^c)\, dz d\mu^c \\
-\sum_i \rho_i s_i^c \int \Lambda_j^c(\theta) \int z\,(I^c)'(z)\, \lambda^c(z,\theta;\mathbf{w}^c)\, dz d\mu^c \\
-\zeta_j \frac{1}{w_j^c} \int \Lambda_j^c \int y^c \lambda^c dz d\mu^c \\
+\sum_i \frac{\zeta_i^c}{w_i^c} \int \Lambda_i^c \Lambda_j^c \int z\,(y^c)'\, \lambda^c(z,\theta;\mathbf{w}^c)\, dz d\mu^c \\
+\sigma \sum_i \frac{\zeta_i^c}{w_i^c} \int \Lambda_i^c(\theta) \left[\mathbf{1}[i=j] - \Lambda_j^c(\theta)\right] \int y^c \lambda^c dz d\mu^c = 0
\end{aligned}$$

Note that

$$
\begin{aligned}
I^c(z) &= y^c(z) - T^c(y^c(z)) \\
(I^c)'(z) &= (y^c)'(z)\left[1-\tau_\ell(z)\right] \\
z\left(1-\tau_\ell(z)\right)\nu^c &= v'\left(\frac{y^c(z)}{z}\right) \\
\frac{1}{z} - \frac{\tau_\ell'(z)}{1-\tau_\ell(z)} &= \frac{v''}{v'}\left(\frac{(y^c)'(z)}{z} - \frac{y^c(z)}{z^2}\right) \\
\frac{1}{z} - \frac{\tau_\ell'(z)}{1-\tau_\ell(z)} &= \frac{v'' y^c/z}{v'}\left(\frac{(y^c)'}{y^c} - \frac{1}{z}\right) \\
\frac{1}{z} - \frac{\tau_\ell'(z)}{1-\tau_\ell(z)} &= \frac{1}{\varepsilon^c(z)}\left(\frac{(y^c)'}{y^c} - \frac{1}{z}\right) \\
\frac{1+\varepsilon^c}{z} - \varepsilon^c\frac{\tau_\ell'}{1-\tau_\ell} &= \frac{(y^c)'}{y^c} \\
y^c\left(1-\tau_\ell\right)\left[\frac{1+\varepsilon^c}{z} - \varepsilon^c\frac{\tau_\ell'}{1-\tau_\ell}\right] &= (I^c)' \\
\frac{\left(1-\tau_\ell\right)\ell^c}{\nu^c(\mathbf{q})} &= (u^c)'
\end{aligned}
$$

Therefore, the above FOC becomes

$$
\begin{aligned}
&\psi^c \int g^c(\theta)\Lambda_j^c(\theta)\int \frac{y^c\left(1-\tau_\ell^c\right)}{\nu^c(\mathbf{q}^c)}\lambda^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c \\
&-\rho\int \Lambda_j^c(\theta)\int y^c\left(1-\tau_\ell\right)\left[1+\varepsilon^c-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c \\
&-\rho_j G_{j,L}^c\frac{1}{w_j^c}\int \Lambda_j^c\int y^c\lambda^c dzd\mu^c \\
&+\sum_i \frac{\rho_i G_{i,L}^c}{w_i^c}\int \Lambda_i^c\Lambda_j^c\int y^c\left[1+\varepsilon^c-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c\left(z,\theta;\mathbf{w}^c\right)dzd\mu^c \\
&+\sigma\sum_i \frac{\rho_i G_{i,L}^c}{w_i^c}\int \Lambda_i^c(\theta)\left[\mathbf{1}\left[i=j\right]-\Lambda_j^c(\theta)\right]\int y^c\lambda^c dzd\mu^c = 0
\end{aligned}
$$

Note that if we choose world prices to be $\hat{p}_i = \rho_i/\rho = q_i^c$, then

$$
\frac{\rho_i G_{i,L}^c}{w_i^c} = \rho\frac{\hat{p}_i G_{i,L}^c}{w_i^c} = \frac{\rho}{1-t_i^c} \tag{16}
$$

Thus, we can write the above as

$$\begin{aligned}\rho\left(\sigma-1\right)\frac{1}{1-t_j^c}\int\Lambda_j^c\int y^c\lambda^c dz d\mu^c\\+\int\frac{\psi^c g^c}{\nu^c}\Lambda_j^c\int y^c\left(1-\tau_\ell^c\right)\lambda^c dz d\mu^c\\-\rho\int\Lambda_j^c\int y^c\left(1-\tau_\ell^c\right)\left[1+\varepsilon^c-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c dz d\mu^c\\+\rho\sum\int\frac{\Lambda_i^c\Lambda_j^c}{1-t_i^c}\int y^c\left[1+\varepsilon^c-\sigma-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c dz d\mu^c=0\end{aligned}\tag{17}$$

Moreover, a uniform increase in $u^c\left(z\right)$ implies the following condition:

$$\int\psi^c g^c\left(\theta\right)d\mu^c=\int\gamma^c\left(z\right)dz=\nu^c\sum_i\rho_i s_i^c=\nu^c\rho\tag{18}$$

where in the above we have used (15). Dividing (17) by ρ and using the above, we can write

$$\begin{aligned}\left(\sigma-1\right)\frac{1}{1-t_j^c}Y_j^c\\+\frac{\int g^c\Lambda_j^c\int y^c\left(1-\tau_\ell^c\right)\lambda^c dz d\mu^c}{\int g^c d\mu^c}\\-\int\Lambda_j^c\int y^c\left(1-\tau_\ell^c\right)\left[1+\varepsilon^c-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c dz d\mu^c\\+\sum\int\frac{\Lambda_i^c\Lambda_j^c}{1-t_i^c}\int y^c\left[1+\varepsilon^c-\sigma-\varepsilon^c\frac{z\tau_\ell'}{1-\tau_\ell}\right]\lambda^c dz d\mu^c=0\end{aligned}$$

The above equation leads to the equation in the statement of the proposition. □

A.6 Proof of Proposition 6

Proof. Given the discussion in the text, it is sufficient to show that under absolute advantage, $\overline{\mathcal{W}}_j^c$ is the same for all sectors. From its definition, we have

$$\overline{\mathcal{W}}_j^c=\frac{\int g^c\Lambda_j^c\int y^c\left(1-\tau_\ell^c\right)\lambda^c dz d\mu^c}{\int g^c d\mu^c\int\Lambda_j^c\int y^c\lambda^c dz d\mu^c}.$$

Absolute advantage implies that $\Lambda_j^c(\theta)$ is independent of θ. As a result, we can write the above as

$$\begin{aligned}\overline{\mathcal{W}}_j^c &= \frac{\int g^c \Lambda_j^c \int y^c (1-\tau_\ell^c)\lambda^c dz d\mu^c}{\int g^c d\mu^c \int \Lambda_j^c \int y^c \lambda^c dz d\mu^c} \\ &= \frac{\Lambda_j^c \int g^c \int y^c (1-\tau_\ell^c)\lambda^c dz d\mu^c}{\Lambda_j^c \int g^c d\mu^c \int \int y^c \lambda^c dz d\mu^c} \\ &= \frac{\int g^c \int y^c (1-\tau_\ell^c)\lambda^c dz d\mu^c}{\int g^c d\mu^c \int \int y^c \lambda^c dz d\mu^c}\end{aligned}$$

which is independent of j. This concludes the proof. □

A.7 Proof of Proposition 7

Proof. Consider the optimization (P) and suppose that $\xi^c(\theta)$ is the costate variable associated with the incentive constraint (7). Then optimality requires the following condition to hold

$$\int \psi^c g^c(\theta)\lambda^c\left(z,\theta;\left\{w_j^c\right\}\right)d\mu^c - \gamma^c(z) + (\xi^c)'(z) = 0$$

together with the limit condition

$$\lim_{z\to\infty}\xi^c(z) = 0$$

From section A.5, we know that

$$\nu^c(\mathbf{q})\,\rho\int_\Theta \lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c = \gamma^c(z)$$

We can replace this in the above optimality condition and use integration to arrive at

$$\xi^c(z) = \int_z^\infty \int \left[\psi^c g^c(\theta) - \rho\nu^c(\mathbf{q})\right]\lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c dz$$

Using the definition of f^c, we can write the above as

$$\begin{aligned}\xi^c(z) &= \rho\nu^c(\mathbf{q})\int_z^\infty \int \left[\frac{\psi^c g^c(\theta)}{\rho\nu^c(\mathbf{q})} - 1\right]\lambda^c(z,\theta;\mathbf{w}^c)\,d\mu^c dz \\ &= \rho\nu^c(\mathbf{q})\int_z^\infty \left[\mathcal{W}^c(z) - 1\right] f^c(z)\,dz\end{aligned}$$

where

$$\begin{aligned}
\mathcal{W}^c(z) &= \frac{\int \frac{\psi^c g^c(\theta)}{\rho\nu^c(\mathbf{q})} \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c}{f^c(z)} \\
&= \frac{\int \psi^c g^c(\theta)\, \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c}{\rho\nu^c(\mathbf{q})\, f^c(z)} \\
&= \frac{\int \psi^c g^c(\theta)\, \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c}{\int \psi^c g^c(\theta)\, d\mu^c f^c(z)} \\
&= \frac{\int g^c(\theta)\, \lambda^c(z,\theta;\mathbf{w}^c)\, d\mu^c}{\int g^c(\theta)\, d\mu^c f^c(z)}
\end{aligned}$$

where in the above we have used (18).

Next, consider the optimality condition for $\ell^c(z)$. We have

$$\sum_j \frac{\zeta_j^c}{w_j^c} \int z\lambda_j^c\left(z,\theta;\left\{w_j^c\right\}\right) d\mu^c - \gamma^c(z)\, v'(\ell^c(z)) + \xi^c(z)\, \frac{1}{z} v'(\ell^c(z)) \left(1+\frac{1}{\varepsilon^c}\right) = 0$$

Using the fact that $\gamma^c(z) = \nu^c(\mathbf{q}^c)\, \rho f^c(z)$ and $\zeta_j^c = \rho_i \frac{\partial}{\partial L} G_i^c$ and equation (16), we can write the above as

$$\begin{aligned}
&z\rho \sum_j \int \frac{1}{1-t_j^c} \lambda_j^c\left(z,\theta;\left\{w_j^c\right\}\right) d\mu^c - \nu^c(\mathbf{q}^c)\, \rho v'(\ell^c(z)) \\
&+\rho\nu^c(\mathbf{q}) \int_z^\infty \left[\mathcal{W}^c(z)-1\right] f^c(z)\, dz \frac{1}{z} v'(\ell^c(z)) \left(1+\frac{1}{\varepsilon^c}\right) = 0
\end{aligned}$$

We can divide the above by $\rho\nu^c(\mathbf{q}^c)\,(v^c)'(\ell^c(z))$ and we have

$$\begin{aligned}
&\frac{z}{\nu^c(\mathbf{q}^c)\,(v^c)'(\ell^c(z))} \sum_j \int \frac{1}{1-t_j^c} \lambda_j^c\left(z,\theta;\left\{w_j^c\right\}\right) d\mu^c - 1 \\
&+ \frac{1}{z}\left(1+\frac{1}{\varepsilon^c}\right) \int_z^\infty \left[\mathcal{W}^c(z)-1\right] f^c(z)\, dz = 0
\end{aligned}$$

From the definition of labor wedge $\tau_\ell(z)$, we know that

$$\left[1-\tau_\ell^c(z)\right] \frac{z}{\nu^c(\mathbf{q}^c)} = v'(\ell^c(z))$$

Therefore, the above equation can be written as

$$\frac{1}{1-\tau_\ell^c(z)} \sum_j \frac{f_j^c(z)}{1-t_j^c} - 1 = \frac{1}{z}\left(1+\frac{1}{\varepsilon^c}\right) \int_z^\infty \left[1-\mathcal{W}^c(z)\right] f^c(z)\, dz$$

This implies the equation in the statement of proposition 7. □

A.8 Proof of Proposition 4

Proof. If $p_i^c = \hat{p}_i \left(1 - t_i^{p,c}\right)$ is the after tax price faced by firms in sector i country c, then optimality condition with respect to labor of type θ is given by

$$p_i^c \frac{\partial G_i^c}{\partial L(\theta)} = w_i^c(\theta)$$

where $w_i^c(\theta)$ is the wage of workers of type θ in sector i country c and $\mathbf{w}^c = \left\{w_j^c(\theta)\right\}_{j,\theta}$. As a result, the sectoral choice of workers is given by

$$j^c(\theta, \boldsymbol{\epsilon}) \in \arg\max_j w_j^c(\theta)\, a_j^c(\theta)\, \epsilon_j.$$

In equilibrium, the distribution of labor productivity among workers of type θ is given by

$$\lambda^c(z, \theta; \mathbf{w}^c) = \frac{d}{dz}\left(e^{-\sum_j \left(w_j^c(\theta) a_j^c(\theta)\right)^{\sigma^c} z^{-\sigma^c}}\right)$$

with

$$\lambda_j^c(z, \theta; \mathbf{w}^c) = \frac{\left(w_j^c(\theta)\, a_j^c(\theta)\right)^{\sigma^c}}{\sum_i \left(w_i^c(\theta)\, a_i^c(\theta)\right)^{\sigma^c}} \lambda^c(z, \theta; \mathbf{w}^c) = \Lambda_j^c(\theta)\, \lambda^c(z, \theta; \mathbf{w}^c).$$

Therefore, the optimal taxation problem becomes

$$\max_{\mathbf{q}, \mathbf{w}^c, \mathbf{p}, \{\ell^c(z), u^c(z), I^c(z)\}} \sum_c \psi^c \int_\Theta \int_0^\infty u^c(z)\, \lambda^c(z, \theta; \mathbf{w}^c)\, dz d\mu^c$$

subject to

$$\sum_c \int_\Theta \int_0^\infty x_j^c(\mathbf{q}; I^c(z))\, \lambda^c(z, \theta; \mathbf{w}^c)\, dz d\mu^c + \sum_c \sum_j Q_{ji}^c = \sum_i G_i^c\left(\left\{L_i^c(\theta)\right\}, \left\{Q_{ij}^c\right\}\right)$$

$$V^c(\mathbf{q}; I^c(z)) - v^c(\ell^c(z)) = u^c(z)$$

$$\frac{\ell^c(z)\,(v^c)'(\ell^c(z))}{z} = (u^c)'(z)$$

$$\Lambda_i^c(\theta) \int_0^\infty z \ell^c(z)\, \lambda^c(z, \theta; \mathbf{w}^c)\, dz = L_i^c(\theta)$$

$$p_i^c \frac{\partial G_i^c\left(\left\{L_i^c(\theta)\right\}, \left\{Q_{ij}^c\right\}\right)}{\partial L(\theta)} = w_i^c(\theta)$$

Now, consider a perturbation of Q_{ij}^c by ε. We know that

$$\frac{\partial}{\partial Q_{ij}} \frac{\frac{\partial G_i^c}{\partial L(\theta)}}{\frac{\partial G_i^c}{\partial L(\theta')}} = 0$$

Therefore

$$\frac{\frac{\partial^2 G_i^c}{\partial Q_{ij} \partial L(\theta)}}{\frac{\partial G_i^c}{\partial L(\theta)}} - \frac{\frac{\partial^2 G_i^c}{\partial Q_{ij} \partial L(\theta')}}{\frac{\partial G_i^c}{\partial L(\theta')}} = 0.$$

This implies that if we perturb p_i^c by $p_i^c \frac{\frac{\partial^2 G_i^c}{\partial Q_{ij} \partial L(\theta)}}{\frac{\partial G_i^c}{\partial L(\theta)}} \varepsilon$ for some θ, then $p_i^c \frac{\partial G_i^c}{\partial L(\theta')} = w_i^c(\theta')$ is satisfied for all θ'. Therefore, that the optimality condition with respect to this perturbation is

$$\rho_i \frac{\partial G_i^c}{\partial Q_{ij}^c} = \rho_j$$

where ρ_i is the lagrange multiplier on the resource constraint for good i. The above is the condition that is equivalent to pseudo-efficiency of production. This concludes the proof. $\square$

B Data and Calibration

We calibrate the model in Section 5 by taking a laissez-faire version of this economy (i.e., one without government policies) to the data. We choose some of the parameters of the model similar to those chosen in the literature; these are mostly parameters related to the behavioral responses: elasticities of labor supply (intensive and extensive margin), trade elasticities, etc. The parameters of technology, productivities, as well as production functions are chosen to match the data.[20]

Data. The data we use to calibrate the model come from the following two sources: 1) World Input-Output Database (WIOD), which includes the World Input-Output Tables (WIOT) and Socio Economic Accounts (SEA), [21] and 2) Groningen Growth and Development Centre (GGDC) Productivity Level Database, which includes data on relative prices and labor productivity across countries for 42 major economies and up to 35 industries in 2005. [22]

We use the World Input-Output Tables to construct bilateral trade shares, and intermediate input and labor shares for each sector and country. The data for employment by sector come

[20]We do this mainly because finding comprehensive data on government policies is difficult, and it is in line with the rest of the international trade literature.

[21]http://www.wiod.org/

[22]https://www.rug.nl/ggdc/productivity/pld/

ISIC rev.3 code	Sector Description	Predicted $\frac{\pi_j^{US,China} \text{in year 2011}}{\pi_j^{US,China} \text{in year 2000}}$
15–16	Food, Beverage & Tobacco	2.82
17–19	Textile, Textile Products & Leather Products	3.12
20	Wood & Products of Wood and Cork	5.09
21–22	Pulp, Paper, Printing & Publishing	3.18
23	Coke, Refined Petroleum & Nuclear Fuel	1.66
24	Chemicals & Chemical Products	4.30
25	Rubber & Plastics	3.41
26	Other Non-Metallic Mineral Products	3.03
27–28	Basic & Fabricated Metal	3.48
29	Machinery (nec)	8.32
30–33	Electrical & Optical Equipment	6.49
34–35	Transport Equipment	8.23
36–37	Manufacturing (nec) & Recycling	3.95
	Non-Manufacturing	4.65

Table 1: List of 14 aggregated sectors. The last column shows the predicted changes in U.S. expenditure shares on Chinese goods between 2011 and 2000.

from the Socio Economic Accounts. Finally, we use price data from the GGDC together with data from the WIOT to back out the parameters of preferences and technology.

Sectors and Countries. We calibrate the model to a sample of 10 countries and 14 aggregated sectors. Our sectors consist of 13 manufacturing industries classified according to the International Standard Industrial Classification (ISIC Rev. 3.1), with Textile and Leather products combined. We also include an aggregated non-manufacturing sector. Table 1 lists the 14 sectors we use. We include the following 10 countries in our sample: Australia, Canada, China, Denmark, Finland, France, Germany, Japan, Spain, and the United States.

Parameters Chosen Independently. In line with many estimates of the intensive elasticity of the labor supply, we choose $\varepsilon_c = 0.5$; see Chetty et al. (2011) for an extensive discussion. We choose the dispersion of productivity to be $\nu = 5$, based on estimated trade elasticities in Head and Mayer (2014). The elasticity of substitution in the final goods production function is $\rho = 4$. Moreover, we choose $\sigma^c = 2$ as estimated by Galle et al. (2017). Finally, we assume that the elasticity of substitution across different consumption goods is 1.

Bilateral Trade Shares, Labor Shares, and Intermediate Input Shares. We use the World Input-Output Tables in 2000 to construct bilateral trade flows, $X_j^{c,c'}$; gross output, X_j^c; and bilateral trade shares, $\pi_j^{c,c'}$. These quantities are used in calibrating consumption shares and trade costs, as we describe below.

We also use the WIOD to construct the share of value added in gross output, χ_j^c, and the intermediate input shares, γ_{jk}^c, across countries and sectors using data on value added, gross output, and intermediate consumption.

Final Consumption Shares. Using value-added data from the WIOD together with sectoral employment data in SEA, we compute the value added per worker in each sector and each country. We assume that the national income in each country is equal to the sum of all value added. We also assume the trade balance at the country level.[23] With these two assumptions, we can use the equation market clearing condition for good j in country c to back out the final consumption shares as

$$\alpha_j^c = \frac{X_j^c - \sum_{k=1}^{N_J} \gamma_{k,j}^c \sum_{c'=1}^{C} X_k^{c',c}}{I^c},$$

where I^c is a the sum of all value added in country c.

Sectoral Choice Parameters. Let $\mu^c(\theta)$ be the share of population in country c who are in group/type θ. Since for all countries other than the U.S. we have only one group, $\mu^c(\theta) = 1$ for all $c \neq USA$. As we described in Section 5, for the U.S. we define group/types based on geographic locations. In other words, we use commuting zones (CZs) to define local labor markets. We split each commuting zone into two education groups based on whether workers hold at least an Associate degree or not. This implies that there are a total of 1,444 types (722 CZs $\times$ 2 skills). We use the crosswalk provided by David Dorn to map Census Public Use Micro-data Areas (PUMAs) into commuting zones. [24] Using Census data, we can calculate the distribution of workers across different groups, $\mu^{USA}(\theta)$. Moreover, using Census data, we can calculate the share of workers of each group in each sector $\Lambda_j^{US}(\theta)$, which, in the model, is given by the following formula:

$$\Lambda_j^c(\theta) = \frac{\left(w_j^c a_j^c(\theta)\right)^{\sigma}}{\sum_i \left(w_i^c a_i^c(\theta)\right)^{\sigma}}.$$

In countries other than the U.S. the share of workers in each sector is given by the Socio Economic Accounts (SEA), which is part of the World Input-Output Database (WIOD).

To calibrate the model, we need to back out the wage rate in each sector, w_j^c, and productivity shifter $a_j^c(\theta)$. To do this, we need data on labor income in each sector and, in the case of U.S., each

[23]Note that this calculation assumes that trade balances are 0. An alternative approach is to include the trade deficit and rebate it to workers in a lump-sum fashion, as in Caliendo and Parro (2015). The resulting parameters are very close to what we compute under the balanced trade assumption and, as a result, we have decided to proceed with balanced trade.

[24]See https://www.ddorn.net/data.htm.

group. We get this information from value added data in the WIOD. For the U.S. we have income data for workers in each sector and group $Y_j^{USA,Census}(\theta)$. We split the value added in sector j from SEA across U.S. groups so that the distribution of income of each sector across different groups is consistent with Census data and the aggregate income in each sector is consistent with value-added data in the WIOD. In other words, we use the following as a measure of labor income in sector j and group θ

$$Y_j^{USA}(\theta) = \frac{Y_j^{USA,Census}(\theta)}{\sum_\theta Y_j^{USA,Census}(\theta)} Y_j^{USA,WIOD},$$

where $Y_j^{USA,WIOD}$ is the value added in sector j in the U.S. from the WIOD data.

Using the properties of the Fréchet distribution in the model, we can derive the equation for labor income per worker in country c and sector j as

$$w_j^c L_j^c = (P^c)^{-\varepsilon} \Gamma\left(1 - \frac{1+\varepsilon}{\sigma}\right) \sum_\theta \mu^c(\theta) \Lambda_j^c(\theta) \left(\Phi^c(\theta)\right)^{1+\varepsilon}, \tag{19}$$

where

$$\Phi^c(\theta) = \left(\sum_i \left(w_i^c a_i^c(\theta)\right)^\sigma\right)^{\frac{1}{\sigma}}$$

and

$$P^c = \prod_{j=1}^{N} \left(\frac{P_j^c}{\alpha_j^c}\right)^{\alpha_j^c}.$$

We can calculate the aggregate price P^c for each country using data on relative sectoral prices, P_j^c, in GGDC's Productivity Level Database. Finally, we make the following normalization with respect to the productivity shifter $a_j^c(\theta)$ in each country

$$\sum_\theta \mu^c(\theta) a_j^c(\theta) = 1 \text{ for all } j \text{ and } c.$$

Using normalization, we can solve the equation (19) for w_j^c and $a_j^c(\theta)$ noting that the left hand side of the equation is given in the data (from the WIOD and Census).

Trade Costs and Efficiency Parameter. The last set of parameters to be calculated are trade costs, $d_j^{c,c'}$, and sectoral productivities, A_j^c.[25] To calculate these values, we use the price data and computed wages to calculate the unit cost ψ_j^c (assuming value-added taxes are zero in the benchmark). The iceberg trade costs can be calculated using the equation for expenditure shares.

[25] Our calibration procedure closely follows that of Lewis et al. (2018).

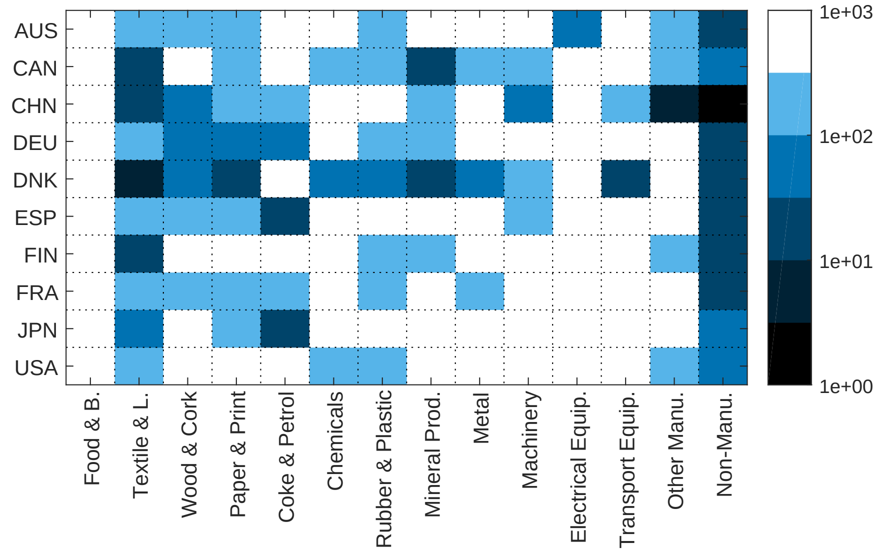


Figure 16: Log of sector TFPs for all sectors and countries of the sample. Lighter shades represent higher TFP.

In particular, we have

$$\pi_j^{c,c'} = \frac{A_j^{c'}\left(\tau_j^{c,c'}\psi_j^{c'}\right)^{-\nu}}{\left(P_j^c\right)^{-\nu}\Gamma\left(1+\frac{1-\rho}{\nu}\right)^{\frac{\nu}{1-\rho}}},$$

where in the above we have used the definition of P_j^c in 14. Therefore,

$$\frac{\pi_j^{c,c'}}{\pi_j^{c',c'}} = \left(d_j^{c,c'}\right)^{-\nu}\left(\frac{P_j^{c'}}{P_j^c}\right)^{-\nu},$$

since $d_j^{c,c} = 1$. Given our assumed value for ν, the equation above can be solved to back out the trade cost $d_j^{c,c'}$.

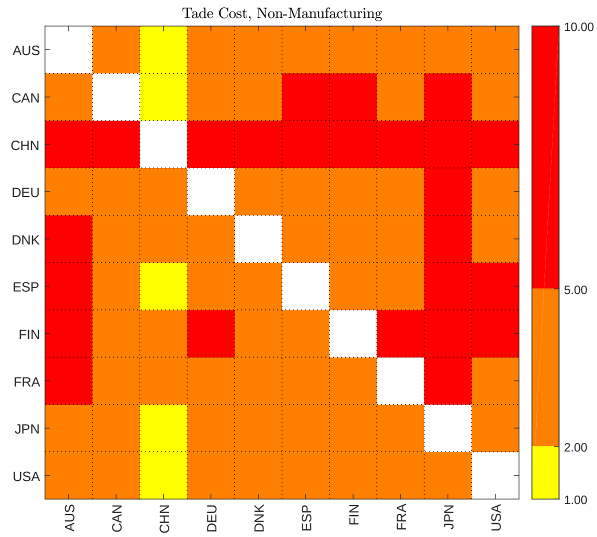
Finally, note that we can combine the equations above and (14) to arrive at the following equation:

$$\pi_j^{c,c}\left(P_j^c\right)^{-\nu} = A_j^c\left(\psi_j^c\right)^{-\nu}\Gamma\left(1+\frac{1-\rho}{\nu}\right)^{\frac{-\nu}{1-\rho}}.$$

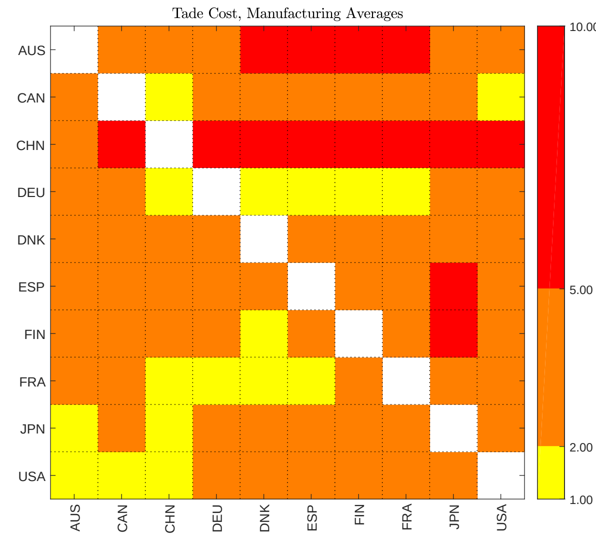
We can then use this equation to solve for sectoral TFP parameters, A_j^c . This concludes the calibration.

Figure 16 shows the heat map of the logarithm of sectoral TFPs in each county. Lighter shades represent higher TFP.

Figure 17 shows the heat map of trade costs in the non-manufacturing sector and average over all manufacturing sectors. The rows (y-axis) represent origin countries, and the columns (x-axis)



(a) Agriculture, forestry & fishing



(b) Manufacturing

Figure 17: Iceberg trade cost. Left panel shows trade costs for non manufacturing sector. Right panel shows the average over all manufacturing sectors. Row (y-axis) countries represent origin. Column (x-axis) represent destination.

represent destinations. Here, white cells represent no trade costs or very small trade costs, and darker cells represent higher trade costs. Note that the U.S. and many other advanced countries face lower trade costs when exporting goods in all sectors. This is consistent with findings in Waugh (2010), among others.