FTG Working Paper Series

Production, Trade, and Cross-Border Data Flows

by

Lin William Cong (??)
Qing Chang
Longtian Zhang
Liyong Wang

Working Paper No. 00135-00

Finance Theory Group

www.financetheory.com

Production, Trade, and Cross-Border Data Flows*

Qing Chang Lin William Cong Liyong Wang Longtian Zhang

First draft: September 2022; this draft: June 2023.

Abstract

We build a tractable general equilibrium model to analyze the effects of cross-border data flows and pre-existing development gaps in data economies on each country's production and international trade. Raw data as byproducts of consumption can be transformed into various types of working data (information) to be used by both domestic and foreign producers. Because data constitute a new production factor for intermediate goods, a large extant divide in data utilization can reduce or even freeze trade. Cross-border data flows mitigate the situation and improve welfare when added to international trade. Data-inefficient countries where data are less important in production enjoy a "latecomer's advantage" with international trade and data flows, contributing more raw data from which the data-efficient countries generate knowledge for production. Furthermore, cross-border data flows can reverse the cyclicity of working data usage after productivity shocks, whereas shocks to data privacy or import costs have opposite effects on domestic and foreign data sectors. The insights inform future research and policy discussions concerning data divide, data flows, and their implications for trade liberalization, the data labor market, among others.

Keywords: Data Sharing, Digital Economy, International Economics, Transition Dynamics

JEL Classification: F15, F29, F43

*The authors are especially grateful to Laura Veldkamp and Shengxing Zhang for their detailed feedback and invaluable guidance. They also thank Cui Hu, Shota Ichihashi, Ting Ji, Ernest Liu, Yue Liu, Jingye Wang, Danxia Xie, Liyan Yang, Zhongchao Yang, and seminar and conference participants at the Peking University National School of Development & Institute of Digital Finance FinTech Workshop, Central University of Finance and Economics & China Agricultural University Workshop, 2023 China Financial Research Conference, 2023 Annual Meeting of Northern Financial Association, and 2023 Asian Meeting of the Econometric Society at Tsinghua University and Nanyang Technological University for helpful comments and suggestions. Lavinia Burchielli and Vishal Prabakaran provided excellent assistance in proofreading the manuscript. The authors acknowledge the financial support from Ripple's University Blockchain Research Initiative (UBRI) and the FinTech at Cornell Initiative, and from the National Natural Science Foundation of China (Grant No. 72273161, 72303261). Chang is at Central University of Finance and Economics School of Statistics and Mathematics; Cong is at Cornell University SC Johnson College of Business and NBER; Wang and Zhang are both at Central University of Finance and Economics School of International Trade and Economics. Send correspondence to Cong at will.cong@cornell.edu.

1. Introduction

With the rapid technological advancement involving the Internet, artificial intelligence, large-scale computation, etc., data not only grow exponentially, but also have become an indispensable production factor in all major economies—a "new oil" in the information age.[1] Data enter the intermediate goods production with long-term effects on innovation, growth, and macroeconomic outcomes (Jones and Tonetti, 2020; Cong et al., 2021; Veldkamp and Chung, 2023), while affecting how firms operate and compete (e.g., Farboodi et al., 2019; Eeckhout and Veldkamp, 2023). The most salient feature of data, non-rivalry, makes their reproduction and sharing much easier than that of other production factors, even across countries. Establishing effective digital connections across countries intuitively facilitates communication and trade, which promote economic growth (Cory, 2017; Jouanjean, 2019; Buera and Oberfield, 2020; van der Marel and Ferracane, 2021). Yet, despite the accelerating the pace of data flows due to the COVID-19 pandemic, many countries have erected barriers to such cross-border flows, e.g., by passing laws or confining data within a country's borders, a concept known as "data localization" that is motivated by mercantilism, protectionism, or national security and privacy concerns.[2]

What roles do data and their cross-border flows play in production and international trade? Which countries benefit from such flows? How do pre-existing differences in the

[1]For example, "The world's most valuable source is no longer oil, but data." *The Economist*, May 6, 2017. The International Data Corporation also predicts in the book "Digital Age 2025" the total quantity of data to reach 175 Zettabytes by 2025.

[2]Based on OECD market regulation data, the Information Technology & Innovation Foundation finds that a 1% increase in a nation's data restrictiveness cuts its gross trade output by 7%, slows its productivity by 2.9%, and hikes the downstream prices by 1.5% over a 5-year-period (Cory and Dascoli, 2021). According to a Brookings Institution study, the cross-border flows of global data contributed as much as 10.1% to global economic growth from 2009 to 2018. In particular, the value contribution of cross-border data flows to global economic growth in 2014 exceeded $2.8 trillion, and this figure is expected to exceed $11 trillion by 2025. The World Economic Forum released in 2019 "A Brief History of Globalization" stating that we are entering "Globalization 4.0," in which cross-border data flows have become a crucial force in shaping international trade. Examples of data barriers are discussed in "China Locks Information on the Country Inside a Black Box" by Wei, Kubota, and Trumpf, *The Wall Street Journal*, April 30, 2023 and "China's Data-Security Laws Rattle Western Business Executives," *The Economist*, May 4, 2023. Chinese platforms supplying information about the Chinese economy and companies, such as WIND, Qichacha, and CNKI, are allegedly coerced by their domestic controllers to curtail their foreign services. Restricting data flow can also take the form of banning foreign companies that are data-intensive. For example, India banned many Chinese apps in 2020; Britain, Canada, and European parliaments have banned TikTok from official devices; and Montana became the first U.S. state to pass legislation in April 2023 banning TikTok on all personal devices.

development of data economy affect trade? How would domestic and foreign data usage and labor markets evolve after shocks to data productivity, privacy cost, and flow frictions? Finally, should a country allow importing or exporting data on top of trade? To understand the tradeoffs involved and design effective national policies for international trade and data sharing, one needs a theory of how data and their cross-border flows interact with production, trade, macroeconomic shocks, and the development of the data economy.

To this end, we build the first general equilibrium model of production and trade in a global economy, where data play crucial roles as input factors in production both domestically and abroad. Our model features representative households, data intermediaries, and production sectors that include final good producers, intermediate good producers, and wholesale good producers in open economies. Our analysis is designed to assess the impact of data and cross-border data flows in various settings qualitatively, ranging from a closed economy, to partially open economies with only goods traded or unilateral data flows, and finally to a fully open economy. We find that: (i) International data flows significantly improve welfare in steady states, especially for countries more backward in its data economy—a latecomer's advantage; (ii) trade liberalization (including goods and data) only happens when the pre-existing divide in effectively utilizing data between the two countries is sufficiently small, and is facilitated by data flows; (iii) with cross-border data flows, more working data are concentrated in the data-efficient country for production whereas the data-inefficient country provides more raw data; and (iv) open economies with data flows experience reversed cyclicity in data usage after a productivity shock, compared with that in a closed economy, when pre-existing data divide is not too large; shocks to data privacy and those to flow costs have opposite effects on domestic and foreign data sectors.

Specifically, we follow the seminal work of Ichihashi (2020), Jones and Tonetti (2020), and Farboodi and Veldkamp (2021) to assume that households generate raw data as a byproduct of their consumption. The data are then sold to data intermediaries in exchange for compensation to offset potential privacy breaches or price discrimination. We innovate by allowing data intermediaries to transform raw data into working data (i.e., useful information) for production, which come in different varieties with potentially different usage in domestic and foreign countries. This assumption is not only realistic but also gives a necessary degree of freedoms for us to determine data flows in the international context. Our model thus provides a new way to deal with nonrival factors like data that need to link a single suppler with multiple and heterogeneous demanders. Intermediate good producers in each country accumulate and purchase new working data from both domestic and foreign data intermediaries to make intermediate goods, which enables cross-border data flows to play an important role. We also explicitly model data accumulation for the first time in the literature to investigate the effects that depreciation has on the volatility of working data generation and accumulation, as well as cyclicity in transition dynamics.

We first characterize the equilibria in steady states. Because a data-inefficient country (i.e., data is not a big augmenting factor in the production) has lower productivity, which makes its goods expensive, trade freezes if the gap in utilization of working data between the two countries is sufficiently large. But with unrestricted trading of data across borders being added to the trading of conventional goods, consumers' raw data are transformed into working data, which are used by producers in both domestic and foreign countries simultaneously. The incentives for trading data across borders facilitate and restore international trade (of goods and data).

That said, a data-efficient counterpart may still refuse to trade if it needs to export much more intermediate goods than it imports to reach trade balance. Various restrictions on cross-border data flows such as unilateral flows increase frictions in importing data, further reducing the feasible interval of trade. The welfare analysis shows that when the pre-existing data economy gap is large between the countries, a data-efficient country's loss due to trading goods can outweigh the benefits of allowing data to flow. In other words, a country that does not keep up the pace of developing the data economy may face a refusal of trade from a foreign country with much more efficient use of data in production.

Moving onto the transition dynamics after shocks to key variables, we observe opposite cyclic patterns of data usage following a productivity shock in an open economy versus in a closed economy under various levels of substitution of data from the two countries, provided that the pre-existing data divide between the countries is not extremely large. The intuition lies in that, unlike in a closed economy where the factors with relatively low costs such as capital and labor substitute data, an open economy allows foreign countries to supplement data, thereby increasing total data usage after a productivity shock. However, as the data divide between the two countries becomes very large, working data are more concentrated in the data-efficient country, and the productivity shock in the data-inefficient country can no longer reverse the data flows.

We next analyze two representative shocks that directly influence the generation and utilization of data. We find that firms tend to demand larger quantities of data with lower costs, and this preference becomes more pronounced with a larger elasticity of substitution between the different sources of data. Consequently, after a positive shock to disutility (privacy) cost of domestically generated data, we observe greater fluctuations in the usage of domestic working data under larger values of the elasticity of substitution. In contrast, the opposite relationship ensues after a shock to the trade costs of data importation. These significant fluctuations in data usage may have implications for the stability of labor incomes in associated sectors, potentially impeding the growth of the data economy.

To our best knowledge, we are the first to model the emerging phenomenon of cross-border data flows by introducing data as a production factor in an international context. The emerging literature on how data affect firms and enter production mostly features domestic settings. For example, Eeckhout and Veldkamp (2023) show how markups measured at different levels of aggregation reflect the impact of data on market power and distinguish data from other intangible investments. Data can also generate positive externality and feedback that give enterprises an edge in competition (Kubina et al., 2015; Cong and Mayer, 2023). Data-driven decision-making tends to be more accurate and effective (e.g., McAfee and Brynjolfsson, 2012; Brynjolfsson and McElheran, 2016), and big data can enhance forecasting, and thereby performance and profitability (e.g., Bajari et al., 2019; Farboodi and Veldkamp, 2021). Existing general equilibrium models in the international context focus on fiscal policy (Bhattarai and Trzeciakiewicz, 2017; Gross, 2021), monetary policy (Clarida et al., 2002; Galí and Monacelli, 2005), capital control (Devereux et al., 2019; Bacchetta et al., 2022), exchange rate (Ca'Zorzi et al., 2017; Adler et al., 2019), trade policy (Caldara et al., 2020; Alessandria et al., 2021), and interactions between goods trade and capital flows and their implications for the speed of convergence (Kleinman et al., 2023).

Our paper thus contributes to the recent literature on the economics of data from a macroeconomic perspective. Jones and Tonetti (2020) emphasize horizontal non-rivalry of data and directly incorporate data into the production process. Cong et al. (2021) introduce data into the innovation process to "distill" knowledge that accumulates and study how dynamic non-rivalry of data affects economic growth. Cong et al. (2022) further highlight the vertical non-rivalry of data, characterizing data usage in both the production and innovation sectors simultaneously. Xie and Zhang (2023) extend the discussion from "consumer data" to "producer data," with implications for production and growth. Importantly, Farboodi and Veldkamp (2020, 2021), Hou et al. (2022), and Veldkamp and Chung (2023) point out that data do not always lead to sustained economic growth. While previous studies have extensively discussed the role of data in innovation and long-term growth, we focus on the direct outcomes of data in production and their flows in the international context.

More generally, our paper is related to studies examining the economics of data in the digital age. Bergemann and Bonatti (2019), Ichihashi (2021a,b), and Acemoglu et al. (2022) study consumer privacy and welfare in the presence of data intermediaries, which lend micro-foundations to how we model data intermediaries in our paper. We also incorporate privacy issues arising from data usage (e.g., Casadesus-Masanell and Hervas-Drane, 2015; Acquisti et al., 2016; Abowd and Schmutte, 2019; Fainmesser et al., 2022; Ichihashi, 2020; Liu et al., 2023). Our discussion of cross-border data flows also adds an international dimension to the debate on data sharing and open banking (e.g., Babina et al., 2022; Goldstein et al., 2022; He et al., 2022; Cong and Mayer, 2023); in particular, allowing cross-border data flows is a pre-requisite for international data sharing. More recently, Sun et al. (2021), Farboodi et al. (2022), and Veldkamp (2023) use field experiments or develop sufficient statistic approaches to value data. We highlight how the ease of data flows can affect their usage and functionality; exploring how this affects data value constitutes interesting future research.

The rest of this paper is structured as follows. Section 2 introduces the baseline closed economy, before modeling (partially) open economies under various policies of cross-border data flows. Section 3 analyzes the steady-state equilibrium and welfare. Section 4 explores the transition dynamics following various shocks in the economy. Section 5 concludes.

2. The Global Data Economy: Model Setups and Solutions

We introduce the modeling ingredients and the variants for the closed economy, the partially open economies, and the fully open economy, which are variants of one another. Each model can be viewed as the outcomes under a particular policy choice, i.e., whether trade, data importing, or data exporting, etc., are allowed. Their comparison offers insights for choosing among various trade and data sharing policies.

Overall, our assumptions are mostly standard in the literature, except for two elements. First, we separate working data from raw data and enrich the settings of data intermediary. This new setting makes us easier to separate the usage of data in different countries, while remaining the non-rival property of data since raw data are used in a non-rival way to transform into working data. Second, while earlier studies assume data are fully depreciated and only focus on the growth rates of the economy, we allow data to accumulate, which is again more realistic, and allows us to understand how data depreciation affects transition dynamics (Jones and Tonetti, 2020; Cong et al., 2021).

2.1 Closed Data Economy

We first consider a simple, closed economy in which data enter the production process as an input factor but do not flow across borders. In this benchmark setting with infinite and discrete time, we introduce a representative household, a final good producer, multiple intermediate good producers, and a data intermediary—the building blocks for later analyses.

Representative household. A representative household maximizes its lifetime utility by choosing consumption (C_t), labor provision (N_t), and raw data contribution (D_t) in each period. As in Jones and Tonetti (2020) and Cong et al. (2021), data are generated as byproducts when households consume final goods. For simplicity and given the recent developments in data ownership and privacy protection (e.g., the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the U.S.), we stipulate that the household can sell raw data D_t to a data intermediary (introduced later) at a competitive price $P_{D,t}$. However, the household incurs a privacy cost due to potential leakages, violations, and risks of abuse or discrimination, which is reflected in a third term in the household's utility optimization:

$$\max_{C_t, N_t, D_t} U_t = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \Omega \frac{N_t^{1+\eta}}{1+\eta} - \Pi(1+b\mathbb{I})\pi_t D_t^2 \right), \tag{1}$$

where β is the discount factor, σ is the relative risk aversion coefficient (also the aversion of the intertemporal elasticity of substitution), Ω is the share of leisure in the household's utility, η is the inverse of Frisch's labor supply elasticity, and Π is a stochastic parameter tuning the disutility of data misuses (including price discrimination) or privacy violations specified as a quadratic cost (following Jones and Tonetti, 2020).[3] π_t captures the household's preference shock for data risk, and we assume that it follows an AR(1) process:

$$\ln \pi_t - \ln \pi = \rho_\pi (\ln \pi_{t-1} - \ln \pi) + \sigma_\pi \varepsilon, \quad \varepsilon \sim N(0,1),$$

where π is the shock in the steady state, and $\rho_\pi < 1$ and σ_π are the persistence and shock parameters. Finally, the indicator function $\mathbb{I}$ is 1 when the country's government allows cross-border data flows and 0 otherwise; b reflects the relative additional disutility caused by the international sharing (exporting) of data.[4] Obviously, in this closed economy, $\mathbb{I} = 0$.

The budget constraint for the household satisfies:

$$C_t + I_t + B_t = w_t N_t + r_t K_t + R_{t-1} B_{t-1} + P_{D,t} D_t,$$

where I_t is the investment, B_t is the household's assets, R_t is the return on assets, K_t is the physical capital, and r_t is the return on capital.[5] For clarity, we normalize the price of final

[3]This specification implicitly assumes a regime in which the households are not close to contributing all their data, otherwise the marginal cost of contributing data turns diminishing because eventually the return to learning through additional data is diminishing (Farboodi and Veldkamp, 2020).

[4]In our model, the last term $(1 + b\mathbb{I})\pi_t D_t^2$ in the household utility assumes that data exported to foreign countries has already been used in the domestic context. Therefore, when data is exported abroad, consumers face higher privacy disutility costs, and the parameter b, whether greater than or less than 1, will lead to this result.

[5]We follow the literature on data economy (e.g., Acquisti et al., 2016; Jones and Tonetti, 2020) to abstract away from the multiple uses of data (Cong et al., 2022) and consider a generic usage of data. Moreover, the compensation to consumers for data provision can manifest itself in the form of perks or reduced service prices in practice.

goods (as well as investments) to 1. The physical capital K_t owned by the household follows a dynamic process:

$$K_{t+1} = (1 - \delta_k)K_t + I_t, \tag{2}$$

where δ_k is the capital depreciation rate.

Final good producer. The final good producer takes in a continuum of intermediate goods to produce the output (the number of varieties is normalized to 1), according to the following CES technology (Dixit and Stiglitz, 1977):

$$Y_t = \left(\int_0^1 Y_{i,t}^{\frac{\rho-1}{\rho}} \mathrm{d}i\right)^{\frac{\rho}{\rho-1}},$$

where Y_t is the final good, $Y_{i,t}$ is the intermediate good of variety i, and ρ is the elasticity of substitution between varieties.

Intermediate good producers. A unit measure of monopolistically competitive intermediate good producers are indexed by $i \in [0, 1]$. They each hire labor $n_{i,t}$, rent capital $k_{i,t}$ from the household, and buy working data φ_t from the data intermediary to generate outputs. The data bought by the intermediate good producers φ_t are not the same as the raw data provided by the consumers D_t. The data intermediary works as a transformer from raw data to the working data usable by the producers, as we discuss shortly.

Data are accumulated according to the following process with the depreciation rate δ_Φ:

$$\Phi_{i,t+1} = (1 - \delta_\Phi)\Phi_{i,t} + \varphi_t. \tag{3}$$

The non-rivalry of data and their homogeneous role in intermediate good production (below) mean that each intermediate good producer buys the same data, which allows us to drop the subscript i for φ_t here.[6] The accumulated data $\Phi_{i,t}$ then act as an augmenting factor in the formulation of a capital composite.[7] Therefore, we specify the production function of

[6]Unlike Jones and Tonetti (2020) and Cong et al. (2021), which stipulate full depreciation of data in every period, we allow data to accumulate, which is realistic and has non-trivial effects in transition dynamics. Section 4.3 discusses this and provides an example for further illustration.

[7]Ichihashi (2021b) provides a potential micro-foundation based on data externalities. In studies such as Erickson and Rothberg (2014), Farboodi et al. (2019), and Sadowski (2019), data are treated as a special type of

intermediate good producer i to be:

$$Y_{i,t} = A_t(\Phi_{i,t}^{\xi} k_{i,t})^{\alpha} n_{i,t}^{1-\alpha}, \tag{4}$$

where ξ is the importance of data in the capital composite, and α is the contribution of this composite factor in the production function. Meanwhile, A_t is the productivity level, which evolves according to the following AR(1) process:

$$\ln A_t - \ln A = \rho_A(\ln A_{t-1} - \ln A) + \sigma_A \varepsilon, \quad \varepsilon \sim N(0,1),$$

where A is the steady state productivity, and $\rho_A < 1$ and σ_A are the corresponding coefficients.

Data intermediary. As in Jones and Tonetti (2020), a data intermediary buys raw data D_t from the household at price $P_{D,t}$ and then sells a quantity of "working data" φ_t at price $P_{i,\varphi,t}$ to the intermediate good producers. The data intermediary also employs labor l_t for collecting and cleaning data, with a data generation function:

$$\varphi_t = B D_t^{\gamma} l_t^{1-\gamma}, \tag{5}$$

where $B > 0$ is the efficiency term, and $0 < \gamma < 1$ describes the contribution of raw data in generating working data. Because of the non-rivalry of data, this data intermediary buys raw data once and sells the working data to all intermediate good producers simultaneously. This makes a perfectly competitive environment unsuitable for this sector. To pin down the prices, we assume the intermediary to be a monopolist which is subject to free-entry and get zero profit in equilibrium.

Equilibrium definition. An equilibrium consists of quantities $\{C_t,\ n_t,\ l_t\ K_t,\ I_t,\ \varphi_t,\ \Phi_t,\ D_t,\ Y_t\}$ as well as prices $\{w_t,\ R_t,\ r_t,\ P_{\varphi,t},\ P_{D,t}\}$, such that:

1. Given $\{w_t,\ r_t,\ P_{D,t}\}$, $\{C_t,\ N_t,\ D_t\}$ maximize the household's utility. Given $\{P_{\varphi,t},\ r_t,\ w_t\}$, $\{\varphi_t,\ K_t,\ n_t\}$ maximize the profit of producers. Given $\{P_{D,t},\ P_{\varphi,t},\ w_t\}$, $\{D_t,\ \varphi_t,\ l_t\}$ maximize the profit of the data intermediary.

capital. Abis and Veldkamp (2021) discuss an alternative way of combining data with capital.

2. Capital accumulation follows $K_{t+1} = (1 - \delta_k)K_t + I_t$, and data accumulation follows $\Phi_{t+1} = (1 - \delta_\Phi)\Phi_t + \varphi_t$.

3. $Y_t = C_t + I_t$ clears the goods market, $l_t + n_t = N_t$ clears the labor market, r_t clears the capital market when the capital supply equals demand, R_t clears the assets market when $B_t = 0$, and $\{P_{\varphi,t}, P_{D,t}\}$ clear the data markets.

2.2 Open Economy with International Trade and Data Flows

We now consider a two-country open economy with a home country and a foreign country. We use the subscripts "H" and "F" to indicate factors or outputs generated in the home and foreign countries, respectively, and we use the superscript "$*$" to indicate factors or outputs employed in the foreign country. Again, each country consists of a representative household, a final good producer, multiple intermediate good producers, and a data intermediary. In this case, the new elements are the wholesale producers, who assemble intermediate goods produced both domestically and imported from abroad. Furthermore, data intermediaries can now also produce working data to be exported and used by foreign intermediate good producers. For simplicity, we only describe the setup for the home country next; that for the foreign country is symmetric.

Representative households. The representative household's utility function is the same as that in the closed model (1), except that now $\mathbb{I} = 1$. We normalize the price of final goods in the home country to 1, and set the corresponding price in the foreign country as P_t^*. The budget constraint becomes:

$$C_t + I_t + B_{H,t} + B_{F,t} = w_t N_t + r_t K_t + R_{t-1} B_{H,t-1} + R_{t-1}^* B_{F,t-1} + P_{D,t} D_t,$$

where $B_{H,t}$ represents the assets held by the household in the home country at time t, and R_{t-1} is the corresponding return; $B_{F,t}$ represents the assets held by the household in the foreign country at time t, and R_{t-1}^* is the return on this asset; $P_{D,t}$ is the price of the raw data provided by the consumer in the home country. Finally, physical capital accumulates similarly as in (2).

Final good producer and wholesale producers. Each country has a representative final good producer using intermediate goods to make final goods for consumption, just as in the closed model. However, following the convention in the literature (Galí and Monacelli, 2005; Caldara et al., 2020), the intermediate goods going to the final production should first be assembled by wholesale producers with goods produced domestically and those imported from abroad (which generally differ as they are produced in different factories under potentially different processes despite being of the same variety) according to the following CES technology:

$$Y_{i,t} = \left(Y_{H,i,t}^{\frac{m-1}{m}} + Y_{F,i,t}^{\frac{m-1}{m}} \right)^{\frac{m}{m-1}}.$$

Here, $Y_{i,t}$ denotes the wholesale goods (which can also be viewed as the composite intermediate goods), $Y_{H,i,t}$ refers to the domestic-produced intermediate goods used in the home country, and $Y_{F,i,t}$ represents the intermediate goods imported from the foreign country. As for the parameter, m is the elasticity of substitution in this combination.

Intermediate good producers. In each country, a unit mass of monopolistically competitive producers is indexed by $i \in [0, 1]$. Each producer generates outputs both for domestic use ($Y_{H,i,t}$ and $Y^*_{F,i,t}$) and for exporting ($Y^*_{H,i,t}$ and $Y_{F,i,t}$), according to the following technology:

$$Y_{H,i,t} + Y^*_{H,i,t} = A_t (\Phi^{\xi}_{i,t} k_{i,t})^{\alpha} n^{1-\alpha}_{i,t}, \tag{6}$$

where the input variables are similar to those shown in the closed model. The data accumulation process is also similar to that in the closed economy:

$$\Phi_{i,t+1} = (1 - \delta_{\Phi})\Phi_{i,t} + \varphi_t. \tag{7}$$

In contrast, φ_t here represents a data composite generated by the intermediate good producers, which combines domestic-generated working data $\varphi_{H,t}$ with foreign-generated working data $\varphi_{F,t}$ in a CES form:

$$\varphi_t = \left[\chi^{\frac{1}{\omega}} \varphi_{H,t}^{\frac{\omega-1}{\omega}} + (1 - \chi)^{\frac{1}{\omega}} \varphi_{F,t}^{\frac{\omega-1}{\omega}} \right]^{\frac{\omega}{\omega-1}}, \tag{8}$$

where χ is the share of the working data in the home country, and ω is the elasticity of substitution between the two different sources of working data.

Moreover, when a country imports working data, it may encounter various restrictions arising from legal gaps in privacy laws or national security policies between different countries. To address these issues, the importing country may need to pay additional fees or comply with certain requirements to obtain foreign data. Considering this, suppose that $P_{\varphi,F,t}$ is the price the foreign country sells its working data to the home country, then the price that the home country should in fact pay is $df_t P_{\varphi,F,t}$. Here, d is the cost multiplier of using imported working data, which captures various frictions. f_t is the shock of this cost, which follows an AR(1) process:

$$\ln f_t - \ln f = \rho_f(\ln f_{t-1} - \ln_f) + \sigma_f \varepsilon, \qquad \varepsilon \sim N(0,1),$$

where f is the shock in the steady state, and $\rho_f < 1$ and σ_f are the persistence and shock parameters, respectively. These two variables are important in cross-border data flows, since they reflect the change in the related laws and policies in different countries. We discuss this issue in detail in the following sections.

Data intermediaries. The data intermediaries in this two-country world work differently from those in the baseline closed economy. The data intermediary in each country buys raw data from households in its own country and then potentially sells the working data to intermediate producers in both countries, transforming the raw data to separate working data sold domestically and abroad. Specifically, we have the following working data generation functions in the home country:

$$\varphi_{H,t} = BD_t^{\gamma} l_{H,t}^{1-\gamma}, \tag{9}$$

and

$$\varphi_{H,t}^{*} = BD_t^{\gamma} (l_{H,t}^{*})^{1-\gamma}. \tag{10}$$

The corresponding functions in the foreign country can be defined similarly.[8] Here, the home country's data intermediary uses the same quantity of raw data collected from domestic consumers to generate different types of working data, using the same technology B but employing different quantities of labor ($l_{H,t}$ and $l^*_{H,t}$).

Equilibrium definition. An equilibrium in this open economy consists of quantities $\{C_t$, N_t, K_t, I_t, $\varphi_{H,t}$, $\varphi_{F,t}$, Φ_t, D_t, $Y_{H,t}$, $Y_{F,t}$, Y_t, n_t, $l_{H,t}$, $l^*_{H,t}\}$ for the home country and $\{C^*_t$, N^*_t, K^*_t, I^*_t, $\varphi^*_{H,t}$, $\varphi^*_{F,t}$, Φ^*_t, D^*_t, $Y^*_{H,t}$, $Y^*_{F,t}$, Y^*_t, n^*_t, $l_{F,t}$, $l^*_{F,t}\}$ for the foreign country, and prices $\{w_t$, R_t, r_t, $P_{\varphi,H,t}$, $P_{\varphi,F,t}$, $P_{D,t}$, $P_{H,t}$, $P_{F,t}\}$ for the home country and prices $\{w^*_t$, R^*_t, r^*_t, $P^*_{\varphi,H,t}$, $P^*_{\varphi,F,t}$, $P^*_{D,t}$, $P^*_{H,t}$, $P^*_{F,t}$, $P^*_t\}$ for the foreign country, such that:

- For the home country, $\{C_t, N_t, D_t\}$ maximize the household's utility given $\{w_t, P_{D,t}\}$, $\{Y_t\}$ maximizes the profit of the final good producer given $\{P_{i,t}\}$, $\{Y_{H,t}, Y_{F,t}\}$ maximize the profit of wholesale producers given $\{P_{H,t}, P_{F,t}\}$, $\{P_{H,t}, P^*_{H,t}\ \varphi_{H,t}, \varphi_{F,t}, K_t, n_t\}$ maximize the profit of intermediate good producers given the demand behavior of final good producer, and $\{D_t, l_{H,t}, l^*_{H,t}\}$ maximize the profit of the data intermediary given $\{P_{D,t}$, $w_t\}$.

- For the foreign country, $\{C^*_t, N^*_t, D^*_t\}$ maximize the household's utility given $\{w^*_t, P^*_{D,t}\}$, $\{Y^*_t\}$ maximizes the profit of the final good producer given $\{P^*_{i,t}\}$, $\{Y^*_{H,t}, Y^*_{F,t}\}$ maximize the profit of wholesale producers given $\{P^*_{H,t}, P^*_{F,t}\}$, $\{P_{F,t}, P^*_{F,t}\ \varphi^*_{H,t}, \varphi^*_{F,t}, K^*_t, n^*_t\}$ maximize the profit of intermediate good producers given the demand behavior of final good producer, and $\{D^*_t, l^*_{F,t}, l_{F,t}\}$ maximize the profit of data intermediary given $\{P^*_{D,t}, w^*_t\}$.

- For the home country, capital accumulation follows (2), and data accumulation follows (7). For the foreign country, there are similar accumulation processes.

- For the home country, the final good market clears, i.e., $Y_t = C_t + I_t$; the wholesale market clears, i.e., $Y_t = \left(Y_{H,t}^{\frac{m-1}{m}} + Y_{F,t}^{\frac{m-1}{m}}\right)^{\frac{m}{m-1}}$; the intermediate good market clears, i.e., $Y_{H,t} + Y^*_{H,t} = A_t(\Phi_t^{\xi} K_t)^{\alpha} n_t^{1-\alpha}$; and the labor market clears, i.e., $l_{H,t} + l^*_{H,t} + n_t = N_t$. Moreover, $\{r_t\}$ clears the capital market; $\{R_t\}$ clears the assets market when $B_{H,t} = 0$;

[8]As has been discussed previously, we also denote $\varphi^*_{F,t}$ as the working data generated and used in the foreign country, and $\varphi_{F,t}$ as the working data generated in the foreign country but used in the home country.

$B^*_{F,t} = 0$, $B^*_{H,t} = B_{F,t}$, and $\{P_{\varphi,H,t}, P^*_{\varphi,H,t}, P_{\varphi,F,t}, P^*_{\varphi,F,t}\}$ clear the data market. For the foreign country, similar market clearing conditions hold.

- Finally, the risk sharing condition between the two countries is $(C^*_t)^{-\sigma}/C_t^{-\sigma} = P^*_t$.

2.3 Solving the Models

2.3.1 Closed Data Economy

First, from the household's utility maximization, we get:

$$\Omega C_t^{\sigma} N_t^{\eta} = w_t, \quad 2\Pi\pi_t D_t C_t^{\sigma} = P_{D,t}, \quad \text{and} \quad \beta\mathbb{E}_t\left(\frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}}\right) = \frac{1}{R_t} = \frac{1}{r_{t+1} + 1 - \delta_k}.$$

Then, given the prices of each individual variety i, $P_{i,t}$, the final good producer maximizes:

$$\max_{Y_{i,t}}\left[Y_t - \int_0^1 P_{i,t}Y_{i,t}\mathrm{d}i\right].$$

Thus, its demand for the intermediate goods can be derived as:

$$Y_{i,t} = P_{i,t}^{-\rho}Y_t. \tag{11}$$

The zero profit condition for the competitive final good producer implies that:

$$\left(\int_0^1 P_{i,t}^{1-\rho}\mathrm{d}i\right)^{\frac{1}{1-\rho}} = 1. \tag{12}$$

The intermediate good producers' optimization is:

$$\max_{P_{i,t},\Phi_{i,t},k_{i,t},n_{i,t}} \mathbb{E}_t\sum_{k=0}^{\infty} Q_{t,t+k}\left(P_{i,t+k}Y_{i,t+k} - P_{\varphi,t+k}\varphi_{t+k} - r_{t+k}k_{i,t+k} - w_{t+k}n_{i,t+k}\right),$$

subject to (3), (4), and (11), where $Q_{t,t+k} = \beta^k\mathbb{E}_t(C_{t+k}^{-\sigma}/C_t^{-\sigma})$ is the discount factor and $P_{\varphi,t}$ is the price of data φ_t. Solving this problem gives the price of intermediate goods and the demand functions for each factor as follows:

$$P_{i,t} = \frac{\rho}{\rho - 1}\mathrm{MC}_t, \quad r_t = \alpha \frac{Y_{i,t}}{k_{i,t}}\mathrm{MC}_t, \quad w_t = (1-\alpha)\frac{Y_{i,t}}{n_{i,t}}\mathrm{MC}_t,$$

and

$$P_{\varphi,t} = Q_{t,t+1}\mathbb{E}_t\left[\alpha\xi\frac{Y_{i,t+1}}{\Phi_{i,t+1}}\mathrm{MC}_{t+1} + (1-\delta_\Phi)P_{\varphi,t+1}\right].$$

Here, MC_t is the marginal cost, which is also the shadow price of this problem.

Finally, the optimization problem faced by the data intermediary is:

$$\max_{P_{i,\varphi,t}, D_t, l_t} \int_0^1 P_{i,\varphi,t}\varphi_t \mathrm{d}i - P_{D,t}D_t - w_t l_t,$$

subject to the data generation function (5) and the zero-profit condition

$$\int_0^1 P_{i,\varphi,t}\varphi_t \mathrm{d}i - P_{D,t}D_t - w_t l_t = 0.$$

In equilibrium, the prices of the working data are equalized among different intermediate good producers, i.e., $P_{i,\varphi,t} \equiv P_{\varphi,t}, \forall i$. Then, we have the demand functions of the raw data and the labor employed in the data intermediary:

$$B\gamma P_{\varphi,t}D_t^{\gamma-1}l_t^{1-\gamma} = P_{D,t},$$

and

$$B(1-\gamma)P_{\varphi,t}D_t^{\gamma}l_t^{-\gamma} = w_t.$$

All the conditions derived above, together with the equilibrium conditions, characterize the system of the closed data economy.

2.3.2 Open Economies

The utility maximization gives the relationship between consumption and various prices:

$$\Omega C_t^{\sigma}N_t^{\eta} = w_t, \quad 2\Pi(1+b)\pi_t D_t C_t^{\sigma} = P_{D,t}, \quad \text{and} \quad \beta\mathbb{E}_t\left(\frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}}\right) = \frac{1}{R_t} = \frac{1}{R_t^*} = \frac{1}{r_{t+1} + 1 - \delta_k}.$$

The problem of final good producer leads to similar conditions as demonstrated in (11) and (12). Meanwhile, the profit maximization problem for the wholesale producers maximizes $Y_{i,t} - P_{H,i,t}Y_{H,i,t} - P_{F,i,t}Y_{F,i,t}$ by choosing $Y_{H,i,t}$ and $Y_{F,i,t}$. We can derive the demands for these two types of intermediate goods as:

$$Y_{H,i,t} = P_{H,i,t}^{-m} Y_{i,t}, \quad \text{and} \quad Y_{F,i,t} = P_{F,i,t}^{-m} Y_{i,t}. \tag{13}$$

Here, $P_{H,i,t}$ and $P_{F,i,t}$ are the prices of domestic and imported intermediate goods, respectively. The price for the wholesale goods satisfies $\left(P_{H,i,t}^{1-m} + P_{F,i,t}^{1-m}\right)^{\frac{1}{1-m}} = 1$, since we have normalized the price of the final good to 1 in the home country.

The intermediate good producers' optimization can be divided into two steps. Given the production decisions in each period and the prices of the two different sources of working data, intermediate good producers decide on the quantities of working data purchased from the two countries, respectively, which reduces to the following static problem:

$$\min_{\varphi_{H,t}, \varphi_{F,t}} P_{\varphi,H,t}\varphi_{H,t} + df_t P_{\varphi,F,t}\varphi_{F,t}, \tag{14}$$

where $P_{\varphi,H,t}$ and $P_{\varphi,F,t}$ are the prices of working data used in the home country, and generated in the home country and the foreign country, respectively.[9] Then, we derive the demands from the two sources of data as follows:

$$\varphi_{H,t} = \chi \left(\frac{P_{\varphi,H,t}}{P_{\varphi,t}}\right)^{-\omega} \varphi_t, \quad \text{and} \quad \varphi_{F,t} = (1-\chi)\left(\frac{df_t P_{\varphi,F,t}}{P_{\varphi,t}}\right)^{-\omega} \varphi_t,$$

where the price index of the data composite is defined as

$$P_{\varphi,t} = \left[\chi P_{\varphi,H,t}^{1-\omega} + (1-\chi)(df_t P_{\varphi,F,t})^{1-\omega}\right]^{\frac{1}{1-\omega}}.$$

With $Q_{t,t+k}$ denoting the discount factor similarly as that in the closed economy, interme-

[9] Similarly, we can also define the prices of working data generated in both countries and used in the foreign country as $P^*_{\varphi,H,t}$ and $P^*_{\varphi,F,t}$, respectively.

diate good producers then solve the following dynamic profit maximization problem:

$$\max_{P_{H,i,t},P^*_{H,i,t},\Phi_{i,t},k_{i,t},n_{i,t}} \mathbb{E}_t \sum_{k=0}^{\infty} Q_{t,t+k} \Big(P_{H,i,t+k}Y_{H,i,t+k} + P^*_{H,i,t+k}Y^*_{H,i,t+k} - P_{\varphi,t+k}\varphi_{t+k}$$
$$-r_{t+k}k_{i,t+k} - w_{t+k}n_{i,t+k} \Big),$$

subject to (6), (7), (13), and $Y^*_{H,i,t} = (P^*_{H,i,t}/P^*_{i,t})^{-m}Y^*_{i,t}$, which comes from the foreign wholesale producers' optimization, where $P^*_{H,i,t}$ is the price of the foreign intermediate goods imported from the home country, $P^*_{i,t}$ is the price of foreign wholesale goods, and $Y^*_{i,t}$ denotes the wholesale goods in the foreign country. We omit the derivation of this equation here since it is similar to that in the case of the home country. Then, we obtain the prices of these two sources of intermediate goods and the demand functions for each factor:

$$P_{H,i,t} = P^*_{H,i,t} = \frac{m}{m-1}\mathrm{MC}_t, \quad r_t = \alpha\frac{Y_{H,i,t} + Y^*_{H,i,t}}{k_{i,t}}\mathrm{MC}_t, \quad w_t = (1-\alpha)\frac{Y_{H,i,t} + Y^*_{H,i,t}}{n_{i,t}}\mathrm{MC}_t,$$

and

$$P_{\varphi,t} = Q_{t,t+1}\mathbb{E}_t\left[\alpha\xi\frac{Y_{H,i,t+1} + Y^*_{H,i,t+1}}{\Phi_{i,t+1}}\mathrm{MC}_{t+1} + (1-\delta_\Phi)P_{\varphi,t+1}\right].$$

Finally, the optimization problem of the data intermediary is:

$$\max_{P_{i,\varphi,H,t},P^*_{i,\varphi,H,t},D_t,l_{H,t},l^*_{H,t}} \int_0^1 P_{i,\varphi,H,t}\varphi_{H,t}\mathrm{d}i + \int_0^1 P^*_{i,\varphi,H,t}\varphi^*_{H,t}\mathrm{d}i - P_{D,t}D_t - w_t(l_{H,t} + l^*_{H,t}), \quad (15)$$

subject to (9), (10), and the zero-profit condition:

$$\int_0^1 P_{i,\varphi,H,t}\varphi_{H,t}\mathrm{d}i + \int_0^1 P^*_{i,\varphi,H,t}\varphi^*_{H,t}\mathrm{d}i - P_{D,t}D_t - w_t(l_{H,t} + l^*_{H,t}) = 0.$$

Solving this problem, we derive the prices of working data as:

$$P_{i,\varphi,H,t} \equiv P_{\varphi,H,t}, \qquad \text{and} \qquad P^*_{i,\varphi,H,t} \equiv P^*_{\varphi,H,t}, \forall i.$$

Meanwhile, the demand functions for the raw data and the labor employed satisfy:

$$B\gamma D_t^{\gamma-1}\left[P_{\varphi,H,t}l_{H,t}^{1-\gamma}+P_{\varphi,H,t}^{*}(l_{H,t}^{*})^{1-\gamma}\right]=P_{D,t},$$

$$B(1-\gamma)P_{\varphi,H,t}D_t^{\gamma}l_{H,t}^{-\gamma}=w_t,$$

and

$$B(1-\gamma)P_{\varphi,H,t}^{*}D_t^{\gamma}(l_{H,t}^{*})^{-\gamma}=w_t.$$

For simplicity, we have only shown the solutions in the home country. In the foreign country, an analogous set of first order conditions hold. Combining the conditions derived in home and foreign countries completes the system of the open economy.

2.4 Partially Open Economies with Trade and Unilateral Data Flows

To isolate the incremental or unilateral effects of international data flows, we consider the solutions in three alternative economies. We first consider a partially open economy in which only goods are traded internationally (the "goods trade model"). Intermediate good producers then only buy data from their domestic data intermediary, and equation (8) becomes $\varphi_t=\varphi_{H,t}$. The data intermediary's optimization in (15) becomes:

$$\max_{D_{H,t},l_{H,t}}\int_0^1 P_{i,\varphi,H,t}\varphi_{H,t}\mathrm{d}i-P_{D,H,t}D_t-w_tl_{H,t}.$$

Two other partially open economies involve only one country allowing data flows, i.e., only the home country imports data from the foreign country (unilateral flow with H importing), and only the foreign country imports data from the home country (unilateral flow with F importing). These setups can be specified similarly as the goods trade model. Equilibrium definitions follow from that of open economy.

We compare equilibrium outcomes in these models with those in the open economy to gain further insights on the effects of international data flows, and to inform and guide domestic policies and strategic responses to policies in foreign countries.

3. Steady-State Equilibria

We conduct quantitative analyses to characterize the equilibria. First, we calibrate model parameters based on historical data and the existing literature. Then, we analyze the steady states of key variables under different forms of data combination, compare the welfare levels in the different models, and extend the models to allow a data divide between two asymmetric countries. We follow the literature (e.g., Clarida et al., 2002; Galí and Monacelli, 2005) to build a parsimonious theory to provide an initial investigation of the data economy in the international context. Future studies when data about cross-border flows are more available will likely offer further quantitative insights.

3.1 Calibration

Table 1 displays the calibration parameters. First, frequently used parameters, e.g., the subjective discount factor β, the reciprocal of elasticity of intertemporal substitution σ, the capital depreciation rate δ_k, and the contribution of labor to good production $(1-\alpha)$, take on standard values. Second, most other parameters follow the literature: the weight on leisure in the utility function Ω comes from Christensen and Dib (2008), the elasticity of substitution among different varieties of intermediate goods ρ comes from Fernández-Villaverde et al. (2015), the elasticity of substitution between domestically produced goods and imported intermediate goods m comes from Alessandria et al. (2021), and the persistence of exogenous shocks ρ_A, ρ_e, and ρ_f comes from Alessandria et al. (2013). Because the reciprocal of Frisch's labor supply elasticity η usually lies between 1 and 2, we set it as 1.3, having checked that our key findings are robust under other values in the range. Third, some data-related parameters are new and are determined under our discretion. For example, the depreciation rate of data δ_Φ takes the same value as that of capital. The share of domestic data χ in the data composite also takes a value of 0.5 to maintain the symmetry between the two countries. Finally, we verify the robustness of our findings under some alternative values and discuss some important parameters in the following subsections, such as the importance of data ξ, the scale of the disutility caused by raw data usage Π, the cost multiplier of imported working data

d, and the elasticity of the substitution of data from different sources ω. All the parameters in the foreign country take the same values as the corresponding parameters in the home country unless otherwise specified.

Table 1: Calibration of Parameters

Parameters	Meaning	Value	Source
β	Subjective discount factor	0.99	Standard
σ	Reciprocal of elasticity of intertemporal substitution	2	Standard
η	Reciprocal of Frisch's labor supply elasticity	1.3	Standard
$1-\alpha$	Contribution of labor in good productions	2/3	Standard
δ_k	Capital depreciation rate	0.025	Standard
Ω	Weight on leisure in the utility function	1.315	Christensen and Dib (2008)
ρ	Elasticity of substitution (varieties)	21	Fernández-Villaverde et al. (2015)
m	Elasticity of substitution (domestic and imported)	5	Alessandria et al. (2021)
ρ_A, ρ_e, ρ_f	Persistence of exogenous shocks	0.95	Alessandria et al. (2013)
δ_Φ	Data depreciation rate	0.025	Discretionary
χ	Share of domestic data in the data composite	0.5	Discretionary
B	Efficiency term in working data generation	1	Discretionary
γ	Contribution of raw data in working data generation	0.5	Discretionary
b	Additional disutility caused by cross-border data flows	1	Discretionary

3.2 Steady-States of Key Variables

After calibrating the model, we first analyze the steady states of the main variables to understand the open economy. Specifically, we focus on different costs surrounding data generated domestically and imported from abroad in the two countries, which lead to different values of the parameters Π in the utility function and the cost multiplier d. In Figures 1 and 2, we outline the relationship between the elasticity of substitution of data from different countries ω and the steady states of the following variables: production-related variables such as wholesale goods Y, capital K, and labor employed in production n, as well as the data-related variables such as the four different directions of working data flows φ_H, φ_F, φ_H^*, and φ_F^*, in addition to the working data φ and data stock Φ. We also show the change in raw data D, and labor employed in generating working data for the home and foreign countries l_H and

l_H^*. In Figure 1, we fix the parameters at $d = d^* = 1$ and $\Pi^* = 2$, and display the relationship in the five different models for a wide range of Π. In Figure 2, we fix the parameters $\Pi = \Pi^* = 1$ and $d^* = 1$, and present the relationship in the five different models for a wide range of d.

From the figures, we see that the steady states of the variables do not change as ω increases when $\Pi = \Pi^* = 2$ or $d = d^* = 1$ given the symmetry of the two countries in the open economy. These lines serve as references as we modify the values of Π or d in the other models. In Figure 1, production Y decreases and data are substituted by labor as Π increases, which means that higher disutility from using data provided by consumers leads to lower outputs. However, we observe little change in production as the elasticity of substitution of data ω increases. As for the flows of data, we find similar patterns in the change in data generated in the home country φ_H and φ_H^*, whereas the relationship becomes reversed for the data generated in the foreign country φ_F and φ_F^* when ω becomes large enough. This occurs because φ_H and φ_H^* are restricted by the high costs of the raw data D, and when ω increases, the imported data φ_F become dominated, which increases the demand for raw data in foreign country D^* and further pushes φ_F^* to a higher level.[10] In contrast, when ω is relatively low, which means that data from the countries are complements, the changes in the four directions of data flows become synchronous, and the total provision of working data turns to a lower stage.

Figure 2 displays further results as we alter the import cost multiplier d. First, production Y becomes lower as d increases, while this negative effect is alleviated as the substitution of data from different countries becomes more flexible. Consistent with this, we observe similar negative effects in other input factors n, K, and Φ. Second, as d increases, which means that importing data in the home country becomes more expensive, the provisions of its domestic data φ_H and φ_H^* become larger as ω increases. In this case, data in the home country become dominated in both countries. Finally, similar to the above analyses of Π, the four directions of data flows are also synchronous when ω is relatively small, due to the imperfect substitution of data from different countries.

[10]The raw data within a country are non-rival, which means they can be used in generating working data used in both countries simultaneously. As a result, the increasing demand for one type of data pushes the usage of raw data higher, and then further pushes the supply of the other type of data higher.

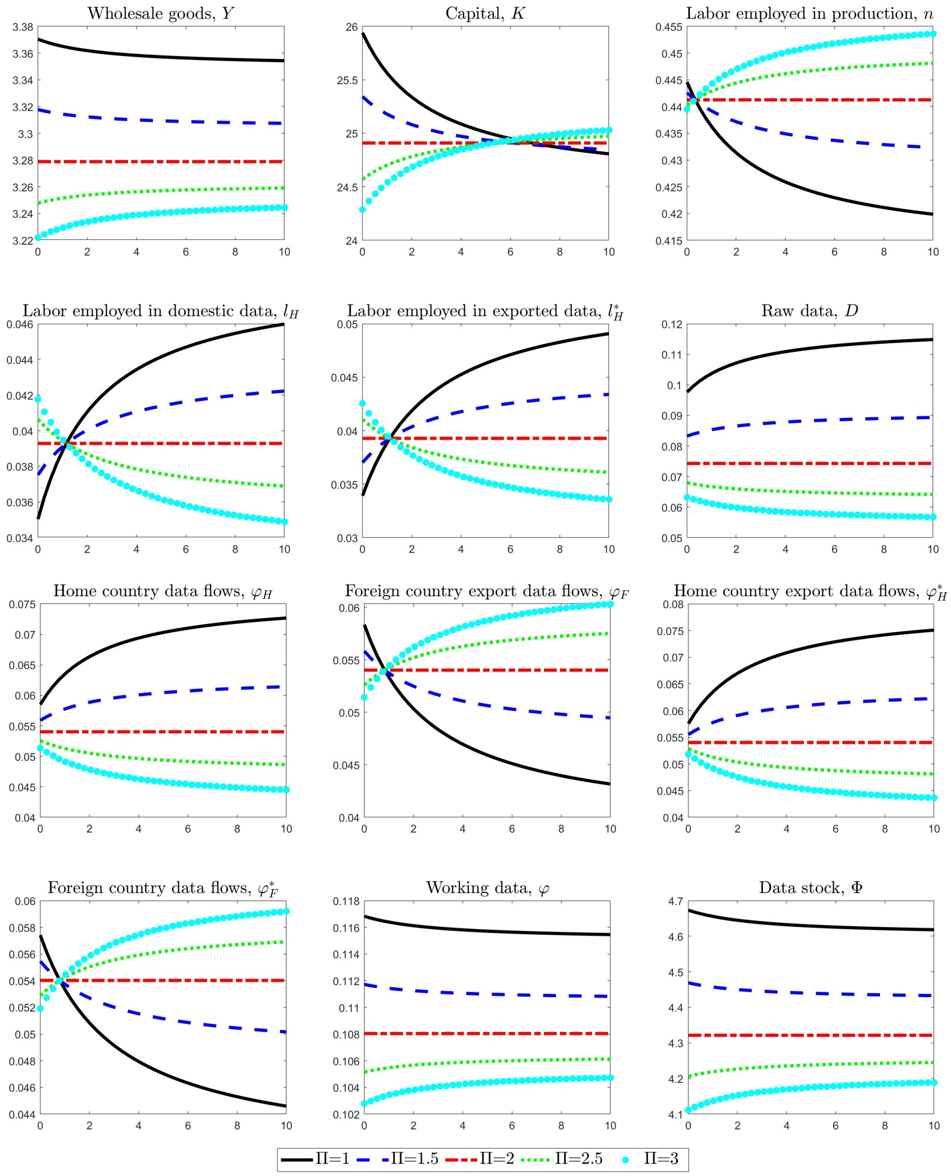

Figure 1: Steady-States of the Main Variables with Different Values of Disutility Parameter Π

Notes. This figure depicts the shifts in the steady states of the main variables with different values of the parameter Π in the open economy, as the relationship between the elasticity of substitution of data from different sources ω (x-axis) increases from 0 (lacking elasticity) to 10 (full of elasticity). Five different models are shown to illustrate the effects of data disutility: $\Pi = 1$ (full black line), $\Pi = 1.5$ (blue dashed line), $\Pi = 2$ (red dashed dotted line), $\Pi = 2.5$ (green dotted line), and $\Pi = 3$ (cyan star line). In the foreign country, we always have $\Pi^* = 2$.

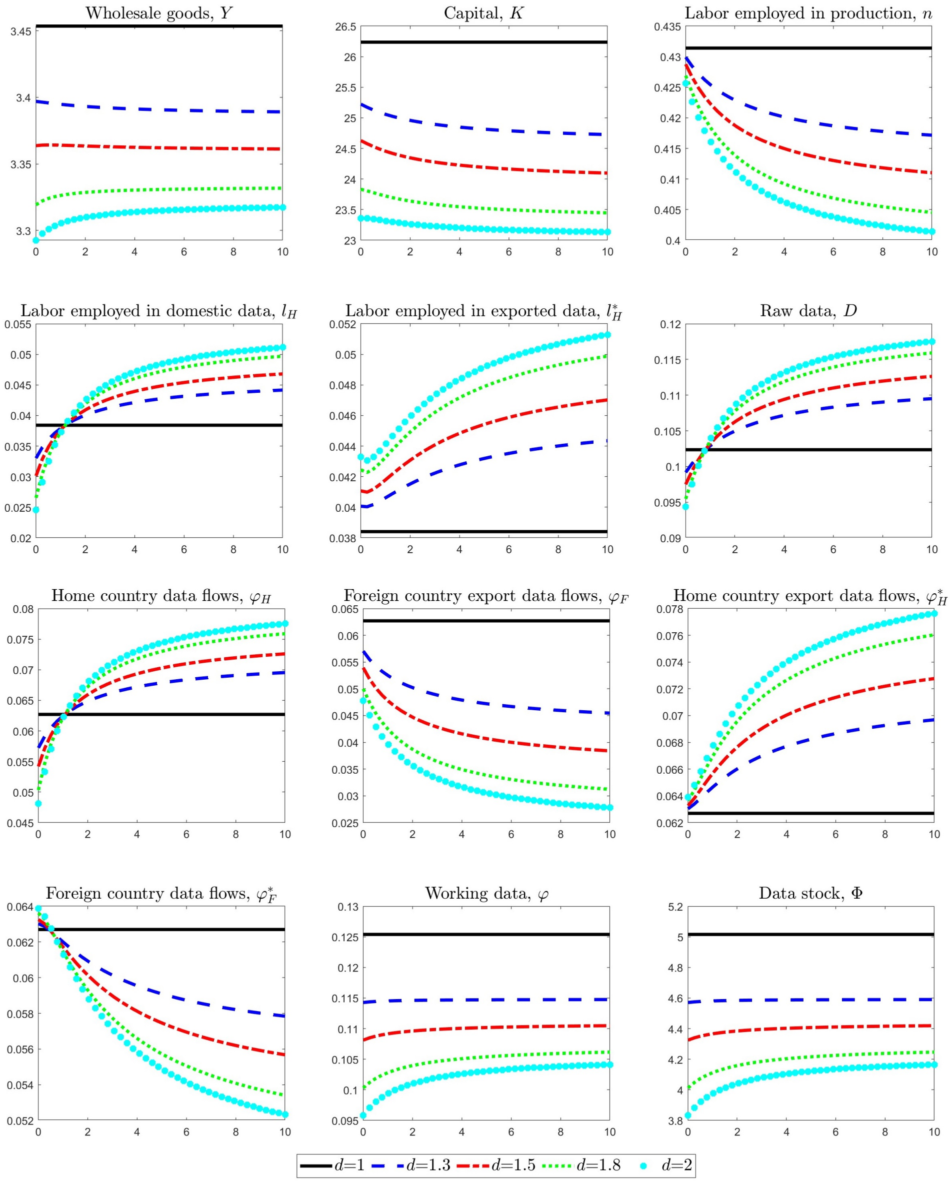

Figure 2: Steady-States of the Main Variables with Different Values of Cost Multiplier d

Notes. This figure depicts the shifts in the steady states of the main variables with different values of the cost multiplier d in the open economy, as the relationship between the elasticity of substitution of data from different sources ω (x-axis) increases from 0 (lacking elasticity) to 10 (full of elasticity). Five different models are shown to illustrate the effects of import friction: $d = 1$ (full black line), $d = 1.3$ (blue dashed line), $d = 1.5$ (red dashed dotted line), $d = 1.8$ (green dotted line), and $d = 2$ (cyan star line). In the foreign country, we always have $d^* = 1$.

3.3 Welfare Analysis: the Role of Data and Data Flows

We now examine the welfare in each country (the steady state of total utility) and its link to ξ in the closed model introduced in Section 2.1, the open economy introduced in Section 2, and the goods trade model and partially open economies with unilateral data flows introduced in Section 2.4. Because the patterns do not change significantly when the elasticity of substitution of data from different countries ω changes, so we only display the case with $\omega = 5$ in Figure 3. We also fix Π and d because they do not change the relative locations of the welfare curves derived in different models much. All the parameters in the foreign country also remain fixed at $\xi^* = 1$, $\Pi^* = 1$, and $d^* = 1$, for illustration.

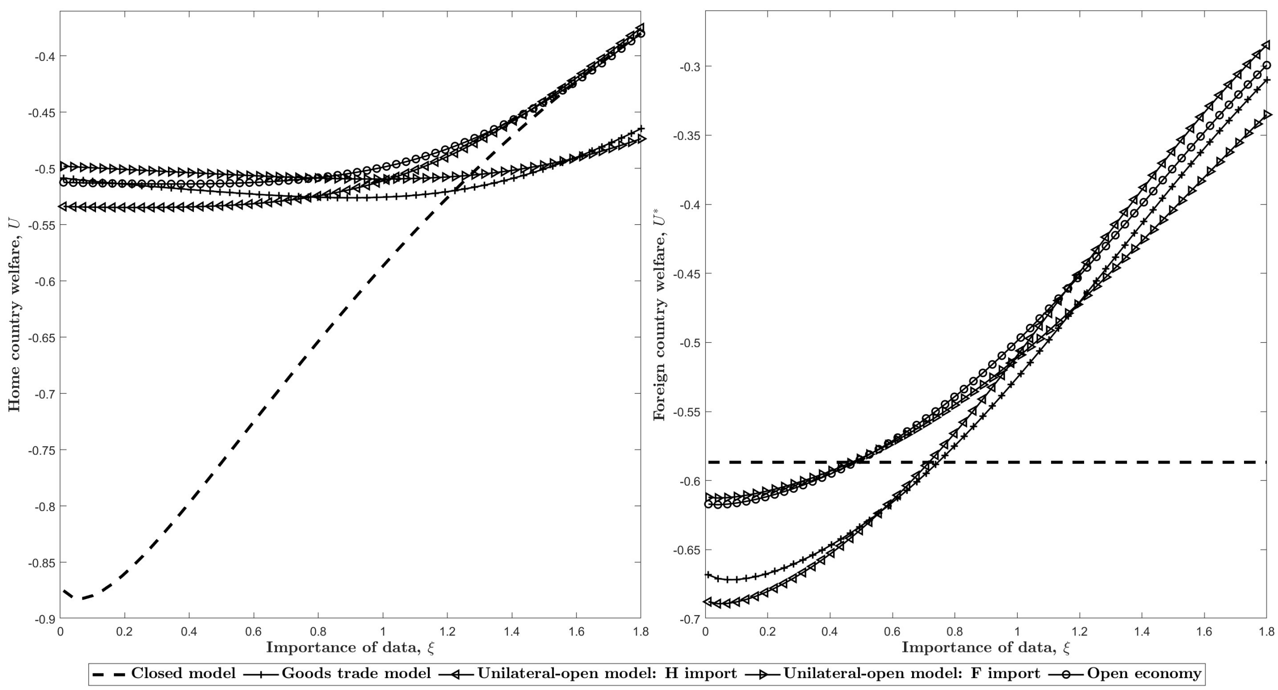

Figure 3: Welfare Levels with Different Importance of Data (ξ) in the Two Countries

Notes. These figures outline the relationship between the welfare levels and the importance of data in the composite factor ξ in the home country in five different models in steady states: the closed model, the model with only goods traded, the two unilateral flow models, and the open economy with bilateral data flows. We only present the models when the elasticity of substitution of the data is $\omega = 5$, and the importance of data in the foreign country is fixed at $\xi^* = 1$. It is important to note that the focus should be on how welfare outcomes differ across the various models, rather than solely on the welfare level.

Productivity augmentation and trade without data flows. Because data are productivity augmenting, welfare increases as data become more important in production (larger ξ). For the home country, the welfare of the closed model increases sharply as ξ increases, and even surpasses that of the open and partially open models when ξ becomes much larger than ξ^*. As a result, a data-inefficient country (i.e., with low data importance) is willing to trade with a data-efficient country (with high data importance), with or without data flows. In contrast, trade liberalization is not always desirable for the data-efficient country, especially when the divergence between ξ and ξ^* is big—a large pre-existing data divide. We show in the upper panels of Figure 4 the welfare improvements due to trade for regions of ξ where both countries are willing to trade. The empty regions correspond to where trade breaks down. While previous studies describe instances where trade liberalization can result in welfare loss, our paper highlights another potential data-related channel where such undesirable situations may arise (Bagwell and Staiger, 1999; Fajgelbaum et al., 2011).

For a country with relatively low ξ, trade liberalization leads to an expansion in the market for its goods, thereby increasing demand and improving welfare. However, for a country with relatively high ξ, trade liberalization may also lead to market expansion and increasing demand, but the goods exported to the country with relatively low ξ (i.e., low productivity) are priced lower than domestically produced goods in that country. As a result, the data-efficient country exports more goods but imports less from its trading partner, reducing its welfare gain from the trade.[11] In extreme cases, when the data divide between the two countries is large (e.g., the home country's ξ is much smaller than the foreign country's ξ^*), the data-efficient country faces welfare loss from trade, which leads to a breakdown of trade.

Cross-border data flows and welfare. Cross-border data flows can mitigate the welfare loss from exporting goods at a low price by expanding the sources of working data for domestic production. As shown in the right panel of Figure 3, open economies have higher welfare than those in the goods trade model for a wide range of parameters, revealing the benefits of cross-border data flows. Goods trade and data flows form a policy bundle in an open

[11] Notice that the domestic and foreign goods do not perfectly substitute in the open economy, though we set the elasticity of substitution ρ to be very large. Nevertheless, in the close model, households only demand domestic goods and can only obtain those goods. This is a friction arise from trade liberalization, and the introduction of data as a promotion factor in production amplifies it and makes the results more complex.

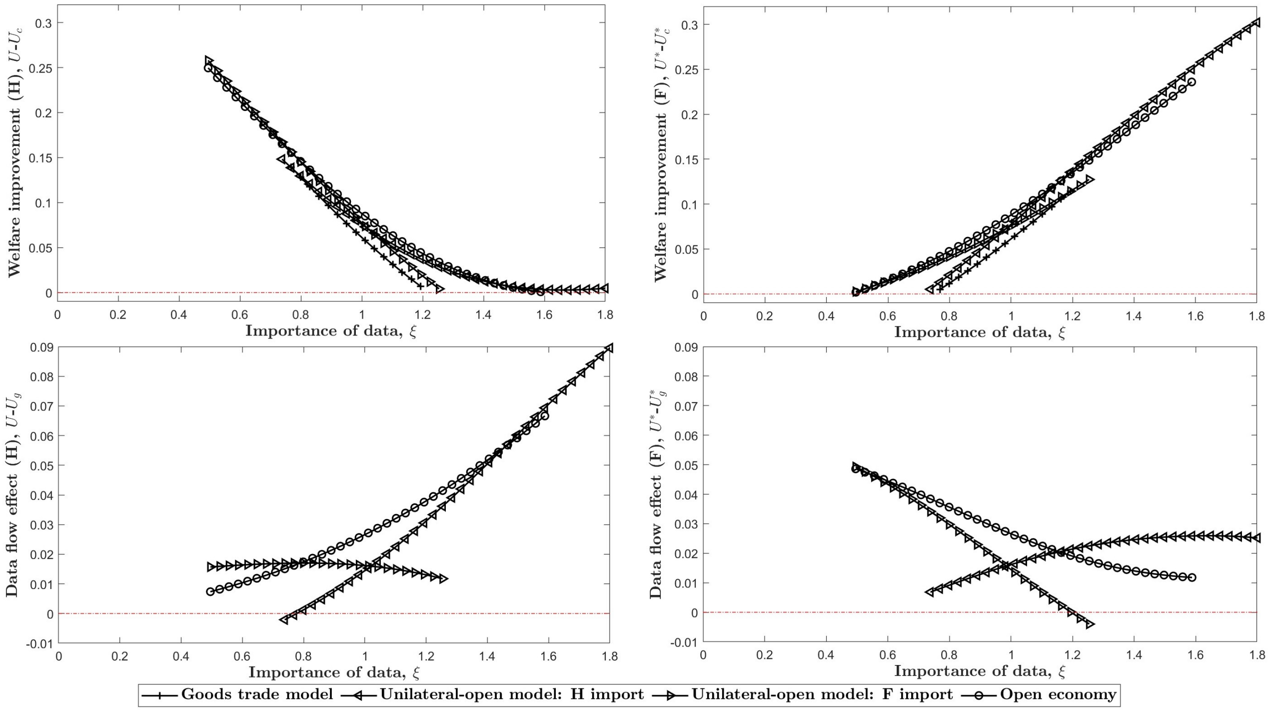

Figure 4: Welfare Improvements from Trade and Data Flows under Different Importance of Data (ξ)

Notes. The figure shows the relationship between the welfare improvements due to trade (the upper two sub-figures), as well as due to data flows (the lower two sub-figures), and the importance of data ξ in the home country in four different models in steady states: the model with only goods traded, the two unilateral flow models, and the open economy with bilateral data flows. We present the models with the elasticity of substitution of the data $\omega = 5$, and the importance of data in the foreign country is fixed at $\xi^* = 1$. We only retain the regions that have positive welfare improvements for both countries (i.e., both are willing to trade).

economy. Data flows can provide additional gains to both countries from enlarging the usage of data, pushing the welfare curve upward and enabling trade that is otherwise infeasible under some large differences in ξ.[12]

We can delve deeper into the impact of cross-border data flows by comparing the goods trade model with the open economies. In the lower panel of Figure 4, the differences in welfare between these two models can be interpreted as the data flow effects. Overall, cross-border data flows tend to improve welfare in most cases for both countries. On the one hand, by comparing the goods trade model with the unilateral flow model in which only

[12]For the country with low ξ, allowing data flows increases its production and thus improve its consumers' welfare. At the same time, for the country with high ξ, its consumers' welfare can not only be improved from this channel, but can also be pushed higher by lowering the usage of raw data in the country and mitigating the privacy costs (more working data are concentrated in this country).

the foreign country imports data, we find that the home country still experiences welfare improvements from trade, even though it only exports data at the cost of an additional utility loss. However, this unilateral flow model only has a very small feasible interval (from about $\xi = 0.5$ to about $\xi = 1.3$ when $\xi^* = 1$), which shows a narrow desirable range of this outward data flow. On the other hand, the welfare improvement curve of the unilateral flow model in which only the home country imports data is close to that of the open economy, highlighting the significant welfare gains from importing foreign data. It is also worth noting that this unilateral flow model with data imports can be a desirable alternative when bilateral data flow (open economy) becomes undesirable, especially when ξ becomes very large.

Decomposition of welfare effects. To further understand the two forces that affect welfare changes moving from a closed model to an open economy with both goods trade and data flows, we decompose welfare improvements and examine the effects of pre-existing data divide. Figure 5 shows that as the divide shrinks, welfare improvements from goods trade decrease while those from data flows increase in the data-inefficient country. Conversely, in the data-efficient country, welfare improvements from goods trade increase while those from data flows decrease. Furthermore, goods trade can lead to welfare loss in the data-efficient country, but this effect is compensated by data flows, which improves welfare.

The migration of labor among different sectors can explain the wane and wax of the two forces. As data divide gets larger, working data in the data-inefficient country becomes less important, which reduces the demand for labor in the data intermediary in that country. Moreover, the data-inefficient country has relatively low productivity, which further reduces labor demand in the production sector. These two factors combined reduce the cost of generating exported working data, which leads to an increase in the data-efficient country's importation of working data from the data-inefficient country, and subsequently increases welfare in the data-efficient country. However, as data divide becomes very large, the welfare loss from goods trade becomes dominant, and the data-efficient country may refuse to trade goods or data.

Feasible intervals of international trade and data flows. We present in Table 2 intervals of ξ in the home country for different levels of ξ^* in the foreign country so that trade (and data

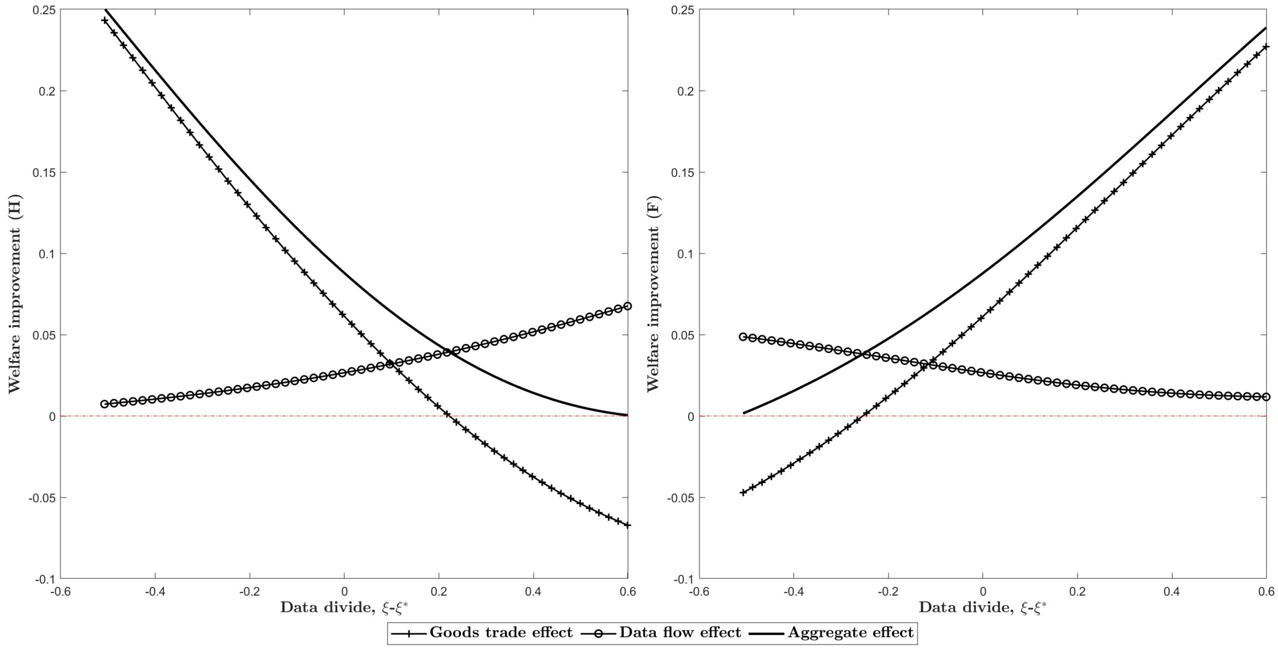

Figure 5: Decomposition of Welfare Improvements under Different Levels of Data Divide

Notes. These figures outline the decomposition of welfare improvements from goods trade and data flows with different levels of data divide ($\xi - \xi^*$) between the two countries. We only present the models when the elasticity of substitution of the data is $\omega = 5$, and the importance of data in the foreign country is fixed at $\xi^* = 1$.

flows) is feasible. In addition to the partially open models that were previously discussed, we also report results for open economies with different levels of data import frictions (by varying d) to demonstrate the variation of the intervals. Appendix A presents the welfare improvements due to trade and data flows with different levels of d. In the table, We focus on ξ and ξ^* within $[0, 1.75]$ for illustration. In general, the feasible intervals of ξ are always around the corresponding level of ξ^*, and these intervals become larger as we extend from the goods trade model to open economies with low import costs in most cases. Data flows can make trade between the two countries more desirable, especially when the data divide is large and yet not too large that trade becomes infeasible.

Gaps in equilibrium outcomes and the impact of trade and data flows. We next compute the raw data, working data, total production, and welfare improvement for the two countries, and plot in Figure 6 how trade and data flows alter the gaps in these outcomes with diff. In the figure, we consider ξ^* at three levels, namely 0.5, 1.0, and 1.5, and adjust ξ continuously.

Table 2: Feasible Intervals of Openness and Data Flows in Different Cases

	Models	**Foreign Country: Importance of Data, ξ^*** $\xi^* = 0.5$	$\xi^* = 1.0$	$\xi^* = 1.5$
Home Country: Feasible Interval of Trade, ξ. ($U - U_c > 0$ and $U^* - U_c^* > 0$)	Goods Trade Model	$[0, 0.79]$	$[0.75, 1.22]$	$[1.31, 1.67]$
	Unilateral Flow Model (F importing)	$[0, 0.80]$	$[0.47, 1.27]$	$[0.88, 1.75]$
	Unilateral Flow Model (H importing)	$[0, 1.03]$	$[0.72, 1.75]$	$[1.23, 1.75]$
	Open Economy: $d = 1$	$[0, 1.02]$	$[0.48, 1.61]$	$[0.93, 1.75]$
	Open Economy: $d = 1.5$	$[0, 0.91]$	$[0.49, 1.39]$	$[0.94, 1.75]$
	Open Economy: $d = 2$	$[0, 0.87]$	$[0.50, 1.33]$	$[0.96, 1.75]$
Home Country: Interval of Positive Data Flow Effect, ξ. ($U - U_g > 0$ and $U^* - U_g^* > 0$)	Unilateral Flow Model (F importing)	$[0, 0.73]$	$[0.47, 1.19]$	$[0.88, 1.61]$
	Unilateral Flow Model (H importing)	$[0.29, 1.03]$	$[0.78, 1.75]$	$[1.37, 1.75]$
	Open Economy: $d = 1$	$[0.13, 1.02]$	$[0.48, 1.61]$	$[0.93, 1.75]$
	Open Economy: $d = 1.5$	$[0.15, 0.91]$	$[0.49, 1.39]$	$[0.94, 1.75]$
	Open Economy: $d = 2$	$[0.17, 0.87]$	$[0.50, 1.33]$	$[0.96, 1.75]$

Notes: This table shows two intervals of the importance of data in the home country ξ given different levels of ξ^*: the interval that is feasible for trade and data flows, and the interval that has positive data flow effects. We only focus on the range that ξ and ξ^* are within $[0, 1.75]$. The import cost multiplier in the foreign country is fixed at $d^* = 1$. The values of other parameters are shown in Table 1.

In general, we observe a negative relationship between the usage of raw data and working data. The data-inefficient country provides more raw data for the generation of working data, while the data-efficient country uses more working data for production. The country with a larger ξ also ends up producing more final goods. The data-inefficient country has a larger welfare improvement from goods trade and data flows, which can indeed be beneficial, as long as its production efficiency from the use of data is high enough for the data-efficient country to agree to trade.

These results align with the prevailing trend in the global digital economy, where the United States, the European Union, and China are the three major economic entities driving cross-border data flows at present. Zheng (2021) documents and compares the different regulatory paradigms of cross-border data flows among these three main economies. The U.S., with its advanced digital economy, attracts a significant amount of global data due

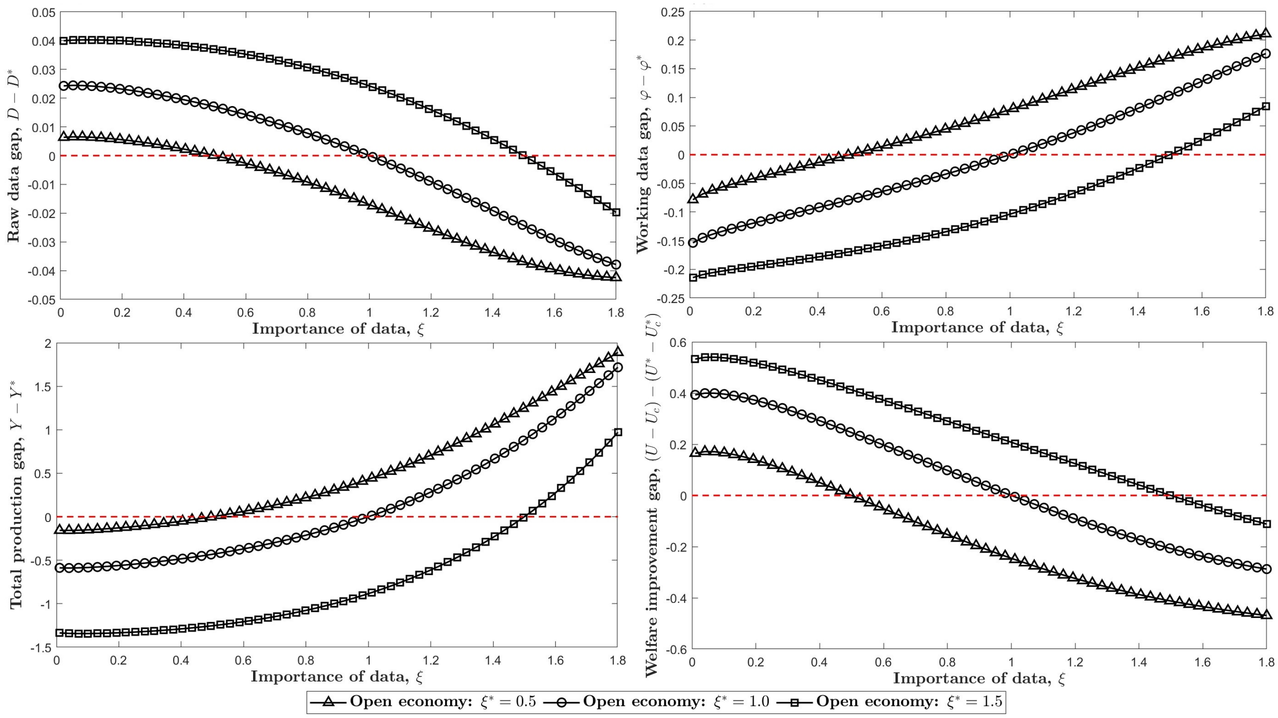

Figure 6: Gaps between Asymmetric Countries with Different Importance of Data (ξ)

Notes. These figures show the gaps in raw data contribution, working data usage, total production, and welfare improvement in the presence of trade and data flows. We only focus on the open economies ξ^* fixed at three values: 0.5, 1.0, and 1.5.

to its relatively permissive data flow policies. In contrast, the EU, with a less developed digital economy, may not receive as much data from other countries due to its stricter data flow policies. China falls somewhere in between, with relatively strong digital economic development and flexible policies, which allows it to attract a certain amount of working data.

4. Transition Dynamics

We turn to the transition dynamics when a country experiences exogenous shocks to key variables. Specifically, we consider productivity shocks A_t (shown in (6)), disutility shocks to the usage of raw data π_t (shown in (1)), and cost shocks to importing data for production f_t (shown in (14)). For clarity and without loss of generality, we examine the responses of the home country to positive shocks and keep the parameters for the foreign country unchanged.

Our analyses are divided into two parts. First, we show the transition dynamics with two ex-ante symmetric countries, and the only difference is that the home country undergoes an exogenous shock. Then, we analyze the case of two asymmetric countries where the importance of data (ξ and ξ^*) differ. For tractability, we do not let ξ to change over the short horizon we focus on. If the use of working data feeds back to how ξ changes, then many of the results should be further amplified. Unless otherwise specified, the parameters take the values shown in Table 1.

4.1 Symmetric Countries in Transition

For illustration, we start with the cases where the two countries are symmetric. Our findings do not depend on the knife-edge case with perfect symmetry and would go through as long as the two countries are very similar, especially in terms of the importance of data in production. We describe the changes in 20 variables of the economies after the shocks. In each of the analyses, we provide the results of six models: the closed model, the goods trade model, along with the open economies with four different levels of elasticity of substitution of data ω, which range from $\omega = 0.01$ (lacking elasticity) to $\omega = 10$ (full of elasticity).

Productivity shocks. We begin by analyzing the effects of exogenous productivity shocks on different models and present the results in Figure 7. In general, the open economy with different levels of elasticity of substitution (ω) and the goods trade model exhibit similar production-related outcomes in response to the shock. In contrast, the closed economy experiences different transition paths. This indicates that while data combinations have a loose connection with final good production in the model, trade liberalization has a significant impact on the models.[13] Among all the variables, the most significant differences arise in the working data flow (φ) and the data stock (Φ). In the open economies, these two data-related variables increase to levels that are higher than the steady state levels before the shock is eliminated. However, in the closed and goods trade models, they decrease sharply before returning to the steady states.

[13]In the closed model, we only have impulse response functions regarding the following variables: Y, C, N, K, φ, n, l_H, D, and Φ. In the following figures, we also present the results in a similar way.

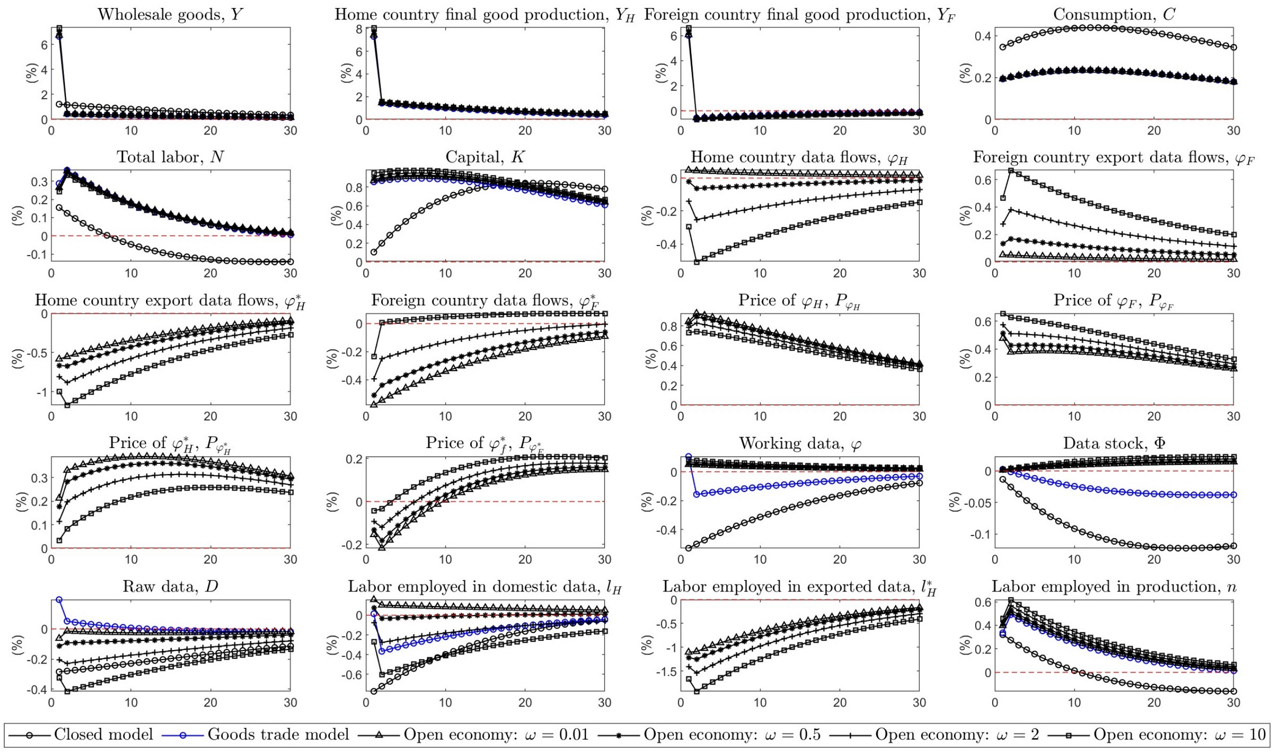

Figure 7: Impulse Responses of the Positive Productivity Shock A_t in Symmetric Countries

Note. This figure illustrates the impulse responses of 20 main variables to a positive productivity shock A_t in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% productivity shock in six different models: the closed model, the model with only goods traded, and the four models of the open economy with different levels of elasticity of substitution of data ($\omega = 0.01$, $\omega = 0.5$, $\omega = 2$, and $\omega = 10$). The lines of the open economies and the goods trade model are very close to each other in some variables.

The variations of data cyclicity observed in the closed model and in the open economies are caused by different sources of the data and goods traded. In the closed model and the goods trade model, data only come from domestic households and are crowded out by other factors, such as capital and labor, when a productivity shock occurs. In open economies, data come from diverse sources, making them cheaper and more flexible. When a productivity shock occurs in the home country, data concentrate in the high-productivity country, leading to higher quantities of working data and data stock than those in the steady states.[14] Notably, the goods trade model exhibits a smoother variation than that in the closed model due to

[14]We can further consider this resource reallocation through the four directions of data flows. The data flow toward the home country φ_F increases sharply, while the decreasing of φ_H is relatively smooth. This leads to the pro-cyclicity of working data in the home country. Meanwhile, as the elasticity of substitution ω increases, φ_H and φ_F go in the opposite way, thus we see insignificant changes as for their combination φ.

the buffering effect of goods trade on data flow variations during shocks. These differences in data cyclicity highlight why different policies may be needed for cross-border data flows to mitigate aggregate shocks.

Further insights can be gained when we examine the distinctions in wholesale goods, denoted as Y, and consumption, referred to as C, within the contexts of closed and open economies. In the absence of international trade, the total production and consumption levels exceed those observed when trade becomes open. The rationale behind this phenomenon lies in the fact that, when a country experiences a positive productivity shock, some of the surplus output is exported to foreign countries in the open economies. This shifting pattern corresponds with changes in working data and data stock, as inputs like data tend to be concentrated in the high-productivity country, while outputs are distributed across both countries.[15]

Disutility shocks and cost shocks. We proceed to analyze the transitions after shocks related to data. First, consider a shock to the disutility of using raw data (Figure 8), denoted as π_t, in the utility function (1), and a shock to the cost multiplier of importing data (Figure 9), denoted by f_t in (14). In addition, we present the transitions in the closed model and the goods trade model using the same notation π_t to represent the disutility shock in the corresponding utility function (1).

In Figure 8, it is evident that the closed model and goods trade model exhibit distinct behaviors compared to the open economies across most variables. The two models without data flows exhibit similar dynamics with respect to data-related variables, such as working data φ and data stock Φ. These findings highlight the implications of allowing cross-border data flows for transition dynamics, in addition to the welfare analysis conducted in the previous section. Specifically, while the fluctuations in working data and data stock after the shocks are subdued in open economies, they are significantly higher in the closed and goods trade models. This reflects the substitution of data from different countries, which mitigates the variation of total working data used in production in the economy.

In Figure 9, we observe that the shock to the price of imported data flows $P_{\varphi,F}$ results

[15]This intuition is consistent with our earlier finding in the welfare analysis that "the country with high data importance (ξ) tends to export more goods."

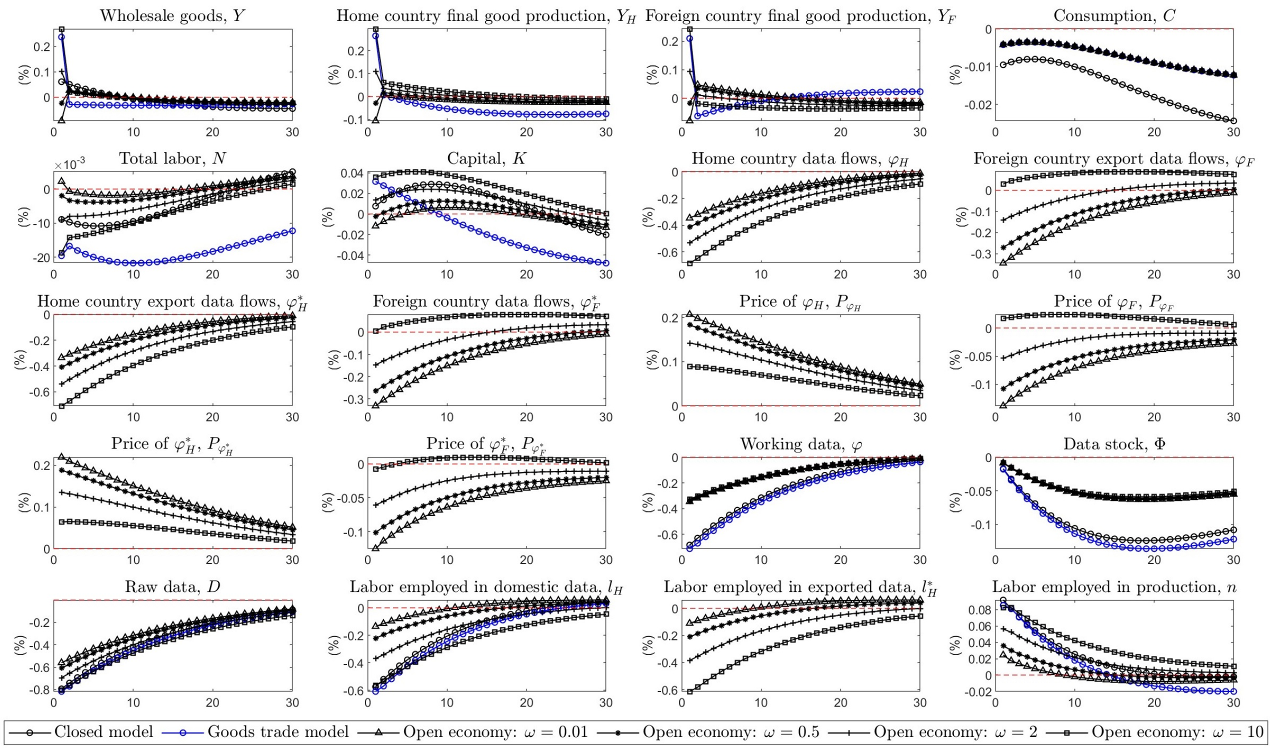

Figure 8: Impulse Responses of the Positive Disutility Shock to Raw Data π_t in Symmetric Countries

Note. This figure shows the impulse responses of 20 main variables to a positive disutility shock π_t to raw data flows. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% disutility shock to the raw data flows in the six different models: the closed model, the model with only goods traded, and the four models of open economies with different levels of elasticity of substitution of data ($\omega = 0.01$, $\omega = 0.5$, $\omega = 2$, and $\omega = 10$).

in a decrease of less than 1%. This reduction in turn leads to a decrease in the quantity of imported data φ_F, which further reduces the home country data flows φ_H when their substitution is not flexible enough. However, as the elasticity of substitution ω increases to around 10, there is a reversal in the trend of φ_H, switching from decreasing to increasing. Similar patterns can be observed in the transitions of data exported to foreign countries φ_H^*, as both types of data are generated from the same raw data.

Comparing the transition dynamics in these two figures, we see that although both shocks increase restrictions on the use of data, they affect the domestic and foreign data flows in different ways. Specifically, a disutility shock leads to a restriction on domestically generated data (φ_H and φ_H^*) due to an increase in the cost of raw data in the home country, whereas

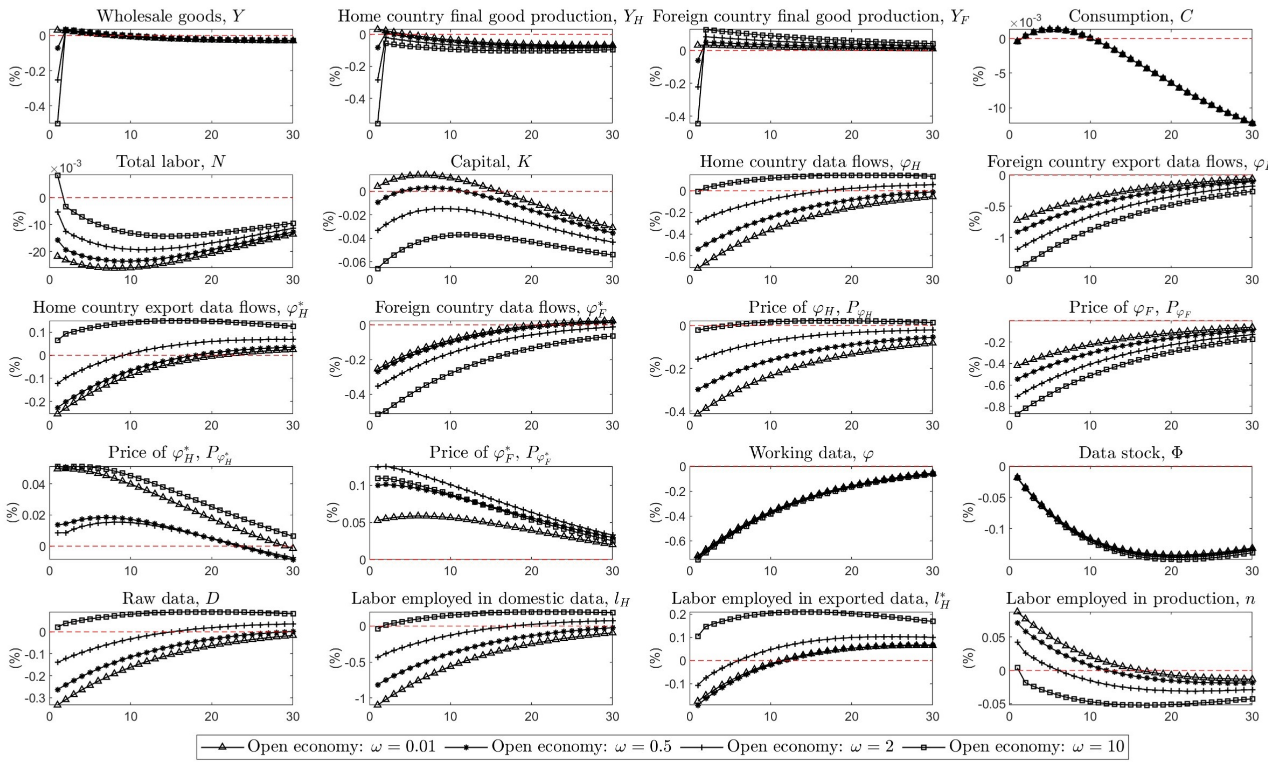

Figure 9: Impulse Responses of the Positive Cost Shock to Imported Data Flows f_t in Symmetric Countries

Notes. This figure depicts the impulse responses of 20 main variables to a positive cost shock to imported data flows f_t in production in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% cost shock to imported data flows in production in the four models of the open economy with different levels of elasticity of substitution of data ($\omega = 0.01$, $\omega = 0.5$, $\omega = 2$, and $\omega = 10$).

an import cost shock affects foreign-generated data (φ_F and φ_F^*). Firms tend to use working data that have relatively low costs; thus, we see that φ_H and φ_H^* fluctuate more severely when the elasticity of substitution ω increases under a disutility shock, and the relationship becomes reversed under a cost shock. This opposite relationship reflects how data flows respond to different types of shocks. In addition to the welfare effects discussed in the previous section, governments should also carefully consider the fluctuation effects of these two types of shocks when making policies. For the former shock, it usually relates to changes in privacy concerns and data regulation, while for the latter one, it can arise from agreements on cross-border data flows or geopolitical tensions among countries. A higher variation of data usage can reduce data labor payoff, which, in turn, may harm the development of the data economy in the long run.

4.2 Asymmetric Countries with Pre-Existing Data Divide

In the following investigation, we focus on data-related variables and explain the changes in the 16 equilibrium variables after the shocks. In each of the analyses, we focus on the open economies with different levels of elasticity of substitution of data ω, consider $\xi = 0.5$, $\xi^* = 1.0$ as well as their reverse, and subject one of the countries to exogenous shocks. Because we do not observe significant differences for non-productivity shocks, we focus on reporting our findings on the transitions after productivity shocks and leave the rest to Appendix B.1.

Figure 10 depicts the transition dynamics of countries with different importance of data after an exogenous productivity shock. We observe that the open economies with different levels of elasticity of substitution ω show similar patterns for the production-related variables C and K, consistent with the analysis in the previous subsection. However, the most significant differences are in the transitions of working data φ and data stock Φ, which represent the usage of data in the country. When the productivity shock occurs in a country with relatively low data importance, φ and Φ decrease before returning to the steady states. In contrast, when this shock occurs in a data-efficient country with high data importance, these two variables increase sharply before returning to the steady state. These opposite directions of transition dynamics widen the data usage gap between the two countries with a pre-existing data divide, which further exacerbates the concentration of data distribution in the country with a greater importance of data.

The asymmetric countries analyzed in this subsection can be considered as a generalization of the transition dynamics observed in the symmetric countries discussed in the previous subsection, where we observe an increase in both working data and data stock following a positive productivity shock. However, under this asymmetric situation, both of these variables decrease after the shock in the case of a country with relatively low ξ. This can be attributed to the fact that the data divide between the two countries is still too large, which places the data-inefficient country in a disadvantaged position, even if that country experiences a positive productivity shock. As a result, the data-efficient country experiences

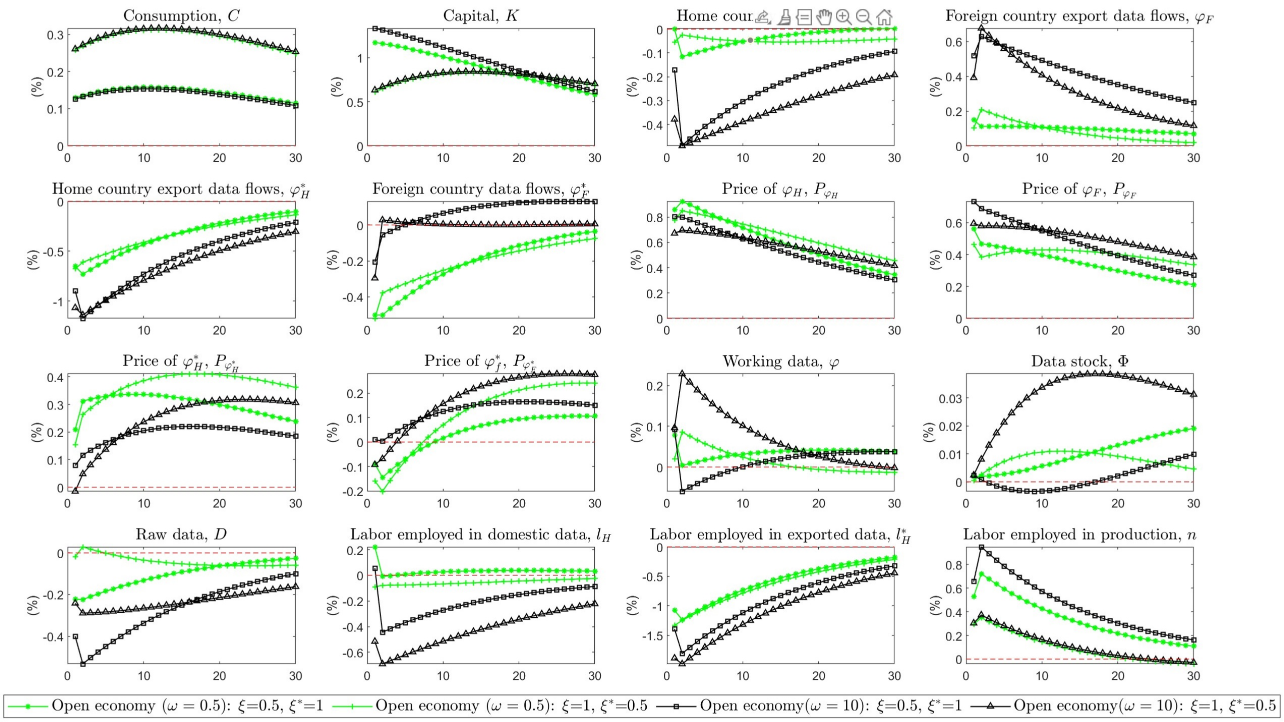

Figure 10: Impulse Responses of the Positive Productivity Shock A_t in Asymmetric Countries

Note. This figure illustrates the impulse responses of 16 main variables to a positive productivity shock A_t in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% productivity shock in four different models: the open economy with different levels of elasticity of substitution of data ($\omega = 0.5$ and $\omega = 10$) and different importance of data ($\xi = 0.5$ and $\xi^* = 1.0$, together with $\xi = 1.0$ and $\xi^* = 0.5$). The productivity shock only happens in the home country.

a production expansion and absorbs more working data from the data-inefficient country, leading to a counter-cyclical pattern of data usage in the latter country. We further support this argument in Appendix B.2, where we demonstrate that as the data divide between the two countries diminishes, working data become pro-cyclical when either of the countries experiences a productivity shock.

Finally, concerning changes in consumption, we observe an increase in consumption in both countries following the shock, albeit with a larger variation in the more data-efficient country. The sensitivity of production to the shock is greater in this country, as data play a more crucial role in promoting production. Put differently, in the event of a positive productivity shock to the more data-efficient country, it becomes less willing to trade with the other country and the gap in production widens even further.

4.3 The Effect of Data Accumulation and Depreciation

Different from existing models on the data economy, we allow data to accumulate. In general, we find that data accumulation moderates the fluctuations in the working data and data stock, as well as their cyclicity after productivity shocks. The reason is that the demand for working data becomes greater when working data depreciate at a higher rate, which in turn alleviates the crowding out effect from other factors when the economy experiences a positive productivity shock.

For illustration, Figure 11 plots the transition dynamics after a positive productivity shock when data are fully depreciated. We focus on the two data-related variables φ and Φ and compare this figure with Figure 7, where data depreciate partially (i.e., accumulate) in our baseline model.

Comparing the figures, we observe that when working data change from full depreciation to full accumulation, the fluctuations of φ and Φ in the open economies become subdued, and the two variables switch from pro-cyclical to counter-cyclical in the closed economy. As for the disutility shock and the cost shock, the depreciation rate of data stock does not have a significant effect and are thus not reported here.

5. Conclusion

We build a general equilibrium model of production, trade, and cross-border data flows. Our findings suggest that international data flows can significantly improve welfare in steady states, especially when the importance of data in a country is smaller than that in other countries—a latecomer's advantage. However, trade liberalization may come to a halt when the data divide between two countries is too large, especially with restricted cross-border data flows. We also find that working data tend to concentrate in the data-efficient country, while raw data primarily come from the data-inefficient country. Finally, we show that, unlike a closed economy, an open economy with data flows experiences a reversed cyclical pattern after an aggregate productivity shock, whereas shocks to data privacy cost or data import costs have the opposite effects on domestic and foreign data sectors.

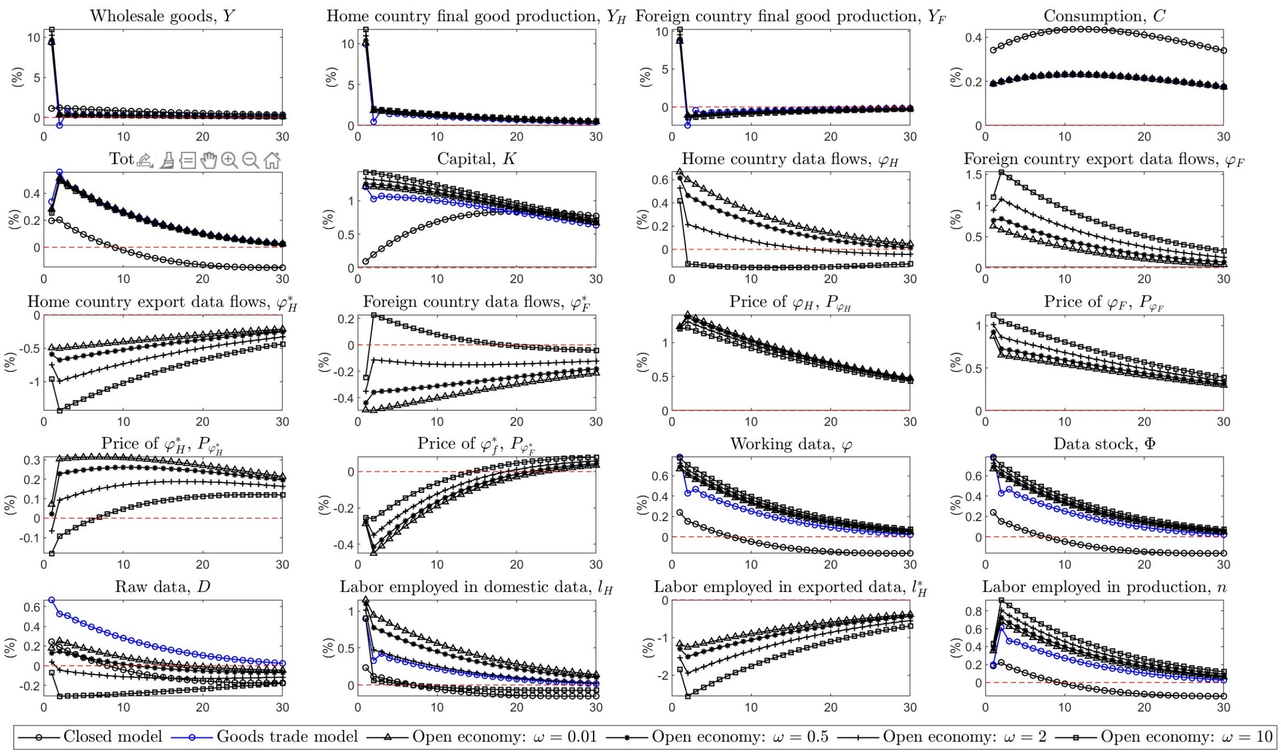

Figure 11: Impulse Responses of the Positive Productivity Shock A_t in Symmetric Countries When Data are Fully Depreciated

Note. This figure illustrates the impulse responses of 20 main variables to a positive productivity shock A_t in the home country when data are fully depreciated. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% productivity shock in six different models: the closed model, the model with only goods traded, and the four models of the open economy with different levels of elasticity of substitution of data ($\omega = 0.01$, $\omega = 0.5$, $\omega = 2$, and $\omega = 10$). The lines of the open economies and the goods trade model are very close to each other in some variables.

Our study contributes to the literature by providing the first analysis of data factor and its cross-border flows in the international context. Despite recent progress in this field, measurements of cross-border data flows and value added are still lacking (e.g., Beraja et al., 2023; Veldkamp, 2023). Our paper provides an initial theoretical benchmark for further research, both theoretical and empirical, on international data flows and their effects on development and international trade. Our model is flexible to admit extensions along multiple dimensions, including data ownership, privacy protection, antitrust, and others. In particular, considering the feedback of working data utilization to the gradual changes in data importance ξ constitutes interesting future research. Extending the model to a multi-country setting to match empirical patterns of cross-border data flows also likely provides further quantitative insights. Finally, our findings hopefully offer policy guidance concerning the development of data-related industries, restricting cross-border data flows, and mitigating aggregate domestic shocks in the global data economy.

References

Abis, S. and L. Veldkamp (2021). The Changing Economics of Knowledge Production. Working Paper, Columbia Business School, New York.

Abowd, J. M. and I. M. Schmutte (2019). An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *American Economic Review 109*(1), 171–202.

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022). Too Much Data: Prices and Inefficiencies in Data Markets. *American Economic Journal: Microeconomics 14*(4), 218–256.

Acquisti, A., C. Taylor, and L. Wagman (2016). The Economics of Privacy. *Journal of Economic Literature 54*(13), 442–492.

Adler, G., R. Lama, and J. P. Medina (2019). Unconventional Policies and Exchange Rate Dynamics. *Journal of International Money and Finance 95*, 402–423.

Alessandria, G., H. Choi, and K. J. Ruhl (2021). Trade Adjustment Dynamics and the Welfare Gains from Trade. *Journal of International Economics 131*, 103458.

Alessandria, G., J. Kaboski, and V. Midrigan (2013). Trade Wedges, Inventories, and International Business Cycles. *Journal of Monetary Economics 60*(1), 1–20.

Babina, T., G. Buchak, and W. Gornall (2022). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. Available at SSRN 4071214.

Bacchetta, P., M. Davenport, and E. Van Wincoop (2022). Can Sticky Portfolios Explain International Capital Flows and Asset Prices? *Journal of International Economics 136*, 103583.

Bagwell, K. and R. W. Staiger (1999). An Economic Theory of GATT. *American Economic Review 89*(1), 215–248.

Bajari, P., V. Chernozhukov, A. Hortaçsu, and J. Suzuki (2019). The Impact of Big Data on Firm Performance: An Empirical Investigation. *AEA Papers and Proceedings 109*, 33–37.

Beraja, M., D. Y. Yang, and N. Yuchtman (2023). Data-Intensive Innovation and the State: Evidence from AI Firms in China. *Review of Economic Studies 90*(4), 1701–1723.

Bergemann, D. and A. Bonatti (2019). Markets for Information: An Introduction. *Annual Review of Economics 11*, 85–107.

Bhattarai, K. and D. Trzeciakiewicz (2017). Macroeconomic Impacts of Fiscal Policy Shocks in the UK: A DSGE Analysis. *Economic Modelling 61*, 321–338.

Brynjolfsson, E. and K. McElheran (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review 106*(5), 133–139.

Buera, F. J. and E. Oberfield (2020). The Global Diffusion of Ideas. *Econometrica 88*(1), 83–114.

Caldara, D., M. Iacoviello, P. Molligo, A. Prestipino, and A. Raffo (2020). The Economic Effects of Trade Policy Uncertainty. *Journal of Monetary Economics 109*, 38–59.

Casadesus-Masanell, R. and A. Hervas-Drane (2015). Competing with Privacy. *Management Science 61*(1), 229–246.

Ca'Zorzi, M., M. Kolasa, and M. Rubaszek (2017). Exchange Rate Forecasting with DSGE Models. *Journal of International Economics 107*, 127–146.

Christensen, I. and A. Dib (2008). The Financial Accelerator in an Estimated New Keynesian Model. *Review of Economic Dynamics 11*(1), 155–178.

Clarida, R., J. Galí, and M. Gertler (2002). A Simple Framework for International Monetary Policy Analysis. *Journal of Monetary Economics 49*(5), 879–904.

Cong, L. W. and S. Mayer (2023). Antitrust, Regulation, and User Union in the Era of Digital Platforms and Big Data. NBER Working Papers 30881, National Bureau of Economic Research.

Cong, L. W., W. Wei, D. Xie, and L. Zhang (2022). Endogenous Growth under Multiple Uses of Data. *Journal of Economic Dynamics and Control 141*, 104395.

Cong, L. W., D. Xie, and L. Zhang (2021). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science 67*(10), 6480–6492.

Cory, N. (2017). Cross-border Data Flows: Where Are the Barriers, and What Do They Cost? Technical report, Information Technology and Innovation Foundation.

Cory, N. and L. Dascoli (2021). How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them. Technical report, Information Technology and Innovation Foundation.

Devereux, M. B., E. R. Young, and C. Yu (2019). Capital Controls and Monetary Policy in Sudden-Stop Economies. *Journal of Monetary Economics 103*, 52–74.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review 67*(3), 297–308.

Eeckhout, J. and L. Veldkamp (2023). Data and Markups: A Macro Information Frictions Perspective. Working Paper.

Erickson, S. and H. Rothberg (2014). Big Data and Knowledge Management: Establishing a Conceptual Foundation. *Electronic Journal of Knowledge Management 12*(2), 108–116.

Fainmesser, I., A. Galeotti, and R. Momot (2022). Digital Privacy. *Management Science 69*(6), 3157–3173.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). Income Distribution, Product Quality, and International Trade. *Journal of Political Economy 119*(4), 721–765.

Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp (2019). Big Data and Firm Dynamics. *AEA Papers and Proceedings 109*, 38–42.

Farboodi, M., D. Singal, L. Veldkamp, and V. Venkateswaran (2022). Valuing Financial Data. NBER Working Papers 29894, National Bureau of Economic Research.

Farboodi, M. and L. Veldkamp (2020). Long-Run Growth of Financial Data Technology. *American Economic Review 110*(8), 2485–2523.

Farboodi, M. and L. Veldkamp (2021). A Model of the Data Economy. NBER Working Papers 28427, National Bureau of Economic Research.

Fernández-Villaverde, J., P. Guerrón-Quintana, K. Kuester, and J. Rubio-Ramírez (2015). Fiscal Volatility Shocks and Economic Activity. *American Economic Review 105*(11), 3352–3384.

Galí, J. and T. Monacelli (2005). Monetary Policy and Exchange Rate Volatility in a Small Open Economy. *Review of Economic Studies 72*(3), 707–734.

Goldstein, I., C. Huang, and L. Yang (2022). Open Banking with Depositor Monitoring. Working Paper.

Gross, T. (2021). Dynamic Optimal Fiscal Policy in a Transfer Union. *Review of Economic Dynamics 42*, 194–238.

He, Z., J. Huang, and J. Zhou (2022). Open Banking: Credit Market Competition When Borrowers Own the Data. NBER Working Papers 28118, National Bureau of Economic Research.

Hou, Y., J. Huang, D. Xie, and W. Zhou (2022). The Limits to Growth in the Data Economy: How Data Storage Constraint Threats. Available at SSRN 4099544.

Ichihashi, S. (2020). Online Privacy and Information Disclosure by Consumers. *American Economic Review 110*(2), 569–595.

Ichihashi, S. (2021a). Competing Data Intermediaries. *RAND Journal of Economics 52*(3), 515–537.

Ichihashi, S. (2021b). The Economics of Data Externalities. *Journal of Economic Theory 196*, 105316.

Jones, C. I. and C. Tonetti (2020). Nonrivalry and the Economics of Data. *American Economic Review 110*(9), 2819–2858.

Jouanjean, M.-A. (2019). Digital Opportunities for Trade in the Agriculture and Food Sectors. OECD Food, Agriculture and Fisheries Papers 122, OECD Publishing.

Kleinman, B., E. Liu, S. J. Redding, and M. Yogo (2023). Neoclassical Growth in an Interdependent World. Working Paper.

Kubina, M., M. Varmus, and I. Kubinova (2015). Use of Big Data for Competitive Advantage of Company. *Procedia Economics and Finance 26*, 561–565.

Liu, J. Z., M. Sockin, and W. Xiong (2023). Data Privacy and Algorithmic Inequality. Working Paper.

McAfee, A. and E. Brynjolfsson (2012). Big Data: The Management Revolution. *Harvard Business Review 90*(10), 60–68.

Sadowski, J. (2019). When Data Is Capital: Datafication, Accumulation, and Extraction. *Big Data & Society 6*(1), 1–12.

Sun, T., Z. Yuan, C. Li, K. Zhang, and J. Xu (2021). The Value of Personal Data in Internet Commerce: A High-stake Field Experiment on Data Regulation Policy. *Management Science, forthcoming*.

van der Marel, E. and M. F. Ferracane (2021). Do Data Policy Restrictions Inhibit Trade in Services? *Review of World Economics 157*(4), 727–776.

Veldkamp, L. (2023). Valuing Data as an Asset. *Review of Finance, forthcoming*.

Veldkamp, L. and C. Chung (2023). Data and the Aggregate Economy. *Journal of Economic Literature, forthcoming*.

Xie, D. and L. Zhang (2023). A Generalized Model of Growth in the Data Economy. Available at SSRN 4033576.

Zheng, G. (2021). Trilemma and Tripartition: The Regulatory Paradigms of Cross-Border Personal Data Transfer in the EU, the U.S. and China. *Computer Law & Security Review 43*, 105610.

Appendix

A. Welfare and Data Flows Under Various Import Frictions

From Figure A.1, we see that the welfare improvements and data flow effects both decrease as the data import friction, d, increases. The feasible interval becomes narrower as d increases.

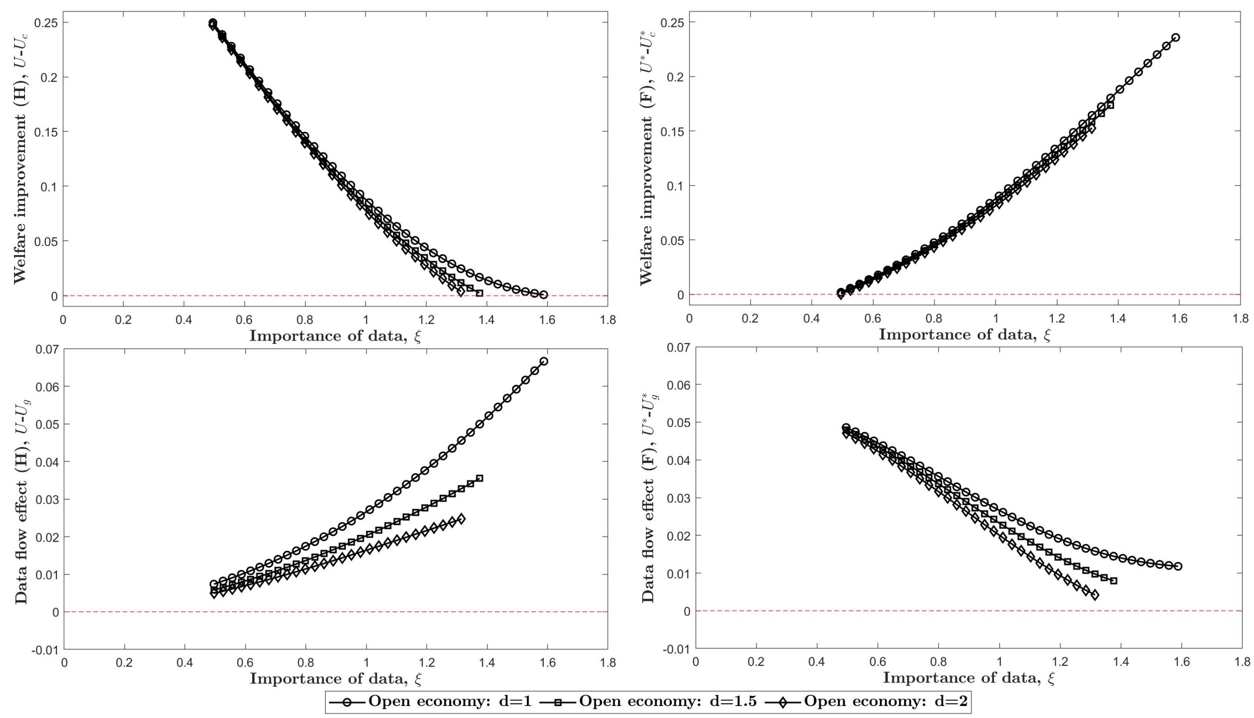

Figure A.1: Welfare Improvements and Data Flows with Different Importance of Data in the Two Countries When d Varies

*Note.*The figure shows the relationship between the welfare improvements due to trade (the upper two sub-figures), as well as due to data flows (the lower two sub-figures), and the importance of data in the composite factor ξ in the home country in open economies with different levels of d. We present the models with the elasticity of substitution of the data $\omega = 5$, and the importance of data in the foreign country is fixed at $\xi^* = 1$. We only retain regions that have positive welfare improvements for both countries (i.e., both are willing to trade).

B. Asymmetric Countries in Transitions: Other Shocks

B.1 Disutility Shock and Cost Shock Under Extant Data Divide

From Figure B.1 which shows the disutility shock, we see that the differences in transition states largely come from different levels of the elasticity of substitution ω, but the different importance of data in ξ and ξ^* do not change the result too much.

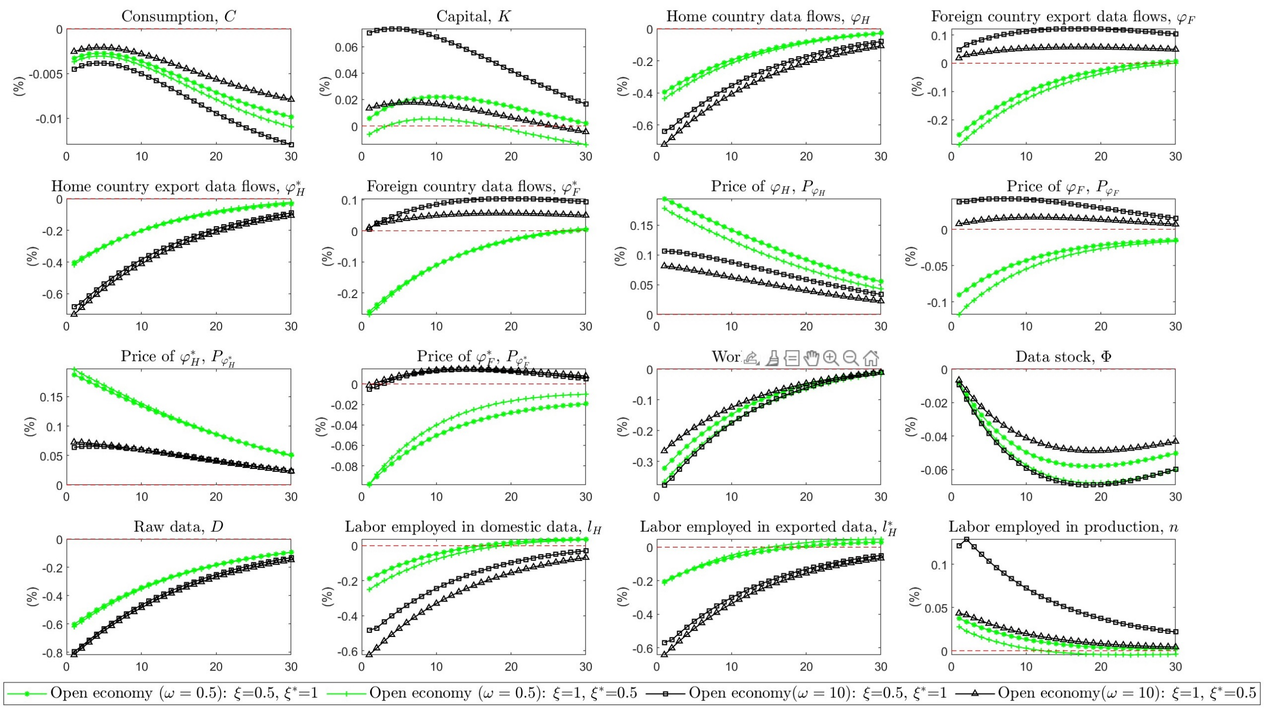

Figure B.1: Impulse Responses of the Positive Disutility Shock π_t in Asymmetric Countries

Note. This figure illustrates the impulse responses of 16 main variables to a positive disutility shock π_t concerning raw data in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% disutility shock in four different models: the open economy with different levels of elasticity of substitution of data ($\omega = 0.5$ and $\omega = 10$) and different importance of data ($\xi = 0.5$ and $\xi^* = 1.0$, together with $\xi = 1.0$ and $\xi^* = 0.5$). The disutility shock only happens in the home country.

From Figure B.2 which shows the cost shock, we see that variations are larger in the country where data are more important when there is a cost shock on imported data.

B.2 Productivity Shocks Under Small Data Divide

From Figure B.3, we see that working data and data stock in the two countries both increase when there is a positive productivity shock and the data divide is not very large, which

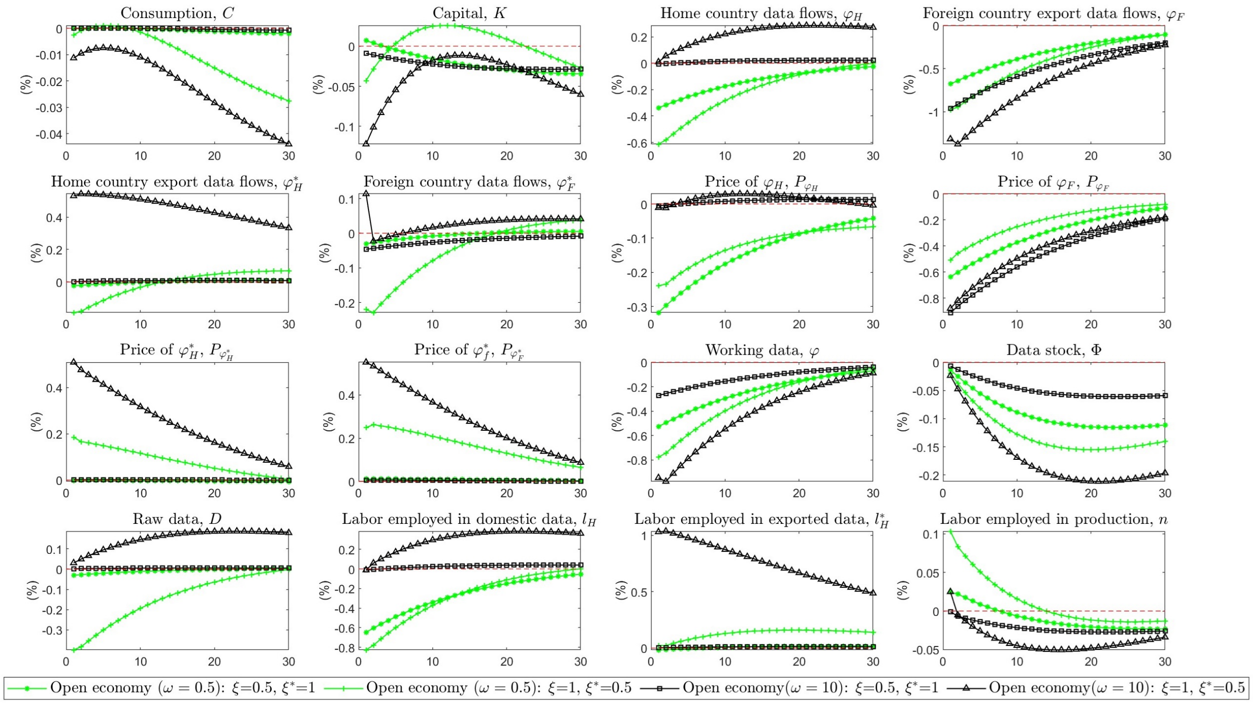

Figure B.2: Impulse Responses of the Positive Cost Shock f_t in Asymmetric Countries

Note. This figure illustrates the impulse responses of 16 main variables to a positive cost shock f_t in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% cost shock in four different models: the open economy with different levels of elasticity of substitution of data (ω = 0.5 and ω = 10) and different importance of data (ξ = 0.5 and ξ^* = 1.0, together with ξ = 1.0 and ξ^* = 0.5). The cost shock only happens in the home country.

supports the results discussed in Section 4.2.

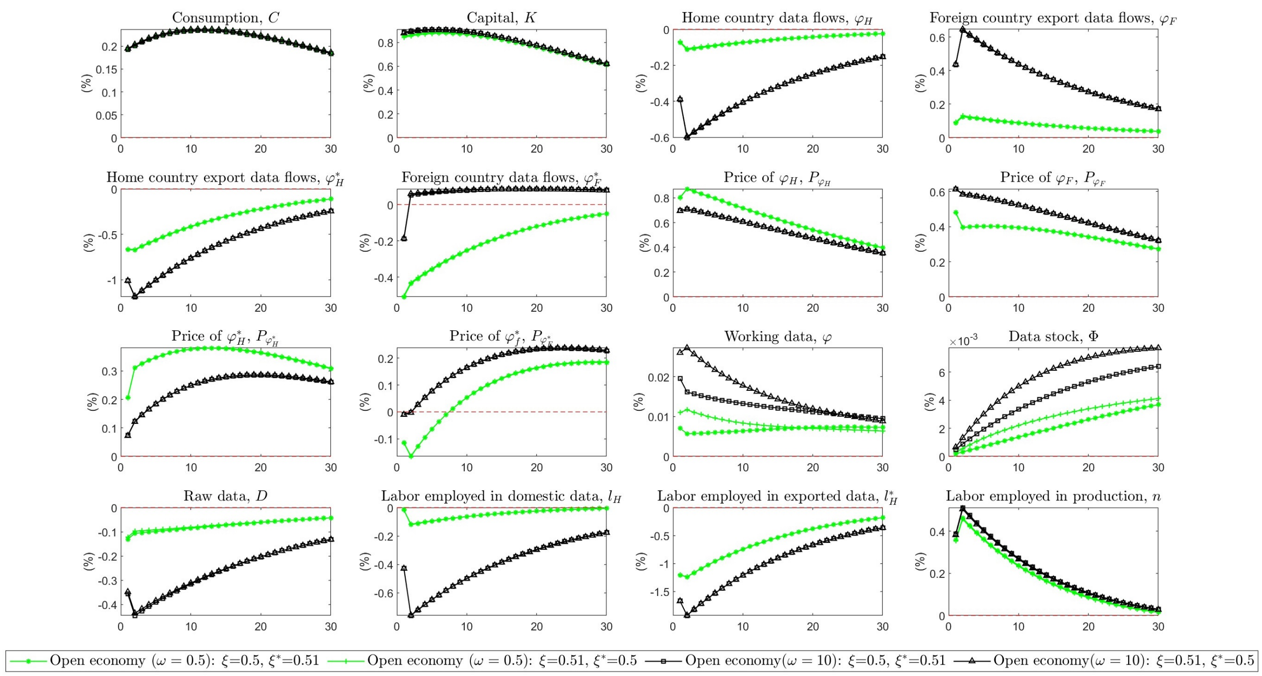

Figure B.3: Impulse Responses of the Positive Productivity Shock A_t in Asymmetric Countries When the Data Divide Is Small

Note. This figure illustrates the impulse responses of 16 main variables to a positive productivity shock A_t in the home country. Each sub-figure represents the steady-state percentage deviation of the corresponding variable (y-axis) over time (x-axis) after a positive 1% productivity shock in four different models: the open economy with different levels of elasticity of substitution of data ($\omega = 0.5$ and $\omega = 10$) and different importance of data ($\xi = 0.5$ and $\xi^* = 0.51$, together with $\xi = 0.51$ and $\xi^* = 0.5$). The productivity shock only happens in the home country.