NETWORKS, BARRIERS, AND TRADE

David Baqaee
Emmanuel Farhi

ABSTRACT

We study a non-parametric class of neoclassical trade models with global production networks. We characterize their properties in terms of sufficient statistics useful for growth and welfare accounting as well as for counterfactuals. We establish a formal duality between open and closed economies and use it to analytically quantify the gains from trade. Accounting for nonlinear (non-Cobb- Douglas) production networks with realistic complementarities in production significantly raises the gains from trade relative to estimates in the literature. We use our general comparative statics results to show how models that abstract away from intermediates, no matter how well calibrated, are incapable of simultaneously predicting the costs of tariff and non-tariff barriers to trade. Given trade volumes and elasticities, accounting for intermediates doubles the losses from tariffs. Better quantitative accuracy demands the use of more complicated, oftentimes computational, models. This paper seeks to help bridge the gap between computation and theory.

David Baqaee
UCLA
315 Portola Plaza
Los Angeles, CA
baqaee@econ.ucla.edu

Emmanuel Farhi
Harvard University
Department of Economics
Littauer Center
Cambridge, MA 02138
and NBER
emmanuel.farhi@gmail.com

1 Introduction

Trade economists increasingly recognize the importance of using large-scale computational general equilibrium models for studying trade policy questions. One of the major downsides of relying on purely computational methods is their opacity: computational models can be a black box, and it is sometimes hard to know which forces in the model drive specific results. On the other hand, simple stylized models, while transparent and parsimonious, can lead to unreliable quantitative predictions when compared to the large-scale models.

This paper attempts to provide a theoretical map of territory usually explored by machines. It studies output and welfare in open economies with disaggregated and interconnected production structures and heterogeneous consumers. We address two types of questions: (i) how to measure and decompose the sources of output and welfare changes, and (ii) how to predict the responses of output, welfare, as well as disaggregated prices and quantities, to changes in trade costs or tariffs. Our analysis is non-parametric and quite general, which helps us to isolate the common forces and sufficient statistics necessary to answer these questions without committing to a specific parametric set up.

We show how accounting for the details of the production structure can theoretically and quantitatively change answers to a broad range of questions in open-economy settings. Simple stylized models, no matter how deftly calibrated, can get both the magnitude and even the direction of effects wrong.

In analyzing the structure of open-economy general equilibrium models, we emphasize their similarities and differences to the closed-economy models used to study growth and fluctuations. To fix ideas, consider the following fundamental theorem of closed economies. For a perfectly-competitive economy with a representative household and inelastically supplied factors,

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log A_i} = \frac{\mathrm{d}\log Y}{\mathrm{d}\log A_i} = \frac{sales_i}{GDP}, \tag{1}$$

where W is real income or welfare (measured as an equivalent variation), Y is real output or GDP, and A_i is a Hicks-neutral shock to some producer i.[1] Equation (1), also known as Hulten's Theorem, shows that the sales share of producer i is a sufficient statistic for understanding the impact of a shock on aggregate welfare, aggregate income, and aggregate output to a first order. Specifically, Hulten's theorem implies that, to a first order, any

---

[1]Equation (1) is fundamental in the sense that it is a consequence of the first welfare theorem. Although versions of this result existed for a long time, at least since Domar (1961), the modern treatment is due to Hulten (1978).

disaggregated information beyond the sales share (the input-output network, the number of factors, the degrees of returns to scale, and the elasticities of substitution) is macroeconomically irrelevant.

In this paper, we examine the extent to which the logic of (1) can be transported into international economics. We provide the open-economy analogues of equation (1), and show that although versions of Hulten's theorem continue to hold in open-economies, the sales shares are no longer such universal sufficient statistics. Ultimately, there are two main barriers to blindly applying Hulten's theorem in an open-economy: first, in an open-economy, output and welfare are no longer the same since welfare depends on terms-of-trade but output does not (see e.g. Burstein and Cravino, 2015); second, much of trade policy concerns the effects of tariffs, which knocks out the foundation of marginal cost pricing and Pareto efficiency that Hulten's Theorem is built on. Our generalizations make clear precisely the conditions under which a naive-application of (1) to an open-economy is valid. Even when not directly applicable, it proves helpful to think in terms of (1), and deviations from it.

Notwithstanding the differences between open and closed economies, we also prove that, under some conditions, there exists a useful isomorphism between the two types of models. In particular, for any open-economy with nested-CES import demand there exists a companion (dual) closed economy, and the welfare effects of iceberg shocks in the open-economy are equal to the output effects of productivity shocks in the closed economy. This means that we can use results from the closed-economy literature, principally Hulten (1978) and Baqaee and Farhi (2017a), to characterize the effects of iceberg shocks on welfare up to the second-order. Our formulas provide a generalization of some of the influential insights of Arkolakis et al. (2012) to environments with disaggregated, non-loglinear (non-Cobb-Douglas) input-output connections. Compared to the loglinear (Cobb-Douglas) production networks common in the literature (e.g. Costinot and Rodriguez-Clare, 2014; Caliendo and Parro, 2015), we find that accounting for nonlinear production networks significantly raises the gains from trade. Accounting for nonlinear input-output networks is as, or more important, as accounting for intermediates in the first place. For example, for the US, the gains from trade increase from 4.5% to 9% once we account for intermediates with a loglinear network, but they increase further to 13% once we account for realistic complementarities in production. The numbers are even more dramatic for more open economies, for example, the gains from trade for Mexico go from 11% in the model without intermediates, to 16% in the model with a loglinear network, to 44.5% in the model with a non-loglinear network.

For most of the paper, we restrict ourselves to efficient economies, but extending the

results to allow for arbitrary distorting wedges (e.g tariffs or markups) is straightforward. To our knowledge, this is the first paper in the literature to derive comparative statics with respect to tariffs (in terms of model primitives) in a general production environment with intermediate goods. We show that, in general, the output losses to the world as a whole, and to the output of each country, from the imposition of tariffs or other distortions can be computed by an appropriate summing up of Harberger triangles, even in the absence of implausible compensating transfers. We provide explicit formulas for what these Harberger triangles are equal to in terms of microeconomic primitives. We explain how to adjust these formulas to obtain welfare losses. We show that the existence of global value chains dramatically increases the costs of protectionism by inflating both the area of each triangle *and* the weight used to aggregate the triangles. We show that simple (non-input-output) models, regardless of how they are calibrated, get either the area of the triangles or their weight wrong.

Intuitively, the weight on each triangle is just the sales share of the taxed good. Since input-output connections inflate sales relative to value-added, that means accounting for intermediates can inflate the weight each Harberger triangle receives. More subtly, the area of each triangle is also increased in the presence of intermediates. There are two reasons for this: first, global value chains mean that tariffs are compounded each time an unfinished good crosses the border, à la Yi (2003); second, in the presence of intermediates, the quantity of traded goods is more elastic with respect to tariffs, since, holding fixed the volume of trade, trade is a smaller portion of each individual agent's basket. Both of these effects combine, in roughly equal magnitudes, to amplify the cost of tariffs to the world economy. As an example, we find that a worldwide increase in import tariffs from zero to ten percent reduces world output by -0.43%. If we ignore input-output connections, this number is halved.

The outline of the paper is as follows. In Section 2, we set up the model and define the objects of interest. In Section 3, we derive some growth-accounting results useful for measurement. In Section 4, we establish the dual relationship between closed and open economies which can be used to generalize some of the results in Arkolakis et al. (2012) and Costinot and Rodriguez-Clare (2014). In Section 5, we derive comparative statics in terms of microeconomic primitives, useful for prediction. In Section 6, we extend our analysis to allow for distortions like tariffs, and we show that analytically, the costs of tariffs are very different to those of iceberg shocks. In Section 7, we use a quantitative model to study the magnitude of the forces we identify. All the proofs are in Appendix L.

Related Literature

At a high-level, this paper is related to the classic papers of Hulten (1978), Harberger (1964), and Jones (1965). We extend Hulten (1978) and prove growth-accounting formulas for open-economies; we extend Harberger (1964) and show that deadweight-loss triangles can be used to measure productivity and welfare losses from tariffs in general equilibrium, even in the absence of compensating transfers; we extend the hat-algebra of Jones (1965) beyond the $2 \times 2 \times 2$ no input-output economies he considered.

More broadly, our paper is related to three literatures: the literature on the gains from trade, the literature on production networks, and the literature on growth accounting. We discuss each literature in turn starting with the one on the gains (or losses) from trade. As far as we are aware, this is the first paper to characterize the comparative static response of income and output to changes in iceberg costs and tariffs non-parametrically in a model with a rich input-output structure. In particular, our results generalize some of the results in Arkolakis et al. (2012) and Costinot and Rodriguez-Clare (2014) to environments with non-linear input-output connections. Our framework generalizes the input-output models emphasized in Caliendo and Parro (2015), Caliendo et al. (2017), Morrow and Trefler (2017), Fally and Sayre (2018), and Bernard et al. (2019). Our results about the effects of trade in distorted economies also relates to Epifani and Gancia (2011), Arkolakis et al. (2015), Berthou et al. (2018), Bai et al. (2018). Our results also relate to work with non-parametric or semi-parametric models of trade like Adao et al. (2017) and Lind and Ramondo (2018) (though our analysis does not rely on the invertibility, or stability, of factor demand systems), as well as Allen et al. (2014), (although we do not impose a gravity equation). Finally, our characterization of how factor shares and prices respond to shocks is related to an incredibly deep literature, for example, Trefler and Zhu (2010), Elsby et al. (2013), Davis and Weinstein (2008), Feenstra and Sasahara (2017), Burstein and Vogel (2017), Artuç et al. (2010), Dix-Carneiro (2014), Galle et al. (2017), among others.

The literature on production networks has primarily been concerned with the propagation of shocks in closed economies, typically assuming a representative agent. For instance, Long and Plosser (1983), Acemoglu et al. (2012), Atalay (2017), Carvalho et al. (2016), Baqaee and Farhi (2017a,b), and Baqaee (2018), among many others. A recent focus of the literature, particularly in the context of open economies, has been to model the formation of links, for example Chaney (2014), Lim (2017), Tintelnot et al. (2018), and Kikkawa et al. (2018). Our approach, which builds on the results in Baqaee and Farhi (2017a,b), is different: rather than modelling the formation of links as a binary decision, we use a Walrasian environment where the presence and strength of links are endogenously determined by cost minimization and input-substitution subject to some produc-

tion technology.

Finally, our growth accounting results are related to closed-economy results like Solow (1957), Hulten (1978), as well as to the literature extending growth-accounting to open economies, including Kehoe and Ruhl (2008) and Burstein and Cravino (2015). Perhaps closest to us are Diewert and Morrison (1985) and Kohli (2004) who introduce output indices which account for terms-of-trade changes. Our real income and welfare-accounting measures share their goal, though our decomposition into pure productivity changes and reallocation effects is different. In explicitly accounting for the existence of intermediate inputs, our approach also speaks to how one can circumvent the double-counting problem and spill-overs arising from differences in gross and value-added trade, issues studied by Johnson and Noguera (2012) and Koopman et al. (2014). Relative to these other papers, our approach has the added bonus of easily being able to handle inefficiencies and wedges.

Our approach is general, and relies heavily on the theory of duality, along the same lines as Dixit and Norman (1980). We differ from the classic analysis, however, in that, in extending Hulten's theorem to open economies, we state our comparative static results in terms of readily observable sufficient statistics: expenditure shares, changes in expenditure shares, properties of the input-output network, and elasticities. Our approach relies heavily on the notion of the allocation matrix, which helps to give a physical interpretation to the theorems, and is also very convenient for extending the results to inefficient economies. In inefficient economies, the abstract approach that relies on macro-level envelope conditions, taken by Dixit and Norman (1980), runs into problems. However, our results, and their interpretation in terms of the allocation matrix, can readily be extended to inefficient economies.

2   Framework

In this section, we do the spadework of setting up the model and defining the key statistics of interest. We assume that there are no distortions. In Section 6, we extend our results to environments with distortions (e.g. markups, tariffs, taxes).[2]

---

[2]Distortions can be represented as wedges (implicit or explicit taxes). Tariffs, markups, and financial frictions are wedges, but iceberg trade costs are not.

2.1 Model

There is a set of countries C with representative households, a set of producers N producing different goods, and a set of factors F. Each producer and each factor is assigned to be within the borders of one of the countries in C. The sets of producers and factors inside country c are N_c and F_c. The set F_c of factors physically located in country c may be owned by any household, and not necessarily the households in country c. We assume a representative agent for each country in order to not clutter the exposition.[3]

Factors

In each country, the representative household c is endowed with some share Φ_{cf} of the supply L_f of each factor f. A factor is simply a non-produced good, and we take the quantities and ownership structure of factors as exogenously given.[4]

Households

The representative household in country c maximizes a homogenous-of-degree-one demand aggregator[5]

$$W_c = \mathcal{W}_c(\{c_{ci}\}_{i \in N}),$$

subject to the budget constraint

$$\sum_{i \in N} p_i c_{ci} = \sum_{f \in F} \Phi_{cf} w_f L_f + T_c,$$

where c_{ci} is the quantity of the good produced by producer i and consumed by household c, p_i is the price of good i, w_f is the wage of factor f, and T_c is an exogenous lump-sum transfer. The lump-sum transfer allows for trade imbalances as in Dekle et al. (2008).

---

[3]Appendix H generalizes the results to cover situations with heterogenous agents within each country.

[4]In Appendix C, we discuss how to endogenize factor supply by using a model à la Roy (1951) and discuss the connection of our results with those in Galle et al. (2017).

[5]In mapping our model to data, we interpret domestic "households" as any agent which consumes resources without producing resources to be used by other agents. Specifically, this means that we include domestic investment and government expenditures in our definition of "households".

Producers

Each producer i in country c produces a different good using a constant-returns-to-scale production function with the associated production function

$$y_i = A_i F_i \left(\{x_{ik}\}_{k \in N}, \{l_{if}\}_{f \in F_c} \right),$$

where y_i is the total quantity of good i produced, x_{ik} is intermediate inputs from k, l_{if} is factor inputs from f, and A_i is an exogenous Hicks-neutral productivity shifter.

Generality

This set up is more general than it might appear at first glance. The assumption that production has constant returns to scale is without loss of generality. As pointed out by McKenzie (1959), neoclassical production functions are constant-returns-to-scale without loss of generality, since any decreasing-returns production function can always be written as a constant returns production function by adding quasi-fixed factors.[6]

The assumption that each producer produces only one output good is without loss of generality. One can always represent a multi-output production function as a single output production function by letting all but one of the outputs enter as negative inputs. Joint production is therefore allowed by the model.

The assumption that productivity shifters are Hicks neutral is also without loss of generality. For example, an input-augmenting technical change for producer i's use of input k can be captured by introducing a fictitious producer buying from k and selling to i and hitting this fictitious producer with a Hicks-neutral productivity change.

Finally, the assumption that there are no shocks to the composition of final demand is without loss of generality, since such shocks can be represented via relabeling as combinations of positive and negative productivity shocks.

Iceberg Trade Costs

We capture changes in iceberg trade costs as Hicks-neutral productivity changes to specialized importers or exporters whose production functions represent the trading technology. The decision of where the trading technology should be located is ambiguous since it generates no value added. It is possible to place them in the exporting country

---

[6]Increasing returns can also in principle be accommodated, but only to some limited extent, by allowing these quasi-fixed factors to be local "bads", i.e. to receive negative payments over some range. However, care must be taken because increasing returns introduce non-convexities in the cost minimization over variable inputs, and our formulas only apply when variable-input demand changes smoothly.

or in the importing country, and this would make no difference in terms of the welfare of agents or the allocation of resources. We do not need to take a precise stand at this stage, but we note that this will matter for our conclusions regarding real country GDP changes (as pointed out by Burstein and Cravino, 2015).

Equilibrium

Given productivities A_i and a vector of transfers satisfying $\sum_{c \in C} T_c = 0$, a general equilibrium is a set of prices p_i, intermediate input choices x_{ij}, factor input choices l_{if}, outputs y_i, and consumption choices c_{ci}, such that: (i) each producer chooses inputs to minimize costs taking prices as given; (ii) each household maximizes utility subject to its budget constraint taking prices as given; and, (iii) the markets for all goods and factors clear so that $y_i = \sum_{c \in C} c_{ci} + \sum_{j \in N} x_{ji}$ for all $i \in N$ and $L_f = \sum_{j \in N} l_{jf}$ for all $f \in F$.

2.2 Definitions

In this subsection, we define the statistics of interest. Although these definitions are standard to national income accountants, and the distinctions we stress may seem tedious, it turns out that they make all the difference for the economics of the model.

Nominal Output and Nominal Expenditure

Nominal output or Gross Domestic Product (GDP) for country c is the total final value of the goods produced in the country. It coincides with the total income earned by the factors located in the country:

$$GDP_c = \sum_{i \in N} p_i q_{ci} = \sum_{f \in F_c} w_f L_f,$$

where $q_{ci} = y_i 1_{\{i \in N_c\}} - \sum_{j \in N_c} x_{ji}$ is the net quantity of good $i \in N$ in the GDP of country c, which can be positive or negative.

Nominal Gross National Expenditure (GNE) for country c, also known as domestic absorption, is the total final expenditures of the residents of the country. In our model, it coincides with nominal Gross National Income (GNI) which is the total income earned by the factors owned by its residents and adjusted for international transfers:

$$GNE_c = \sum_{i \in N} p_i c_{ci} = \sum_{f \in F} \Phi_{cf} w_f L_f + T_c.$$

Nominal output or GDP and nominal GNE are *not* the same at the country level in general because they account for the value created by different sets of factors: the factors in the country vs. the factors owned by the residents of the country.

Of course, these differences vanish at the world level:

$$GDP = GNE = \sum_{f \in F} w_f L_f = \sum_{i \in N} p_i q_i = \sum_{i \in N} p_i c_i,$$

where $c_i = q_i$ with $c_i = \sum_{c \in C} c_{ci}$, $q_i = \sum_{c \in C} q_{ci}$, $T = 0$ with $T = \sum_{c \in C} T_c$.

We let world GDP be the numeraire, so that $GDP = GNE = 1$. All prices and transfers are expressed in units of this numeraire.

Real Output and Real Expenditure

We now define changes in real output and real expenditures at different levels of aggregation. We use Divisia indices throughout to separate quantity and price changes, and rely on their convenient aggregation properties.

The change in real output (real GDP) of country c and the corresponding deflator are

$$\mathrm{d} \log Y_c = \sum_{i \in N} \chi_i^{Y_c} \mathrm{d} \log q_{ci}, \quad \mathrm{d} \log P_{Y_c} = \sum_{i \in N} \chi_i^{Y_c} \mathrm{d} \log p_i,$$

where $\chi_i^{Y_c} = p_i q_{ci} / GDP_c$.[7],[8]

The change in real expenditure or welfare (real GNE) of country c and the corresponding deflator are

$$\mathrm{d} \log W_c = \sum_{i \in N} \chi_i^{W_c} \mathrm{d} \log c_{ci}, \quad \mathrm{d} \log P_{W_c} = \sum_{i \in N} \chi_i^{W_c} \mathrm{d} \log p_i, \tag{2}$$

where $\chi_i^{W_c} = p_i c_{ci} / GNE_c$. The fact that (2) measures the welfare of country c is a consequence of Shephard's lemma.

Changes in real output and in real expenditure are *not* the same at the country level in general. The difference comes from two sets of reasons. First, with border cross-border

---

[7]We slightly abuse notation since $q_{ci} \leq 0$ for $i \notin N_c$, in which case we define $\mathrm{d} \log q_{ci} = \mathrm{d} q_{ci} / q_{ci}$.

[8]Note that country real output is only defined in changes, and these changes cannot be integrated to recover a real GDP function. This means that the any discrete change in real output depends on the path of the change. The precise way to proceed is to index the economy by a continuous index (say time t), which indexes all the relevant shifters and all the equilibrium variables. We can then compute changes in real output between the initial period ($t = 0$) and some final period ($t = \tau$) as the integral of the infinitesimal real output changes along the resulting path. The differential change stated in the theorem is the real output change which obtains in the limit of small time intervals $\tau \to 0$: it is independent of the particular path of integration. The same goes at the world level for real output and real expenditure.

factor holdings and international transfers, changes in nominal output and in nominal expenditure are not the same in general. Second, changes in the price deflators for GDP and welfare are not the same in general.

Of course, these differences vanish at the world level so that $d \log Y = d \log W$ and $d \log P_Y = d \log P_W = d \log P$, where

$$d \log Y = \sum_{i \in N} \chi_i^Y d \log q_i, \quad d \log P_Y = \sum_{i \in N} \chi_i^Y d \log p_i,$$

$$d \log W = \sum_{i \in N} \chi_i^W d \log c_i, \quad d \log P_W = \sum_{i \in N} \chi_i^W d \log p_i,$$

with $\chi_i^Y = p_i q_i / GDP$ and $\chi_i^W = p_i c_i / GNE$.[9] Conveniently, changes in country real GDP and real GNE aggregate up to their world counterparts.[10]

Finally, the infinitesimal changes that we have defined for real output and real expenditure or welfare can be integrated or *chained* into discrete changes by updating the corresponding shares along the integration path. We denote the corresponding discrete changes by $\Delta \log Y$, $\Delta \log Y_c$, $\Delta \log W$, and $\Delta \log W_c$. In the case of GDP, this is how these objects are typically measured in the data, and in the case of welfare, this coincides with the way the welfare of each agent c changes in the model.

Input-Output Concepts

We define the Heterogenous-Agent Input-Output (HAIO) matrix to be the $(C + N + F) \times (C + N + F)$ matrix Ω whose ijth element is equal to i's expenditures on inputs from j as a share of its total revenues/income

$$\Omega_{ij} = \frac{p_j x_{ij}}{p_i y_i}.$$

The HAIO matrix Ω includes the factors of production and the households, where factors consume no resources (zero rows), while households produce no resources (zero columns). The Leontief inverse matrix is

$$\Psi = (I - \Omega)^{-1} = I + \Omega + \Omega^2 + \dots.$$

---

[9]Though, we must tread carefully since the change in real expenditure for the world, unlike the one for each country, is no longer a legitimate global measure of welfare, in the sense that it cannot be integrated to recover a social welfare function. However, there does exist a welfare function that, to a first order, coincides with changes in real world GNE/GDP. We discuss this issue in more detail in Section 6.

[10]Namely, $d \log Y = \sum_{c \in C} \chi_c^Y d \log Y_c$ and $d \log W = \sum_{c \in C} \chi_c^W d \log W_c$. This makes use of the following definitions $\chi_c^Y = GDP_c / GDP$, $\chi_c^W = GNE_c / GNE$.

The input-output matrix Ω records the *direct* exposures of one agent or producer to another, whereas the Leontief inverse matrix Ψ records instead the *direct and indirect* exposures through the production network.

It will sometimes be convenient to treat goods and factors together and index them by $k \in N + F$ where we use the plus symbol to denote the union of these two sets. To this effect, we must slightly extend our definitions. We also write interchangeably y_k and p_k for L_k and w_k when $k \in F$. To capture the fact that the household endowment of the goods are zero, we define $\Phi_{ck} = 0$ for $(c, k) \in (C, N)$.

We define the *exposures* of real expenditure (welfare) and real output to each good and each factor. The exposures of country c's real expenditure (welfare) and real output to a good or factor k are

$$\lambda_k^{W_c} = \sum_{i \in N} \chi_i^{W_c} \Psi_{ik}, \quad \lambda_k^{Y_c} = \sum_{i \in N} \chi_i^{Y_c} \Psi_{ik},$$

where recall that $\chi_i^{W_c} = p_i c_{ci}/GNE_c$ and $\chi_i^{Y_c} = p_i q_{ci}/GDP_c$. The exposures of world real expenditure or welfare and real output to a good or factor k are

$$\lambda_k^{W} = \sum_{i \in N} \chi_i^{W} \Psi_{ik}, \quad \lambda_k^{Y} = \sum_{i \in N} \chi_i^{Y} \Psi_{ik},$$

where $\chi_i^{W} = p_i c_i/GNE$ and $\chi_i^{Y} = p_i q_i/GDP$.

Exposures of real output to good or factor k at the country and world levels have a direct connection to the sales of the producer:

$$\lambda_k^{Y_c} = 1_{\{k \in N_c + F_c\}} \frac{p_k y_k}{GDP_c}, \quad \lambda_k^{Y} = \frac{p_k y_k}{GDP}.$$

Hence, for example, λ_k^{Y} is just the sales share (or *Domar weight*) of k in world output $\lambda_k = p_k y_k/GDP$. Similarly $\lambda_k^{Y_c} = 1_{\{k \in N_c + F_c\}}(GDP/GDP_c)\lambda_k$ is the *local Domar weight* of k in country c.

We also define the following factor *income shares* as the shares in income. The share of a factor f in the income of country c and of the world are given by

$$\Lambda_f^c = \frac{\Phi_{cf} w_f L_f}{GNE_c}, \quad \Lambda_f = \frac{w_f L_f}{GNE},$$

where, from now on, we sometimes denote exposures to factors or factor shares with capital Λ to distinguish them from sales shares and exposures to non-factor producers λ. In other words, when $f \in F$, we write $\Lambda_f^{Y_c} = \lambda_f^{Y_c}$ and $\Lambda_f^{W_c} = \lambda_f^{W_c}$.

In general, the exposures of welfare and real output to a good or factor k are *not* the

same at the level of a country. Intuitively, $\lambda_k^{W_c}$ measures household c's total (direct and indirect) reliance on good k in its consumption basket, where $\lambda_k^{Y_c}$ is simply the sales share of k in the GDP of country c. Similarly, when applied to a factor f, these exposures are *not* the same as the income share of that factor at the level of a country. These differences disappear at the world level so that $\lambda_i^Y = \lambda_i^W = \lambda_i = p_i y_i / GDP$ for a good $i \in N$ and $\Lambda_f^Y = \Lambda_f^W = \Lambda_f = w_f L_f / GDP$ for a factor $f \in F$.

3 Comparative Statics: Ex-Post Sufficient Statistics

In this section, we characterize the response of real output and welfare to shocks at the country and world levels. Since iceberg trade costs can be represented as productivity shocks, these characterizations extend to iceberg trade shocks.

We introduce the concept of the allocation matrix, which helps to give a physical interpretation to the theorems, and which is also very convenient for extending the results to inefficient economies. Following Baqaee and Farhi (2017b), define the $(C + N + F) \times (C + N + F)$ *allocation matrix* $\mathcal{X}$ as follows: $\mathcal{X}_{ij} = x_{ij}/y_j$ is the share of the quantity y_j of good j used by some agent i, where the indices i and j the households, factors, and producers. Every feasible allocation is defined by a feasible allocation matrix $\mathcal{X}$, a vector of productivities A, and a vector of factor supplies L. In particular, the equilibrium allocation gives rise to an allocation matrix $\mathcal{X}(A, L, T)$ which, together with A, and L, completely describes the equilibrium.[11]

3.1 Output-Accounting

We start with our output-accounting result: the response of real output (real GDP) to shocks. We state the result at the level of a country c and explain how to translate it to the level of the world. Let $\mathcal{Y}_c(A, L, \mathcal{X}(A, L, T))$ be the value of the GDP of country c using prices in the initial equilibrium. Differentiating yields a decomposition into two components: the direct or "pure" effect of changes in technology $d \log A$ and $d \log L$, holding the distribution of resources $\mathcal{X}$ constant; and the indirect effects arising from the equilibrium changes in the allocation of resources $d \mathcal{X}$. In other words, this decomposition breaks down changes in real GDP into: "pure" technology effects capturing changes from increased production of each good, holding fixed the allocation of resources; and reallocation effects capturing changes in real GDP from changes in the distribution of resources.

---

[11]Since there may be multiplicity of equilibria, technically, the competitive equilibrium gives a correspondence from A to $\mathcal{X}$. In this case, we restrict attention to perturbations of isolated equilibria.

The following theorem characterizes this decomposition.

Theorem 1 (Output-Accounting). *The change in real output (real GDP) of country c to productivity shocks, factor supply shocks, and transfer shocks, can be written as*

$$\mathrm{d}\log Y_c = \underbrace{\frac{\partial \log \mathcal{Y}_c}{\partial \log L}\,\mathrm{d}\log L + \frac{\partial \log \mathcal{Y}_c}{\partial \log A}\,\mathrm{d}\log A}_{\Delta Technology} + \underbrace{\frac{\partial \log \mathcal{Y}_c}{\partial \mathcal{X}}\,\mathrm{d}\,\mathcal{X}}_{\Delta Reallocation},$$

where

$$\frac{\partial \log \mathcal{Y}_c}{\partial \log L}\,\mathrm{d}\log L + \frac{\partial \log \mathcal{Y}_c}{\partial \log A}\,\mathrm{d}\log A = \sum_{i\in N}\lambda_i^{Y_c}\,\mathrm{d}\log A_i + \sum_{f\in F}\Lambda_f^{Y_c}\,\mathrm{d}\log L_f,$$

and

$$\frac{\partial \log \mathcal{Y}_c}{\partial \mathcal{X}}\,\mathrm{d}\,\mathcal{X} = 0.$$

The change $\mathrm{d}\log Y$ *of world real output (GDP) can be obtained by simply suppressing the country index c.*

Theorem 1 is an adaptation of Hulten's theorem to open economies. It implies that to a first order, a unit productivity shock to i moves real output in a country c by an amount equal to producer i's local Domar weight $\lambda_i^{Y_c}$. A counterintuitive implication of this equation is that to a first order, productivity shocks to foreign producers have *no* effect on domestic real output.[12] Since discrete changes in real output are obtained by chained integration of infinitesimal changes, the same counterintuitive implication actually holds globally.[13]

As emphasized by Burstein and Cravino (2015), productivity-accounting à la Hulten (1978) is the same in an open economy as it is in a closed economy. Country c's aggregate productivity change can be measured by its Solow residual and is equal to the local Domar-weighted sum of productivity shocks of domestic producers:

$$\mathrm{d}\log Y_c - \sum_{f\in F}\Lambda_f^{Y_c}\,\mathrm{d}\log L_f = \sum_{i\in N}\lambda_i^{Y_c}\,\mathrm{d}\log A_i.$$

Since shocks to iceberg costs are just shocks to the productivities of trading technolo-

---

[12]There are two important caveats to this statement: (i) shocks to foreign producers may change domestic local Domar weights, and thereby change the way local shocks affect the domestic economy (a nonlinear interaction), (ii) it is ceases to be true when the economy is no longer efficient. We shall discuss both of these issues at length in Sections 4 and 6.

[13]The same reasoning applies to factor supply shocks. Transfer shocks, no matter where they occur, have no effect on domestic real output to the first order and globally (holding fixed factor quantities).

gies, iceberg trade shocks outside of the borders of country c have no effect on its real output or on its real productivity. In a small-open economy, with exogenous world prices, shocks to the terms of trade (the relative price of exports and imports) can also be modeled as shocks to the productivity of a trading technology. Folk wisdom and naive intuition suggest that shocks to the terms of trade should have the same effect as negative domestic productivity shocks. Our result reinforces the observation by Kehoe and Ruhl (2008) that this intuition is invalid. Holding fixed factor quantities, real output (real GDP) and aggregate productivity only respond to shocks inside a country's borders.

The fact that reallocations have no effect on real GDP $\partial \log \mathcal{Y}_c / \partial \log \mathcal{X} \, d \, \mathcal{X} = 0$ is a consequence of the first-welfare theorem, and fails whenever the initial equilibrium is inefficient (see Appendix F).

3.2 Welfare-Accounting

Theorem 1 shows that a straightforward extension of Hulten's theorem holds in open economies for changes in real output. However, this is no longer true for changes in welfare (or real expenditure). Let $\mathcal{W}_c(A, L, \mathcal{X}(A, L, T))$ denote the equilibrium welfare of household c.

Theorem 2 (Welfare-Accounting, Reallocation). *The change in real expenditure or welfare of country c in response to productivity shocks, factor supply shocks, and transfer shocks can be written as:*

$$
d \log W_c = \underbrace{\frac{\partial \log \mathcal{W}_c}{\partial \log L} \, d \log L + \frac{\partial \log \mathcal{W}_c}{\partial \log A} \, d \log A}_{\Delta \textit{Technology}} + \underbrace{\frac{\partial \log \mathcal{W}_c}{\partial \mathcal{X}} \, d \, \mathcal{X}}_{\Delta \textit{Reallocation}},
$$

where the "pure" technology effects are given by

$$
\frac{\partial \log \mathcal{W}_c}{\partial \log L} \, d \log L + \frac{\partial \log \mathcal{W}_c}{\partial \log A} \, d \log A = \sum_{f \in F} \Lambda_f^{W_c} \, d \log L_f + \sum_{i \in N} \lambda_i^{W_c} \, d \log A_i,
$$

and the reallocation effects are given by

$$
\frac{\partial \log \mathcal{W}_c}{\partial \mathcal{X}} \, d \, \mathcal{X} = \sum_{f \in F} (\Lambda_f^c - \Lambda_f^{W_c}) \, d \log \Lambda_f + (1/\chi_c^W) \, d \, T_c.
$$

The change $d \log W$ *of world real expenditure can be obtained by simply suppressing the country*

15

index c.[14]

Importantly, we can see that at the country level, welfare, unlike real output, *does* respond to productivity shocks outside the country. To better understand this result, consider for example a unit change in the productivity of producer i. Intuitively, for given factor wages, the "pure" technology effect of the shock is given by the exposure $\lambda_i^{W_c}$ of the real expenditure of country c to this producer. The productivity shock also leads to endogenous changes in the wages of the different factors $d \log w_f$, which, given that factor supplies are fixed, coincide with the changes in their factor income shares $d \log \Lambda_f$.[15] The reallocation effect depends, for each factor f, on the change $d \log \Lambda_f$ in the wage of that factor, and on the difference $\Lambda_f^c - \Lambda_f^{W_c}$ between the share of a that factor in the country's income and of the country's exposure of real expenditure to that factor.

We can define the change in the *factoral terms of trade* to be $\sum_{f \in F} (\Lambda_f^c - \Lambda_f^{W_c}) d \log w_f$. The previous discussion makes clear that with fixed factor supplies and in the absence of transfers, the reallocation effect is given by the change in the factoral terms of trade.[16] Intuitively, the factoral terms of trade weighs the change in each factor's price by the households "net" position to the price of that factor, since each factor may contribute to the household's earnings Λ_f^c as well as that household's expenditures $\Lambda_f^{W_c}$. For instance, if household c owns factor f and is the only consumer of services produced by factor f, $\Lambda_f^c = \Lambda_f^{W_c}$ and changes in the price of factor f are irrelevant for welfare.

Once we aggregate to the level of the world, of course, there are no reallocation effects.[17] Furthermore, the "pure" technology effect and the reallocation effect at the country level aggregate up to their world counterparts. This implies that the effects of country reallocations sum up to zero:

$$\sum_{c \in C} \chi_c^W (d \log \mathcal{W}_c / d \mathcal{X}) d \mathcal{X} = (d \log \mathcal{W} / d \mathcal{X}) d \mathcal{X} = 0.$$

Reallocation effects can therefore be interpreted as zero-sum distributive changes.

If all production functions and all demand aggregators are Cobb-Douglas, then we can

---

[14]At the world level, and with a slight abuse of notation, the interpretation of the decomposition goes through provided we define the "pure" technology effects $(d \log \mathcal{W} / d \log L) d \log L + (d \log \mathcal{W} / d \log A) d \log A$ as changes in real expenditure at fixed prices holding the allocation matrix constant, and reallocation effects $(d \log \mathcal{W} / d \mathcal{X}) d \mathcal{X}$ as the residual.

[15]The formula actually still applies with endogenous factor supplies.

[16]Our definition can be seen as a formalization and a generalization of the "double factoral terms of trade" (in changes) discussed in Viner (1937). Our reallocation decomposition then provides a formal connection, missing in the analysis of Viner, between the changes in the factoral terms of trade and the change in welfare.

[17]This follows immediately from the since $\Lambda_f^Y = \Lambda_f^W = \Lambda_f$ for all factors f and since $d T = 0$.

anticipate that factor income shares do not respond to productivity shocks $d \log \Lambda_f = 0$. Theorem 2 implies that in such an economy, the allocation matrix is constant.Therefore, as long as there are no shocks to transfers, a Cobb-Douglas economy only experiences "pure" technology effects, and therefore provides a useful Hulten-like benchmark without reallocation effects.

Since Theorem 2 depends on endogenous movements in factor income shares, it cannot be used directly to make predictions. However, despite this fact, Theorem 2 is useful for three reasons: (i) it provides intuition about why and how Hulten's theorem fails to describe welfare in open economies, (ii) it can be used to measure and decompose changes in welfare into different sources *conditional* on observing the changes in factor shares, and (iii) it can be combined with the results in Section 5 to perform counterfactuals.[18]

Outline of the Rest of the Paper

Theorem 2 shows that changes in welfare, unlike changes in output, depend on changes in factor shares. In Section 5, we provide a full characterization of how factor shares change in terms of microeconomic primitives (ex-ante sufficient statistics).

Before doing so, in Section 4, we consider a simple case where the changes in factor shares can be deduced from changes in import shares. For such economies, we establish a duality result between open and closed economies, which allows us study gains from trade without solving for changes in world factor shares. Doing this also allows us to introduce a key concept, which we will use repeatedly in Sections 5 and Section 6: the input-output covariance operator. Section 4 therefore also serves as a good way to build intuition for the rest of the paper.

4   Duality Between Open and Closed Economies

In Section 3.1, we showed that shocks to a country's terms of trade do not act like domestic productivity shocks in the sense that they do not affect its real output or productivity. In this section, we show that such shocks do act like productivity shocks on the country's real expenditure or welfare.[19]

---

[18]Crucially, our reallocation decomposition is not the same as the conventional (non-factoral) terms-of-trade decomposition common in the literature. For a detailed theoretical and empirical comparison of our decomposition relative to the terms-of-trade decomposition, see Appendix J. In particular, the terms-of-trade decomposition leads to a different Hulten-like benchmark with no terms-of-trade effects (instead of no reallocation effects): a small open economy taking world prices as given.

[19]Our results are related in spirit, but different, to those of Deardorff and Staiger (1988).

We do so by establishing a useful duality between the effects of foreign shocks to iceberg trade shocks in an open economy and the effects of domestic productivity shocks in a closed one. This allows us to shed light on the gains from trade in an open economy by leveraging the characterizations of the linear and nonlinear effects of productivity shocks on real output in closed economies provided respectively in Hulten (1978) and Baqaee and Farhi (2017a). These duality results build on a formula in Feenstra (1994) and can be seen as a generalization of some of results in ACR (Arkolakis et al., 2012). Unlike the other results in the paper, they do rely on the nested-CES parametric restriction.

4.1 Setup

We start by specializing the model and defining some new input-output concepts.

Nested-CES Economies

Throughout this section, we restrict our attention to the class of models that belong to the nested-CES class, where each production function and each demand aggregator is a nested-CES function, with an arbitrary number of nests and arbitrary elasticities.

We adopt the following *standard form* representation. Since we restrict our attention to nested-CES models, we can relabel the network and rewrite the input-output matrix in such a way that: each producer corresponds to a single CES nest, with a single elasticity of substitution; the representative household in each country c consumes a single specialized good which, with some abuse of notation, we also denote by c. Importantly, note that this procedure, while it keeps the set of factors F unchanged, changes the set of producers, which, with some abuse of notation we still denote by N.

Input-Output Concepts

We use the following matrix notation throughout. For a matrix X, we define $X^{(i)}$ to be its ith row and $X_{(j)}$ to be its jth column. We define the *input-output covariance operator* to be

$$Cov_{\Omega^{(k)}}(\Psi_{(i)}, \Psi_{(j)}) = \sum_{l \in N+F} \Omega_{kl} \Psi_{li} \Psi_{lj} - \left(\sum_{l \in N+F} \Omega_{kl} \Psi_{li} \right) \left(\sum_{l \in N+F} \Omega_{kl} \Psi_{lj} \right).$$

It is the covariance between the ith and jth columns of the Leontief inverse using the kth row of the input-output matrix as the probability distribution. We make extensive use of the input-output covariance operator throughout the rest of the paper.

4.2 Duality Mapping

Consider an open economy c of the nested-CES form written in standard form. Each node of the network is then a producer $i \in N_c$ with a simple CES production function with a single elasticity of substitution θ_i with associated unit-cost function[20]

$$p_i = \frac{1}{A_i} \left(\sum_{j \in N + F_c} \Omega_{ij} p_j^{1-\theta_i} \right)^{\frac{1}{1-\theta_i}}.$$

We construct a dual closed economy with the same set of producers $i \in N_c$ with CES production functions with the same set of elasticities θ_i and a HAIO matrix $\check{\Omega}$ given by $\check{\Omega}_{ij} = \Omega_{ij} / \Omega_{ic}$, where $\Omega_{ic} = \sum_{j \in N_c} \Omega_{ij}$ is the *domestic input share* of i. The unit-cost function of producer i in the dual closed economy is given by

$$\check{p}_i = \frac{1}{\check{A}_i} \left(\sum_{j \in N_c + F_c} \check{\Omega}_{ij} \check{p}_j^{1-\theta_i} \right)^{\frac{1}{1-\theta_i}}.$$

In words, the closed dual economy has the same set of producers as the open economy with the same elasticities, except the expenditure shares of each producer on foreign goods has been set to zero, and its domestic expenditures have been rescaled. Variables with "inverted-hats" are the closed-economy counterparts of the original variable.

We denote by $M_c \subseteq N_c$ the set of importing producers: the domestic producers which directly use foreign inputs in non-zero amounts. For such an importing producer $i \in M_c$, we sometimes use the notation $\epsilon_i = \theta_i - 1$ since this also corresponds to the trade elasticity of this producer.

4.3 Duality Results

To facilitate the exposition, we restrict ourselves to the case where the country c of interest has only one primary factor, which we call labor, with no tariffs. We extend all the results to the case of multiple domestic factors and tariffs in Appendix B. In Appendix C, we also show that duality can even be extended to Roy models with endogenous factor supply, along the lines of Galle et al. (2017).

Denote by $\check{W}_c$ the welfare of the dual closed economy. Since the "inverted-hat" economy is closed, welfare is equal to real output $\check{W}_c = \check{Y}_c$.

---

[20]Our results go through even when producers which do not directly use foreign inputs do not have CES production functions, but we assume for simplicity that they do.

Theorem 3 (Exact Duality). *The discrete change in welfare $\Delta \log W_c$ of the original open economy in response to discrete shocks to iceberg trade costs or productivities outside of country c is equal to the discrete change in real output $\Delta \log \check{Y}_c$ of the dual closed economy in response to discrete shocks to productivities $\Delta \log \check{A}_i = -(1/\varepsilon_i)\Delta \log \Omega_{ic}$.*

It is useful to introduce the mapping T, which to every vector of price changes $\Delta \log \check{p}$ for the goods of the dual closed economy, associates a new vector of price changes $\Delta \log \check{p}' = T(\Delta \log \check{p}; \Delta \log \check{A})$ given by:

$$\Delta \log \check{p}'_i = -\Delta \log \check{A}_i + \frac{1}{1-\theta_i} \log \left(\sum_{j \in N_c + F_c} \check{\Omega}_{ij} e^{(1-\theta_i)\Delta \log \check{p}_j} \right).$$

It is easy to verify that $T(\cdot; \Delta \log \check{A})$ is a contraction mapping, the fixed-point of which gives the response of prices to productivity shocks in the dual closed economy: $\Delta \log \check{p} = \lim_{n \to \infty} T^n(\Delta \log \check{p}_{\text{init}}; \Delta \log \check{A})$, for all $\Delta \log \check{p}_{\text{init}}$. The response of welfare in the original economy is equal to the response of real output in the dual closed economy and is given by $\Delta \log W_c = \Delta \log \check{Y}_c = -\Delta \log \check{p}_c$, where recall that we denote by c the final good consumed by the representative agent of the dual closed economy.

This expression gives a representation of the exact response of output to productivity shocks via an infinite iteration of a nonlinear contraction mapping. Of course, when there is no reproducibility in the dual closed economy, then the fixed-point problem has a finite recursive structure moving from upstream producers to downstream producers, and leads to a closed-form expression.

This duality allows us to leverage results from the literature on the real output effects of productivity shocks in closed-economy models to characterize the welfare effects of trade shocks in open economy models.

Corollary 1 (First-Order Duality). *A first-order approximation to the change in welfare of the original open economy is:*

$$\Delta \log W_c = \Delta \log \check{Y}_c \approx \sum_{i \in M_c} \check{\lambda}_i \Delta \log \check{A}_i,$$

where applying Hulten's theorem, $\check{\lambda}_i$ is the sales share or Domar weight of producer i in the dual closed economy.

Conditional on the size of the associated productivity shocks, the presence of intermediate inputs amplifies the effects of trade shocks much in the same way that the effect of intermediate inputs amplifies the effects of productivity shocks in closed economies.

This is because (gross) sales shares are greater than (net) value-added shares, reflecting and intermediate-input multiplier discussed by, among others, Jones (2011). This observation is behind the findings of Costinot and Rodriguez-Clare (2014) that allowing for intermediate inputs significantly increases gains from trade.

Corollary 2 (Second-Order Duality). *A second-order approximation to the change in welfare of the original open economy is:*

$$\Delta \log W_c = \Delta \log \check{Y}_c \approx \sum_{i \in M_c} \check{\lambda}_i \Delta \log \check{A}_i + \frac{1}{2} \sum_{i,j \in M_c} \frac{d^2 \log \check{Y}_c}{d \log \check{A}_j \, d \log A_i} \Delta \log \check{A}_j \Delta \log \check{A}_i,$$

where applying Baqaee and Farhi (2017a),

$$\frac{d^2 \log \check{Y}_c}{d \log \check{A}_j \, d \log \check{A}_i} = \frac{d \check{\lambda}_i}{d \log \check{A}_j} = \sum_{k \in N_c} (\theta_k - 1) \check{\lambda}_k Cov_{\check{\Omega}^{(k)}} \left(\check{\Psi}_{(i)}, \check{\Psi}_{(j)} \right).$$

We can re-express the change in welfare in the original open economy as

$$\Delta \log W_c = \Delta \log \check{Y}_c \approx \sum_{i \in M_c} \check{\lambda}_i \Delta \log \check{A}_i + \frac{1}{2} \sum_{k \in N_c} (\theta_k - 1) \check{\lambda}_k Var_{\check{\Omega}^{(k)}} \left(\sum_{i \in M_c} \check{\Psi}_{(i)} \Delta \log \check{A}_i \right).$$

We start by discussing the second equation. It follows from Hulten's theorem that $d \log \check{Y}_c / d \log \check{A}_i = \check{\lambda}_i$. This immediately implies that $d^2 \log \check{Y}_c / (d \log \check{A}_j \, d \log \check{A}_i) = d \check{\lambda}_i / d \log \check{A}_j$. The term $(\theta_k - 1) \check{\lambda}_k Cov_{\check{\Omega}^{(k)}} (\check{\Psi}_{(i)}, \check{\Psi}_{(j)})$ captures the direct and indirect increase in expenditure on i in response to a shock to j because of substitution by k across its inputs. The term $\check{\lambda}_k$ is the total sales of k. The term $\theta_k - 1$ determines how much k substitutes expenditure towards ($\theta_k > 1$) or away from ($\theta_k < 1$) inputs which get relatively cheaper. The vector $\check{\Psi}_{(j)}$ captures the change in the input-price vector in response to the shock to j. The vector $\check{\Psi}_{(i)}$ captures how much an increase in expenditure on each input increases expenditure on i. These effects must be summed over producers k to determine the change $d \check{\lambda}_i / d \log \check{A}_j$ in the sales share of i in response to a shock to j.

The third equation in the corollary indicates that the ultimate impact of the shock depends on how heterogeneously exposed each producer k is to the average productivity shock via its different inputs as captured by the term $Var_{\check{\Omega}^{(k)}} \left(\sum_{i \in M_c} \check{\Psi}_{(i)} \Delta \log \check{A}_i \right)$, and on whether these different inputs are complements ($\theta_k < 1$), substitutes ($\theta_k > 1$) or Cobb-Douglas $\theta_k = 1$. It indicates that complementarities lead to negative second-order terms which amplify negative shocks and mitigate positive shocks. Conversely, substitutabilities lead to positive second-order terms which amplify negative shocks and mitigate pos-

itive shocks. Of course, there are no second-order terms in the Cobb-Douglas case.

Duality with an Industry Structure

To discuss these results further, it is useful to assume that there is an *industry structure*: producers are grouped into industries and the goods produced in any given industry are aggregated with a CES production function; and all other agents only use aggregated industry goods. In this case, all domestic producers in a given industry are uniformly exposed to any other given domestic producer. This implies that in Corollary 2, only the elasticities of substitution across industries receive non-zero weights. The elasticities of substitution across producers within industries receive a zero weight, and they only matter via their influence on the productivity shocks through the trade elasticities.

In fact, the matrix $\check{\Omega}$ of the dual closed economy can be specified entirely at the industry level where the different producers are the different industries $\iota \in \mathcal{N}_c$. Given the productivity shocks $\Delta \log \check{A}_\iota$ to the importing industries $\iota \in \mathcal{M}_c$, Theorem 3 and Corollaries 1 and 2 can then be applied at the industry level, with this industry level input-output matrix, and with only elasticities of substitution across industries.

Many cases considered in the literature have such an industry structure, and impose the additional assumption that all the elasticities of substitution across industries (and with the factor) in production and in consumption are unitary (but those within industries are above unity). This makes the dual closed economy Cobb-Douglas. Such assumptions are made for example by ACR, Costinot and Rodriguez-Clare (2014), and Caliendo and Parro (2015). In this Cobb-Douglas case, the dual closed economy is exactly log-linear in the dual productivity shocks $\Delta \log \check{A}_\iota$. The effects of shocks to iceberg trade costs or to productivities outside of the country then coincide with the first-order effects of the dual shocks given by Corollary 1. Their second-order effects given by Corollary 2 are zero, and the same goes for their higher-order effects.

For example, we can recover the basic result of ACR by assuming that there is a single industry ι producing only from labor so that $\check{\lambda}_\iota = 1$. In this case, we get $\Delta \log W_c = \Delta \log \check{A}_\iota$ as an exact expression. We can also recover the extension of the ACR result by Costinot and Rodriguez-Clare (2014) to allow for multiple industries and input-output linkages, but restricting elasticities of substitution across industries to be unitary. In this case, we get $\Delta \log W_c = \sum_{\iota \in \mathcal{M}_c} \check{\lambda}_\iota \Delta \log \check{A}_\iota$ as an exact expression.

Our results therefore generalize some of the insights of ACR and of Costinot and Rodriguez-Clare (2014) to models with input-output linkages and where elasticities of

substitution across industries (and with the factor) are not unitary.[21] In such models, the dual closed economy is no longer Cobb-Douglas. Deviations from Cobb Douglas generate nonlinearities, which can either mitigate or amplify the effects of the shocks depending on whether there are complementarities or substituabilities, and with an intensity which depends on how heterogeneously exposed the different producers are to the shocks.

Corollary 3 (Exact Duality and Nonlinearities with an Industry Structure). *For country c with an industry structure, we have the following exact characterization of the nonlinearities in welfare changes of the original open economy.*

(i) *(Industry Elasticities) Consider two economies with the same initial input-output matrix and industry structure, the same trade elasticities and changes in domestic input shares, but with lower elasticities across industries for one than for the other so that $\theta_\kappa \leq \theta_\kappa'$ for all industries κ. Then $\Delta \log W_c = \Delta \log \check{Y}_c \leq \Delta \log W_c' = \Delta \log \check{Y}_c'$ so that negative (positive) shocks have larger negative (smaller positive) welfare effects in the economy with the lower industry elasticities.*

(ii) *(Cobb-Douglas) Suppose that all the elasticities of substitution across industries (and with the factor) are equal to unity ($\theta_\kappa = 1$), then $\Delta \log W_c = \Delta \log \check{Y}_c$ is linear in $\Delta \log \check{A}$.*

(iii) *(Complementarities) Suppose that all the elasticities of substitution across industries (and with the factor) are below unity ($\theta_\kappa \leq 1$), then $\Delta \log W_c = \Delta \log \check{Y}_c$ is concave in $\Delta \log \check{A}$, and so nonlinearities amplify negative shocks and mitigate positive shocks.*

(iv) *(Substituabilities) Suppose that all the elasticities of substitution across industries (and with the factor) are above unity ($\theta_\kappa \geq 1$), then $\Delta \log W_c = \Delta \log \check{Y}_c$ is convex in $\Delta \log \check{A}$, and so nonlinearities mitigate negative shocks and amplify positive shocks.*

(v) *(Exposure Heterogeneities) Suppose that industry κ is uniformly exposed to the shocks as they unfold, so that $Var_{\check{\Omega}_s^{(\kappa)}} \left(\sum_{\iota \in \mathcal{M}_c} \check{\Psi}_{(\iota),s} \Delta \log \check{A}_\iota \right) = 0$ for all s where s indexes the dual closed economy with productivity shocks $\Delta \log \check{A}_{\iota,s} = s \Delta \log \check{A}_\iota$, then $\Delta \log W_c = \Delta \log \check{Y}_c$*

---

[21]Costinot and Rodriguez-Clare (2014) show that the gains from trade are higher in multi-sector economies without input-output linkages when sectors and complements in consumption. Corollary 3 generalizes these results to economies with input-output linkages. As we shall see, this matters quantitatively given that most empirical evidence points to the presence of much more important complementarities in production than in consumption. Furthermore, even if the production and consumption elasticities were the same, Corollary 2 shows that given the size of the trade shocks $\Delta \log \check{A}_i$, the nonlinear effects of non-unitary elasticities $\theta_k - 1$ scale with the size of the Leontief inverse and the Domar weights. Therefore, even if the elasticities are identical in both consumption and production, input-output linkages amplify the gains from trade.

is independent of θ_κ. Furthermore[22]

$$\Delta \log W_c = \Delta \log \check{Y}_c = \sum_{\iota \in \mathcal{M}_c} \check{\lambda}_\iota \Delta \log \check{A}_\iota$$

$$+ \int_0^1 \sum_{\kappa \in \mathcal{N}_c} (\theta_\kappa - 1) \check{\lambda}_{\kappa,s} Var_{\check{\Omega}_s^{(\kappa)}} \left(\sum_{\iota \in \mathcal{M}_c} \check{\Psi}_{(\iota),s} \Delta \log \check{A}_\iota \right) (1-s) ds.$$

These results (ii), (iii), and (iv) follow immediately from Corollary 2 applied at the industry level, which allows us to determine that at every point, the Hessian of the function $\Delta \log \check{Y}_c(\Delta \log \check{A})$ is null in case (ii), negative semi-definite in case (iii), and positive semi-definite in case (iv). The same logic can be used to prove a local version of (i) since the Hessians of the two economies at the original point are ordered (using the semi-definite condition partial ordering). Similar arguments can be used to derive a local version of (v).

These results can also be derived using $\Delta \log W_c = \Delta \log \check{Y}_c = -\Delta \log \check{p}_c$ with $\Delta \log \check{p} = \lim_{n \to \infty} T^n(\Delta \log \check{p}_{\text{init}}; \Delta \log \check{A})$, where the mapping $\Delta \log \check{p}' = T(\Delta \log \check{p}; \Delta \log \check{A})$ is applied at the industry level. Indeed, the mapping is always linear in $\Delta \log \check{A}$. If all the elasticities of substitution across industries are below (above) unity, then the mapping is convex (concave) in $\Delta \log p$, and hence the mapping preserves convexity (concavity) in $\Delta \log A$. This immediately implies (ii), (iii), and (iv). The result (i) also follows from the fact that the T mapping for the high-elasticity and low-elasticity economy are monotonically ordered and the ordering is preserved by iterating on T.

The formula in (v) can be obtained by integrating by parts the path-integral version of Hulten's theorem $\Delta \log W_c = \Delta \log \check{Y}_c = \int_0^1 \check{\lambda}_{\iota,s} \, d\log \check{A}_{\iota,s}$ using Corollary 2 to get an expression for $d \log \check{\lambda}_{\iota,s}$. The formula immediately implies the associated result.

Example: Critical Foreign Inputs

To see the importance of nonlinearity in the production network, consider the simple example of a country c depicted in Figure 1. The only traded good is energy E.[23] The representative household in the country consumes domestic goods 1 through to N with some elasticity of substitution θ_0, with equal sales shares $1/N$ at the initial point. Some

---

[22]In the Cobb-Douglas case, the structure of the domestic network of the dual closed economy is irrelevant since the sales shares are sufficient statistics. Outside of this case, the structure of the network matters in general beyond the first order. There is one non-Cobb-Douglas case where it does not and where we get a network-irrelevant closed-form solution: when all the elasticities across industries are uniform ($\theta_\kappa = \theta$) and when the domestic upstream supply chains of the different importing industries are disjointed set, we get $\Delta \log W_c = \Delta \log \check{Y}_c = \left(\sum_{\iota \in \mathcal{M}_c} \check{\lambda}_\iota e^{(\theta-1)\Delta \log \check{A}_\iota} \right) / (\theta - 1)$.

[23]This example is an open-economy version of an example in Baqaee and Farhi (2017a).

fraction of goods, goods 1 through to M, are made via labor L and a composite energy good E with an elasticity of substitution θ_1, with an initial energy share $(N/M)\check{\lambda}_E$. The composite energy good is a CES aggregate of domestic and foreign energy with elasticity of substitution $\theta_E > 1$. Domestic energy E, as well as the rest of the consumption goods, goods $M+1$ through to N, are made using only domestic labor. We assume that the elasticity of substitution in production $\theta_1 < 1$ and that production has stronger complementarities than consumption $\theta_1 < \theta_0$.
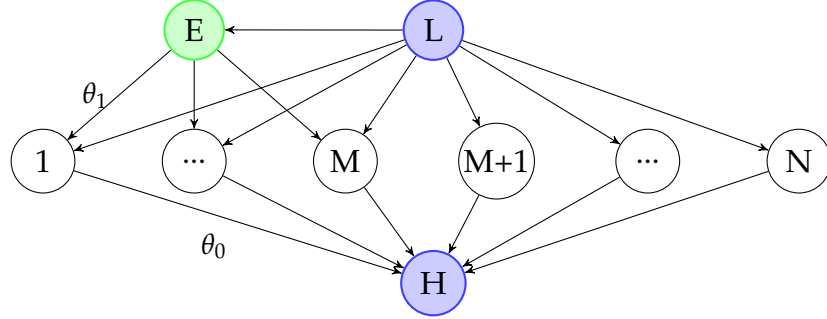

Figure 1: An illustration of the economy with a key energy input E. Each industry has different shares of labor and energy and substitutes across labor and energy with elasticity $\theta_1 < 1$. The household can substitute across goods with elasticity of substitution $\theta_0 > \theta_1$. Energy is a traded good, which can either be produced domestically or sourced from the rest of the world, with an elasticity of substitution $\theta_E > 1$ between the two.

Consider an increase in iceberg trade costs which increases the cost of import of foreign energy. The welfare effect of this trade shock is the same as that of a negative productivity shock to the energy producer of the dual closed economy

$$\Delta \log \check{A}_E = -\frac{1}{\varepsilon_E} \Delta \log \check{\Omega}_{Ec} < 0,$$

where $\varepsilon_E = \theta_E - 1$ is the trade elasticity of the energy composite good E and $\Delta \log \Omega_{Ec}$ is the change of its domestic expenditure share.

The second-order expression for welfare from Corollary 2 is more transparent than the closed-form expression (given in Appendix B):

$$\Delta \log W_c \approx \check{\lambda}_E \Delta \log \check{A}_E + \frac{1}{2} \sum_{k \in N_c} (\theta_k - 1) \check{\lambda}_k Var_{\check{\Omega}^{(k)}} \left(\check{\Psi}_{(E)} \Delta \log \check{A}_E \right),$$

$$= \check{\lambda}_E \Delta \log \check{A}_E + \frac{1}{2} \check{\lambda}_E \left((\theta_0 - 1) \check{\lambda}_E (\frac{N}{M} - 1) + (\theta_1 - 1)(1 - \frac{N}{M} \check{\lambda}_E) \right) (\Delta \log \check{A}_E)^2.$$

When $M = N$, energy becomes a universal input, and the elasticity of substitution in consumption θ_0 drops out of the expression because $Var_{\check{\Omega}^{(0)}}(\check{\Psi}_E) = 0$. This is because all

consumption goods are uniformly exposed to the trade shock, and so substitution by the household is irrelevant. Since $\theta_1 < 1$, nonlinearities captured by the second-order term amplify the negative welfare effects of the trade shock. This is because complementarities between energy and labor imply that the sales share of energy $\check{\lambda}_E$ increases with the shock, thereby amplifying its negative effect.

When $M < N$, the elasticity of substitution in consumption θ_0 matters. Since $\theta_0 > \theta_1$, the nonlinear adverse effect of the trade shock is reduced compared to the case $M = N$ when we keep the initial sales share of energy $\check{\lambda}_E$ constant. This is true generally but the effect is easiest to see when $\theta_0 > 1$ since the household can now substitute away from energy-intensive goods, which mitigates the increase of the sales share of energy $\check{\lambda}_E$, and hence the negative welfare effects of the shock. These effects are stronger, the lower is M, i.e. the more heterogeneous are the exposures of the different goods to energy.

These effects are absent in the cases analyzed by ACR and Costinot and Rodriguez-Clare (2014) who make the Cobb-Douglas assumption $\theta_0 = \theta_1 = 1$, which renders the model log-linear and eliminates all nonlinearities.

Quantitative Gains from Trade: Intermediate Inputs and Nonlinearities

We apply our duality results using the World Input-Output Database (WIOD) (see Timmer et al., 2015) to study the gains from trade. We compare the welfare losses from moving different countries to autarky. The WIOD contains the expenditures of each industry in each country on intermediate input purchases from every other industry in every other country. It also contains data on final consumption demand. We use data from 2008, which is the final year in the 2013 release of the data. The dataset has 41 countries, one of which is an aggregate Rest-of-World country, and each country has 30 industries. See Appendix A for more details.

We assume that production takes a nested CES form, where θ_0 is the elasticity of substitution across industries in consumption, θ_1 is the elasticity of substitution between value-added and intermediate inputs, θ_2 is the elasticity of substitution across industries in intermediate input use. The dual productivity shocks to the importing producers corresponding to a move to autarky of the original open economy are given by $\Delta \log \check{A}_i = -(1/\varepsilon_i) \log \Omega_{ic}$, since we know that in Autarky, all domestic shares must go to one.

To calibrate the trade elasticities ε_i, we use the estimates from Caliendo and Parro (2015). For the rest of the elasticities $(\theta_0, \theta_1, \theta_2)$, our benchmark sets the elasticity of substitution across industries $\theta_2 = 0.2$, the one between value-added and intermediates $\theta_1 = 0.5$, and the one in consumption $\theta_0 = 0.9$. These elasticities are broadly consistent with the estimates of Atalay (2017), Boehm et al. (2015), Herrendorf et al. (2013), and

Oberfield and Raval (2014). Overall, the evidence suggests that these elasticities are all less than one (sometimes significantly so).

$(\theta_0, \theta_1, \theta_2)$	VA $(1,1,1)$	$(1,1,1)$	$(1,0.5,0.6)$	$(0.9,0.5,0.2)$
FRA	9.8%	18.5%	24.7%	30.2%
JPN	2.4%	5.2%	5.5%	5.7%
MEX	11.5%	16.2%	21.3%	44.5%
USA	4.5%	9.1%	10.3%	13.0%

Table 1: Gains from trade for a selection of countries. The first column is a multi-sector value-added economy with no intermediate inputs and with the Cobb-Douglas assumption. The second column allows for intermediate inputs but maintains the Cobb-Douglas assumption in the direction of complementarities. The other columns allow for intermediate inputs and relax the Cobb-Douglas assumption. The microeconomic trade elasticities are the same across all columns and taken from Caliendo and Parro (2015), so the size of the trade shock to each industry is the same across all columns.

The gains from trade are in Table 1 for different values of the elasticities of substitution $(\theta_0, \theta_1, \theta_2)$. The first column replicates the results of a multi-sector value-added model without intermediate inputs and with the Cobb-Douglas assumption $(\theta_0, \theta_1, \theta_2) = (1,1,1)$, reported in Costinot and Rodriguez-Clare (2014).[24] The second column replicates the results of an a model which allows for intermediate inputs but maintains the Cobb-Douglas assumption, also reported in Costinot and Rodriguez-Clare (2014). As expected, allowing for intermediate inputs increases gains from trade. This is because of the first-order or log-linear effect captured by Corollary 1: it reflects the fact that abstracting away from intermediate inputs reduces the volume of imports relative to GDP. The other columns continue to allow for intermediate inputs, but deviates from the Cobb-Douglas assumption, giving rise to nonlinearities. Moving across columns towards more complementarities increases the gains from trade. This is because of the nonlinear effect captured by Corollary 2: more complementarities magnify gains from trade by increasing nonlinearities. Our benchmark calibration is the one on the far right.

The magnitudes of these different effects are different across countries. The importance of accounting for intermediate inputs is largely independent of the degree of openness of the country. By contrast, the importance of accounting for nonlinearities does depend on the degree of openness: the more open the country, the larger are the dual

---

[24]Since the value-added version of the model has no intermediate inputs, the production elasticities θ_1 and θ_2 are irrelevant. Unlike elasticities of substitution in production θ_1 and θ_2, most estimates of elasticities of substitution in consumption θ_0 are close to one. Hence even if we were to allow for realistic complementarities in consumption (say $\theta_0 = 0.9$) in the value-added version of the model, it would make little difference compared to the Cobb-Douglas benchmark ($\theta_0 = 1$) reported in the first column of the table.

productivity shocks, and the more nonlinearities matter. Overall, it seems that nonlinearities are as important as intermediate goods to the study of gains from trade.

5   Comparative Statics: Ex-Ante Sufficient Statistics

In Section 3, we showed that welfare responses to shocks depend on changes in factor income shares (ex-post sufficient statistics). In Section 4 we studied a special case where changes in welfare can be predicted without solving directly for changes in factor shares.

In this section we characterize the responses of factor income shares to shocks as a function of sufficient-statistic microeconomic primitives: the HAIO matrix and elasticities of substitution in production and in consumption (ex-ante sufficient statistics). The results of this section can then be combined with Theorem 2 to answer counterfactual questions. We also characterize the responses to shocks of all prices and quantities. We focus on productivity shocks because shocks to factor supplies and to iceberg costs are special cases of productivity shocks. Transfer shocks are covered in Appendix I.

Throughout this section, we restrict our attention to the class of models that belong to the nested-CES class written in standard-form. The reason for this is clarity not tractability. We refer the reader to Appendix G for a discussion of how to generalize all of our results and intuitions to arbitrary economies with non-nested-CES production functions and demand aggregators.

5.1   Comparative Statics

We define two matrices. The first is the $(N + F) \times (N + F)$ "propagation-via-substitution" matrix Γ whose ijth element makes use of the input-output covariance operator defined in Section 4

$$\Gamma_{ij} = \sum_{k \in N} (\theta_k - 1) \frac{\lambda_k}{\lambda_i} Cov_{\Omega^{(k)}} \left(\Psi_{(i)}, \Psi_{(j)} \right),$$

and which encodes substitutions by all producers discussed in Section 4. The second is the $(N + F) \times F$ "propagation-via-redistribution" matrix Ξ whose ifth element is

$$\Xi_{if} = \sum_{c \in C} \frac{\lambda_i^{W_c} - \lambda_i}{\lambda_i} \Phi_{cf} \Lambda_f,$$

where we write λ_i and Λ_i interchangeably when $i \in F$ is a factor, and which encodes the redistribution of income across the different households in the different countries and its effects given their different expenditure patterns.

Factor Shares and Sales Shares/Domar Weights

We start with a characterization of the responses of factor income shares.

Theorem 4 (Factor Shares and Sales Shares/Domar Weights). *The changes in the factor income shares (factor Domar weights) in response to a productivity shock to producer i are solve the linear system*

$$\frac{d \log \Lambda_f}{d \log A_i} = \Gamma_{fi} - \sum_{g \in F} \Gamma_{fg} \frac{d \log \Lambda_g}{d \log A_i} + \sum_{g \in F} \Xi_{fg} \frac{d \log \Lambda_g}{d \log A_i}.$$

Given these changes in factor income shares, the changes in producer sales shares (producer Domar weights) in response to a productivity shock to producer i are:

$$\frac{d \log \lambda_j}{d \log A_i} = \Gamma_{ji} - \sum_{g \in F} \Gamma_{jg} \frac{d \log \Lambda_g}{d \log A_i} + \sum_{g \in F} \Xi_{jg} \frac{d \log \Lambda_g}{d \log A_i}.$$

Consider for example the response $d \log \Lambda_f / d \log A_i$ of the income share of factor f to a positive unit shock to the productivity of producer i. It helps to write out the expression explicitly

$$\frac{d \log \Lambda_f}{d \log A_i} = \underbrace{\sum_{k \in N} \frac{\lambda_k}{\Lambda_f} (\theta_k - 1) Cov_{\Omega^{(k)}} \left(\Psi_{(i)}, \Psi_{(f)} \right)}_{\text{Substitution Impulse } \Gamma_{fi}}$$

$$- \underbrace{\sum_{g \in F} \sum_{k \in N} \frac{\lambda_k}{\Lambda_f} (\theta_k - 1) Cov_{\Omega^{(k)}} \left(\Psi_{(g)}, \Psi_{(f)} \right) \frac{d \log \Lambda_g}{d \log A_i}}_{\text{Additional Substitution } \sum_{g \in F} \Gamma_{fg} \frac{d \log \Lambda_g}{d \log A_i}}$$

$$+ \underbrace{\sum_{g \in F} \sum_{c \in C} \frac{\Lambda_f^{W_c} - \Lambda_f}{\Lambda_f} \Phi_{cg} \Lambda_g \frac{d \log \Lambda_g}{d \log A_i}}_{\text{Income Redistribution } \sum_{g \in F} \Xi_{fg} \frac{d \log \Lambda_g}{d \log A_i}}.$$

For fixed factor prices, every producer k will substitute across its inputs in response to this shock. Suppose that $\theta_k > 1$, so that producer k substitutes (in expenditure shares) *towards* those inputs j that are more reliant on producer i, captured by Ψ_{ji}, the more so, the higher is $\theta_k - 1$. Now, if those inputs are also more reliant on factor f, captured by a high $Cov_{\Omega^{(k)}} \left(\Psi_{(f)}, \Psi_{(i)} \right)$, then substitution by k will increase demand for factor f and hence the income share of factor f. These substitutions, which happen at the level of each

producer k, must be summed across producers. This leads to a an impulse change in factor prices captured by the propagation-via-substitution term Γ_{fi}.

These impulse changes in factor prices then set off additional rounds of substitution in the economy that we must account for. The change in the price of each factor g is given by $\mathrm{d}\log w_g / \mathrm{d}\log A_i = \mathrm{d}\log \Lambda_g / \mathrm{d}\log A_i$. The effect on the share of factor f is the same as that of a set of equivalent negative productivity shock to the different factors, leading to the second propagation-via-substitution term $-\sum_{g \in F} \Gamma_{fg} \, \mathrm{d}\log \Lambda_g / \mathrm{d}\log A_i$.

These changes in factor prices also change the distribution of income across households in different countries. This in turn affects the demand for the factor f since the different households are differently exposed, directly and indirectly, to factor f. The overall effect can be found by summing over countries c the increase $\sum_{g \in F} \Phi_{cg} \Lambda_g \, \mathrm{d}\log \Lambda_g / \mathrm{d}\log A_i$ in the factor share of country c multiplied by its relative welfare exposure $(\Lambda_f^{W_c} - \Lambda_f)/\Lambda_f$ to factor f. This leads to the propagation-via-redistribution term $\sum_{g \in F} \Xi_{fg} \, \mathrm{d}\log \Lambda_g / \mathrm{d}\log A_i$.

The intuition for how the Domar weight of goods respond to shocks is near identical. These formulas show that Cobb-Douglas assumptions, prevalent in the literature which incorporates production networks in trade models for their analytical convenience, are also special (see e.g. Costinot and Rodriguez-Clare, 2014; Caliendo and Parro, 2015). Whenever $\theta_k = 1$, the term accounting for expenditure substitution by producer k in the propagation-via-substitution matrix Γ is equal to zero.

Real Output and Real Expenditure or Welfare

Theorem 4 gives the endogenous responses of factor shares to shocks as a function of microeconomic primitives. These were left implicit in Theorem 2. Theorem 4 can therefore be used in conjunction with Theorem 2 to characterize the response of welfare to shocks as a function of microeconomic primitives, up to the first order.

Theorem 4 also gives the endogenous responses to shocks of producer sales shares or Domar weights. Since the response of real output to productivity shocks is given by the corresponding local Domar weight, Theorem 4 can also be used to give the response of real output to shocks, up to the second order:

$$\frac{\mathrm{d}\log Y_c}{\mathrm{d}\log A_j} = \lambda_j^{Y_c}, \quad \frac{\mathrm{d}^2 \log Y_c}{\mathrm{d}\log A_j \, \mathrm{d}\log A_i} = \frac{\mathrm{d}\lambda_j^{Y_c}}{\mathrm{d}\log A_i} = \lambda_j^{Y_c} \left(\frac{\mathrm{d}\log \lambda_j}{\mathrm{d}\log A_i} - \sum_{f \in N_c} \Lambda_f^{Y_c} \frac{\mathrm{d}\log \Lambda_f}{\mathrm{d}\log A_i} \right),$$

where $\mathrm{d}\lambda_j / \mathrm{d}\log A_i$ and $\mathrm{d}\log \Lambda_f / \mathrm{d}\log A_i$ are given by Theorem 4.[25]

---

[25]The expression for $\mathrm{d}^2 \log Y_c / (\mathrm{d}\log A_j \, \mathrm{d}\log A_i)$ is a gross abuse of notation and must be handled with care. We do not dwell on the subtleties in this paper, but technically, the change in real output from one

Prices and Quantities

Armed with Theorem 4, it is straightforward to characterize the response of prices and quantities to shocks, where prices are expressed in the world GDP numeraire.

Corollary 4. *(Prices and Quantities) The changes in the wages of factors and in the prices and quantities of goods in response to a productivity shock to producer i are given by:*

$$\frac{\mathrm{d}\log w_f}{\mathrm{d}\log A_i} = \frac{\mathrm{d}\log \Lambda_f}{\mathrm{d}\log A_i},$$

$$\frac{\mathrm{d}\log p_j}{\mathrm{d}\log A_i} = -\Psi_{ji} + \sum_{g \in F} \Psi_{jg} \frac{\mathrm{d}\log w_g}{\mathrm{d}\log A_i},$$

$$\frac{\mathrm{d}\log y_j}{\mathrm{d}\log A_i} = \frac{\mathrm{d}\log \lambda_j}{\mathrm{d}\log A_i} - \frac{\mathrm{d}\log p_j}{\mathrm{d}\log A_i},$$

where $\mathrm{d}\log \Lambda_f / \mathrm{d}\log A_i$ *is given in Theorem 4.*

These results on the responses of prices and quantities to productivity shocks, and hence by implication to shocks to factor supplies and to iceberg trade costs, generalize the classic results of Stolper and Samuelson (1941) and Rybczynski (1955).[26]

Other Uses of Theorem 4

Other than helping to characterize the response of output, welfare, prices and quantities, Theorem 4 has many other uses which we do not discuss in detail for brevity. In Appendix E, we provide some additional applications of Theorem 4.

In particular, Appendix E.1 uses Theorem 4 to write trade elasticities at any level of aggregation as a linear combination of underlying microeconomic elasticities of substitution with weights that depend on the input-output table.[27]

Theorem 4 can also be used as a recipe for analytically characterizing the behavior of fairly complicated general equilibrium models. For example, in Appendix E.2, we show

---

allocation to another in general depends on the *path* taken. Hulten's theorem guarantees that changes in real output are a path integral of the vector field defined by the local Domar weights along a path of productivity changes. Hence, the expression $\mathrm{d}^2 \log Y_c / (\mathrm{d}\log A_j \, \mathrm{d}\log A_i)$ is really the derivative of the vector field defined by the local Domar weights. Conditional on the path taken from one allocation to the next, it can be used to compute the second derivative of the change in output at any point along that path.

[26]See Appendix I for a discussion of how our, by taking a limit, our results can be applied to economies where traded goods are perfect substitutes as assumed by these theorems.

[27]In Appendix K, we even show that the effect of supply chains on the trade elasticity, emphasized by Yi (2003), are formally identical to the issues of reswitching and capital reversing identified in the Cambridge Capital Controversy of the 1950s and 60s.

how this theorem can be used to study the way opening up to trade can increase growth by showing how export-led growth can help escape Baumol's cost disease.

6   Tariffs and Other Distortions

So far, we have maintained the assumption of no distortions. In this section, we extend our results to allow for tariffs, or more generally for all distortions that can be modeled as explicit or implicit taxes.

We start by summarizing how to generalize the ex-post and ex-ante comparative statistic results of Sections 3 and 5 in the presence of large tariffs or other distortions. As far as we are aware, this is the first time such comparative statics have been derived in the literature, and so we consider them to be an important contribution of this paper. However, in order to avoid repetition, we relegate the formal statement of the results and the underlying analysis to Appendix F.

We focus instead on a different perspective showing that the losses from small tariffs or other distortions are given, up to a second order, by a Domar-weighted sum of deadweight-loss Harberger triangles. As usual, we present our result in two ways, using ex-post and ex-ante sufficient statistics.

6.1   Allowing for Distortions

We denote by μ the $N \times 1$ matrix of tax wedges, where the ith element is an ad valorem tax on the output of producer i. To state our results, we assume that the revenue generated by the wedge μ_i are *included* in the revenue of producer i (the producer collects the tax revenue as part of its revenue, and then pays the government). This is merely an accounting convention, and it is straightforward to convert our results for situations where the revenue generated by the wedge are not included in revenues. We can capture tariffs on imports (exports) as taxes on specialized importers (exporters) who buy foreign (domestic) goods and sell them across borders. These wedges can also capture other distortions such as markups, financial constraints, or quotas.

6.2   Generalizing the Comparative Statics in Sections 3 and 5

Theorems 1, 2, and 4 are generalized to allow for arbitrary distortions in Appendix F. The analysis reveals important and interesting differences. For example, in the presence of distortions such as tariffs or markups, shocks to productivities, factor supplies, and

iceberg trade costs outside of a country now typically have first-order effects on its real output and aggregate productivity. Nevertheless, our output- and welfare-accounting results can be fully generalized to cover these cases, along with the accompanying decomposition into pure technology and reallocation effects.

6.3 Costs of Tariffs and Other Distortions: Ex-Post Sufficient Statistics

In what follows, instead of restating all the results when there are distortions, we provide an alternative characterization of the effect of distortions based on a different perspective.

We write tariffs or other distortions as $\exp(\Delta \log \mu_i)$ and provide approximations for small tariffs $\Delta \log \mu_i$ around an efficient equilibrium with no tariffs or other distortions.

Real Output

We start by characterizing changes in real output.

Theorem 5 (Reduced-Form Output Loss). *Starting at an efficient equilibrium, up to the second order, in response to the introduction of small tariffs or other distortions:*

(i) changes in world real output and real expenditure are given by

$$\Delta \log Y = \Delta \log W \approx \frac{1}{2} \sum_{i \in N} \lambda_i \Delta \log y_i \Delta \log \mu_i;$$

(ii) changes in the real output of country c are given by:

$$\Delta \log Y_c \approx \frac{1}{2} \sum_{i \in N_c} \lambda_i^{Y_c} \Delta \log y_i \Delta \log \mu_i.$$

Hence, for both the world and for each country, the reduction in real output from tariffs and other distortions is given by the sum of all the deadweight-loss triangles $1/2 \Delta \log y_i \Delta \log \mu_i$ weighted by their corresponding local Domar weights.[28] Harberger (1964) argues that the classic welfare deadweight-loss triangle intuition from partial equilibrium models could be applied to general equilibrium models to measure changes in

---

[28]Although Theorem 5 holds in general equilibrium, it also has a partial equilibrium counterpart. For a small open economy operating in a perfectly competitive world market, the introduction of import tariffs reduces the welfare of that country's representative household by $\Delta W \approx (1/2) \sum_i \lambda_i \Delta \log y_i \Delta \log \mu_i$, where μ_i is the ith gross tariff (no tariff is $\mu_i = 1$), y_i is the quantity of the ith import, and λ_i is the corresponding Domar weight (see Appendix D for details). This argument is reminiscent of Paul Krugman's 2018 essay *Does Trade in Intermediate Goods Alter the Logic of Costs From Protectionism?* Theorem 5 shows that this type of intuition can be applied in general equilibrium as well.

welfare as long as there are compensating transfers between households. Theorem 5 shows that, in fact, a similar formula can be used in general equilibrium, in the presence of income effects and when there are no compensating transfers, to measure changes in real output; to the best of our knowledge, Theorem 5 is a new result.

Theorem 5 shows that we only need to track changes in those quantities which are subject to a wedge — if a good is untaxed, or taxed but not included in real GDP (like a tax on imported consumption), then changes in that quantity are not directly relevant for real GDP.

Conditional on matching both the Domar weights λ_i and $\lambda_i^{Y_c}$ and the changes in the quantities $\Delta \log y_i$ of all goods, the details of the production structure are irrelevant. In particular, conditional on these sufficient statistics, we do not need to know anything about whether or not there are international (or domestic) production networks. However, as we shall show, input-output linkages do affect Domar weights and changes in the quantities of all goods, and so accounting for input-output linkages does matter. As we shall see, accounting for global value chains matters a great deal for the quantitative effects of tariffs: the triangles $1/2 \Delta \log y_i \Delta \log \mu_i$ are larger, and they are also aggregated with larger weights given by sales shares λ_i and $\lambda_i^{Y_c}$ rather than value-added shares.

To give some intuition for Theorem 5, we focus on the country level result for simplicity. Starting at an efficient equilibrium, the introduction of tariffs or other distortions leads to changes $\Delta \log y_i$ in the quantities of goods $i \in N_c$ in country c and to changes in the wedges $\Delta \log \mu_i$ between prices and marginal costs. The price-cost margin $p_i \Delta \log \mu_i$ measures the wedge between the marginal contribution to country real output and the marginal cost to real output of increasing the quantity of good i by one unit. Hence, $\lambda_i^{Y_c} \Delta \log \mu_i$ is the marginal proportional increase in real output from a proportional increase in the output of good i. Integrating from the initial efficient point to the final distorted point, we find that $(1/2)\lambda_i^{Y_c} \Delta \log y_i \Delta \log \mu_i$ is the contribution of good i to the change in real output.

Changes $\Delta \log y_i$ in the quantities of goods $i \in N_c$ in country c can be driven by changes in tariffs or other distortions in the country or in other countries. This shows that changes in tariffs or other distortions outside of the country affect the aggregate productivity of the country. This is in a sharp contrast to the results that we derived earlier for efficient economies.

It is instructive to compare the costs of tariffs to the costs of an increase in iceberg costs. In response to a change in iceberg costs outside of the country , the change in country real output is zero at any order of approximation. At the world level, in response to a change $\Delta \log(1/A_i)$ in iceberg trade costs, the change in real output or real expenditure is given

up to a second-order by the sum of trapezoids rather than triangles:

$$\Delta \log Y = \Delta \log W \approx - \sum_{i \in N} \lambda_i \left(1 + \frac{1}{2} \Delta \log \lambda_i \right) \Delta \log(1/A_i).$$

In contrast to equivalent shocks to tariffs, shocks to iceberg trade costs have nonzero first-order effects. This is a way to see why the costs of non-tariff trade barriers are typically so much higher than tariff trade barriers in trade models.

Real Expenditure and Welfare

Theorem 5 shows how real output responds to changes in tariffs or other distortions. These results do not apply to welfare. At the country level, changes in tariffs and other distortions typically lead to first-order changes (due to terms of trade/reallocation effects). But even at the world level where these effects wash out, changes in real expenditure no longer coincide with changes in welfare, since changes in world real expenditures $d \log W$ cannot be integrated to arrive at a well-defined social welfare function.[29]

To proceed, we introduce a homothetic Bergson-Samuelson social welfare function

$$W^{BS}(W_1, \ldots, W_C) = \sum_c \overline{\chi}_c^W \log W_c,$$

where $\overline{\chi}_c^W$ is the initial income share of country c at the efficient equilibrium. These welfare weights are chosen so that there is no incentive to redistribute across agents at the initial equilibrium. Even though this welfare function has no desire to redistribute across agents at the initial point, distributive effects across households do appear at the second-order in response to shocks.

We measure the change in welfare by asking what fraction of consumption would society be prepared to give up to avoid the imposition of the tariffs. Formally, we measure changes in welfare by $\Delta \log \delta$, where δ solves the equation

$$W^{BS}(\delta \overline{W}_1, \ldots, \delta \overline{W}_C) = W^{BS}(W_1, \ldots, W_C),$$

where $\overline{W}_c$ and W_c are the values at the initial and final equilibrium. We use a similar definition for country level welfare δ_c

Corollary 5 (Reduced-Form Welfare). *Starting at an efficient equilibrium in response to the introduction of small tariffs or other distortions:*

---

[29]This has to do with the fact that individual household preferences across all countries are non-aggregable.

(i) changes in world welfare are given up to the second order by

$$\Delta \log \delta \approx \Delta \log W + Cov_{\chi_c^W} \left(\Delta \log \chi_c^W, \Delta \log P_{W_c} \right);$$

(ii) changes in country real expenditure or welfare are given up to the first order by

$$\Delta \log \delta_c \approx \Delta \log W_c \approx \Delta \log \chi_c^W - \Delta \log P_{W_c}.$$

The change in world welfare is the sum of the change in world real expenditure (output) and a redistributive term. The redistributive term is positive whenever the covariance between the changes in household income shares and the changes in consumption price deflators is positive. It captures a familiar deviation from perfect risk sharing. It would be zero if households could engage in perfect risk sharing before the introduction of the tariffs or other distortions. In our applications, this redistributive effect is quantitatively small and so changes in world welfare are approximately equal to changes in world real output.

6.4 Costs of Tariffs and Other Distortions: Ex-Ante Sufficient Statistics

Theorem 5 and Corollary 5 express the effects of tariffs and other distortions in terms of endogenous individual output changes up to the second order. In this subsection, we provide formulas for these individual output changes, and hence for the effects of tariffs and other distortions, in terms of primitives: microeconomic elasticities of substitution and the HAIO matrix.

Our results will make use of the following generalization of Theorem 4 (see Appendix F for more information). Changes in factor shares are given up to the first order by the system of linear equations

$$\Delta \log \Lambda_f \approx - \sum_{i \in N} \Gamma_{fi} \Delta \log \mu_i - \sum_{g \in F} \Gamma_{fg} \Delta \log \Lambda_g + \sum_{g \in F} \Xi_{fg} \Delta \log \Lambda_g$$
$$- \sum_{i \in N} \frac{\lambda_i}{\Lambda_f} \Psi_{if} \Delta \log \mu_i + \sum_{i \in N} \Xi_{fi} \Delta \log \mu_i,$$

where the definition of Ξ is extended for $f \in F$ and $i \in N$ by $\Xi_{fi} = \frac{1}{\Lambda_f} \sum_{c \in C} (\Lambda_f^{W_c} - \Lambda_f) \Phi_{ci} \lambda_i$, and Φ_{ci} is the share of the revenue raised by the tariff or other distortion on good i which accrues to country c. Changes in country income shares are given up to the

first order by

$$\chi_c^W \Delta \log \chi_c^W \approx \sum_{g \in F} \Phi_{cf} \Lambda_g \Delta \log \Lambda_g + \sum_{i \in N} \Phi_{ci} \lambda_i \Delta \log \mu_i.$$

Changes in country real expenditure deflators are given up to the first order by

$$\Delta \log P_{W_c} \approx \sum_{i \in N} \lambda_i^{W_c} \Delta \log \mu_i + \sum_{g \in F} \Lambda_g^{W_c} \Delta \log \Lambda_g.$$

In these equations, changes in tariffs and other distortions play three distinct roles. First, as described in Section 5, they act via prices like negative productivity shocks, and the changes in factor wages that they trigger also have similar effects. Second, they raise revenues. Third, for given sales and revenues, and for given factor wages, they reduce input (and hence ultimately factor) demand.

The system of linear equations for the changes in factor income shares $\Delta \log \Lambda_f$ is similar to the one that we encountered in Theorem 4, with two new terms on the second line, $-\sum_{i \in N} \frac{\lambda_i}{\Lambda_f} \Psi_{if} \Delta \log \mu_i$ representing reductions in factor demand, and $\sum_{i \in N} \Xi_{fi} \Delta \log \mu_i$ representing distribution effects arising from differential ownerships across households with different expenditure patterns of the revenues raised by the tariffs or other distortions. The formula for the changes in the country income shares $\chi_c^W \Delta \log \chi_c^W$ is also similar to the one that we encountered in Section 5, with a new second term $\sum_{i \in N} \Phi_{ci} \lambda_i \Delta \log \mu_i$ representing the part of the revenue raised by the tariff or other distortion on good j which accrues to country c. The formula for the changes in the real expenditure deflators $\Delta \log P_{W_c}$ is again similar to the one that we encountered in Section 5.

The main results of this section, Theorem 6 (real output) and Corollary 6 (welfare) below, build on these structural characterizations of changes in factor shares, country income shares, and country price deflators.

Theorem 6 (Structural Output Loss). *Around an efficient equilibrium, changes in world real output/expenditure in response to changes in tariffs or other distortions are given, up to the second order, by*

$$\Delta \log Y = \Delta \log W \approx -\frac{1}{2} \sum_{l \in N} \sum_{k \in N} \Delta \log \mu_k \Delta \log \mu_l \sum_{j \in N} \lambda_j \theta_j Cov_{\Omega^{(j)}}(\Psi_{(k)}, \Psi_{(l)})$$

$$-\frac{1}{2} \sum_{l \in N} \sum_{g \in F} \Delta \log \Lambda_g \Delta \log \mu_l \sum_{j \in N} \lambda_j \theta_j Cov_{\Omega^{(j)}}(\Psi_{(g)}, \Psi_{(l)})$$

$$+\frac{1}{2} \sum_{l \in N} \sum_{c \in C} \chi_c^W \Delta \log \chi_c^W \Delta \log \mu_l (\lambda_l^{W_c} - \lambda_l).$$

Changes $\Delta \log Y_c$ in the real output of country c are similar and in Appendix L.

First, all the terms scale with the square of the tariffs or other distortions $\Delta \log \mu$. There is therefore a sense in which misallocation increases with the tariffs and other distortions. Second, all the terms scale with the elasticities of substitution θ of the different producers. There is therefore a sense in which elasticities of substitution magnify the costs of these tariffs and other distortions. Third, all the terms also scale with the sales shares λ of the different producers and with the square of the Leontief inverse matrix Ψ. There is therefore also a sense in which accounting for intermediate inputs magnifies the costs of tariffs and other distortions. Fourth, all the terms mix the tariffs and other distortions, the elasticities of substitution, and of properties of the network. Hence, in general, the costs of tariffs and other distortions depends on how they are distributed over the network.

For a given producer $l \in N$, there are terms in $\Delta \log \mu_l$ on the three lines. Taken together, these terms sum up to the Harberger triangle $(1/2)\lambda_l \Delta \log \mu_l \Delta \log y_l$ corresponding to good l in terms of microeconomic primitives. The three lines break it down into three components, corresponding to three different effects responsible for the change in the quantity $\Delta \log y_l$ of good l.

The term $-(1/2) \sum_{k \in N} \Delta \log \mu_k \Delta \log \mu_l \sum_{j \in N} \lambda_j \theta_j Cov_{\Omega^{(j)}}(\Psi_{(k)}, \Psi_{(l)})$ on the first line corresponds to the change $\Delta \log y_l$ in the quantity of good l coming from *substitutions* by all producers j in response to changes in all tariffs and other distortions $\Delta \log \mu_k$, holding factor wages constant.

The term $-(1/2) \sum_{g \in F} \Delta \log \Lambda_g \Delta \log \mu_l \sum_{j \in N} \lambda_j \theta_j Cov_{\Omega^{(j)}}(\Psi_{(g)}, \Psi_{(l)})$ on the second line corresponds to the change $\Delta \log y_l$ in the quantity of good l coming from *substitutions* by all producers j in response to the endogenous changes in factor wages $\Delta \log w_g = \Delta \log \Lambda_g$ brought about by all the changes in tariffs and other distortions.

The term $\sum_{c \in C} \chi_c^W \Delta \log \chi_c^W \Delta \log \mu_l (\lambda_l^{W_c} - \lambda_l)$ on the third line corresponds to the change $\Delta \log y_l$ in the quantity of good l coming from *redistribution* across agents with different spending patterns, in response to the endogenous changes in factor wages brought about by all the changes in tariffs and other distortions.

Note that in contrast to the expressions for changes in sales shares, the two substitution terms feature, for each producer j, the elasticity of substitution θ_j and not $\theta_j - 1$. This is because they characterize the change in the *quantity* of good l, and not the change in its *sales*. The neutral case leading to no changes is no longer Cobb Douglas but Leontief: when all the elasticities of substitution are zero, there are no changes in the quantity of good l, and no effect of tariffs or other distortions on real output.

Corollary 6 (Structural Welfare). *Starting at an efficient equilibrium, changes in world and*

country welfare $\Delta \log \delta$ *and* $\Delta \log \delta_c \approx \Delta \log W_c$ *are given via Corollary 5, respectively up to the second order (world) and up to the first order (country).*
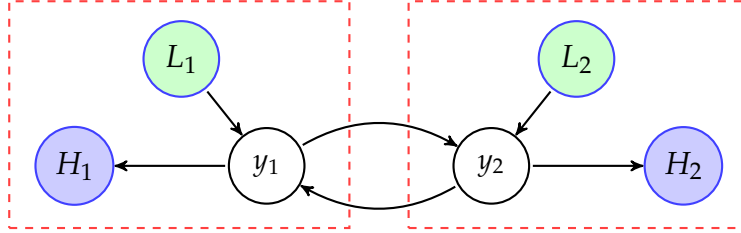
Example: Tariffs in a Round-About World Economy

Figure 2: The solid lines show the flow of goods. Green nodes are factors, purple nodes are households, and white nodes are goods. The boundaries of each country are denoted by dashed box.

To see intuitively why intermediates matter in Theorem 6, consider the example in Figure 2. Assume the two countries are symmetric. Let Ω be imports as a share of sales, and θ be the elasticity of substitution between intermediates and labor. Suppose that each country introduces a symmetric tax $\Delta \log \mu$ on its imports from the other country. Because of symmetry, changes in country real output, country welfare, world real output, and world welfare are all the same. Hence, using Theorem 6, the reduction of any of these variables in response to tariffs on any of these variables is given by

$$-\frac{1}{2}(\lambda_{12}\Delta \log y_{12}\Delta \log \mu + \lambda_{21}\Delta \log y_{21}\Delta \log \mu) = \theta\frac{\Omega}{2(1-\Omega)^2}(\Delta \log \mu)^2,$$

where y_{12} is the quantity of imports from country 2 by country 1, λ_{12} is the corresponding sales share, and y_{21} and λ_{21} are defined similarly. Because of symmetry $y_{12} = y_{21}$ and $\lambda_{12} = \lambda_{21}$.

The losses increase with the elasticity of substitution θ and with the intermediate input share Ω. This is both because the relevant sales shares $\lambda_{12} = \lambda_{21} = \Omega/[2(1-\Omega)]$ become larger and because the reductions in the quantities of imports $-\Delta \log y_{12} = -\Delta \log y_{21} = [\theta/(1-\Omega)]\Delta \log \mu$ become larger. The latter effect occurs because when the intermediate input share increases, goods effectively cross borders more times, and hence get hit by the tariffs more times, which increases the relative price of imports more and leads to a larger reduction in their quantity.

7 Quantitative Example

In this section, we use a multi-factor production network model calibrated to fit world input-output data. We quantify the way increasing trade costs (tariffs or iceberg) affect output, welfare, and factor rewards, and use our analytical results to give intuition for our findings.

Benchmark Calibration with Input-Output Linkages

The benchmark model consists of 40 countries as well as a "rest-of-the-world" composite country, each with four factors of production: high-skilled, medium-skilled, low-skilled labor, and capital. Each country has 30 industries each of which produces a single industry good.

The model has a nested-CES structure. Each industry produces output by combining its value-added (consisting of the four domestic factors) with intermediate goods (consisting of the 30 industries). The elasticity of substitution across primary factors is θ_3, and across intermediates is θ_1, the elasticity of substitution between factors and intermediate inputs is θ_2, the elasticity of substitution in consumption goods across industries is θ_0. When a producer or the household i in country c purchases inputs of some industry j, it consumes a CES aggregate of goods from this industry sourced from various countries with some trade elasticity ε_j. We use data from the WIOD to calibrate the CES share parameters to match expenditure shares in the year 2008.

To calibrate the elasticities of substitution, we use the same elasticities as in Section 4. That is, we set the elasticity of substitution across industries $\theta_1 = 0.2$, the one between value-added and intermediates $\theta_2 = 0.5$, the elasticity of substitution in consumption $\theta_0 = 0.9$, and the trade elasticities ε_j following the estimates from Caliendo and Parro (2015). Finally, we set the elasticity of substitution among primary factors $\theta_3 = 0.5$. See Appendix A for more details.

Value-Added Calibrations

The benchmark model features input-output linkages. To emphasize the importance of taking input-output linkages into account and to connect with our analytical results, we compare the benchmark model with input-output linkages to two alternative value-added calibrations which are common in the trade literature. These alternative calibrations both assume that all production takes place with value-added production functions (no intermediates) but trivialize the input-output connections in two different ways. We

call these two calibrations the low-trade value-added (LVA) and the high-trade value-added (HVA) economies. As we shall see, these two value-added calibrations are problematic, because they are not exact representations of the benchmark economy.[30]

Low-trade value-added (LVA): the value-added produced by each producer matches the one in the data. It is then assumed that the fraction of the value added of i which is sold to each country is equal to the corresponding fraction of the sales of i in the data. This calibration matches trade a share of total sales, and therefore, it lowers the volume of trade relative to GDP. We call this the low-trade value-added economy for this reason. This is the procedure used by Costinot and Rodriguez-Clare (2014) in the handbook chapter for their value-added calibration, and it is also how ACR mapped their model to the data.

High-trade value-added (HVA): the value-added produced by each producer matches the one in the data. However, it is then assumed that the value of i which is sold to each foreign country is equal to the corresponding sales of i in the data. The residual value-added is sold in the domestic country of i. This calibration preserves trade as a share of GDP, so we call this the high-trade value-added economy.

Results

In Table 2, we report the impact on the welfare of a few countries, as well as the effect on world welfare, of a universal 10% increase in either the iceberg costs of trade or a 10% increase in import tariffs. We compare the response of our benchmark economy to those of the LVA and HVA economies (which do not have intermediate goods).

Across the board, and as suggested by the discussion of trapezoids versus triangles in Section 6.3, an increase in iceberg trade costs (or other non-tariff barriers to trade) is significantly more costly than an equivalent increase in tariffs. For example, US welfare actually increases by 0.1% in response to increases in tariffs, but decreases by 1.1% in response to increases in trade costs. World welfare decreases by 0.4% in response to in-

---

[30]The only correct way of representing this economy with intermediate inputs as a value-added economy is to follow Adao et al. (2017) by assuming that only factors are traded within and across borders, and that households have preferences over factors. For some welfare counterfactuals, this representation can be put to use by directly specifying and estimating a parsimoniously-parameterized factor demand system. This parsimony advantage must of course be traded of against the cost of misspecification, and, as our formulas in Section 5 make clear, the "true" functional form of the factor demand system is likely to be complex for realistic economies. Furthermore, using this approach is more difficult in the presence of intermediate goods, trade costs, and tariffs. For example, shocks to iceberg costs of trading goods between two countries must first be translated into shocks to costs of trading the factors used directly and indirectly to produce them, which must be handled by expanding the set of factors to reflect the fact that the same factor may be crossing boundaries a different number of times along different global supply chains. These difficulties are compounded in the presence of tariffs. In Appendix E, we use Theorem 4 to explicitly characterize the factor-demand system.

	AUS	BEL	BGR	CAN	CHN	DEU	IRL	LUX	USA	World
Benchmark Iceberg	-2.0%	-6.7%	-5.9%	-3.3%	-1.3%	-3.0%	-8.3%	-16.9%	-1.1%	-2.26%
HVA Iceberg	-2.1%	-5.2%	-5.5%	-3.9%	-1.5%	-3.4%	-7.1%	-10.8%	-1.0%	-2.34%
LVA Iceberg	-0.8%	-2.3%	-1.9%	-1.4%	-0.5%	-1.2%	-3.6%	-5.0%	-0.3%	-0.86%
Benchmark Tariff	-0.7%	-2.3%	-0.7%	-1.2%	-0.2%	-0.9%	-3.7%	-5.4%	0.1%	-0.43%
HVA Tariff	-0.5%	-0.6%	-0.1%	-1.5%	-0.1%	-0.7%	-2.1%	-5.3%	0.3%	-0.23%
LVA Tariff	-0.4%	-0.3%	0.3%	-0.7%	0.1%	-0.5%	-1.3%	-0.7%	0.1%	-0.17%

Table 2: Percentage change in real income for a subset of the countries in response to a universal 10% change in iceberg trade costs or import tariffs. We compare the results from the benchmark economy with intermediate goods and input-output linkages with economies that assume only value-added production functions (HVA and LVA).

creases in tariffs, but decreases by 2.3% in response to increases in trade costs. Hence, drawing inferences about increases in tariffs by studying increases in iceberg trade costs, as sometimes happens in the literature, can be highly misleading.

Next, we compare the benchmark model to the value-added economies. The reduction in world welfare from increases in iceberg trade costs is 2.3% for the benchmark economy, it is also 2.3% for the HVA economy, but it is only 0.9% for the LVA economy. The HVA economy does a better job than the LVA economy because it preserves the volume of trade, and hence, by Theorems 1 and 2, the response of world welfare in that model is, at the first order, identical to that of the benchmark model.[31] The response of country welfare is different at the first order, but for the shock that we consider, these differences seem to be relatively small for most (but not all) countries. The LVA economy, which is the much more common calibration in the literature, is hopeless. Since LVA reduces the volume of trade to GDP, it greatly understates, at the first order, the welfare effects of shocks to iceberg trade costs.

The reduction in world welfare from an increase in tariffs is 0.43% for the benchmark economy, but it is only 0.23% for the HVA economy, and it is even less at 0.17% for LVA economy. In this case, neither the LVA nor the HVA economy does a good job of replicating the benchmark model.

Theorem 5 helps explain why: the losses from tariffs are given by $(1/2) \sum_i \lambda_i \Delta \log y_i \times \log \mu_i$, where λ_i is the sales share, $\log y_i$ is the quantity, and $\log \mu_i$ is the (gross) tax for good i. Since the HVA economy preserves the volume of trade, λ_i are the same for the benchmark and the HVA economy. Nevertheless, the response of the HVA economy is half that of the benchmark. This is because in the HVA economy, the reduction in export

---

[31]That means, as long as the shocks are sufficiently small (ruling out nonlinearities), we should expect the benchmark and HVA economies to deliver similar welfare results for the world as a whole.

quantities $\Delta \log y_i$ in response to tariffs is significantly lower. The LVA economy is still hopeless, since it gets both the output elasticity $\Delta \log y_i$ wrong and the trade volumes λ_i wrong.

There are two reasons why the HVA economy underestimates the reduction in exports $\Delta \log y_i$ in response to tariffs. First, the impact of the tariffs on the price of traded goods is smaller. In the true economy, the universal tariff affects global value chains, and the tariffs are compounded each time unfinished goods cross borders, as in the round-about example of Section 6.4. This effect is necessarily absent in HVA where unfinished goods only cross borders once.

The second reason is more subtle. In the HVA economy, imported goods are a larger share of each agent's basket. Intuitively, the higher is the share of imports in the consumer's consumption basket, the lower is the elasticity of that consumer's demand with respect to the tax. The reason is that increases in the price of imports increase the overall price index by more, and hence reduce substitution away from imports.

Conceptually, one way to separate the double-marginalization effect (where tariffs are paid multiple times) from the elasticity effect is to tax traded goods based only on the domestic content of their exports.[32] For the HVA and LVA economies, since each good crosses a border at most once, this would have no effect. However, for the benchmark economy, taxing only the domestic content of exported goods would reduce global output by -0.31% instead of the benchmark -0.43%. This suggests there is a significant degree of re-exporting in world trade, but since the loss of -0.31% is still higher than the -0.23% losses in the HVA, it suggests both the elasticity effect and the double marginalization effect are quantitatively important.

Intuition from a Round-About Economy

To see these two mechanisms in more detail, we formally work through the round-about economy depicted in Figure 2. In Section 6, we showed that the welfare and output losses from a universal increase in tariffs in that economy were given by

$$-\frac{1}{2}\sum_{i\neq j}\lambda_{ji}\Delta \log y_{ji} \times \Delta \log \mu_i = \lambda_{ji}\underbrace{\frac{\theta}{1-\Omega}\Delta \log \mu}_{-\Delta \log y_{ji}}\underbrace{\Delta \log \mu}_{\Delta \log \mu_i},$$

---

[32]Formally, for each traded good i produced in country c, define $\Omega_{ij}^c = \Omega_{ij}\mathbf{1}(i \in c)$ and $\Psi^c = (I - \Omega^c)^{-1}$. Let $\delta_i = \sum_{j\notin c}\Psi_{ij}^c$ and impose the tax $\mu_i = (1+0.10^{1-\delta_i})$ on each i. If the traded good i does not rely on foreign inputs in its supply chain, then $\delta_i = 0$, if the traded good i contains no domestic value-added (directly or indirectly) then $\delta_i = 1$.

with the sales share of each traded good given by $\lambda_{ji} = \Omega/[2(1-\Omega)]$, and where i and j index the origin and destination of the traded good, Ω is the share of traded goods in sales, and θ is the elasticity of substitution between traded and non-traded goods.

The expression above follows from the fact that

$$-\Delta \log y_{ji} = \theta \Delta \log p_{ji} = \theta \frac{1}{1-\Omega} \Delta \log \mu.$$

The term $1/(1-\Omega)$ captures the fact that global value chains amplify the effect of the tariff on the price — each time the good crosses the border, the tariff must be paid again.

Now, imagine that we take data from this economy and calibrate it using the HVA and LVA structures, keeping the elasticity of substitution between traded and non-traded goods (the trade elasticity) constant and equal to θ throughout.[33] The HVA economy has the same value $\lambda_{ji} = \Omega/[2(1-\Omega)]$ for the sales share of traded goods as the round-about economy, but the LVA economy has a lower value for $\lambda_{ji} = \Omega/2$.[34]

Both for the HVA and LVA economies, the reduction in the quantity of traded goods in response to tariffs is given by

$$-\Delta \log y_{ji} = \theta(\Delta \log p_{ji} - \Delta \log P_c) = \theta(1 - \lambda_{ji})\Delta \log \mu,$$

where P_c is the consumer price index in both countries. Combining these two facts, the welfare and output losses from tariffs are given by

$$-\frac{1}{2}\sum_{i \neq j} \lambda_{ji} \Delta \log y_{ji} \Delta \log \mu_i = \lambda_{ji} \underbrace{\theta(1 - \lambda_{ji})\Delta \log \mu}_{-\Delta \log y_{ji}} \underbrace{\Delta \log \mu}_{\Delta \log \mu_i}, \tag{3}$$

where $\lambda_{ji} = \Omega/[2(1-\Omega)]$ for the HVA economy and $\lambda_{ji} = \Omega/2$ for the LVA economy.

Since for a given tariff $\Delta \log \mu$, the loss is proportional to the product of λ_{ji} and $-\Delta \log y_{ji}$, the HVA and LVA economies will give the wrong answer to the extent that they fail to match the values of these two variables in the round-about economy.

By construction, the HVA economy matches λ_{ji}, whereas the LVA has a lower λ_{ji}. This leaves the change in the quantity $\Delta \log y_{ji}$. We already mentioned that the HVA economy underpredicts the reduction in imports for two different reasons. From these simple expressions, we can see both effects operating. First, holding fixed the consumer price

---

[33]In this context, we define the trade elasticity to be the elasticity of expenditures on foreign goods relative to domestic goods to an iceberg trade shock, holding factor and import prices (before the iceberg cost) constant. This is the notion used by Caliendo and Parro (2015) in the context of economies with IO linkages.

[34]This implies that to a first order, the welfare and output losses from increases in iceberg trade costs in the HVA economy are the same as for the round-about economy, but they are lower in the LVA economy.

index, the effect of the tariff on the price of the traded good in both the HVA and the LVA economy is 1 instead of $1/(1-\Omega)$ — since there are no longer global value chains amplifying the effect of the tariff on the price via multiple border crossings. Second, and more subtly, in the round-about economy, what matters is the change in the price of traded goods, whereas in the HVA and LVA economies, what matters is the change in the price of the traded good relative to the change in the consumer price index because the traded good competes with the domestic good in final consumption. This is the reason that in the value-added economy the tariff is multiplied by $(1-\lambda_{ji})$. Therefore, for LVA and HVA, the higher is the volume of trade λ_{ji}, the lower is the reduction in trade $-\Delta \log y_{ji}$ from the imposition of the tariff, because when λ_{ji} is high, the domestic price index moves by more in response to trade shocks, effectively reducing the elasticity of output with respect to the tax. Therefore, $-\Delta \log y_{ji}$ is lower in the HVA economy than in the LVA economy, and both are lower than in the round-about economy.

To separate these two effects, we neutralize the first one by considering taxing only the foreign content of imports. That is, instead of a $\Delta \log \mu$ tax, we impose a tax of $(1-\Omega)\Delta \log \mu$. The losses to welfare and output in the round-about economy are then $\theta \lambda_{ji}\Delta \log \mu$ instead of $\lambda_{ji}\theta/(1-\Omega)\Delta \log \mu$. In the HVA and LVA calibrations, the losses do not change and are still equal to $\lambda_{ji}\theta(1-\lambda_{ji})$, since in these economies exports are made from primary factors only. This shows that even when we only tax the foreign content of trade, removing the compounding effects of goods crossing borders multiple times, the HVA and LVA calibrations still undershoot the round-about model.

In Table 2, we saw that both the LVA and HVA economies undershoot the true effect of the tariffs by about the same amount. Equation (3) gives some intuition for why this happens: there, the losses (3) are proportional to $\lambda_{ji}(1-\lambda_{ji})$, so that when the volume of trade to GDP λ_{ji} is close to the empirically realistic value of a half, both the HVA and LVA give broadly similar results, and both undershoot the round-about economy.

8   Conclusion

This paper establishes a unified framework for studying output and welfare in efficient and distorted open-economies. We provide ex-post sufficient statistics for measurement and ex-ante sufficient statistics for conducting counterfactuals. Our formulas bring together results from the open and closed-economy literatures, and provide new characterizations of the gains from trade and the losses from trade protectionism. As discussed in the appendix, these results also have implications for the aggregation of trade elasticities, and the distributional consequences of trade policy.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review 107*(3), 633–89.

Allen, T., C. Arkolakis, and Y. Takahashi (2014). Universal gravity. *NBER Working Paper* (w20787).

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2015). The elusive pro-competitive effects of trade. *The Review of Economic Studies*.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Artuç, E., S. Chaudhuri, and J. McLaren (2010). Trade shocks and labor adjustment: A structural empirical approach. *American economic review 100*(3), 1008–45.

Atalay, E. (2017). How important are sectoral shocks? *American Economic Journal: Macroeconomics (Forthcoming)*.

Bai, Y., K. Jin, and D. Lu (2018). Misallocation under trade liberalization. *Unpublished Working Paper, University of Rochester*.

Baqaee, D. R. (2018). Cascading failures in production networks. *Econometrica (Forthcoming)*.

Baqaee, D. R. and E. Farhi (2017a). The macroeconomic impact of microeconomic shocks: Beyond Hulten's Theorem.

Baqaee, D. R. and E. Farhi (2017b). Productivity and Misallocation in General Equilibrium. NBER Working Papers 24007, National Bureau of Economic Research, Inc.

Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2019). The origins of firm heterogeneity: A production network approach. Technical report, National Bureau of Economic Research.

Berthou, A., J. J. Chung, K. Manova, and C. S. D. Bragard (2018). Productivity,(mis) allocation and trade. Technical report, Mimeo.

Boehm, C., A. Flaaen, and N. Pandalai-Nayar (2015). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake.

Borusyak, K. and X. Jaravel (2018). The distributional effects of trade: Theory and evidence from the united states.

Burstein, A. and J. Cravino (2015). Measured aggregate gains from international trade. *American Economic Journal: Macroeconomics 7*(2), 181–218.

Burstein, A. and J. Vogel (2017). International trade, technology, and the skill premium. *Journal of Political Economy 125*(5), 1356–1412.

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies 82*(1), 1–44.

Caliendo, L., F. Parro, and A. Tsyvinski (2017, April). Distortions and the structure of the world economy. Working Paper 23332, National Bureau of Economic Research.

Carvalho, V. M., M. Nirei, Y. Saito, and A. Tahbaz-Salehi (2016). Supply chain disruptions: Evidence from the great east japan earthquake. Technical report.

Chaney, T. (2014). The network structure of international trade. *American Economic Review 104*(11), 3600–3634.

Costinot, A. and A. Rodriguez-Clare (2014). Trade theory with numbers: quantifying the consequences of globalization. *Handbook of International Economics 4*, 197.

Davis, D. R. and D. E. Weinstein (2008). *The Factor Content of Trade*, Chapter 5, pp. 119–145. Wiley-Blackwell.

Deardorff, A. V. and R. W. Staiger (1988). An interpretation of the factor content of trade. *Journal of International Economics 24*(1-2), 93–107.

Dekle, R., J. Eaton, and S. Kortum (2008). Global rebalancing with gravity: measuring the burden of adjustment. *IMF Staff Papers 55*(3), 511–540.

Diewert, W. E. and C. J. Morrison (1985). Adjusting output and productivity indexes for changes in the terms of trade.

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica 82*(3), 825–885.

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

Domar, E. D. (1961). On the measurement of technological change. *The Economic Journal 71*(284), 709–729.

Elsby, M. W., B. Hobijn, and A. Şahin (2013). The decline of the us labor share. *Brookings Papers on Economic Activity 2013*(2), 1–63.

Epifani, P. and G. Gancia (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics 83*(1), 1–13.

Fally, T. and J. Sayre (2018). Commodity trade matters. Technical report, National Bureau of Economic Research.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, 157–177.

Feenstra, R. C. and A. Sasahara (2017). The 'china shock', exports and us employment: A global input-output analysis. Technical report, National Bureau of Economic Research.

Galle, S., A. Rodriguez-Clare, and M. Yi (2017). Slicing the pie: Quantifying the aggregate and distributional effects of trade. Technical report, National Bureau of Economic Research.

Harberger, A. C. (1964). The measurement of waste. *The American Economic Review 54*(3), 58–76.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013). Two perspectives on preferences and structural transformation. *American Economic Review 103*(7), 2752–89.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, 511–518.

Johnson, R. C. and G. Noguera (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of international Economics 86*(2), 224–236.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics*, 1–28.

Jones, R. W. (1965). The structure of simple general equilibrium models. *Journal of Political Economy 73*(6), 557–572.

Kehoe, T. J. and K. J. Ruhl (2008). Are shocks to the terms of trade shocks to productivity? *Review of Economic Dynamics 11*(4), 804–819.

Kikkawa, A. K., G. Magerman, E. Dhyne, et al. (2018). Imperfect competition in firm-to-firm trade. Technical report.

Kohli, U. (2004). Real gdp, real domestic income, and terms-of-trade changes. *Journal of International Economics 62*(1), 83–106.

Koopman, R., Z. Wang, and S.-J. Wei (2014). Tracing value-added and double counting in gross exports. *American Economic Review 104*(2), 459–94.

Lim, K. (2017). Firm-to-rm trade in sticky production networks.

Lind, N. and N. Ramondo (2018). Trade with correlation. Technical report, National Bureau of Economic Research.

Long, J. B. and C. I. Plosser (1983). Real business cycles. *The Journal of Political Economy*, 39–69.

McKenzie, L. W. (1959). On the existence of general equilibrium for a competitive market. *Econometrica: journal of the Econometric Society*, 54–71.

Morrow, P. M. and D. Trefler (2017). Endowments, skill-biased technology, and factor prices: A unified approach to trade. Technical report, National Bureau of Economic Research.

Oberfield, E. and D. Raval (2014). Micro data and macro technology. Technical report, National Bureau of Economic Research.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

Rybczynski, T. M. (1955). Factor endowment and relative commodity prices. *Economica 22*(88), 336–341.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312–320.

Stolper, W. F. and P. A. Samuelson (1941). Protection and real wages. *The Review of Economic Studies 9*(1), 58–73.

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*(3), 575–605.

Tintelnot, F., A. K. Kikkawa, M. Mogstad, and E. Dhyne (2018). Trade and domestic production networks. Technical report, National Bureau of Economic Research.

Trefler, D. and S. C. Zhu (2010). The structure of factor content predictions. *Journal of International Economics 82*(2), 195–207.

Viner, J. (1937). Studies in the theory of international trade.

Yi, K.-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of political Economy 111*(1), 52–102.