Ricardian Business Cycles*

Lorenzo Bretscher
HEC Lausanne, SFI, CEPR, and Enterprise for Society (E4S)†

Jesús Fernández-Villaverde
University of Pennsylvania, NBER, and CEPR ‡

Simon Scheidegger
HEC Lausanne and Enterprise for Society (E4S) §

November 16, 2022

Abstract

This paper presents a dynamic stochastic general equilibrium model of Ricardian business cycles. Our model is Ricardian because countries (or, equivalently, regions) trade to take advantage of their comparative advantages. Their relative efficiencies, however, change over time stochastically. Similarly, country-specific shocks to demand, supply, and investment efficiency induce countries to engage in intra- and intertemporal substitutions in non-durable consumption, investment, services, and trade, generating business cycles. Finally, all agents have rational expectations about the stochastic components of the model. We solve the model globally using deep neural networks and calibrate it to the U.S., Europe, and China. Our quantitative results highlight the role of trading costs in shaping the responses of the economy to different shocks.

JEL classification: C45, C63, F10, F40

Keywords: International Trade, Business Cycles, General Equilibrium, Comparative Advantage, Deep Learning.

*We thank Jonathan Eaton for early encouragement into this research project and Artem Kuriksha for superb research assistance. This work was supported by the Swiss National Science Foundation (SNF), under project IDs "Can Economic Policy Mitigate Climate-Change?", "New methods for asset pricing with frictions", and the Enterprise for Society (E4S) for research support, and the Swiss National Supercomputing Center (CSCS) under project ID 995.

†Department of Finance, Email: lorenzo.bretscher@unil.ch.

‡Department of Economics, Email: jesusfv@econ.upenn.edu.

§Department of Economics, Email: simon.scheidegger@unil.ch.

1 Introduction

This paper presents a dynamic stochastic general equilibrium model of Ricardian business cycles. Our model is Ricardian because countries (or, equivalently, economic regions) trade to take advantage of their comparative advantages: within each economic sector of a country, there are firms with high relative production efficiency, which export, and firms with low relative production efficiency, which either do not export (due to trade costs) or do not produce. The firms' relative efficiencies, however, change over time stochastically. Similarly, country-specific shocks to demand, supply, and trade costs induce countries to engage in intra- and intertemporal substitutions in non-durable consumption, investment, services, and trade, generating business cycles (notice that we do not impose balanced trade period by period, only that the intertemporal budget constraints of each country is satisfied). Finally, all agents have rational expectations about the stochastic components of the model and act accordingly.

Analyzing the interaction between time-varying comparative advantages and demand, supply, and trade cost shocks is key to understanding many questions in international trade and macroeconomics. Among many others: How do supply shocks change the composition between consumption and investment when each activity has different international trade structures? How do trade patterns react to demand shocks such as those triggered by fiscal and monetary policies? What are the global effects of country-specific shocks? What is the role of trade costs in the transmission of shocks? What are the effects of uncertainty shocks on world output and trade?

However, relatively little progress has been made in answering these questions quantitatively because considering both intra- and intertemporal aspects of international trade in a stochastic environment is challenging.[1] First, we face a problem with high dimensionality. Even considering just three countries, our model has 32 states. Furthermore, the states are random, there are non-linearities in the equilibrium functions, and the ergodic sets of states present irregular geometries. Thus, this is an environment where classical solution methods, such as perturbations and projections, are likely to fail or be of limited accuracy.

[1] There has been some important work using perfect-foresight equilibria, such as in Eaton et al. (2016a) and Álvarez and Lucas (2007), but not with stochastic shocks. We will return to these papers below.

To get around this challenge, we use deep equilibrium nets (Azinovic et al., 2022), a deep-learning-based method to compute the recursive equilibrium of the model as characterized by the associated social planner's problem of the world economy. Nothing essential depends on approximating the social planner's problem instead of the competitive equilibrium. At the cost of some extra notation, we could solve directly for the competitive equilibrium.

Intuitively, deep neural networks are a series of geometric operations (linear translations and transformations by activation functions) that move the functional approximation problem from its original formulation into a more convenient geometric space (Bronstein et al., 2021). By doing so, deep neural networks can tame the "curse of dimensionality" and approximate functions arbitrarily well, even in the most challenging scenarios that defeat alternative methods. Furthermore, deep neural networks are easy to implement using standard libraries such as (in this paper) `TensorFlow` and easily amenable to multicore programming and co-processor acceleration (Cheela et al., 2022).

We calibrate our model to three regions: the U.S., China, and the Eurozone (which, for convenience, we will often refer to as "Europe"). Jointly, these three regions produce nearly 50% of the world's output. We will study two alternative calibrations. In the first calibration, trade costs are symmetric across all three regions. This calibration captures the behavior of a highly integrated world economy. In the second calibration, trade costs are asymmetric, with China nearly excluded from trade with the U.S. and Europe due to high trade costs. This calibration aims at simulating an environment where China disengages from the world economy, a possibility suggested by current events (trade war between the U.S. and China, China's zero COVID policy, etc.). However, in general, this calibration allows us to understand how asymmetries in the world economy operate.

Our main findings when trade costs are symmetric are as follows. First, we show how a productivity shock in one region has asymmetric effects on other regions, even though trade costs are symmetric. In particular, a shock to productivity in the U.S. has a more persistent effect in the Eurozone than in China. The reason is that, since the internal shocks in the Eurozone are more persistent than in China, the representative household in the Eurozone also responds differently to outside shocks: an outside shock today will be followed, in distribution, by some inside shock tomorrow. In comparison, a shock to productivity in China has roughly the same effects in the U.S. and the Eurozone.

Second, a shock that shifts relative world demand toward the U.S. increases output in the U.S., but it reduces output in China and the Eurozone.[2] Interestingly, under the symmetric calibration, this shock reduces total world output. This somewhat surprising result has, however, a simple explanation. In our calibration, China is the economy that can produce the cheapest goods due to its large population that drives local wages down. By shifting the world demand toward the U.S. (and given positive trade costs), production in the world economy becomes, on average, more expensive.

Third, a shock to one sector in the U.S. lowers output in the U.S. (as it makes it harder to equate production across different sectors), but increases output in China and the Eurozone, as these two regions can export more to the U.S.

The main findings change when trade costs become asymmetric. For example, an increase in productivity in the U.S. still expands output in the Eurozone (in fact, even more than with symmetric trade costs), but the effect is minimal in China, which is largely isolated from the rest of the world. Conversely, a productivity shock in China has a large effect on Chinese output but a very small effect on the U.S. and European GDP. We document similar results for other shocks in this calibration. In other words: trade costs matter for the transmission of shocks in the world economy.

Beyond our quantitative results, our methodological contribution opens the door to a much larger set of quantitative experiments than those we can report in this paper due to space constraints. The use of deep neural networks allows researchers to tackle a wide set of questions in international economics that were, so far, hard to analyze. Additionally, our methodology opens the door to the validation of Ricardian models of trade in terms of their responses to different shocks; again, a task that is beyond the scope of our paper.

Our paper is related most closely to Eaton et al. (2016a), who build a Ricardian model in the tradition of Eaton and Kortum (2002) (see also Eaton et al., 2016b). Our main difference is the use of rational expectations instead of perfect foresight. By doing so, we can document, for instance, the important role of the persistence of inside shocks in gauging the effect of an outside shock. We are also close to Álvarez and Lucas (2007), who analyze the determinants of the cross-country distribution of trade volumes in a general equilibrium version of Eaton

[2] One can interpret this shock as an increase in U.S. demand, for example, triggered by expansionary fiscal or monetary policies that we do not model explicitly.

and Kortum (2002) with balanced trade and deterministic transition dynamics. Caliendo et al. (2019) use a related environment to study how dynamic trade interacts with spatially distinct labor markets facing varying exposure to international trade. Ravikumar et al. (2019) compute welfare gains from trade in a dynamic, multicountry model with capital accumulation and trade imbalances by computing the exact transition paths following trade liberalization.

The models in Eaton et al. (2016a), Álvarez and Lucas (2007), and in our paper depart from older literature that relies on dynamic two-country models. Important milestones in this literature include Stockman and Tesar (1995) and Alessandria et al. (2010), among many others. One key advantage of having more than two countries is that we can explore the asymmetric responses to the same shocks across different regions. As we document in the paper, these asymmetric responses tell us much about the dynamics of the world economy. For instance, the effects of a U.S. domestic shock in Europe are very different than in China.

In terms of business cycles, we follow the tradition of models of international business cycles driven by real shocks and where countries exchange goods (Backus et al., 2020). Our main advance is the focus on multiple countries and the modeling of Ricardian comparative advantage. We also connect with the literature on uncertainty shocks in international business cycles. Although our model does not have uncertainty shocks, we could extend it to incorporate them, and the results in Fernández-Villaverde et al. (2011) suggest that uncertainty, in itself, is an important driver of the international transmission of shocks.

We are also related to a booming literature that shows how to use deep learning to solve highly-dimensional models in economics. Among others, see Duffy and McNelis (2001), Norets (2012), Maliar et al. (2021), Duarte (2018), Fernández-Villaverde et al. (2019), Villa and Valaitis (2019), Han et al. (2021), and Ebrahimi Kahou et al. (2021). Our main methodological contribution is to solve a "real-life" application with as many as 32 economic state variables, not a "proof-of-concept" scenario.

The remainder of this article is organized as follows. In Section 2, we outline the model. In Section 3, we briefly describe our deep-learning-based global solution method to tackle the trade model. Thereafter, Section 4 details the calibration of the model, whereas Sections 5 and 6 present the results of our numerical experiments. Section 7 concludes. An Appendix presents further details about the model and its implementation.

2 A model of Ricardian Business Cycles

In this section, we present a dynamic stochastic general equilibrium model of Ricardian business cycles. We build on previous work by Eaton et al. (2016a) and use a similar framework. Our main point of departure with respect to the existing literature is that we endow our agents with rational expectations about the random demand and supply shocks that hit the economy instead of staying with the assumption of perfect foresight.

Our model works in discrete time and features an arbitrary number of countries or regions. In what follows, we use "regions" and "countries" interchangeably. The key goal is to capture the notion that economic activity is located in different locations in space. We index countries by $n = 1, \ldots, \mathcal{N}$. In each country, there is a representative household and three sectors: a non-durable sector (N), a service sector (S), and a durable sector (D). We understand the output of the durable sector as a good that can be added to the stock of (physical) capital, either in the home country or abroad. Capital itself can be used to provide durable goods services (e.g., cars, dwellings) or be employed as input in the production of other goods and services (e.g., delivery vehicles, office space). For later convenience, let $\Omega = \{N, S, D\}$ denote the set of these three sectors.[3]

The economy will be hit by demand, supply, and trade shocks. We model demand shocks as i) shocks to the discount factor of the representative household in each country and ii) shocks to the relative weight of tradable vs. non-tradable goods in the preferences of the representative household within a country. We model supply shocks as i) shocks to the productivity of firms in each sector and ii) a shock to investment. We model trade shocks as shocks to the iceberg transportation costs of moving goods across countries. These five shocks will give us a wide range of aggregate dynamics and time-varying patterns of trade.

2.1 Preferences

In each country n, there is a representative consumer with a weight ω_n in the world economy, which can be read as a relative population size.

[3] Our methodology is very general and, in principle, allows us to incorporate many sectors within each country. However, three sectors strike a good balance between the richness of mechanisms at work and transparency.

At each date t, this representative household consumes the final output of the non-durable and services sectors in amounts $C^N_{n,t}$ and $C^S_{n,t}$, respectively. The household also consumes the services yielded by the stock of capital in the amount $K^{H,D}_{n,t}$. The utility function aggregates these flows of consumption with positive Cobb-Douglas weights, where $\psi^N_{n,t} + \psi^S_{n,t} + \psi^D_n = 1$. Thus, the lifetime utility of the representative household in country n is given by:

$$U_{n,0} = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \rho^t \phi_{n,t} \left(\psi^N_{n,t} \ln C^N_{n,t} + \psi^S_{n,t} \ln C^S_{n,t} + \psi^D_n \ln K^{H,D}_{n,t} \right) \right], \tag{1}$$

where $\mathbb{E}_0$ is the expectation operator as of time 0, $\rho < 1$ is a constant discount factor, and $\phi_{n,t}$ is a shock to inter-temporal preferences for the representative household in country n at date t, which we interpret as a stand-in aggregate demand shock encoding more complex shocks that we do not model explicitly (e.g., fiscal and monetary policy shocks, demographic shocks, financial frictions shocks).

We restrict aggregate demand shocks to have no global component. In particular, for $\mathcal{N} - 1$ countries $\phi_{n,t}$ evolves as:

$$\ln \phi_{n,t} = \rho_{\phi_n} \ln \phi_{n,t-1} + \varepsilon_{\phi_n,t},$$

where $\varepsilon_{\phi_n,t} \sim N(0, \sigma_{\phi,n})$. Moreover, for the $\mathcal{N}$-th country $\phi_{N,t}$ is determined as:

$$\phi_{\mathcal{N},t} = \mathcal{N} - \sum_{n=1}^{\mathcal{N}-1} \omega_n \phi_{n,t}.$$

In such a way, we analyze differential variations in aggregate demand across countries, instead of focusing on worldwide changes to aggregate demand (although incorporating the latter into our model would be straightforward). Additionally, the calibrated value of $\sigma_{\phi,n}$ will be small enough that in the simulation, we do not need to worry about the possibility that $\phi_{\mathcal{N},t} < 0$.

Furthermore, the stochastic Cobb-Douglas weights on non-durables and services allow for country-specific shifts between these two sectors over time. Since ψ^D_n is a constant parameter and $\psi^N_{n,t} + \psi^S_{n,t} + \psi^D_n = 1$ needs to hold, specifying a law of motion for either $\psi^N_{n,t}$

or $\psi^S_{n,t}$ is sufficient. Thus, we assume that $\psi^N_{n,t}$ follows:

$$\psi^N_{n,t} = \left(1 - \rho_{\psi_n}\right) \frac{1 - \psi^D_n}{2} + \rho_{\psi_n} \psi^N_{n,t-1} + \varepsilon_{\psi_n,t},$$

where $\varepsilon_{\psi_n,t} \sim N(0, \sigma_{\psi_n})$. We will also calibrate the standard deviation σ_{ψ_n} to be small enough for $\psi^N_{n,t}$ to always be in the interval $[0, 1 - \psi^D_n]$ in our simulations.

Two points are worth mentioning. First, the utility function (1) includes a rational expectation $\mathbb{E}_0$ because we have random shocks in the economy. While accounting for the expectation operator is computationally non-trivial, algebraically[4] , the equilibrium conditions of the model do not change much. In fact, only the investment Euler equation changes (see equation (9) below). Second, we allow for country-specific shifts between non-durables and services over time, but we treat the weights on household capital services as fixed across countries and over time. As we will see below, the goods in the non-durable sector are tradable, but not those in the services sector. Thus, we can trace the consequences of changing preferences (or policies) between tradable and non-tradable goods within a country and the spillovers of these changes to other countries.

2.2 Technology

At period t, country n has a labor endowment L_n and a stock $K^D_{n,t}$ of capital. Following the tradition of trade models since Ricardo (1821), these endowments of labor and capital are not traded internationally.

There is a perfectly competitive firm in each sector $j \in \Omega$ that produces a final (sector) output by aggregating a continuum of sector-specific goods using a CES production function with an elasticity of substitution σ:

$$x^j_{n,t} = \left(\int_0^1 x^j_{n,t}(z)^{(\sigma-1)/\sigma} dz \right)^{\sigma/(\sigma-1)},$$

where $x^j_{n,t}(z)$ can either be produced locally or imported.

The good z in sector j in country n is produced by a perfectly competitive firm that

[4] Appendix B.3 introduces a Deep Equilibrium Net based method on how to compute the high-dimensional integrals occurring in our model.

combines the services of labor, $L_n^j(z)$, the services of capital, $K_{n,t}^{jD}(z)$, and –as intermediates– the final output from each of the three sectors. We call $M_{n,t}^{jj'}(z)$ the intermediates from sector j' used to make good z in sector j. When no ambiguity occurs, we use z to denote both the firm and the good.

More concretely, firm z in sector j in country n at period t has access to a Cobb-Douglas function with constant returns to scale:

$$y_{n,t}^j(z) = a_{n,t}^j(z) \left(\frac{L_n^j(z)}{\beta_n^{L,j}} \right)^{\beta_n^{L,j}} \left(\frac{K_{n,t}^{jD}(z)}{\beta_n^{K,jD}} \right)^{\beta_n^{K,jD}} \prod_{j' \in \Omega} \left(\frac{M_{n,t}^{jj'}(z)}{\beta_n^{M,jj'}} \right)^{\beta_n^{M,jj'}}.$$

The output elasticities in country n and sector j of labor, capital, and intermediates from sector j' are given by $\beta_n^{L,j}$, $\beta_n^{K,jD}$, and $\beta_n^{M,jj'}$ for $j, j' \in \Omega$.

The efficiency of the firm, $a_{n,t}^j(z)$, is the realization of a random variable $a_{n,t}^j$ with a distribution:

$$F_{n,t}^j(a) = \Pr\left[a_{n,t}^j \leq a \right] = \exp\left[-\left(\frac{a}{\gamma A_{n,t}^j} \right)^{-\theta} \right], \tag{2}$$

which is drawn independently for each j, z, and t across countries n. Here, $A_{n,t}^j > 0$ reflects country n's overall productivity in sector j. In turn, $A_{n,t}^j = (1/\gamma) \left(T_{n,t}^j \right)^{1/\theta}$, where:

$$\ln T_{n,t}^j = \rho_{Tjn} \ln T_{n,t-1}^j + \varepsilon_{T_n^j,t}$$

and $\varepsilon_{Tjn,t} \sim N(0, \sigma_{T_n^j})$. The variable $T_{n,t}^j$ captures temporary differences in the comparative advantage of countries in each sector, with possible differences in persistence and volatility across sectors.

The parameter θ is an inverse measure of the dispersion of efficiencies, and γ is related to the gamma function:

$$\gamma = \left[\Gamma\left(\frac{\theta - \sigma + 1}{\theta} \right) \right]^{-1/(\sigma - 1)}.$$

The goods z from the non-durable and durable sectors can be traded internationally.

Thus, we call these two sectors the tradable sectors, $j \in \{N, D\}$. This trade incurs standard (finite) iceberg trade costs so that delivering one unit of a good from country i to country n requires shipping $d^j_{ni,t} \geq 1$ units. Services cannot be traded, which implicitly means that corresponding iceberg costs, $d^S_{ni,t}$, are infinite for $i \neq n$. The iceberg trade costs of sector $l \in \{D, N\}$ for $i \neq n$ evolve as:

$$\ln d^l_{ni,t} = \left(1 - \rho_{d^l_{ni}}\right) \bar{d}^l_{ni} + \rho_{d^l_{ni}} \ln d^l_{ni,t-1} + \varepsilon_{d^l_{ni},t}$$

where $\varepsilon_{d^l_{ni},t} \sim N(0, \sigma_{d^l_{ni}})$ and for $i = n$, we set

$$\ln d^l_{nn,t} = 0, \tag{3}$$

that is, there is no iceberg cost of trading with itself. Given our calibrated $\bar{d}^l_{ni}$ and $\sigma_{d^l_{ni}}$, in our simulations we will have that, indeed, $d^j_{ni,t} \geq 1$ holds. Additionally, notice that we allow the trade costs to be non-symmetric (i.e., $d^j_{ni,t} \neq d^j_{in,t}$). This non-symmetry might reflect, for instance, different tariffs across countries.

The final output of the non-durable and service sectors can be consumed by the representative household or used as an intermediate input. The output of the durable sector can be used as investment $I^D_{n,t}$ to accumulate capital and as an intermediate input. In particular, the law of motion for capital is $K^D_{n,t+1} = \chi^D_{n,t} \left(I^D_{n,t}\right)^{\alpha^D} \left(K^D_{n,t}\right)^{1-\alpha^D} + \left(1 - \delta^D\right) K^D_{n,t}$, where $\chi^D_{n,t}$ is an investment-specific technological shock. This shock evolves as:

$$\ln \chi^D_{n,t} = \rho_{\chi^D_n} \ln \chi^D_{n,t-1} + \varepsilon_{\chi^D_n,t},$$

where $\varepsilon_{\chi^D_n,t} \sim N(0, \sigma_{\chi^D_n})$.

While, in the interest of space, we are not explicit about the intertemporal budget constraint of each country (and the corresponding transversality condition), it is important to notice that countries can engage in intertemporal trade. In other words, we are not enforcing balanced trade in each period.

2.3 Equilibrium Relationships

In what follows, we briefly discuss the solution of the model and associated equilibrium conditions. In addition, Appendix A contains more detailed derivations of the model solution.

2.3.1 Prices and Trade Shares

Given a wage $w_{n,t}$, a rental rate of capital, $r_{n,t}$, and prices for intermediaries $p^j_{n,t}$, the cost $c^j_{n,t}$ of a bundle of inputs in country n for producing good z in sector j is:

$$c^j_{n,t} = (w_{n,t})^{\beta^{L,j}_n} (r_{n,t})^{\beta^{K,jD}_n} \prod_{j' \in \Omega} \left(p^{j'}_{n,t}\right)^{\beta^{M,jj'}_n} . \tag{4}$$

The associated price index for sector j in country n, combining production costs in each country equals:

$$p^j_{n,t} = \left[\sum_{i=1}^{N} \left(\frac{c^j_{i,t} d^j_{ni,t}}{A^j_{i,t}} \right)^{-\theta} \right]^{-1/\theta} . \tag{5}$$

The share of country n's absorption of sector j imported from country i reads:

$$\pi^j_{ni,t} = \left(\frac{c^j_{i,t} d^j_{ni,t}}{A^j_{i,t} p^j_{n,t}} \right)^{-\theta} . \tag{6}$$

2.3.2 Household Spending

The household spending on consumption of good $h \in \{N, S\}$ is:

$$p^h_{n,t} C^h_{n,t} = \omega_n \phi_{n,t} \psi^h_{n,t}, \tag{7}$$

whereas the household spending on renting capital k is:

$$r_{n,t} K^{H,D}_{n,t} = \omega_n \phi_{n,t} \psi^D_n . \tag{8}$$

Summing these two expressions across all sectors and countries, our restriction on global aggregate demand shocks, together with our normalization of the ψ's, implies that the value of world consumption is 1, which serves as our numéraire.

2.3.3 Investment

The investment in capital satisfies the Euler equation:

$$\frac{p^D_{n,t}}{\chi^D_{n,t}}\left(\frac{I^D_{n,t}}{K^D_{n,t}}\right)^{1-\alpha^D} = \rho\alpha^D\mathbb{E}_t\left[r_{n,t+1} + \frac{\left(1-\alpha^D\right)p^D_{n,t+1}I^D_{n,t+1}}{\alpha^D K^D_{n,t+1}} + \frac{\left(1-\delta^D\right)p^D_{n,t+1}}{\alpha^D\chi^D_{n,t+1}}\left(\frac{I^D_{n,t+1}}{K^D_{n,t+1}}\right)^{1-\alpha^D}\right]. \quad (9)$$

The left-hand side in (9) is the sacrifice in period t required to attain another unit capital in period $t+1$. The right-hand side is the expected benefit of another unit of capital in period $t+1$. Notice the role of the investment-specific shock, $\chi^D_{n,t}$, in moving the left- and right-hand side of this equation.

2.3.4 Market Clearing

Next, we define the value of country n's spending on sector j as $X^j_{n,t} = p^j_{n,t}x^j_{n,t}$. Defining $Y^j_{n,t}$ as the value of country n's gross production in sector j, world goods-market clearing implies that:

$$Y^j_{n,t} = \sum_{m=1}^{\mathcal{N}}\pi^j_{mn,t}X^j_{m,t}. \quad (10)$$

We denote final spending on sector h in country n as $X^{F,h}_{n,t} = p^h_{n,t}C^h_{n,t}$ for $h \in \{N,S\}$ and $X^{F,D}_{n,t} = p^D_{n,t}I^D_{n,t}$ for sector D. Total spending on sector j output is the sum of country n's final spending on sector j plus the use of sector j output as intermediates by each sector j', i.e.,

$$X^j_{n,t} = X^{F,j}_{n,t} + \sum_{j'\in\Omega}\beta^{M,j'j}Y^{j'}_{n,t}. \quad (11)$$

Market clearing for country n's labor implies that labor income equals labor demand

across sectors:

$$w_{n,t}L_n = \sum_{j\in\Omega} \beta_n^{L,j} Y_{n,t}^j, \tag{12}$$

whereas clearing in the market for capital implies that:

$$r_{n,t}K_{n,t}^D = \sum_{j\in\Omega} \beta^{K,jD} Y_{n,t}^j + \frac{\psi_n^D}{1-\psi_n^D}\left(X_{n,t}^{F,N} + X_{n,t}^{F,S}\right). \tag{13}$$

The equations in Section 2.3 determine paths of the endogenous variables, which include wages $w_{n,t}$, rental rates $r_{n,t}$, trade shares $\pi_{ni,t}^l$ for sectors $l \in \{N, D\}$, prices $p_{n,t}^j$, total spending $X_{n,t}^j$, final spending $X_{n,t}^{F,j}$, and output $Y_{n,t}^j$ for sectors $j \in \Omega$. We write the law of motion for the capital stocks $K_{n,t}^D$ as:

$$K_{n,t+1}^D = \chi_{n,t}^D \left(X_{n,t}^{F,D}/p_{n,t}^D\right)^{\alpha^D} \left(K_{n,t}^D\right)^{1-\alpha^D} + \left(1-\delta^D\right) K_{n,t}^D, \tag{14}$$

where $I_{n,t}^D = X_{n,t}^{F,D}/p_{n,t}^D$.

2.3.5 Competitive Equilibrium

A competitive equilibrium is characterized by a collection of prices, $\{(p_{n,t}^j, w_{n,t}, r_{n,t})_{n=1}^{\mathcal{N}}\}_{t=0}^{\infty}$ for $j \in \Omega$, and controls $\{(C_{n,t}^h, K_{n,t}^{H,D}, I_{n,t}^D)_{n=1}^{\mathcal{N}}\}_{t=0}^{\infty}$ for $h \in \{N, D\}$, such that equations (4)-(9) hold and goods, labor, and capital markets clear, i.e., equations (10)-(14) are satisfied.

2.4 The Social Planner's Problem

Since we have assumed that markets are complete and perfectly competitive, we can follow Lucas and Prescott (1971) and solve for the market allocation implied by the previous definition of competitive equilibrium as the solution to a world planner's problem. The social planner's problem also makes some of the aggregate constraints more transparent and, therefore, of interest in itself.

The world planner is utilitarian and uses the population weight ω_n to aggregate the welfare of each representative household. The weights ω_n encode the social planner's

relative preferences across households. In both interpretations, the planner's objective at date 0 is to maximize:

$$W = \sum_{n=1}^{\mathcal{N}} \omega_n U_{n,0},$$

where the planner takes as given the initial stocks of each type of capital in each country n, $K^D_{n,0}$.

The social planner is subject to the following nine constraints (which, given the utility function we assume, will hold with equality):

1. The sum of labor assigned to production for each good z in each sector j in country n at date t, $L^j_n(z)$, must be equal to the labor endowment L_n, that is:

$$\sum_{j\in\Omega} \int_0^1 L^j_n(z)dz = L_n.$$

2. The sum of capital assigned to production of each good z in each sector j in country n at date t, $K^{jD}_{n,t}(z)$, along with capital used by the households to render durable goods services, must be equal to the capital endowment $K^D_{n,t}$, that is:

$$\sum_{j\in\Omega} \int_0^1 K^{jD}_{n,t}(z)dz + K^{H,D}_{n,t} = K^D_{n,t}.$$

3. In each country n at each date t, the output $y^j_{n,t}(z)$ of good z in sector j is given by:

$$y^j_{n,t}(z) = a^j_{n,t}(z) \left(\frac{L^j_n(z)}{\beta^{L,j}_n}\right)^{\beta^{L,j}_n} \left(\frac{K^{jD}_{n,t}(z)}{\beta^{K,jD}_n}\right)^{\beta^{K,jD}_n} \prod_{j'\in\Omega} \left(\frac{M^{jj'}_{n,t}(z)}{\beta^{M,jj'}_n}\right)^{\beta^{M,jj'}_n},$$

where $M^{ij'}_{n,t}(z)$ are intermediates from sector j' used to make good z in sector j.

4. The world's use of the output of good z in sector j from country n at date t must be

equal to what n produces, that is:

$$\sum_{m=1}^{N} d^{j}_{mn,t} x^{j}_{mn,t}(z) = y^{j}_{n,t}(z),$$

where $x^{j}_{mn,t}(z)$ is country m 's absorption of good z in sector j from country n. Notice that the social planner still needs to pay the iceberg costs, since these are physical transportation costs.

5. Country n's total absorption of good z in sector j, $x^{j}_{n,t}(z)$, must be equal to what it absorbs from each source i:

$$x^{j}_{n,t}(z) = \sum_{i=1}^{N} x^{j}_{ni,t}(z).$$

6. Absorption from each sector j in country n (for final use as investment or consumption or for intermediate use) $x^{j}_{n,t}$ aggregates across the goods for that sector:

$$x^{j}_{n,t} = \left(\int_{0}^{1} x^{j}_{n,t}(z)^{(\sigma-1)/\sigma} dz \right)^{\sigma/(\sigma-1)}.$$

7. The consumption of households and the usage of intermediates in sector $h \in \Omega$ in country n must be equal to country n's absorption from the same sector:

$$\sum_{j \in \Omega} \int_{0}^{1} M^{jh}_{n,t}(z) dz + C^{h}_{n,t} = x^{h}_{n,t}.$$

8. What firms invest and what they use as intermediates must equal country n's absorption from sector D, that is,

$$\sum_{j \in \Omega} \int_{0}^{1} M^{jD}_{n,t}(z) dz + I^{D}_{n,t} = x^{D}_{n,t}.$$

9. The capital available in country n at date $t+1$ follows its law of motion:

$$K^D_{n,t+1} = \chi^D_{n,t} \left(I^D_{n,t}\right)^{\alpha^D} \left(K^D_{n,t}\right)^{1-\alpha^D} + \left(1-\delta^D\right) K^D_{n,t}.$$

For a more detailed discussion of the solution to the planner's problem as a problem of intertemporal constrained optimization, we refer the interested reader to Appendix A. The first-order conditions, together with the distributional assumption in equation (2), deliver expressions for sectoral allocations and shadow prices on the constraints outlined above. In the last step, we replace the relevant shadow prices with corresponding competitive prices. In particular, we let $p^j_{n,t} = \lambda^j_{n,t}$, $w_{n,t} = \lambda^L_{n,t}$, and $r_{n,t} = \lambda^K_{n,t}$.

3 Solution Method

We numerically approximate a recursive equilibrium (Stokey et al., 1989; Ljungqvist and Sargent, 2004) of our model, that is, a function mapping the states of the economy to the optimal choices that fulfill the equilibrium conditions derived in Section 2.3. Following Spear (1988), we call this a "functional rational expectations equilibrium" (FREE). In such an equilibrium, a potentially high-dimensional *state variable* $\mathbf{x} \in X \subset \mathbb{R}^M$ represents the state of the economy, M is the dimensionality of the state space, and a time-invariant *optimal policy function* $\mathbf{p} : X \to Y \subset \mathbb{R}^K$, the desired unknown, captures the model dynamics and can be characterized as the solution to a functional equation (Fernández-Villaverde et al., 2016):

$$\mathcal{H}(\mathbf{p}) = \mathbf{0}. \tag{15}$$

This abstract description nests various characterizations of recursive equilibria, and in particular, the widespread case where the operator $\mathcal{H}$ captures discrete-time first-order equilibrium conditions. In our formulation of the model, the state of the economy is:

$$\mathbf{x} \in \mathbb{R}^M := \left(K^D_{n,t}, T^N_{n,t}, T^S_{n,t}, T^D_{n,t}, \chi^D_{n,t}, \phi_{n,t}, \psi^N_{n,t}, d^N_{ni,t}, d^D_{ni,t}\right), \tag{16}$$

whereas the policy vector is:

$$\begin{aligned}\mathbf{p} \in \mathbb{R}^K \quad := \quad & \Big(p^N_{n,t}, p^S_{n,t}, p^D_{n,t}, C^h_{n,t}, r_{n,t}, K^{H,D}_{n,t}, Y^N_{n,t}, Y^S_{n,t}, Y^D_{n,t}, w_{n,t}, \\ & X^{F,N}_{n,t}, X^{F,S}_{n,t}, c^N_{n,t}, c^S_{n,t}, c^D_{n,t}, \pi^N_{ni,t}, \pi^D_{ni,t}, X^N_{n,t}, X^S_{n,t}, X^D_{n,t}, I^D_{n,t}\Big), \end{aligned} \tag{17}$$

and where both for the states and policies, $1 \leq n, i \leq \mathcal{N}$ are the indices over countries/regions, and where both M and K scale $\propto \mathcal{N}^2$. For example, with just three countries/regions (as we will have in our calibration), the economy has 32 states.

There are substantial computational challenges in solving this model globally because of a) the presence of random shocks, b) a high-dimensional state space, c) strong non-linearities in the equilibrium functions, and d) irregular geometries of the ergodic set of states.[5] In the presence of these four features, the curse of dimensionality (Bellman, 1961) imposes a considerable roadblock.

While there are methods that can handle a subset of a), b), c), and d), most of them fail at matching all four requirements. Perturbation methods (see, e.g., Schmitt-Grohé and Uribe, 2004; Adjemian et al., 2018, and references) do not allow us to study large shocks or obtain global solutions. Computational methods designed to approximate policy functions in high-dimensional nonlinear settings, such as Smolyak sparse grids and adaptive sparse grids, fail if the dimensionality of the state space exceeds around 20 and require regular geometries of the ergodic set of states.[6] Finally, standard Gaussian processes (e.g., Scheidegger and Bilionis, 2019) face problems with strong non-linearities.

Currently, the only tractable method that jointly addresses all four features relies on neural networks. To this end, we describe in the following Subsection 3.1 how the recently introduced solution method called "Deep Equilibrium Nets" (DEQNs, see Azinovic et al., 2022) can be adopted to compute global solutions for our model. By doing so, we follow previous applications of neural networks in economics (see, e.g., Duffy and McNelis (2001), Norets (2012), Maliar et al. (2021), Duarte (2018), Fernández-Villaverde et al. (2019), and Villa and Valaitis (2019)).

[5] A *global solution* adheres to the model equilibrium conditions throughout the entire state space, that is, the computational domain, whereas a *local solution* is only concerned with the local approximation around a point, typically the deterministic steady state of the model.

[6] For more details on Smolyak and adaptive sparse grids, see for instance Krueger and Kubler (2004); Fernández-Villaverde et al. (2015); Judd et al. (2014); Brumm and Scheidegger (2017); Brumm et al. (2021).

3.1 Using "Deep Equilibrium Nets" to solve our Model

We apply the recently developed DEQN solution method to compute the FREE, which is characterized by the equations (4)-(14) and which formally correspond to the functional equation (15).[7] From an abstract perspective, DEQNs are deep neural networks that approximate the policy function $\mathbf{p}$ and are trained in an unsupervised fashion to satisfy all equilibrium conditions along simulated paths of the economy.[8]

At its core, the DEQN algorithm consists of four building blocks: i) deep neural networks for approximating the equilibrium policies; ii) a suitable loss function measuring the quality of a given approximation at a given state of the economy; iii) an updating mechanism to improve the quality of the approximation; and iv) a sampling method for choosing states for updating and evaluating of the approximation quality. We next outline each of the four components separately for clarity, thereby closely following Azinovic et al. (2022). For more details on the concrete implementation, we refer the interested reader to Appendix B.

3.1.1 Deep Neural Networks as Function Approximators

The objective of the DEQN algorithm is to approximate the equilibrium functions by a deep neural network, that is, $\mathbf{p}(\mathbf{x}) \approx \mathcal{F}_\nu(\mathbf{x})$. In particular, we use so-called densely connected feed-forward neural networks for approximating the equilibrium policies because they combine a set of desirable qualities. Neural networks are universal function approximators (Hornik et al., 1989); that is, they can resolve highly non-linear features and can handle a large amount of high-dimensional input data.[9] The intuition behind these properties is that deep neural networks transform the functional approximation problem from its original complex geometry to a much simpler geometric space thanks to the successive application of linear translations and nonlinear transformations.

More precisely, given the hyper-parameters $\left\{\kappa, \{m_i\}_{i=1}^{\kappa}, \{\boldsymbol{\mu}_i(\cdot)\}_{i=1}^{\kappa}\right\}$ and the trainable pa-

[7] The corresponding equality conditions used in the DEQN algorithm are summarized by the equations (25)-(43) below. Furthermore, the corresponding dynamics for the state transition is given in Appendix C.3.

[8] We implemented DEQN in Tensorflow v2.4, which can be downloaded from `tensorflow.org`.

[9] See Goodfellow et al. (2016) for an excellent introduction to deep learning, and (Bronstein et al., 2021) for a more advanced treatment focused on the geometric properties of the deep networks.

rameters $\boldsymbol{\nu}$, a neural network $\mathcal{F}_{\boldsymbol{\nu}}$ encodes the map:

$$\mathbf{x} \to \mathcal{F}_{\boldsymbol{\nu}}(\mathbf{x}) = \boldsymbol{\mu}_\kappa(\mathbf{W}_\kappa ... \boldsymbol{\mu}_2(\mathbf{W}_2 \boldsymbol{\mu}_1(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) ... + \mathbf{b}_\kappa),$$

where $\mathbf{W}_i \in \mathbb{R}^{m_{i+1} \times m_i}$ are the weight matrices, and $\mathbf{b}_i \in \mathbb{R}^{m_{i+1}}$ are the bias vectors.

Let us discuss each of the previous elements. The hyper-parameter κ is the number of layers of the neural network, and m_i is the number of nodes in layer i. The nonlinear functions μ_i are the activation functions and are applied element-wise to each entry of a vector, that is, $\boldsymbol{\mu}_i(\mathbf{x}) = [\boldsymbol{\mu}_i(x_1), \ldots, \boldsymbol{\mu}_i(x_{m_{i+1}})]^T$. Finally, the vector $\boldsymbol{\nu}$ collects all entries of the weight matrices and the bias vectors. As we pointed out before, a densely connected feedforward neural network is, hence given by a sequence of matrix-vector multiplications followed by the application of an activation function.

3.1.2 An "Economic" Loss Function to train the Deep Neural Network

The selection of hyper-parameters $\left\{\kappa, \{m_i\}_{i=1}^{\kappa}, \{\boldsymbol{\mu}_i(\cdot)\}_{i=1}^{\kappa}\right\}$ is known as the architecture selection. This is usually done by exploring some common alternatives and checking how the resulting deep neural network approximates the function of interest. Approaches to determine these hyper-parameters include using prior experience, manual, random, or grid search, as well as methods such as Bayesian optimization (see, e.g., Bergstra et al., 2011). The determination of the values of $\boldsymbol{\nu}$ is known as training the network. This training is implemented by minimizing a loss function that measures how close $\mathcal{F}_{\boldsymbol{\nu}}(\mathbf{x})$ is to $\mathbf{p}(\mathbf{x})$.

If the neural network $\mathcal{F}_{\boldsymbol{\nu}}$ encoded the true equilibrium functions $\mathbf{p}$ exactly, equations (25)-(43) would evaluate to zero for all states $\mathbf{x}$ of the economy (cf. equation (15)). Hence, a natural loss function quantifies the extent to which the equations (25)-(43) differ from zero when evaluated at a given state $\mathbf{x}$. The points at which we perform this evaluation are commonly referred to as the *training set*, which we denote as $\mathcal{D}_{\text{train}}$.[10]

Given parameters $\boldsymbol{\nu}$ and a set of states $\mathcal{D}_{\text{train}}$, the loss function in the DEQN is defined as the mean-squared error of all equilibrium conditions whose policies are parameterized by

[10] Checking the Euler equation also requires the evaluation of the expectation operator. We do so by either implementing a simple monomial quadrature rule that scales linearly with the number of continuous state variables (see, e.g., Judd, 1998, and references therein) or by using an efficient, "DEQN-compatible" Monte-Carlo approach that we introduce in Appendix B.3.

the deep neural network with trainable parameters $\boldsymbol{\nu}$, and which are evaluated at a given state $\mathbf{x}_i$:

$$\ell_{\mathcal{D}_{\text{train}}}(\boldsymbol{\nu}) := \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{\mathbf{x}_j \in \mathcal{D}_{\text{train}}} \left(\left(\text{eq. } (25)_{\mathbf{x}_j}(\boldsymbol{\nu}) \right)^2 + \cdots + \left(\text{eq. } (43)_{\mathbf{x}_j}(\boldsymbol{\nu}) \right)^2 \right). \tag{18}$$

3.1.3 Updating Mechanism

The loss function above is defined such that smaller values correspond to lower mean-squared errors in the equilibrium conditions. Consequently, trainable parameters are deemed "good" if they minimize the loss function. Due to the functional structure of deep neural networks, variants of gradient descent are typically used to optimize the parameters $\boldsymbol{\nu}$. Gradient descent updates the parameters step-wise in the direction in which the loss function decreases:

$$\boldsymbol{\nu}_k^{\text{new}} = \boldsymbol{\nu}_k^{\text{old}} - \alpha^{\text{learn}} \frac{\partial \ell_{\mathcal{D}_{\text{train}}}(\boldsymbol{\nu}^{\text{old}})}{\partial \boldsymbol{\nu}_k^{\text{old}}} \; \forall k \in \{1, \ldots, length(\boldsymbol{\nu})\}. \tag{19}$$

The *learning rate* $\alpha^{\text{learn}} > 0$ governs by how much the parameters are adjusted with each gradient descent step. Some popular variants of gradient descent can speed up the learning process. In particular, we use `Adam` (see Kingma and Ba, 2014) together with mini-batch gradient descent (for more details, see Appendix B below).

3.1.4 Sampling

We generate our training set $\mathcal{D}_{\text{train}}$ by sampling from the ergodic set of states of the economy. In that way, we will train the DEQN in the region where the economy is going to spend most of its time, and our approximated equilibrium functions will be most accurate where we need them to be, e.g., the states visited on the simulated path of the economy.

First, we start with an arbitrary (but feasible given the technology and preferences of the economy) state $\mathbf{x}_1^0$, and randomly initialized $\boldsymbol{\nu}$. Then, we simulate $T-1$ periods forward based on the approximated equilibrium functions given by the neural network, and the model-implied dynamics of the endogenous and exogenous states (see, e.g., Section C.3). Since DEQN approximates the equilibrium functions directly, simulating the evolution of

the economy is computationally cheap. The resulting T simulated states of the economy constitute our dataset $\mathcal{D}^0_{\text{train}} = \{\mathbf{x}^0_1 \ldots, \mathbf{x}^0_T\}$. We call the set of T simulated periods $\mathcal{D}_{\text{train}}$ an *episode*.

Second, we split this input data into mini-batches –smaller subsets of size m with random membership– and perform gradient descent steps on each subset as given in equation (19). Consequently, each simulated state is only used for a single gradient descent step, which guards our algorithm against overweighting one particular data point. A completion of an *epoch* is defined as when the whole dataset $\mathcal{D}_{\text{train}}$ is passed through the algorithm. Per epoch, the neural network parameters are updated T/m times.

Third, we set $\mathbf{x}^1_1 = \mathbf{x}^0_T$ and use the updated parameters of the neural network to simulate $T - 1$ periods forward, generate a new training set $\mathcal{D}^1_{\text{train}} = \{\mathbf{x}^1_1, \ldots, \mathbf{x}^1_T\}$, and repeat the process. The error on a new set of simulated states is an out-of-sample error and can be used to judge the out-of-sample quality of the approximation. Since we can generate a large amount of training data in this setting, data availability is not a primary concern. In fact, our method allows us to train the neural network on more than a billion simulated states. As the neural network learns better parameter values, the simulated states become better representatives of the ergodic set of states of the economy.

Finally, notice that Ebrahimi Kahou et al. (2022) show how deep learning algorithms yield solutions that fulfill the transversality conditions of equilibrium models. Deep learning has a built-in "inductive bias" toward smooth solutions (in the appropriate sense of functional approximation smoothness), which are precisely the solutions that satisfy those transversality conditions.

Algorithm 1 provides the pseudo-code of the DEQN we just described. For notational simplicity, the pseudo-code implements gradient descent rather than mini-batch gradient descent.

Algorithm 1: Algorithm for training deep equilibrium nets.

Data:
T (length of an episode), N^{epochs} (number of epochs on each episode), N^{iter} (maximum number of iterations),
τ^{mean}, $\tau^{\max}$ (the desired threshold for mean and max error, respectively),
$\epsilon^{\text{mean}} = \infty$, $\epsilon^{\max} = \infty$ (starting value for current mean and max error, respectively),
$\boldsymbol{\nu}^0$ (initial parameters of the neural network), $\mathbf{x}_1^0$ (initial state to start simulations from), $i = 0$ (set counter),
α^{learn} (learning rate)
Result:
success (boolean if thresholds were reached)
$\boldsymbol{\nu}^{\text{final}}$ (final neural network parameters)
while $((i < N^{iter}) \wedge ((\epsilon^{mean} \geq \tau^{mean}) \vee (\epsilon^{max} \geq \tau^{max})))$ **do**
  $\mathcal{D}^i_{\text{train}} \leftarrow \{\mathbf{x}^i_1, \mathbf{x}^i_2, \ldots, \mathbf{x}^i_T\}$ (generate new training data)
  $\mathbf{x}^{i+1}_0 \leftarrow \mathbf{x}^i_T$ (set new starting point)
  $\epsilon_{\max} \leftarrow \max\left\{\max_{\mathbf{x}\in\mathcal{D}^i_{\text{train}}} |e^{\cdot\cdot}_{\mathbf{x}}(\boldsymbol{\nu})|\right\}$, $\epsilon_{\text{mean}} \leftarrow \max\left\{\frac{1}{T}\sum_{\mathbf{x}\in\mathcal{D}^i_{\text{train}}} |e^{\cdot\cdot}_{\mathbf{x}}(\boldsymbol{\nu})|\right\}$ (update errors)
  for $j \in [1, ..., N^{epochs}]$ **do**
    (learn N^{epochs} on data)
    for $k \in [1, ..., length(\boldsymbol{\nu})]$ **do**

$$\boldsymbol{\nu}^{i+1}_k = \boldsymbol{\nu}^i_k - \alpha^{\text{learn}} \frac{\partial \ell_{\mathcal{D}^i_{\text{train}}}(\boldsymbol{\nu}^i)}{\partial \boldsymbol{\nu}^i_k}$$

      (do a gradient descent step to update the network parameters)
    end
  end
  $i \leftarrow i + 1$ (update episode counter)
end
if $i = N^{iter}$ **then return** (success $\leftarrow False, \boldsymbol{\nu}^{\text{final}} \leftarrow \boldsymbol{\nu}^i$) ;
else return (success $\leftarrow True, \boldsymbol{\nu}^{\text{final}} \leftarrow \boldsymbol{\nu}^i$) ;

4 Calibration

We calibrate our model to replicate a world economy with three economic regions, that is, $\mathcal{N} = 3$: the U.S. (US), China (CH), and the Eurozone (EU). These are the three largest economic regions of the world, accumulating among them nearly 50% of the world GDP. Additionally, by having three regions, we can evaluate the asymmetric effects of shocks among different regions while keeping the model sufficiently tractable for building intuition about our economic findings. Table 1 lists all parameters in our model calibration.

In terms of moments to match, we target the relative size of each of the three regions as well as their corresponding productivity processes. First, we ensure that the three model regions' relative GDP is consistent with the data. To that end, we calculate country-specific GDP in the model, which is defined as follows

$$Y_{n,t} = w_{n,t}L_n + r_{n,t}\left(K^D_{n,t} - K^{H,D}_{n,t}\right),$$

where labor income $w_{n,t}L_n$ is given by (12) and capital income $r_{n,t}K^D_{n,t}$ by (13). To avoid valuation problems, we exclude rental payments by households on durables from our measure of GDP. We then calibrate the relative size of the local labor markets based on the number of employed persons aged over 16 years in 2021 (153 million in the U.S., 747 million in China, and 195 million in the Euro area). By normalizing L^{US} to one, we set $L^{CH} = 4.88$ and $L^{EU} = 1.27$.

Second, we calibrate the productivity parameters in line with the empirical evidence of country-specific TFP. For the U.S. and the Eurozone, we align the persistence and shock size of country-specific TFP with estimated counterparts in Smets and Wouters (2007) and Smets and Wouters (2010). For China, we calibrate productivity parameters based on time series RTFPNACNA632NRUG from the St.Louis Federal Reserve. While productivity in all three countries is fairly persistent, it is most persistent in the United States. In contrast, shocks to the persistent TFP process are the largest by far for China, while the United States and the Eurozone are comparable in size.

Third, the rest of the parameters are set to standard values from the literature or, alternatively, to very conservative values. For instance, we set the discount factor, ρ, to 0.987 to

Parameter	Description	Value	Source
A. Preferences			
ρ	quarterly discount factor	0.987	Eaton et al. (2016a)
for $\mathcal{N}-1=2$ countries: ρ_{ϕ_n}	persistence demand shocks	0.84	Primiceri and Jusiniano (2008)
$\sigma_{\phi n}$ (in %)	volatility demand shocks	3.13	Primiceri and Jusiniano (2008)
ψ_n^D		1/3	
ρ_{ψ_n}	persistence sector preference	0.84	
σ_{ψ_n} (in %)	volatility sector preference	1.57	
ω_{EU}	weight social planner	0.21 (1/3)	Population share
ω_{US}	weight social planner	0.15 (1/3)	Population share
ω_{CH}	weight social planner	0.64 (1/3)	Population share
B. Technology			
θ	trade elasticity	2.00	Eaton et al. (2016a)
σ	elasticity of substitution	2.50	Eaton et al. (2016a)
$\beta^{L,N}, \beta^{L,S}, \beta^{L,D}$	output elasticities of labor	2/3	
$\beta^{K,ND}, \beta^{K,SD}, \beta^{K,DD}$	output elasticity of capital	1/12	
$\forall j, j' \in \{N,S,D\}$: $\beta^{M,jj'}$	output elasticities intermediates	1/12	
α^D	capital adjustment cost	0.55	Eaton et al. (2016a)
δ	quarterly depreciation rate	0.026	Eaton et al. (2016a)
C. Trade			
d_{nn}^l		1	
$\rho_{d_{ni}}$ where $i \neq n$	persistence trade cost shocks	0.85	
$\sigma_{d_{ni}^D}, \sigma_{d_{ni}^N}$ (in %) where $i \neq n$	volatility trade cost shocks	1.00	
Case 1: Symmetric Trade Costs:			
$\bar{d}_{ni}^D, \bar{d}_{ni}^N$ where $i \neq n$	iceberg trade costs	ln(1.2)	Irarrazabal et al. (2015)
Case 2: Asymmetric Trade Costs:			
$\bar{d}_{US\_EU}^l, \bar{d}_{EU\_USA}^l$	iceberg trade costs	ln(1.2)	Irarrazabal et al. (2015)
$\bar{d}_{CH\_EU}^l, \bar{d}_{EU\_CH}^l$	iceberg trade costs	ln(5)	
$\bar{d}_{US\_CH}^l, \bar{d}_{CH\_US}^l$	iceberg trade costs	ln(5)	
D. Stochastic Processes			
ρ_{χ_n}	persistence investment efficiency	0.81	Primiceri et al. (2011)
$\sigma_{\chi n}$ (in %)	volatility investment efficiency	5.78	Primiceri et al. (2011)
$\rho_{TD_{US}}, \rho_{TN_{US}}, \rho_{TS_{US}}$	persistence sector productivity U.S.	0.95	Smets and Wouters (2007)
$\rho_{TD_{EU}}, \rho_{TN_{EU}}, \rho_{TS_{EU}}$	persistence sector productivity EU	0.83	Smets and Wouters (2010)
$\rho_{TD_{CH}}, \rho_{TN_{CH}}, \rho_{TS_{CH}}$	persistence sector productivity China	0.81	TFP Data
$\sigma_{TD_{US}}, \sigma_{TN_{US}}, \sigma_{SD_{US}}$ (in %)	volatility sector productivity U.S.	0.45	Smets and Wouters (2007)
$\sigma_{TD_{EU}}, \sigma_{TN_{EU}}, \sigma_{SD_{EU}}$ (in %)	volatility sector productivity EU	0.61	Smets and Wouters (2010)
$\sigma_{TD_{CH}}, \sigma_{TN_{CH}}, \sigma_{SD_{CH}}$ (in %)	volatility sector productivity China	1.21	TFP Data

Table 1: Calibration of the three country model.

get a net annual return on capital of around 5% (given the structure of our model, we want to match the net return on all capital, not the return on bonds). Similarly, the persistence of demand shocks, $\rho_{\phi_n} = 0.85$, and its volatility, $\sigma_{\psi_n} = 0.03$ is taken from Primiceri and Jusiniano (2008).

However, it is important to notice that, unless there is a compelling reason to do otherwise, we set identical parameter values across countries. In that way, we can focus on

the heterogeneity in terms of GDP and productivity across countries triggered by different shocks instead of the heterogeneity caused by *ex-ante* differences.

Fourth, we will have two different calibrations of the trade costs: one with symmetric costs and one with asymmetric costs. In the symmetric costs case, $\bar{d}_{ni}^{D} = \bar{d}_{ni}^{N} = ln(1.2)$, that is, we have iceberg costs of around 18% for all international exchanges. This number is based on the results of Irarrazabal et al. (2015). These authors report a median iceberg cost of 14%, but they emphasize that their methodology can only identify additive trade costs relative to multiplicative costs. Thus, their estimates are a lower bound of the true value of additive costs. We correct this bias by increasing the costs by another 4%.

In the asymmetric cost case, trade between the U.S. and the Eurozone still has iceberg costs of around 18% but now trade with China is prohibitively expensive, with iceberg costs of 61%. In comparison, the Trump administration *ad valorem* tariffs on Chinese goods ranged between 7.5% and 25%. Thus, our goal is not to capture, with this second calibration, an observed situation but to have a formal comparison between a scenario where the three regions are integrated and a scenario where one region is (mostly) excluded from the trade block formed by the other two.

5 Quantitative Results: Symmetric Trade Costs

Since we solve our model globally, we can study its stochastic properties and the model-implied dynamics.[11] In this section, we provide some summary statistics to get a sense of the behavior of the model and, next, report the impulse-response functions (IRFs) to a variety of different three-standard-deviation shocks. All of our results come from the case with symmetric trade costs. In Section 6, we will repeat these exercises in the asymmetric cost case.

[11] We compute the "FREE" on a Laptop with 12 GB Ram and 4 CPUs (Intel(R) Core(TM) i7-5600U CPU @ 2.60GHz) and obtain solutions with average Euler Equation errors as low as 0.004% within approximately 2 hours of run-time. This run-time, however, is only indicative if one solves a model from scratch, where no prior knowledge nor a good initial guess for the policy functions exists. In contrast, one can benefit from pre-training if one wants to re-run a model with slightly changed parameters. That is, by re-starting the model (with changed parameter settings) from a known solution, solution times can reduce to a few minutes. See Appendix B for implementation details.

5.1 Summary Statistics

The left panel of Figure 1 plots a histogram of the U.S. output along a simulated path for the symmetric trade cost case (simulation length is 10,000 periods), while the right panel plots 200 periods of this simulation. The average output is 0.281, with a standard deviation of 0.008, or 2.7 percentage points, close to the historical volatility of the observed U.S. output after World War II (Stock and Watson, 2003).

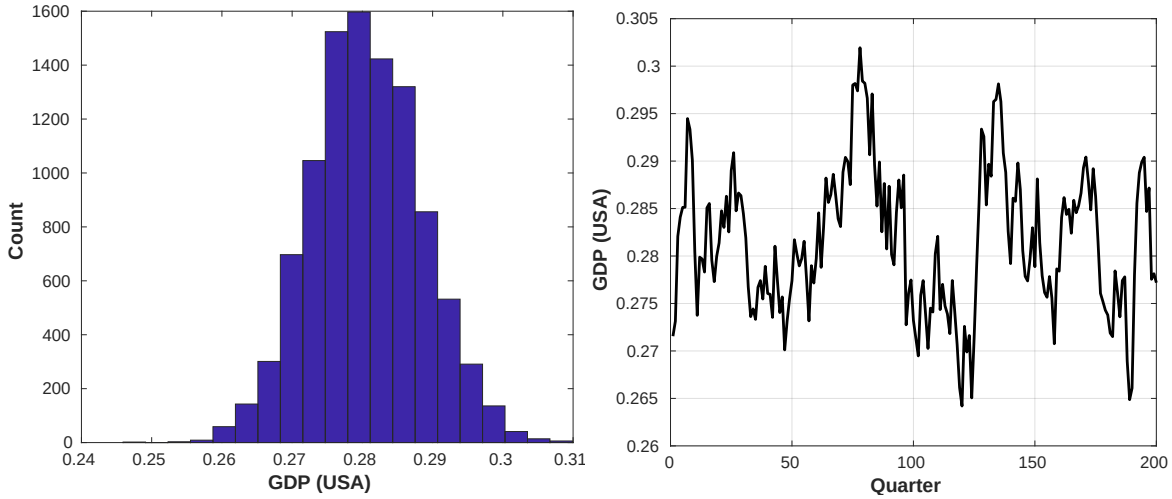

Figure 1: Left panel: histogram of the U.S. output over a 10,000-period simulation. Right panel: the evolution of the U.S. output along a simulated path (simulation length is 200 quarters).

It is instructive to inspect the international trade block of the model. We start with the histogram of import shares of non-durables. The left panel of Figure 2 plots the histogram of $\pi^N_{EU,US}$ over the same simulation of 10,000 periods that we used above. The mean is 29.1%, and the standard deviation is 0.9%. The right panel repeats the same exercise for the case $\pi^N_{CH,US}$, with a mean of 29.2% and standard deviation of 0.9%.

Similarly, Figure 3 plots the histograms of the import shares $\pi^D_{n,USA}$ of durables. In terms of basic statistics, for European imports from the U.S., we have a mean of 29.1% and a standard deviation of 0.9%, while for Chinese imports from the U.S., the mean is 29.2% and a standard deviation of 0.9%

For completeness, Table 2 reports some of the other summary statistics of the model (columns two and three for the case with symmetric trade costs) for the sector of durables. As expected, the production costs are lowest in China due to its larger endowment of labor

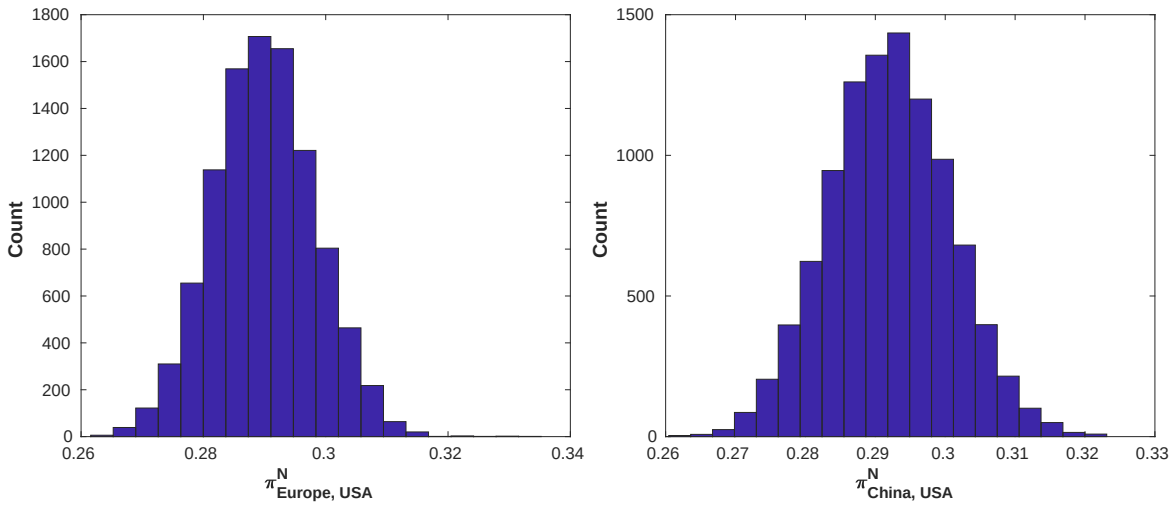

Figure 2: Histogram of import shares $\pi^N_{n,USA}$ of non-durables (EU,USA) (left panel) and (CH, USA) (right panel) along a simulated path (simulation length is 10,000 periods).

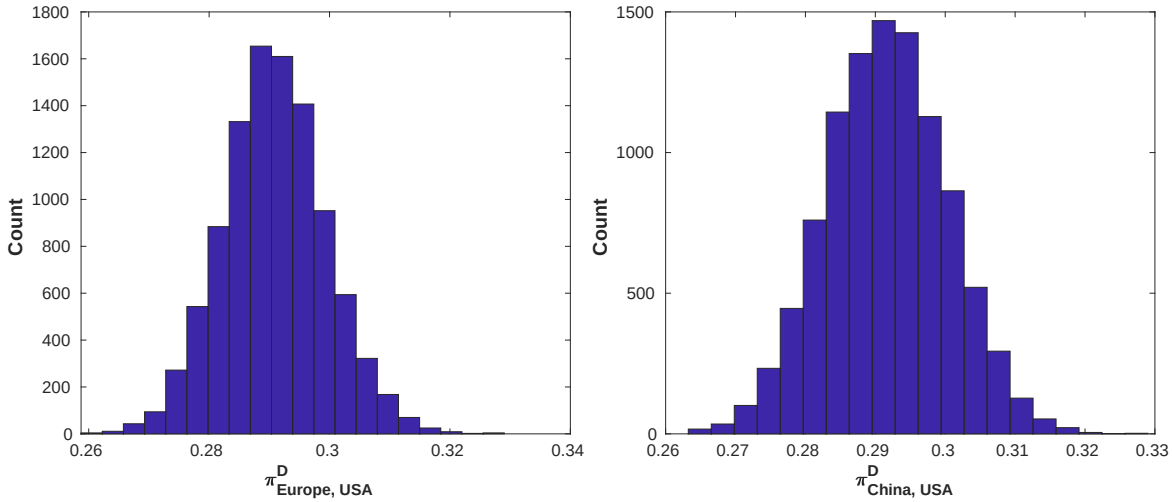

Figure 3: This figure shows the histogram of the evolution of import shares of durables (EU,USA) (left panel) and (CH, USA) (right panel) along a simulated path (simulation length is 10,000 periods).

and highest in the U.S. Not surprisingly, the patterns of trade inherit the symmetry of trade costs.

Table 2: Model-implied Moments

Variable	Symm. Trade Cost		Asymm. Trade Cost	
	Mean	Std	Mean	Std
A. Prices in sector D				
p^D_{US}	0.233	0.005	0.304	0.007
p^D_{EU}	0.195	0.004	0.255	0.006
p^D_{CH}	0.073	0.003	0.123	0.007
B. Costs in sector D				
c^D_{US}	0.153	0.003	0.169	0.004
c^D_{EU}	0.128	0.003	0.142	0.003
c^D_{CH}	0.048	0.002	0.054	0.003
B. Import shares in sector D				
$\pi^D_{US,US}$	0.419	0.007	0.576	0.008
$\pi^D_{US,EU}$	0.291	0.009	0.401	0.009
$\pi^D_{US,CH}$	0.291	0.009	0.023	0.001
$\pi^D_{EU,US}$	0.291	0.009	0.400	0.009
$\pi^D_{EU,EU}$	0.418	0.006	0.576	0.008
$\pi^D_{EU,CH}$	0.290	0.009	0.023	0.001
$\pi^D_{CH,US}$	0.292	0.009	0.037	0.001
$\pi^D_{CH,EU}$	0.290	0.008	0.037	0.001
$\pi^D_{CH,CH}$	0.419	0.007	0.924	0.004

5.2 Impulse-Response Functions

We move now to analyze the IRFs of the model. Our first experiment traces the consequences of a three-standard-deviation productivity shock in the U.S.[12] In Figure 4, we see that *GDP* increases in the U.S. by 0.12% (solid line).[13] In comparison, the effect at impact is only about half in Europe (dashed line) and China (dotted line). The intuition is simple: higher productivity in the U.S. mainly benefits the U.S. but, through international trade, induces growth in output in its trading partners. The effect is more persistent in Europe than in China. Since a potential future productivity shock in Europe is more persistent than in China, higher productivity in the U.S. creates more incentives for investment in Europe than in China.

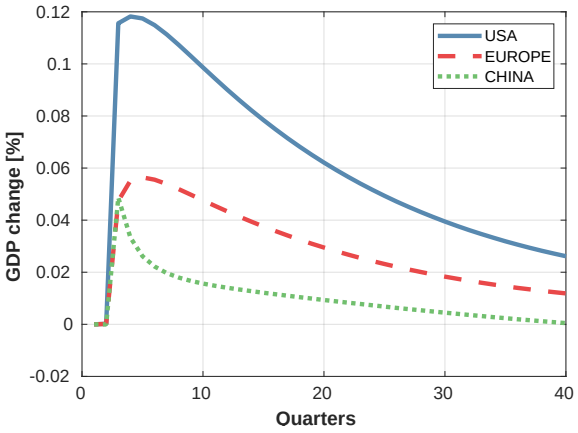

Figure 4: Country-specific IRFs of *GDP* to an unexpected productivity shock across all three sectors (durables, non-durables, services) in the U.S.

Interestingly, this asymmetric response is quite different when the productivity shock hits another region. For example, Figure 5 plots the country-specific IRFs to a three-standard-deviation productivity shock in China. In this case, China's *GDP* increases by about 0.33%. Moreover, the differences in the persistence of the shocks are much less pronounced compared to Figure 4. These two figures show that, for the dynamics of our model, the region where the productivity shock hits is crucial.

[12] Following the convention in the literature on uncertainty shocks, we plot IRFs to three-standard-deviation shocks because they highlight better the non-linear aspects of the model. Technically speaking, we are dealing with generalized impulse-response functions (GIRFs) because their shape depends on the size and sign of the shock as well as where the shock hits. See Andreasen et al. (2018) for details.

[13] In the interest of brevity, we focus on the impulse responses of the most extreme countries in terms of productivity calibrations, that is, in the case of the shock size and persistence, the U.S. and China.

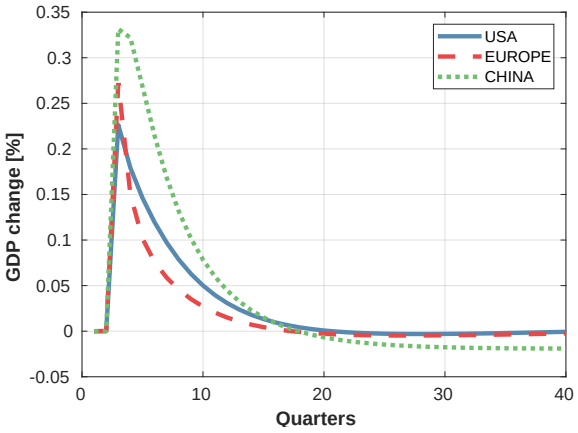

Figure 5: Country-specific IRFs of GDP to an unexpected productivity shock across all three sectors (durables, non-durables, services) in China.

Additionally, GDP numbers in Europe and in the U.S. increase by roughly two-thirds of the magnitude observed for China. That is, a productivity shock in China drives a much smaller output wedge between the three regions compared to the productivity shock in the U.S. The larger magnitude can be explained by the larger productivity shocks in China compared to the U.S., but also by the difference in productivity shock persistence.

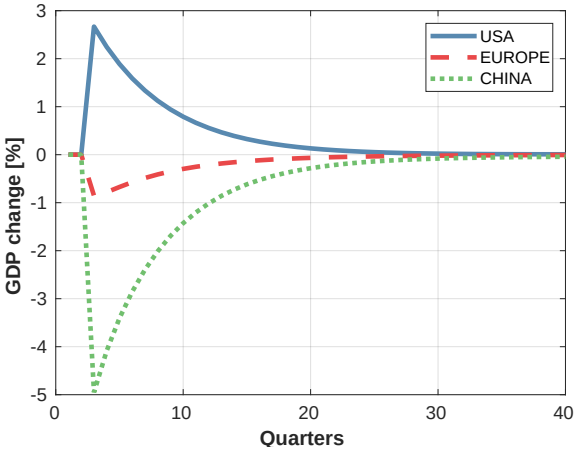

Figure 6: Country-specific IRFs of GDP to an unexpected increase in the demand for the U.S., $\phi_{US,t}$.

Next, we assess the consequences of an unexpected shock that directly increases demand for the U.S. and, indirectly, for Europe. As we restrict aggregate demand shocks to have no global component, and due to the way we set up the model, an increase in $\phi_{U.S.,t}$ will not affect the second country's (Europe) demand coefficient. However, it will lead to a reduction in the third country's (China) demand coefficient. We can think about this shock as a move

in the world demand away from Chinese goods into U.S. goods.

In this case, one would expect the GDP in the U.S. to go up and in China to go down. The response of GDP in Europe, however, is not *ex-ante* clear. There will be some negative effects from the relative increase in the demand for U.S. goods and some opposite signed effects due to the relatively higher demand for European over Chinese goods.

Figure 6 shows the resulting impulse responses of GDP. As expected, output in the U.S. (China) sharply increases (decreases), whereas the output in Europe slightly decreases. Interestingly, given that the three regions are of very similar size due to the symmetric setup, the global effect of a large demand shock is negative. This implies that GDP global GDP goes down upon the arrival of a large demand shock, which can be seen directly from Figure 7.

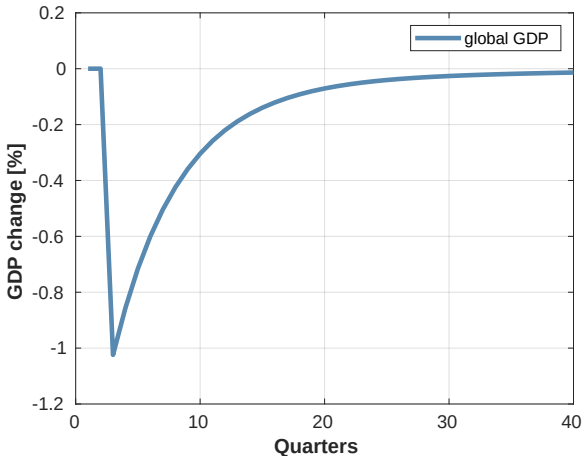

Figure 7: Response of global GDP to an unexpected increase in the demand for the U.S., $\phi_{US,t}$.

The intuition for this finding is simple: China is the cheapest producer of goods, given its relative size. A demand shock that lowers the relative demand for its goods makes, in some sense, the world economy less productive. To the best of our understanding, this is a novel result that shows the advantages of having a rich model with three regions and stochastic dynamics.

Our next experiment studies the effect of a sector-specific demand shift between the services and non-durables sectors within a given country. In particular, we look at the effects of an increase in the demand for non-durables at the expense of services in the U.S. Since any deviation from the equal-weighting scheme of sectors is detrimental to output in

the U.S. (elasticities of substitution are finite), Figure 8 documents that output decreases in the U.S., as it is harder to produce relatively more non-durables, but increases in Europe and China as the latter countries benefit through an increase in their exports to the U.S.

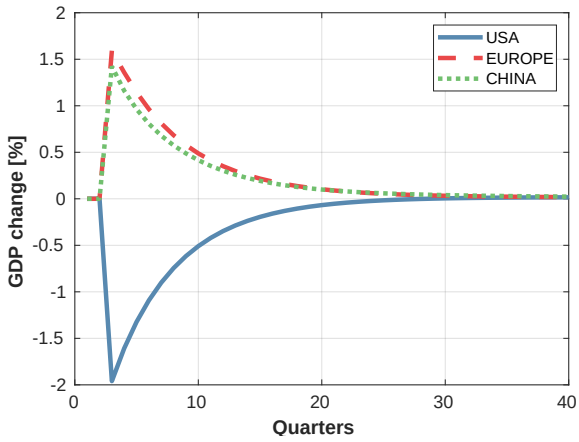

Figure 8: Country-specific IRFs of GDP to an unexpected sector-specific demand shift in the U.S., $\psi^N_{US,t}$.

In particular, the U.S. now wants to invest more (see Figure 9), which requires more inputs from Europe and China. We can see these increased imports in Figure 10. Both Europe and China mildly increase their absorption of U.S. goods in the durables sector. That is, part of the U.S.-specific investment boom spills over to China and Europe.

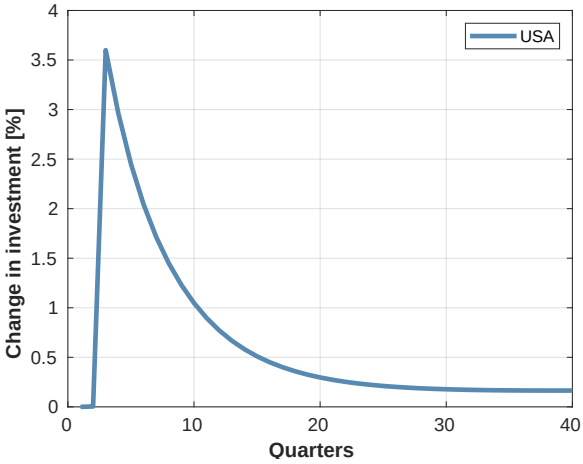

Figure 9: Response of investment in the U.S. to an unexpected sector-specific demand shift in the U.S., $\psi^N_{US,t}$.

The previous IRFs show that access to international trade helps to counteract this sudden change in the consumption composition and, therefore, changes the dynamics of the economy after shocks arrive.

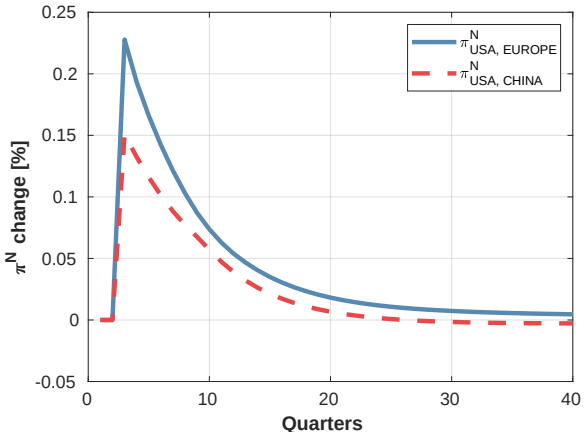

Figure 10: Responses of U.S. specific import shares in the non-durables sector with respect to Europe and China to an unexpected sector-specific demand shift in the U.S., $\psi^N_{US,t}$.

Finally, we investigate the consequences of an investment-specific technological shock in the U.S. As expected, a more efficient technology that converts investment into capital leads to a local investment boom, as reported in Figure 11.

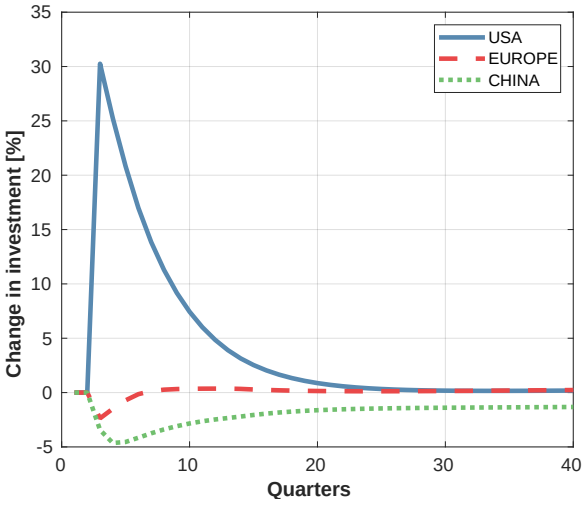

Figure 11: Country-specific IRFs of investment to an unexpected investment-specific technological shock in the U.S., $\chi^D_{US,t}$.

In contrast, investment in China and Europe drops slightly as these countries can indirectly benefit from the investment-specific shock in the U.S. by means of international trade. Therefore, China and Europe find it more beneficial to rely more heavily on imports from the U.S. in the durables sector as opposed to producing the durables goods themselves. Ultimately, the abundance of durable capital is manifested in an increase in output in the U.S. As barriers to trade are pretty low and trade flows are symmetric in this model calibration,

both Europe and China benefit to a similar extent from the improved conditions for local investment in the U.S as reported in Figure 12. In other words, a local investment-specific shock has comparable effects as a local productivity shock.

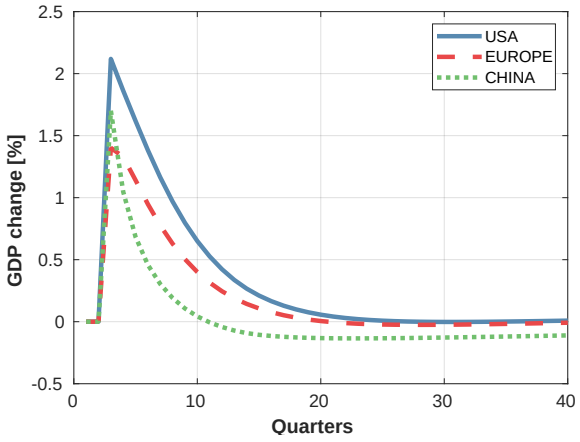

Figure 12: Country-specific IRFs of GDP to an unexpected investment-specific technological shock in the U.S., $\chi^D_{US,t}$.

6 Quantitative Results: Asymmetric Trade Costs

This section discusses the model calibration with asymmetric trade costs reported as "Case 2" in Panel C of Table 1. Similar to Section 5, we start by providing some summary statistics to get a sense of the behavior of the model and then report the IRFs for analogous experiments to those presented for the symmetric trade costs case in Section 5.

6.1 Summary Statistics

The left panel of Figure 13 plots a histogram of the U.S. output along a simulated path for the asymmetric trade cost case (simulation length is 10,000 periods), while the right panel plots 200 periods of this simulation. The average output is 0.283, with a standard deviation of 0.009, or 3.04 percentage points. Both numbers are very close to those observed in the model with symmetric trade costs. As it is well-known in models of international trade, the aggregate output effect of trade is minor when an economy is large (a different point might be the size of the welfare effects, although the results in Arkolakis et al., 2012 suggest that

the welfare effects are also likely to be small). In this case, the U.S. can still trade cheaply with Europe, even if not with China.

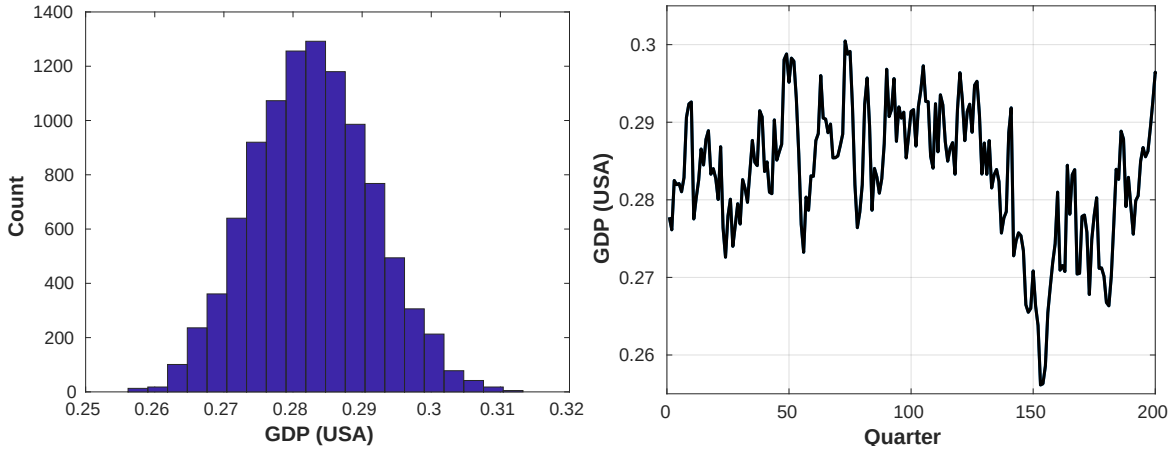

Figure 13: Left panel: histogram of the U.S. output over a 10,000-period simulation. Right panel: the evolution of the U.S. output along a simulated path (simulation length is 200 quarters).

To understand the detailed effects of asymmetric trade costs, it is instructive to inspect the international trade block of the model. We start with the histogram of import shares of non-durables. The left panel of Figure 14 plots the histogram of $\pi^N_{EU,US}$ over the same simulation of 10,000 periods that we used above.

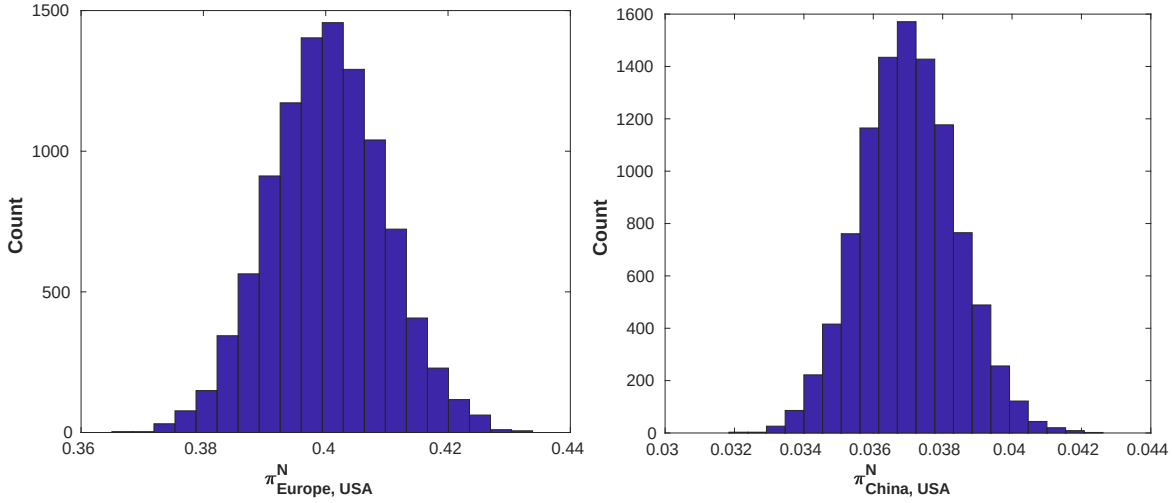

Figure 14: Histogram of import shares $\pi^N_{n,USA}$ of non-durables (EU,U.S.) (left panel) and (CH, U.S.) (right panel) along a simulated path (simulation length is 10,000 periods).

The mean import share is 40%, and the standard deviation 0.9%. The right panel repeats the same exercise for $\pi^N_{CH,US}$, with a mean of 3.7% and standard deviation of 0.1%.

Comparing the absorption of U.S. imports in Europe with the absorption of U.S. imports in China reveals that, unsurprisingly, high barriers to trade render imports from the U.S. very unattractive for China. At the same time, the relatively low barriers to trade between the U.S. and Europe foster a very active trade relationship between the two countries. In fact, the trade exposure of Europe to the U.S. and vice versa is significantly higher in the asymmetric case compared to the symmetric case.

Similarly, Figure 15 plots the histograms of the import shares $\pi^D_{n,USA}$ of durables originating in the U.S. both for Europe (left panel) and China (right panel). As we already saw in the case of the non-durables sector, there is a stark difference in the trade flows between the U.S. and China compared to the flows between the U.S. and Europe due to differential barriers to trade.

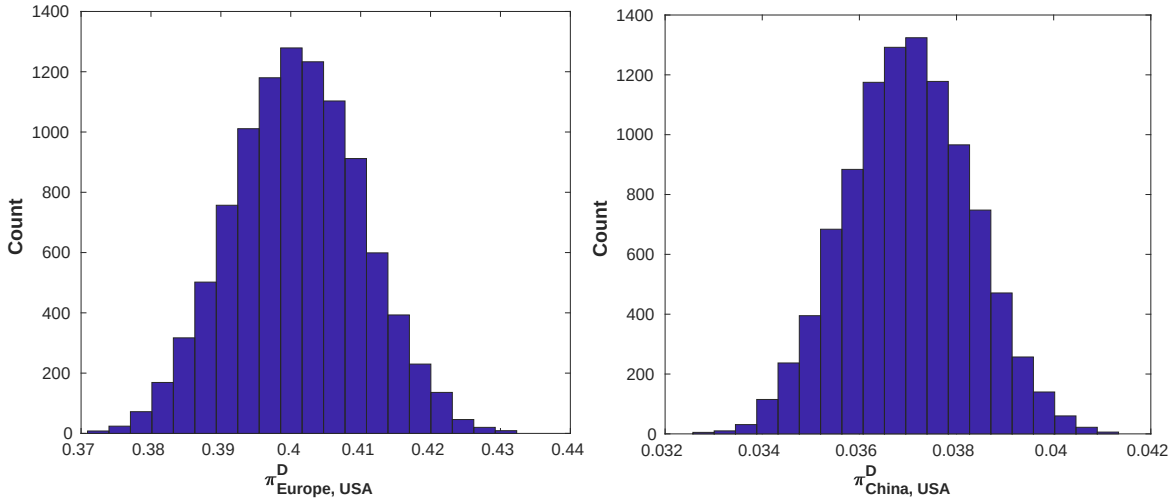

Figure 15: This figure shows the histogram of the evolution of import shares of durables (EU,U.S.) (left panel) and (CH, U.S.) (right panel) along a simulated path (simulation length is 10,000 periods).

For completeness, columns four and five of Table 2 report some of the other summary statistics for the case with asymmetric trade costs for the sector of durables. While the ranking in production costs is the same compared to the case of symmetric trade costs, the average costs are now higher across all sectors. The reason for this increase in costs is the lack of international trade flows and the diversification of shocks due to higher barriers to trade.

6.2 Impulse-Response Functions

We now move to analyze the IRFs of the model with asymmetric trade costs. We follow the same structure as in Section 5 to maximize comparability across the two model calibrations. Therefore, we start by studying the consequences of output across the three regions after a productivity shock in the U.S. in Figure 16.

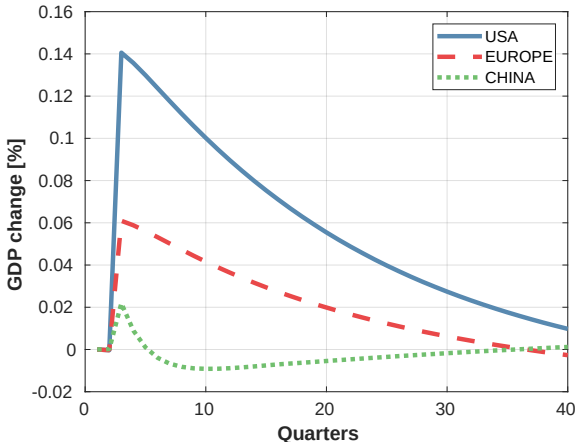

Figure 16: Country-specific IRFs of GDP to an unexpected productivity shock across all three sectors (durables, non-durables, services) in the U.S.

After the shock, output in the U.S. increases by about 0.14% (solid line). In comparison, the effect is slightly less than half this magnitude in Europe (dashed line) and essentially zero in China (dotted line). The intuition is simple: higher productivity in the U.S. mainly benefits the U.S. (direct effect) but, through international trade, induces growth in output in its main trading partner: Europe (indirect effect). In contrast, China finds itself almost in autarky, and the spillovers from other countries are very limited.

The flip side of Figure 16 can be seen in Figure 17, which reports the responses of country-specific output to a productivity shock in China. As expected, China's output increases sharply after the shock (direct effect). The global propagation of this shock through international trade channels, however, is very limited as this is prevented by the high iceberg costs that shipments from or to China are subject to.

Next, we discuss the responses to an unexpected demand shift towards the U.S., that is, $\phi_{US,t}$ increases unexpectedly by three-standard deviations. As we require aggregate demand shocks to have no global component, this will directly increase the demand for the U.S. and indirectly for Europe. Similar to the results in the symmetric trade costs calibration,

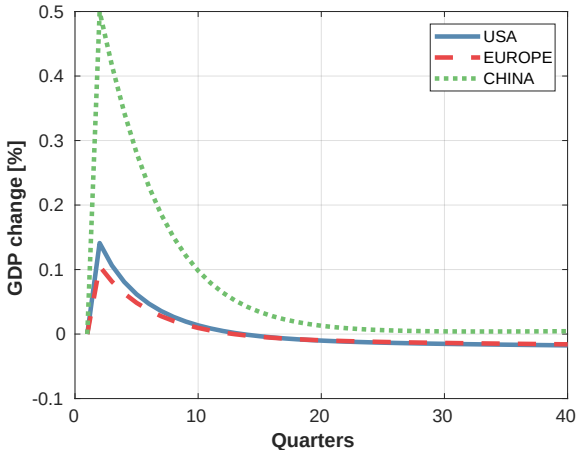

Figure 17: Country-specific IRFs of GDP to an unexpected productivity shock across all three sectors (durables, non-durables, services) in China.

Figure 18 reports that output in the U.S. (China) increases (decreases), and output in Europe decreases slightly. In comparison to Figure 6, the magnitudes are now more extreme. China's output drops by more than one percentage point, the U.S. GDP increases by roughly one-half a percentage point more. This difference can be explained by ex-ante exposures to the U.S.

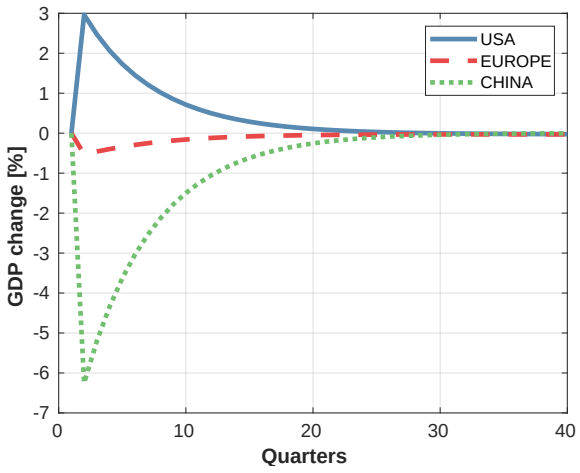

Figure 18: Country-specific IRFs of GDP to an unexpected increase in the demand for the U.S., $\phi_{US,t}$.

If there is close to no trade between China and the U.S. (not even indirectly through Europe) due to high barriers to trade, output in China suffers relatively more if demand shifts towards the U.S. At the same time, the exposure of the U.S. GDP to domestic shocks is larger the lower the international trade flows. While there is a very strong link between Europe and the U.S. (which is the reason for the relatively mild drop in European output),

the overall trade activity is lower in the asymmetric compared to the symmetric trade costs calibration.

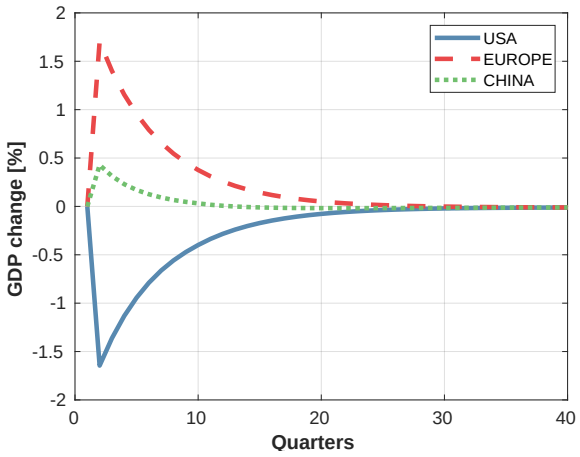

Figure 19: Country-specific IRFs of GDP to an unexpected sector-specific demand shift in the U.S., $\psi^N_{US,t}$.

We move now to explore the consequences of a sector-specific demand shift from the services to the non-durables sector in the U.S. Figure 19 reports that U.S. output decreases, while output in Europe increases upon arrival of the shock. Not surprisingly, the increase in output in China is relatively small as spillovers from the U.S. to China are very limited due to high barriers to trade. Moreover, Figure 20 shows that the sector-specific demand shift leads to a local investment boom as not only the demand for the non-durables sector, but also the demand for the durables sector, has relatively (and indirectly) increased.

The main difference compared to the results previously reported in Figures 8 and 9 is the lower magnitudes of the responses in the case of asymmetric trade costs. The reason for this lies in the *ex-ante* reliance of the U.S. on goods that are produced within the U.S. When trade costs are symmetric, the U.S. relies unconditionally more heavily on foreign goods (about 58% of absorption due to foreign goods) than in the case of the asymmetric costs (about 42% of absorption due to foreign goods). The reallocation of resources within-country across sectors is larger if the reliance on foreign goods is low. Moreover, any increase in investment is internalized to a larger extent, which ultimately results in the need for a smaller increase in investment.

Finally, we study the responses to an investment-specific shock in the U.S. Figures 21 and 22 report the IRFs of investment and output across the three regions. As in Section 5,

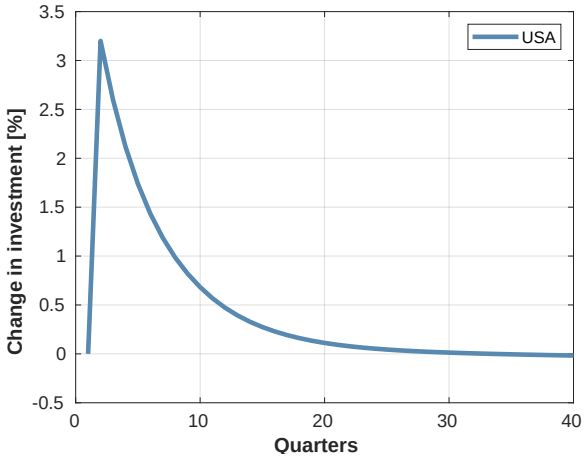

Figure 20: Response of investment in the U.S. to an unexpected sector-specific demand shift in the U.S., $\psi^N_{US,t}$.

an increase in the efficiency with which investment is converted into capital leads to a local investment boom and an increase in output for the U.S., but also frequent trade partners. Two main differences emerge when trade costs are asymmetric as opposed to symmetric.

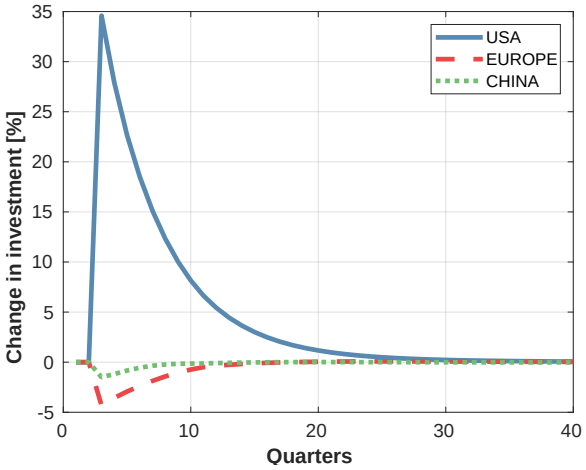

Figure 21: Country-specific IRFs of investment to an unexpected investment-specific technological shock in the U.S., $\chi^D_{US,t}$.

First, the magnitudes of the responses in U.S. investment and output are considerably larger when trade costs are asymmetric. The difference in magnitudes can be explained by the extent to which the positive investment-specific shock is internalized by the U.S. As the reliance on U.S. goods in the durables sector in the U.S. is higher in the asymmetric compared to the symmetric trade costs case (the mean of $\pi_{US,US,t}$ are 57.78% and 41.93%, respectively), the U.S. benefits to a larger extent from the improved local conditions.

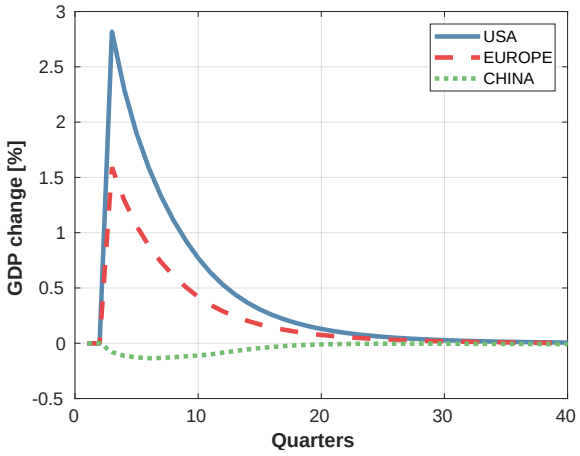

Figure 22: Country-specific IRFs of GDP to an unexpected investment-specific technological shock in the U.S., $\chi^D_{US,t}$.

Second, the responses of investment and output in China are very different across the two trade cost calibrations. As China relies heavily on the output produced internally ($\pi_{CH,CHI,t}$ = 92.41%), an increase in investment efficiency in the U.S. leaves investment and output in China almost unaffected.

7 Conclusion

We postulate and globally solve a model of Ricardian cycles, where shocks to different countries propagate through international trade in different sectors. The model features potentially many countries that each consist of three different sectors; durables, non-durables, and services. While durable and non-durable goods are tradable, services are not. Finally, all agents have rational expectations about the stochastic components of the model.

We solve two versions of the model, one with symmetric trade costs and one with asymmetric trade costs. Comparing the two model versions reveals that the relative strength of trade links across countries is indicative of the extent to which local shocks are transmitted to other countries. Country-specific shocks to demand, supply, and investment efficiency induce countries to engage in intra- and intertemporal substitutions in non-durable consumption, investment, services, and trade, generating business cycles.

With our methodology, many questions can be answered in future research: What is the role of endogenous technological change in Ricardian business cycles? What is the role

of possible nominal rigidities? And of financial frictions? What is the role of fiscal and monetary policy? We can also enrich the model with a climate change block and explore how international trade and shocks matter for climate change mitigation policies. We hope to see many papers addressing these issues in the near future.

References

Adjemian, Stéphane, Houtan Bastani, Michel Juillard, Frédéric Karamé, Junior Maih, Ferhat Mihoubi, George Perendia, Johannes Pfeifer, Marco Ratto, and Sébastien Villemot (2018) "Dynare: Reference Manual Version 4," Dynare Working Papers 1, CEPREMAP.

Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan (2010) "Inventories, Lumpy Trade, and Large Devaluations," *American Economic Review*, Vol. 100, No. 5, pp. 2304–39, December.

Álvarez, Fernando and Robert Lucas (2007) "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade," *Journal of Monetary Economics*, Vol. 54, No. 6, pp. 1726–1768.

Andreasen, Martin, Jesús Fernández-Villaverde, and Juan Rubio-Ramírez (2018) "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *Review of Economic Studies*, Vol. 85, pp. 1–49.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare (2012) "New Trade Models, Same Old Gains?" *American Economic Review*, Vol. 102, No. 1, pp. 94–130, February.

Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger (2022) "Deep Equilibrium Nets," *International Economic Review*, Vol. n/a, No. n/a.

Backus, David K., Patrick J. Kehoe, and Finn E. Kydtand (2020) "International Business Cycles: Theory and Evidence," in Thomas F. Cooley ed. *Frontiers of Business Cycle Research*: Princeton University Press, pp. 331–356.

Bellman, Richard E. (1961) *Adaptive Control Processes: A Guided Tour*: Princeton University Press.

Bergstra, James S, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl (2011) "Algorithms for Hyper-Parameter Optimization," in *Advances in neural information processing systems*, pp. 2546–2554.

Bronstein, Michael M., Joan Bruna, Taco Cohen, and Petar Velickovic (2021) *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*: https://arxiv.org/abs/2104.13478.

Brumm, Johannes and Simon Scheidegger (2017) "Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models," *Econometrica*, Vol. 85, No. 5, pp. 1575–1612.

Brumm, Johannes, Christopher Krause, Andreas Schaab, and Simon Scheidegger (2021) "Sparse Grids for Dynamic Economic Models,"Technical report, SSRN 3979412.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019) "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, Vol. 87, No. 3, pp. 741–835.

Cheela, Bhagath, André DeHon, Jesús Fernández-Villaverde, and Alessandro Peri (2022) "Programming FPGAs for Economics: An Introduction to Electrical Engineering Economics,"Technical Report 29936, National Bureau of Economic Research.

den Haan, Wouter J. and Albert Marcet (1990) "Solving the Stochastic Growth Model by Parameterizing Expectations," *Journal of Business & Economic Statistics*, Vol. 8, No. 1, pp. 31–34.

Duarte, Victor (2018) "Machine Learning for Continuous-Time Economics,"Technical report, SSRN. Working paper.

Duffy, John and Paul D. McNelis (2001) "Approximating and Simulating the Stochastic Growth Model: Parameterized Expectations, Neural Networks, and the Genetic Algorithm," *Journal of Economic Dynamics and Control*, Vol. 25, No. 9, pp. 1273–1303, September.

Eaton, Jonathan and Samuel Kortum (2002) "Technology, Geography, and Trade," *Econometrica*, Vol. 70, No. 5, pp. 1741–1779.

Eaton, Jonathan, Sam Kortum, and Brent Neiman (2016a) "Illustrating the Methodology in EKNR (2016): Some Simple Examples." Mimeo.

Eaton, Jonathan, Samuel Kortum, and Brent Neiman (2016b) "Obstfeld and Rogoff'ss international macro puzzles: a quantitative assessment," *Journal of Economic Dynamics and Control*, Vol. 72, pp. 5–23. Trade and Macroeconomics: New Approaches to Enduring Puzzles.

Ebrahimi Kahou, Mahdi, Jesús Fernández-Villaverde, Jesse Perla, and Arnav Sood (2021) "Exploiting Symmetry in High-Dimensional Dynamic Programming," Working Paper 28981, National Bureau of Economic Research.

Ebrahimi Kahou, Mahdi, Jesús Fernández-Villaverde Sebastián Gómez-Cardona, Jesse Perla, and Jan Rosa (2022) "Spooky Boundaries at a Distance: Exploring Transversality and Stability with Deep Learning,"Technical report, University of Pennsylvania. Mimeo.

Fernández-Villaverde, Jesús, Grey Gordon, Pablo Guerrón-Quintana, and Juan F Rubio-Ramirez (2015) "Nonlinear Adventures at the Zero Lower Bound," *Journal of Economic Dynamics and Control*, Vol. 57, pp. 182–204.

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Juan F. Rubio-Ramírez, and Martin Uribe (2011) "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, Vol. 101, No. 6, pp. 2530–2561.

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuño (2019) "Financial Frictions and the Wealth Distribution," Working Paper 26302, National Bureau of Economic Research.

Fernández-Villaverde, J., J.F. Rubio-Ramírez, and F. Schorfheide (2016) "Solution and Estimation Methods for DSGE Models," in John B. Taylor and Harald Uhlig eds. *Handbook of Macroeconomics*, Vol. 2: Elsevier, pp. 527–724.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016) *Deep Learning*: MIT Press. http://www.deeplearningbook.org.

Han, Jiequn, Yucheng Yang et al. (2021) "DeepHAM: A Global Solution Method for Heterogeneous Agent Models with Aggregate Shocks,"Technical report, arXiv preprint arXiv:2112.14377.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989) "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, Vol. 2, No. 5, pp. 359 – 366.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla (2015) "The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs," *Review of Economics and Statistics*, Vol. 97, No. 4, pp. 777–792.

Judd, Kenneth L (1998) *Numerical Methods in Economics*: The MIT press.

Judd, Kenneth L, Lilia Maliar, Serguei Maliar, and Rafael Valero (2014) "Smolyak Method for Solving Dynamic Economic Models: Lagrange Interpolation, Anisotropic Grid and Adaptive Domain," *Journal of Economic Dynamics and Control*, Vol. 44, pp. 92–123.

Kingma, Diederik P. and Jimmy Ba (2014) "Adam: A Method for Stochastic Optimization."

Krueger, Dirk and Felix Kubler (2004) "Computing Equilibrium in OLG models with Stochastic Production," *Journal of Economic Dynamics and Control*, Vol. 28, No. 7, pp. 1411 – 1436.

Ljungqvist, Lars and Thomas J Sargent (2004) *Recursive Macroeconomic Theory*: Mit Press.

Lucas, Robert E. and Edward C. Prescott (1971) "Investment Under Uncertainty," *Econometrica*, Vol. 39, No. 5, pp. 659–681.

Maliar, Lilia, Serguei Maliar, and Pablo Winant (2021) "Deep Learning for Solving Dynamic Economic Models," *Journal of Monetary Economics*, Vol. 122, pp. 76–101.

Montanelli, Hadrien. and Qiang. Du (2019) "New Error Bounds for Deep ReLU Networks Using Sparse Grids," *SIAM Journal on Mathematics of Data Science*, Vol. 1, No. 1, pp. 78–92.

Norets, Andriy (2012) "Estimation of Dynamic Discrete Choice Models Using Artificial Neural Network Approximations," *Econometric Reviews*, Vol. 31, No. 1, pp. 84–106.

Press, William H, Saul A Teukolsky, William T Vetterling, and Brian P Flannery (2007) *Numerical Recipes: The art of Scientific Computing*: Cambridge University Press, 3rd edition.

Primiceri, Giorgio and Alejandro Jusiniano (2008) "The Time Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, Vol. 98, pp. 604–641.

Primiceri, Giorgio, Alejandro Jusiniano, and Andrea Tambalotti (2011) "Investment Shocks and the Relative Price of Investment," *Review of Economic Dynamics*, Vol. 14, pp. 101–121.

Ravikumar, B., Ana Maria Santacreu, and Michael Sposi (2019) "Capital accumulation and dynamic gains from trade," *Journal of International Economics*, Vol. 119, No. C, pp. 93–110.

Ricardo, David (1821) *On the Principles of Political Economy and Taxation*: John Murray, 3rd edition.

Scheidegger, Simon and Ilias Bilionis (2019) "Machine Learning for High-Dimensional Dynamic Stochastic Economies," *Journal of Computational Science*, Vol. 33, pp. 68 – 82.

Schmitt-Grohé, Stephanie and Martın Uribe (2004) "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of economic dynamics and control*, Vol. 28, No. 4, pp. 755–775.

Smets, Frank and Rafael Wouters (2007) "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, Vol. 97, pp. 586–606.

——— (2010) "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, Vol. 1, pp. 1123–1175.

Spear, Stephen E. (1988) "Existence and Local Uniqueness of Functional Rational Expectations Equilibria in Dynamic Economic Models," *Journal of Economic Theory*, Vol. 44, No. 1, pp. 124–155, February.

Stock, James H. and Mark W. Watson (2003) "Has the Business Cycle Changed and Why?," in *NBER Macroeconomics Annual 2002, Volume 17*: National Bureau of Economic Research, Inc, pp. 159–230.

Stockman, Alan C. and Linda L. Tesar (1995) "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements," *American Economic Review*, Vol. 85, No. 1, pp. 168–185.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott (1989) *Recursive Methods in Economic Dynamics*: Harvard University Press.

Sutton, Richard S and Andrew G Barto (2018) *Reinforcement Learning: An Introduction*: MIT press.

Villa, Alessandro T and Vytautas Valaitis (2019) "Machine Learning Projection Methods for Macro-Finance Models,"Technical report, Duke University.

APPENDIX

This appendix includes additional details about our model and its implementation.

A Model

We have the regions in this model: USA – 1, EU – 2, China – 3. If not stated otherwise, the three regions are encoded with the above-mentioned numbering, with L_{USA} =2, L_{EU} =1.425, L_{CH} =1.1.

$$
\begin{aligned}
\mathcal{L} = \sum_{n=1}^{\mathcal{N}} \mathbb{E}_0 \Bigg\{ & \sum_{t=0}^{\infty} \rho^t \left[\omega_n \phi_{n,t} \left(\sum_{j \in \{S,N\}} \psi_{n,t}^j \ln C_{n,t}^j + \psi_n^D \ln K_{n,t}^{H,D} \right) \right] \\
& + \lambda_{n,t}^L \left(L_n - \sum_{j \in \Omega} \int_0^1 L_n^j(z) dz \right) + \lambda_{n,t}^K \left(K_{n,t}^D - \sum_{j \in \Omega} \int_0^1 K_{n,t}^{jD}(z) dz - K_{n,t}^{H,D} \right) \\
& + \sum_{j \in \Omega} \int_0^1 \lambda_{n,t}^j(z) \left(a_{n,t}^j(z) \left(\frac{L_n^j(z)}{\beta_n^{L,j}} \right)^{\beta_n^{L,j}} \left(\frac{K_{n,t}^{jD}(z)}{\beta_n^{K,jD}} \right)^{\beta_n^{K,jD}} \prod_{j' \in \Omega} \left(\frac{M_{n,t}^{jj'}(z)}{\beta_n^{M,jj'}} \right)^{\beta_n^{M,jj'}} - y_{n,t}^j(z) \right) dz \\
& + \sum_{j \in \Omega} \int_0^1 \hat{\lambda}_{n,t}^j(z) \left(y_{n,t}^j(z) - \sum_{m=1}^{\mathcal{N}} d_{mn,t}^j x_{mn,t}^j(z) \right) dz \\
& + \sum_{j \in \Omega} \int_0^1 \tilde{\lambda}_{n,t}^j(z) \left(\sum_{i=1}^{\mathcal{N}} x_{ni,t}^j(z) - x_{n,t}^j(z) \right) dz \\
& + \sum_{j \in \Omega} \lambda_{n,t}^j \left(\left(\int_0^1 x_{n,t}^j(z)^{(\sigma-1)/\sigma} dz \right)^{\sigma/(\sigma-1)} - x_{n,t}^j \right) \\
& + \sum_{h \in \{S,N\}} \tilde{\lambda}_{n,t}^h \left(x_{n,t}^h - \sum_{j \in \Omega} \int_0^1 M_{n,t}^{jh}(z) dz - C_{n,t}^h \right) + \tilde{\lambda}_{n,t}^D \left(x_{n,t}^D - \sum_{j \in \Omega} \int_0^1 M_{n,t}^{jD}(z) dz - I_{n,t}^D \right) \\
& + \lambda_{n,t}^{V^D} \left(\chi_{n,t}^D \left(I_{n,t}^D \right)^{\alpha^D} \left(K_{n,t}^D \right)^{1-\alpha^D} + \left(1 - \delta^D \right) K_{n,t}^D - K_{n,t+1}^D \right) \\
& + \sum_{j \in \Omega} \int_0^1 \bar{\lambda}_{n,t}^j(z) y_{n,t}^j(z) dz + \sum_{j \in \Omega} \sum_{i=1}^{\mathcal{N}} \int_0^1 \bar{\lambda}_{ni,t}^j(z) x_{ni,t}^j(z) dz \Bigg\},
\end{aligned}
$$

where each λ is the Lagrange multiplier associated with the corresponding constraint. The constraints include the sets discussed below together with non-negativity constraints on the $y_{n,t}^j(z)$'s and the $x_{ni,t}^j(z)$'s. The transversality conditions are $\lim_{t \to \infty} \rho^t \lambda_{n,t}^{V^D} K_{n,t+1}^D = 0$, for each $n = 1, \ldots, \mathcal{N}$.

A.1 Specialization of Production Goods

We start by deriving which countries produce each good, and to which other countries they ship it. The first-order condition with respect to shipments $x^j_{ni,t}(z)$ of good z in sector j from country i to n gives $\tilde{\lambda}^j_{n,t}(z) + \bar{\lambda}^j_{ni,t}(z) = \hat{\lambda}^j_{i,t}(z)d^j_{ni,t}$.

We need to consider two possibilities. If $\bar{\lambda}^j_{ni,t}(z) > 0$ then $\tilde{\lambda}^j_{n,t}(z) < \hat{\lambda}^j_{i,t}(z)d^j_{ni,t}$ and $x^j_{ni,t}(z) = 0$ while if $x^j_{ni,t}(z) > 0$, then $\bar{\lambda}^j_{ni,t}(z) = 0$ and $\tilde{\lambda}^j_{n,t}(z) = \hat{\lambda}^j_{i,t}(z)d^j_{ni,t}$. Since country n will obtain this good from somewhere, looking across all source countries i:

$$\tilde{\lambda}^j_{n,t}(z) = \min_i \left\{ \hat{\lambda}^j_{i,t}(z)d^j_{ni,t} \right\}.$$

The first-order condition with respect to production $y^j_{n,t}(z)$ of good z in sector j by country n is $\hat{\lambda}^j_{n,t}(z) + \bar{\lambda}^j_{n,t}(z) = \lambda^j_{n,t}(z)$.

Thus, $\lambda^j_{i,t}(z) \geq \hat{\lambda}^j_{i,t}(z)$ for all countries i, with equality if $y^j_{i,t}(z) > 0$. Since $x^j_{ni,t}(z) > 0$ implies $y^j_{i,t}(z) > 0$, we can rewrite the equation above as

$$\tilde{\lambda}^j_{n,t}(z) = \min_i \left\{ \lambda^j_{i,t}(z)d^j_{ni,t} \right\}. \tag{20}$$

Country i produces good z in sector j if and only if it achieves this minimum in some destination n.

A.2 The Cost of Production

Suppose country n does produce good z in sector j so that $y^j_{n,t}(z) > 0$. The first-order conditions for inputs of labor, capital, and intermediates to produce it gives us, for each $j \in \Omega$:

$$\lambda^j_{n,t}(z)\beta^{L,j}_n \frac{y^j_{n,t}(z)}{L^j_n(z)} - \lambda^L_{n,t} = 0$$
$$\lambda^j_{n,t}(z)\beta^{K,jD}_n \frac{y^j_{n,t}(z)}{K^{jk}_{n,t}(z)} = \lambda^{K^D}_{n,t}$$

and

$$\lambda^j_{n,t}(z)\beta^{M,jj'}_n \frac{y^j_{n,t}(z)}{M^{jj'}_{n,t}(z)} = \lambda^{j'}_{n,t}$$

for $j' \in \Omega$.

We can relate the shadow cost of producing a good to the shadow costs of the inputs used to

produce it. Multiplying the production function by the associated shadow value of output, we get:

$$Y^j_{n,t}(z) = \lambda^j_{n,t}(z) y^j_{n,t}(z) = \lambda^j_{n,t}(z) a^j_{n,t}(z) \left(\frac{L^j_n(z)}{\beta^{L,j}_n}\right)^{\beta^{L,j}_n} \left(\frac{K^{jD}_{n,t}(z)}{\beta^{K,jD}_n}\right)^{\beta^{K,jD}_n} \prod_{j' \in \Omega} \left(\frac{M^{jj'}_{n,t}(z)}{\beta^{M,jj'}_n}\right)^{\beta^{M,jj'}_n}.$$

Inserting the first-order conditions given above for inputs implies:

$$Y^j_{n,t}(z) = \lambda^j_{n,t}(z) a^j_{n,t}(z) \left(\frac{Y^j_{n,t}(z)}{\lambda^L_{n,t}}\right)^{\beta^{L,j}_n} \left(\frac{Y^j_{n,t}(z)}{\lambda^{K^D}_{n,t}}\right)^{\beta^{K,jD}_n} \prod_{j' \in \Omega} \left(\frac{Y^j_{n,t}(z)}{\lambda^{j'}_{n,t}}\right)^{\beta^{M,jj'}_n}.$$

Constant returns to scale implies that $Y^j_{n,t}(z)$ cancels, giving us the shadow value of good z in sector j in country n:

$$\lambda^j_{n,t}(z) = \frac{c^j_{n,t}}{a^j_{n,t}(z)}, \tag{21}$$

where the term $c^j_{n,t} = \left(\lambda^L_{n,t}\right)^{\beta^{L,j}_n} \left(\lambda^{K^D}_{n,t}\right)^{\beta^{K,jD}_n} \prod_{j' \in \Omega} \left(\lambda^{j'}_{n,t}\right)^{\beta^{M,jj'}_n}$ bundles the shadow costs of labor, capital, and intermediates in producing any good in sector j in country n. Applying equation (21) allows us to write equation (20) as:

$$\tilde{\lambda}^j_{n,t}(z) = \min_i \left\{ \frac{c^j_{i,t}}{a^j_{i,t}(z)} d^j_{ni,t} \right\}.$$

A.3 Demand for Goods

Now we take the first-order condition with respect to $x^j_{n,t}(z)$ to get $\tilde{\lambda}^j_{n,t}(z) = \lambda^j_{n,t} \left(x^j_{n,t}\right)^{1/\sigma} x^j_{n,t}(z)^{-1/\sigma}$, which we can rearrange as:

$$x^j_{n,t}(z) = \left(\frac{\tilde{\lambda}^j_{n,t}(z)}{\lambda^j_{n,t}}\right)^{-\sigma} x^j_{n,t}.$$

Letting $X^j_{n,t}(z) = \tilde{\lambda}^j_{n,t}(z) x^j_{n,t}(z)$ and $X^j_{n,t} = \lambda^j_{n,t} x^j_{n,t}$, we obtain:

$$X^j_{n,t}(z) = \left(\frac{\tilde{\lambda}^j_{n,t}(z)}{\lambda^j_{n,t}}\right)^{-(\sigma-1)} X^j_{n,t}. \tag{22}$$

We can aggregate over the absorption of individual goods using $x^j_{n,t} = \left(\int_0^1 x^j_{n,t}(z)^{(\sigma-1)/\sigma} dz\right)^{\sigma/(\sigma-1)}$. In combination with equation (22), we get:

$$X^j_{n,t} = \int_0^1 X^j_{n,t}(z) dz. \tag{23}$$

Integrating both sides of equation (22) and applying equation (23), we also get:

$$\lambda^j_{n,t} = \left(\int_0^1 \tilde{\lambda}^j_{n,t}(z)^{-(\sigma-1)} dz\right)^{-1/(\sigma-1)}. \tag{24}$$

To obtain sharper results for aggregates, we need to exploit our assumption on the distribution of good-level production efficiency.

A.4 International Trade

We now view the problem from the perspective of not knowing the individual realizations of efficiency $a^j_{n,t}(z)$, but only the parameters of the distribution from which they are drawn:

$$F^j_{n,t}(a) = \Pr\left[a^j_{n,t}(z) \leq a\right] = \exp\left[-\left(\frac{a}{\gamma A^j_{n,t}}\right)^{-\theta}\right].$$

We can derive the probability distribution function $G^j_{n,t}(x)$ of the $\tilde{\lambda}^j_{n,t}(z)$ s as:

$$\begin{aligned} G^j_{n,t}(x) &= \Pr\left[\tilde{\lambda}^j_{n,t}(z) \leq x\right] = 1 - \Pr\left[\min_i \left\{\frac{c^j_{i,t} d^j_{ni,t}}{a^j_{i,t}(z)}\right\} > x\right] \\ &= 1 - \prod_i \Pr\left[a^j_{i,t}(z) < \frac{c^j_{i,t} d^j_{ni,t}}{x}\right] = 1 - \prod_i \exp\left[-\left(\frac{c^j_{i,t} d^j_{ni,t}}{\gamma A^j_{i,t} x}\right)^{-\theta}\right] \\ &= 1 - e^{-\Phi^j_{n,t}(\gamma x)^\theta}, \end{aligned}$$

where $\Phi^j_{n,t} = \sum_{i=1}^N \left(\frac{c^j_{i,t} d^j_{ni,t}}{A^j_{i,t}}\right)^{-\theta}$.

We can use this distribution to simplify the integral in equation (24):

$$\lambda^j_{n,t} = \left(\int_0^\infty x^{-(\sigma-1)} dG^j_{n,t}(x)\right)^{-1/(\sigma-1)} = \left(\Phi^j_{n,t}\right)^{-1/\theta}.$$

In the case of the non-tradable service sector, S, this equation becomes:

$$\lambda_{n,t}^{S} = \left(\int_0^{\infty} x^{-(\sigma-1)} dG_{n,t}^{S}(x) \right)^{-1/(\sigma-1)} = \left(\frac{c_{n,t}^{S}}{A_{n,t}^{S}} \right)^{-1/\theta}.$$

This expression for the shadow value of sector j absorption is the same as that for the price index in Eaton and Kortum (2002). Following the derivation there, the fraction of goods for which country i achieves the minimum in country n is:

$$\pi_{ni,t}^{j} = \frac{\left(c_{i,t}^{j} d_{ni,t} / A_{i,t}^{j} \right)^{-\theta}}{\Phi_{n,t}^{j}} = \left(\frac{c_{i,t}^{j} d_{ni,t}}{A_{i,t}^{j} \lambda_{n,t}^{j}} \right)^{-\theta}.$$

We define the shadow value of all deliveries to country n of sector j goods from country i as:

$$X_{ni,t}^{j} = \int_0^1 \tilde{\lambda}_{n,t}^{j}(z) x_{ni,t}^{j}(z) dz.$$

Since the distribution of $\tilde{\lambda}_{n,t}^{j}(z)$ is the same regardless of the country i from which the goods are shipped, and the fraction of goods shipped from i is $\pi_{ni,t}^{j}$:

$$X_{ni,t}^{j} = \pi_{ni,t}^{j} \int_0^1 \tilde{\lambda}_{n,t}^{j}(z) x_{n,t}^{j}(z) dz = \pi_{ni,t}^{j} \int_0^1 X_{n,t}^{j}(z) dz = \pi_{ni,t}^{j} X_{n,t}^{j}.$$

Integrating over all sector j goods produced in n, we define the value of production as $Y_{n,t}^{j} = \int_0^1 Y_{n,t}^{j}(z) dz$.

Summing across destinations, we can connect the value of production and the value of deliveries to each country $Y_{n,t}^{j} = \sum_{m=1}^{\mathcal{N}} X_{mn,t}^{j} = \sum_{m=1}^{\mathcal{N}} \pi_{mn,t}^{j} X_{m,t}^{j}$.

A.5 Consumption and Investment

The first-order condition for absorption $x_{n,t}^{j}$ of sector $j \in \Omega$ output in country n at date t is simply $\tilde{\lambda}_{n,t}^{j} = \lambda_{n,t}^{j}$. Hence, we drop $\tilde{\lambda}_{n,t}^{j}$ and replace it with $\lambda_{n,t}^{j}$ in the expressions for consumption and investment. The first-order condition for consumption $C_{n,t}^{h}$ for $h \in \Omega_K^*$ can be written as $\lambda_{n,t}^{h} C_{n.t}^{h} = \omega_n \phi_{n,t} \psi_{n,t}^{h}$, while the first-order condition for household capital services $K_{n,t}^{H,D}$ gives $\lambda_{n,t}^{K} K_{n,t}^{H,D} = \omega_n \phi_{n,t} \psi_{n}^{D}$.

Turning to investment, the first-order condition for $I^D_{n,t}$ is:

$$\lambda^{V^D}_{n,t} = \frac{\lambda^D_{n,t}}{\alpha^D \chi^D_{n,t}} \left(\frac{I^D_{n,t}}{K^D_{n,t}} \right)^{1-\alpha^D} .$$

A.6 GDP Calculations

In our model, global GDP is $Y_t = \sum_{n=1}^{\mathcal{N}} \left[w_{n,t} L_n + r_{n,t} \left(K^D_{n,t} - K^{H,D}_{n,t} \right) \right]$, where labor income $w_{n,t} L_n$ is given by (12) and capital incomes $r_{n,t} K^D_{n,t}$ by equation (13). To avoid valuation problems, we exclude rental payments by households on durables in our measure of GDP. Moreover, country-specific GDP is then trivially given by $Y_{n,t} = w_{n,t} L_n + r_{n,t} \left(K^D_{n,t} - K^{H,D}_{n,t} \right)$.

B Implementation Details

We briefly describe some of the challenges our DEQN solution method faces in practical applications and how we deal with them. A drawback of deep neural networks is that the theoretical convergence rates and why and when deep neural networks can ameliorate or break the curse of dimensionality are poorly understood. Even though substantial progress has recently been made (Montanelli and Du, 2019), our method still relies on a non-trivial amount of numerical experimentation.

B.1 Neural Network Architecture

In our model, the dimensionality of the "economic"state vector is given by expression (16) and is of dimensionality $M = 6\mathcal{N} + (\mathcal{N} - 1) + 2(\mathcal{N}^2 - \mathcal{N})$, where $\mathcal{N} \in \mathbb{N}$, is the number of countries/regions in the model (cf. Appendices C.1 and C.2). Thus, if we want to work with three regions—that is, the USA, Europe, and China—as we do in our quantitative exercise in Sections 5 and 6, there are 32 state variables. The policy vector $\mathbf{p}$ on the other hand, is given by (17), and is of dimensionality $K = 23\mathcal{N} + 2\mathcal{N}^2$. Hence, if we stay with the three-regions example, a 32-dimensional state-space maps into a 87-dimensional output vector.

We represent the mapping of states to policies with a deep neural network of the following architecture: After the M-dimensional input layer, the neural network features three hidden layers with $128, 128$, and 128 selu-activated, and batch-normalized hidden nodes. The output layer consists of K nodes, activated with softplus functions to ensure the non-negativity of the policies (see Appendix C.1).

B.2 Auxiliary State Variables to Normalize the Nonlinear Set of Equations

One issue when training the neural network is that the different components that enter the loss function (18) can substantially vary in magnitude, causing the optimizer to focus too much on getting a particular set of policies right while neglecting others. To mitigate this issue, we normalize the equations by introducing one auxiliary state variable per nonlinear equation such that the equations enter the loss function in percent units.

The basic idea is as follows. Suppose that we have a simple set of nonlinear equations:

$$\begin{aligned} a \cdot x_1^2 &= b \cdot y_1^2 + c, \\ d \cdot x_2^2 &= e \cdot y_2^2 + f, \end{aligned}$$

with a two-dimensional state $x = (x_1, x_2)$, and a policy $y = (y_1, y_2)$. The expressions one would naively stick into the loss function thus would read:

$$0 = \underbrace{\left(b \cdot y_1^2 + c - a \cdot x_1^2\right)}_{:=l_1} + \underbrace{\left(e \cdot y_2^2 + f - d \cdot x_2^2\right)}_{:=l_2}.$$

However, the magnitudes of l_1 and l_2 could vary substantially and be unrelated to their economic importance. For example, even if $l_1 = 0.01$ and $l_2 = 0.001$, the relative error of l_2 can be much worse than that of l_1 in terms of its effect on the difference between the computed policy y and the exact one. To get around this issue, we can use the normalizations:

$$\begin{aligned} 0 &= \frac{1}{A}(b \cdot y_1^2 + c - a \cdot x_1^2), \\ 0 &= \frac{1}{B}(e \cdot y_2^2 + f - d \cdot x_2^2), \end{aligned}$$

where $A := a \cdot x_1^2$, and $B := d \cdot x_2^2$.

With this basic idea at hand, the question now is how to implement it within the DEQN algorithm. In this simulation-based setting, a "normalizing"constant has to be the mean along the simulated path. To this end, we introduced one extra state variable per nonlinear equation that enters the loss function. The latter is updated with exponential smoothing from one episode to the next (cf. Algorithm 1). In the concrete case of the stylized example of this section, A and B would be extra state variables and updated, for instance, as $A = \gamma \cdot A_{old} + (1 - \gamma) \cdot (a \cdot x_1^2)$, where A_{old} stems from the

previous episode, $0 < \gamma < 1$, and were in practice, $\gamma = 0.99$.

Normalizing the equations entering the overall loss function helps the convergence process of our solution substantially. However, it comes at the price that the effective dimensionality of the neural network is not M, but rather $M + K$ (since for all K equations, an auxiliary state is added), which increases the memory consumption moderately. In the case of a 3-region model, the effective input of the neural network thus is given by a $32 + 87 = 119$-dimensional vector.

B.3 A DEQN-based Integration Scheme

Evaluating the expectation operator in the Euler equations (9) and (43) can be a daunting task due to the curse of dimensionality. One way to mitigate this issue (depending on the underlying distribution of the shocks) is to apply a simple monomial rule that scales linearly with the number of continuous state variables (see, e.g., Judd, 1998, and references therein).

We, however, propose an alternative and generic way to compute expectations which leverages the algorithmic properties of DEQN and combines ideas from Appendix B.2 with "Monte-Carlo integration" (Press et al., 2007) and the "Parametrized Expectations" literature (see den Haan and Marcet, 1990, for the original method, and Duffy and McNelis, 2001, and Villa and Valaitis, 2019, for extensions with neural networks).

The core idea is as follows: Evaluating the expectation operator accurately with a Monte Carlo method typically requires many sample points as its convergence rate is proportional to $1/\sqrt{M}$, where M is the number of observations. However, as we run the DEQN for many episodes, we are able to "learn" the expectation as time evolves, having only a few samples per episode (Algorithm 1). To do so, we proceed similarly as in section B.2. Abstractly, the Euler equation is $\mathcal{E} = \mathbb{E}(\cdot)$. Next, we re-write it as $0 = \frac{1}{\mathcal{E}}(\mathbb{E}(\cdot) - \mathcal{E})$, where $\mathcal{E}$ is again an auxiliary state variable that is updated with an exponentially weighted moving average:

$$\mathcal{E} = \alpha \cdot \mathcal{E}_{old} + (1 - \alpha) \cdot \mathbb{E}(\cdot),$$

where $0 < \alpha < 1$, and $\mathcal{E}_{old}$ is the pseudo-state from the previous episode. By setting α to a relatively large value, for example, $\alpha = 0.999$, while at the same time generating a small number of Monte Carlo samples per simulated state in each episode to compute $\mathbb{E}(\cdot)$, say $M = 20$, the actual sample size for $\mathcal{E}$, the "true" expectation, is still large. As a result, the computation in each episode is slowed down much less compared to traditional and expensive quadrature rules.

B.4 Simulating and Training of the DEQN

We train the network by simulating in each episode 512 individual states for 128 steps. The minibatch size is 256, and we use the `Adam` optimizer (Kingma and Ba, 2014). To ensure that we do not leave the ergodic distribution in the impulse-response experiments we carry out, we must ensure that the simulated training data covers a reasonably-wide domain. To this end, we generated the simulation paths with 2.5× the volatilities that are used in the calibration of the model—a common practice in reinforcement learning (see, e.g., Sutton and Barto, 2018, chapter 5). However, the actual model equations are still solved with the correct specification of the random shocks occurring in the model.

C Code

C.1 Variables (*code variables: states end with 'x', controls with 'y'*)

1. The set of endogenous state variables has the following elements:

   KD_n_t

   CODE: KD_n_x (> 0).

2. The set of exogenous state variables has the following elements (n: country index):

   $T^D_{n,t}$, $T^N_{n,t}$, $T^S_{n,t}$, $\phi_{n,t}$, $\psi^N_{n,t}$, $d^D_{ni,t}$, $d^N_{ni,t}$.

   CODE: TD_n_x, TN_n_x, TS_n_x, phi_n_x, psi_N_x, $dD\_n\_i\_x$, $dN\_n\_i\_x$.

3. The set of control variables has the following elements (n: country index):

   $A^D_{n,t}$, $A^N_{n,t}$, $A^S_{n,t}$, $\psi^S_{n,t}$, $p^D_{n,t}$, $p^N_{n,t}$, $p^S_{n,t}$, $C^h_{n,t}$, $r_{n,t}$, $K^{H,D}_{n,t}$, $Y^D_{n,t}$, $Y^S_{n,t}$, $Y^N_{n,t}$, $w_{n,t}$, $X^{F,N}_{n,t}$, $X^{F,S}_{n,t}$, $c^D_{n,t}$, $c^N_{n,t}$, $c^S_{n,t}$, $\pi^D_{ni,t}$, $\pi^N_{ni,t}$, $X^D_{n,t}$, $X^N_{n,t}$, $X^S_{n,t}$, $I^D_{n,t}$.

   CODE: AD_n_y (> 0), AN_n_y (> 0), AS_n_y (> 0), $psi\_S\_n\_y$ (> 0), $p\_D\_n\_y$ (> 0), $p\_N\_n\_y$ (> 0), $p\_S\_n\_y$ (> 0), $C\_h\_n\_y$ (> 0), r_n_y (> 0), $K\_HD\_n\_y$ (> 0), YD_n_y (> 0), YS_n_y (> 0), YN_n_y (> 0), w_n_y (> 0), $XF\_N\_n\_y$ (>0), $XF\_S\_n\_y$ (>0), $cost\_D\_n\_y$ (>0), $cost\_N\_n\_y$ (>0), $cost\_S\_n\_y$ (>0), $piD\_n\_i\_y$ (≥ 0; this policy is going to be a complicated one), $piN\_n\_i\_y$ (≥ 0; this policy is going to be a complicated one), XD_n_y (>0), XN_n_y (>0), XS_n_y (>0), ID_n_y (>0).

C.2 Equations

Legend: states in green (given), controls in orange (given by policy guess from current states), next period's exogenous states in magenta (to be integrated over), and next period's controls in blue (given by policy guess at next period's states), parameters are black.

$N \times 3$ equations. Sectoral productivity (in code: **sectoral_productivity**):

$$0 = (1/\gamma)\left(T^{j}_{n,t}\right)^{1/\theta} - A^{j}_{n,t}. \tag{25}$$

The $A^{j}_{n,t}$ are organized in a $N \times 3$ matrix $\mathbf{A_t}$.

N equations. Household spending on consumption good of sector S (in code: **hh_spending_S**):

$$0 = p^{h=S}_{n,t} C^{h=S}_{n,t} - \omega_n \phi_{n,t} \psi^{S}_{n,t}. \tag{26}$$

and for sector N accordingly (in code: **hh_spending_N**):

$$0 = p^{h=N}_{n,t} C^{h=N}_{n,t} - \omega_n \phi_{n,t} \psi^{N}_{n,t}. \tag{27}$$

N equations. Where (in code: **Psi_S_Def**):

$$\psi^{S}_{n,t} = 1 - \psi^{D}_{n} - \psi^{N}_{n,t}. \tag{28}$$

N equations.

Household spending on capital of sector D (in code: **hh_spending_cap_D**)

$$0 = r_{n,t} K^{H,D}_{n,t} - \omega_n \phi_{n,t} \psi^{D}_{n}. \tag{29}$$

N equations.

Wage (in code: **wage**)

$$0 = w_{n,t} L_n - \sum_{j \in \Omega} \beta^{L,j}_{n} Y^{j}_{n,t}. \tag{30}$$

The $w_{n,t}$ are organized in a $N \times 1$ vector $\mathbf{w_t}$.

N equations. Rental rate (in code: **rental_rate**):

$$0 = r_{n,t} K^D_{n,t} - \sum_{j \in \Omega} \beta^{K,jD} Y^j_{n,t} - \frac{\psi^D_n}{1-\psi^D_n} \left(X^{F,N}_{n,t} + X^{F,S}_{n,t} \right). \tag{31}$$

The $r_{n,t}$ are organized in a $N \times 1$ vector $\mathbf{r_t}$.

$N \times 3$ equations. The cost of a bundle of factors (in code: **cost_bundle_D,N,S**):

$$0 = c^j_{n,t} - (w_{n,t})^{\beta^{L,j}_n} (r_{n,t})^{\beta^{K,jD}_n} \prod_{j' \in \Omega} \left(p^{j'}_{n,t} \right)^{\beta^{M,j,j'}_n}. \tag{32}$$

The $c^j_{n,t}$ are organized in a $N \times 3$ matrix $\mathbf{c_t}$.

$N \times 2$ equations. The price index of the tradable sectors $l \in \{D, N\}$ (in code: **price_index_D,N**):

$$0 = p^l_{n,t} - \left(\sum_{i=1}^{N} \left(\frac{c^l_{i,t} d^l_{ni,t}}{A^l_{i,t}} \right)^{-\theta} \right)^{-\frac{1}{\theta}}. \tag{33}$$

And N equations for the service sector:

$$0 = p^S_{n,t} - \left(\frac{c^S_{n,t}}{A^S_{n,t}} \right)^{-1/\theta}. \tag{34}$$

The fraction of goods from tradable sector $l \in \{D, N\}$ that country n obtains as imports from country i (in code: **imports_D,N**):

$$0 = \pi^l_{ni,t} - \left(\frac{c^l_{i,t} d^l_{ni,t}}{p^l_{n,t} A^l_{i,t}} \right)^{-\theta}. \tag{35}$$

These fractions imported form a sector-specific $N \times N$ matrix $\mathbf{\Pi^l_t}$.

$N \times 2$ equations. The gross production of the tradable good in sector $l \in \{D, N\}$ in a country equals (in code: **absoprtion_D,N**):

$$0 = \mathbf{\Pi^l_t X^l_t} - \mathbf{Y^l_t}. \tag{36}$$

N equations. On the other hand, the gross production for the non-tradable good in sector S in a

country equals its absorption (in code: **gross-prod_S**):

$$0 = \mathbf{X_t^S} - \mathbf{Y_t^S}. \tag{37}$$

$N\times 2$ equations. Value of country n's final spending on sector $h \in \{S, N\}$ (in code: **final_spending_S,N**):

$$0 = X_{n,t}^{F,h} - p_{n,t}^{h} C_{n,t}^{h}. \tag{38}$$

N equations. Value of country n's final spending on sector D (in code: **final_spending_D**):

$$0 = X_{n,t}^{F,D} - p_{n,t}^{D} I_{n,t}^{h}. \tag{39}$$

N equations. Total spending on sector j output is the sum of country n' s final spending on sector j plus the use of sector j output as intermediates by each sector j' (in code: **total_spending_D,N,S**):

$$0 = X_{n,t}^{j} - X_{n,t}^{F,j} - \sum_{j' \in \Omega} \beta^{M,j'j} Y_{n,t}^{j'}. \tag{40}$$

N equations. The law of motion of capital (multiply with p in code to avoid a singularity, in code: **law_of_motion_D**):

$$0 = K_{n,t+1}^{D} - \chi_{n,t}^{D} \left(\frac{X_{n,t}^{F,D}}{p_{n,t}^{D}} \right)^{\alpha^D} {K_{n,t}^{D}}^{1-\alpha^D} - \left(1 - \delta^D\right) K_{n,t}^{D}. \tag{41}$$

N equations. The Euler equations (in code: **EE**):

$$\begin{aligned} 0 \;=\; & \frac{p_{n,t}^{D}}{\alpha^D \chi_{n,t}^{D}} \left(\frac{X_{n,t}^{F,D}}{p_{n,t}^{D} K_{n,t}^{D}} \right)^{1-\alpha^D} \\ & - \rho \frac{p_{n,t+1}^{D}}{\alpha^D \chi_{n,t+1}^{D}} \left(\frac{X_{n,t+1}^{F,D}}{p_{n,t+1}^{D} K_{n,t+1}^{D}} \right)^{1-\alpha^D} \left[\chi_{n,t+1}^{D} (1-\alpha^D) \left(\frac{X_{n,t+1}^{F,D}}{p_{n,t+1}^{D} K_{n,t+1}^{D}} \right)^{\alpha^D} + (1-\delta^D) \right] - \rho r_{n,t+1}. \end{aligned} \tag{42}$$

N equations. The Euler equations with rational expectations (in code: **EE**):

$$0 = \frac{p^D_{n,t}}{\alpha^D \chi^D_{n,t}} \left(\frac{X^{F,D}_{n,t}}{p^D_{n,t} K^D_{n,t}} \right)^{1-\alpha^D} - \rho \mathbb{E} \Bigg[r_{n,t+1} + \frac{p^D_{n,t+1}}{\alpha^D \chi^D_{n,t+1}} \left(\frac{X^{F,D}_{n,t+1}}{p^D_{n,t+1} K^D_{n,t+1}} \right)^{1-\alpha^D} \left[\chi^D_{n,t+1} (1-\alpha^D) \left(\frac{X^{F,D}_{n,t+1}}{p^D_{n,t+1} K^D_{n,t+1}} \right)^{\alpha^D} + (1-\delta^D) \right] \Bigg]. \quad (43)$$

C.3 Laws of Motion for Exogenous States

$N \times 3$ equations:

$$\ln T^j_{n,t} = \rho_T \ln T^j_{n,t-1} + \varepsilon_{T^j_n,t}. \quad (44)$$

$N - 1$ equations:

$$\ln \phi_{n,t} = \rho_{\phi_n} \ln \phi_{n,t-1} + \varepsilon_{\phi_n,t}. \quad (45)$$

1 equation (is put as definition; one state less!):

$$\phi_{N,t} = N_{country} - \sum_{i=1}^{N-1} \phi_{i,t}. \quad (46)$$

N equations:

$$\ln \chi^D_{n,t} = \rho_{\chi^D_n} \ln \chi^D_{n,t-1} + \varepsilon_{\chi^D_n,t}. \quad (47)$$

$N \times (N-1) \times 2$ equations (for $l \in \{D, N\}$):

$$\ln d^l_{ni,t} = \left(1 - \rho_{d^l_{ni}}\right) \bar{d}^l_{ni} + \rho_{d^l_{ni}} \ln d^l_{ni,t-1} + \varepsilon_{d^l_{ni},t}. \quad (48)$$

N equations (for $l \in \{D, N\}$):

$$\ln d^l_{nn,t} = 0. \quad (49)$$

N equations:

$$\psi^N_{n,t} = \left(1 - \rho_{\psi_n}\right) \frac{1 - \psi^D_n}{2} + \rho_{\psi_n} \psi^N_{n,t-1} + \varepsilon_{\psi_n,t}. \tag{50}$$