Production, Trade, and Cross-Border Data Flows*

Qing Chang† Lin William Cong‡ Liyong Wang§ Longtian Zhang¶

First draft: September 2022; this draft: March 2026.

Abstract

We build a tractable general equilibrium model to analyze the effects of cross-border data flows and pre-existing development gaps in digital 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-light countries where data are less important in production enjoy a "latecomer's advantage" with international trade and data flows, by contributing more raw data that the data-intensive countries process into working data used to enhance production efficiency. Furthermore, cross-border data flows can reverse the cyclicality 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 research and policy discussions concerning data divide, data flows, and their implications for trade liberalization, the data labor market, data protection, among others.

Keywords: Data Economy, Data Exchange, Digitization, International Trade, Transition Dynamics

JEL Classification: F15, F29, F43

*The authors are especially grateful to Chad Jones and Laura Veldkamp for their detailed feedback and invaluable advice. They also thank Maryam Farboodi, Liyan Yang, the seminar and conference participants at the Peking University National School of Development & Institute of Digital Finance FinTech Workshop, 2023 Asian Meeting of the Econometric Society at Tsinghua University and Nanyang Technological University, 2024 North American Winter Meeting of the Econometric Society, 76th European Meeting of the Econometric Society, and the American Finance Association 2026 Annual Conference for helpful comments and suggestions. The authors acknowledge the financial support from the Ripple's University Blockchain Research Initiative (UBRI) and the FinTech at Cornell Initiative, the China Postdoctoral Science Foundation (Grant No. 2025M771818) and the Postdoctoral Fellowship Program of CPSF (Grant No. GZC20251253), and the National Natural Science Foundation of China (Grant No. 72303261, 72550001, 72273161). The content of this publication is solely the responsibility of the authors. Send correspondence to Cong at will.cong@ntu.edu.sg.

†Institute of Economics, School of Social Sciences, Tsinghua University.

‡Nanyang Technological University, ABFER, CEPR, and NBER.

§School of International Trade and Economics, Central University of Finance and Economics.

¶School of International Trade and Economics, Central University of Finance and Economics.

1. Introduction

With the rapid technological advancement involving the Internet, artificial intelligence, large-scale computation, etc., data not only grow exponentially, but also become an indispensable production factor in all major economies. 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, 2024), while affecting how firms operate and compete (e.g., Farboodi et al., 2019; Eeckhout and Veldkamp, 2024). 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).[1] Yet, despite the accelerating 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 development of data economy affect trade? Concerning policies, should a country allow importing or exporting data on top of trade, and how do policies governing data flows and foreign trade interact? How would domestic and foreign data usage and labor markets evolve after shocks to data productivity, privacy cost, and flow frictions? How do key economic outcomes evolve along the transition path when economies open up to trade and cross-border data flows? 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

[1] 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. 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 crucial in shaping international trade.

[2] 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 or fining 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; Montana became the first U.S. state to pass legislation in April 2023 banning TikTok on all personal devices; and the European Union fined Meta 1.2 billion euros and restricted it from sending data to American servers.

development of the data economy.

To this end, we build a 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. Due to the lack of data on the data economy to fully calibrate a structural model, 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 their 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, as further confirmed by the transition dynamics; (iii) with cross-border data flows, more working data are concentrated in the data-intensive country for production whereas the data-light country provides more raw data; (iv) the optimal data flow policy is largely shaped by and complements a country's international trade policies; and (v) open economies with data flows experience reversed cyclicality 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.

Our results are specific to data rather than to technology or generic production inputs for three reasons. First, because privacy costs enter household utility directly, a country may rationally refuse data exports even when aggregate output rises—a channel absent when the traded input is technology or capital. Second, importing and exporting data generate asymmetric welfare effects: importing working data raises domestic productivity without any privacy cost to domestic households, whereas exporting requires domestic households to supply raw data and bear disutility even though the productivity gain accrues abroad. Third, what crosses borders in our framework is not raw household records but intermediary-processed working data. Because data flows thereby reallocate productivity rather than output, they can compress productivity gaps and restore trade feasibility in environments where goods trade alone breaks down due to large data divides—something technology diffusion cannot achieve over the short horizons we study.

Specifically, we follow the seminal work of Ichihashi (2020), Jones and Tonetti (2020), and Far-

boodi and Veldkamp (2026) 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 supplier 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 first characterize the equilibria in steady states. Because a data-light 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-intensive counterpart may still refuse to trade if it needs to export much more intermediate goods than it imports to reach trade balance, in other words, if it faces real depreciation due to differing goods prices between the two countries. 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-intensive 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 positive productivity shock. However, as the data divide between the two countries becomes very large, working data are more concentrated in the data-intensive country, and the productivity shock in the data-light 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 the best of our 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. Adding to studies on intangible and nonrival goods (e.g., patents and software), we analyze a finer category of intangible assets which involves the critical dimension of privacy and underdeveloped sharing protocols (Ding, 2023; Ridder, 2024).[3] The emerging literature on how data affect firms and enter production mostly features domestic settings. For example, Eeckhout and Veldkamp (2024) 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, 2026). 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., 2024).

[3]There is a growing literature on the measurement and capitalization of data in the national accounts; see Corrado et al. (2025) and the literature review by Bontadini et al. (2022). This literature emphasizes that economically valuable data arise from investment-like activities and are only imperfectly captured under existing intangible asset categories. Our framework is consistent with this perspective in that working data are produced through a resource-using process and accumulate as a productive stock, but we abstract from specific accounting classifications and focus instead on economic mechanisms.

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, 2026), Hou et al. (2025), and Veldkamp and Chung (2024) 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. (2024), Farboodi et al. (2024), 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. The appendices contain derivations, results under alternative specifications and robustness tests, and extended discussions.

2. Modeling the Global Data Economy

We introduce the modeling ingredients for the closed economy, the partially open economies, and the fully open economy, each of which can be viewed as the equilibrium outcome under a particular policy regime governing whether goods trade, data importing, or data exporting is permitted. Comparing across these models offers insights for choosing among various trade and data-sharing policies. Appendix A provides a step-by-step overview of the data lifecycle in the model, tracing the process from the generation of raw records to the accumulation of productive information stocks. Appendix B presents the derivations.

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 it easier to separate the usage of data in different countries, while preserving the non-rival property of data, since raw data are used in a non-rival way to transform into working data. Second, while much of the existing literature abstracts from data accumulation or assumes that data are fully depreciated and focuses on balanced-growth outcomes, we allow data to accumulate over time (Jones and Tonetti, 2020; Cong et al., 2021). This assumption is more realistic and allows us to study how data depreciation affects transition dynamics. We note that there are recent exceptions that consider data accumulation, such as Farboodi and Veldkamp (2026), but our focus is on a multi-country general equilibrium framework with cross-border data flows.

2.1 A Motivating Example

To build intuition before the formal structure, consider two stylized economies. The first is a developing manufacturing economy—think of Vietnam or Mexico—with a large consumer base whose digital activities generate abundant raw data as byproducts of consumption. Domestic firms, however, have limited capacity to transform these records into analytics or AI-driven tools: data play only a minor role in local production. The second is a technologically advanced economy—think of the United States—where processed data are essential inputs in production, with firms relying on working data to improve forecasting, inventory management, and decision-making.

When goods trade is liberalized but cross-border data flows remain restricted, the productivity gap creates a real exchange rate imbalance. Because the developing country's goods are expensive, the advanced economy must export far more than it imports; when the data divide is large enough, this

proves sufficiently costly that the advanced economy prefers autarky, causing trade to freeze entirely.[4] Allowing cross-border data flows changes this: raw data generated by consumers in the developing economy are transmitted to and processed by the advanced economy's platforms into working data, part of which flows back to improve domestic manufacturing. The developing economy becomes a net supplier of raw data; the advanced economy a net user of working data. This reallocation relaxes the trade imbalance, restores the feasibility of goods trade, and raises welfare in both countries—with especially large gains for the data-light economy, which benefits from contributing raw data rather than competing head-to-head in data-intensive production. We call this the latecomer's advantage.

This example also clarifies the role of our assumptions. If raw data were collected by multinational firms directly rather than by domestic consumers, data sovereignty constraints would not bind and the trade-enabling mechanism above would be weaker. If data affected demand rather than productivity, the welfare implications of cross-border flows would differ, though the basic tension between data-intensive and data-light economies would remain. The formal model takes the baseline assumptions as its starting point and we discuss robustness to these alternatives in Section 3 and the appendices.

2.2 Closed Data Economy

We now formalize the story above, beginning with a closed economy in which data enter production as an input factor but do not flow across borders. In this benchmark setting with infinite and discrete time, the key agents—representative household, final good producer, intermediate good producers, and data intermediary—correspond directly to the roles described in the example, and serve as 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 owns and can sell raw data D_t to a data

[4]When we say that trade "freezes" or becomes "infeasible," we mean infeasible under voluntary participation rather than technologically impossible: the open-trade arrangement is not implementable if at least one country's equilibrium welfare under openness is below its autarky welfare, so it would choose to remain closed. With richer fiscal instruments (e.g., flexible transfers/taxes), some of these cases could feature distorted but positive trade; we abstract from such instruments here.

intermediary (introduced later) at a competitive price $P_{D,t}$.[5] 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.[6] Specifically, the household's utility optimization problem is shown as:

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

where β is the discount factor, σ is the relative risk aversion coefficient (also the inverse 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. 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.[7] This is a reduced-form way of capturing institutional and regulatory distance. Obviously, in this closed economy, $\mathbb{I} = 0$.

[5] Concretely, such raw data take the form of digital records generated at the interface between households and the platforms or devices through which consumption takes place, including browsing histories, purchase records, and usage logs. Modern data governance frameworks such as GDPR and CCPA establish households' legal rights to consent to and receive compensation for the commercial use of these records. Although a few studies consider data generated as byproducts of the production process (e.g., Farboodi and Veldkamp, 2026; Xie and Zhang, 2023), models featuring data derived from consumer activities are more standard. Moreover, as noted above, laws and regulations related to data predominantly focus on the consumer side. We therefore concentrate on consumer-generated data and leave the examination of production-generated data to future research.

[6] Following Jones and Tonetti (2020), we specify this term as a quadratic cost, implicitly assuming that households are not contributing all of their data; otherwise, the marginal cost would begin to diminish as the return to learning from additional data decreases (Farboodi and Veldkamp, 2020). We interpret the implied disutility in a Becker-style framework as reflecting households' taste for privacy, which does not contradict voluntary data sharing: data are provided when the benefits of disclosure outweigh the privacy loss. More concretely, this disutility may reflect anticipated losses from data breaches and unauthorized access, as well as downstream exploitation by third parties who profile consumers and engage in price discrimination, a dynamic captured by the concept of "enshittification." This feature distinguishes data from related concepts such as ideas, patents, and software and underlies the distinctive cost structure in our welfare results.

[7] In our model, the term $\Pi(1 + b\mathbb{I})\pi_t D_t^2$ assumes that exported data have already been used domestically, so cross-border transfer generates an additional privacy cost governed by b. This specification reflects the idea that international data transfers may create an added sense of loss of control over personal data, since individuals may be less certain about how their data will be handled once it leaves the domestic jurisdiction. Rather than assuming that foreign environments are inherently less secure, this formulation captures the possibility that cross-border transfers, by increasing uncertainty about data governance, legal protection, enforcement, and downstream use, may generate a trust deficit and thereby raise the expected cost of disclosure.

The budget constraint for the household satisfies:

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

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, F_t is the total profit received from the intermediate good producers (household owns the domestic firms), and r_t is the return on capital.[8] For clarity, we normalize the price of final 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:

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

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

[8] 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, such as access to free digital services, personalized recommendations that improve match quality, loyalty rewards, or enhanced platform functionality.

that each intermediate good producer buys the same data, which allows us to drop the subscript i for φ_t here.[9] As intermediate good producers repeatedly utilize working data, processed information becomes progressively embedded in their operational systems, algorithms, and organizational routines, augmenting the productivity of physical capital. Hence, the accumulated data $\Phi_{i,t}$ act as an augmenting factor in the formulation of a capital composite.[10] Therefore, we specify the production function 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. When $\xi = 0$, data plays no role in production and the model reduces to a standard economy, which serves as a clean benchmark for evaluating the contribution of data. 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. These producers may have non-zero profits, which is denoted as $F_{i,t}$. Obviously, we have $F_t = \int_0^1 F_{i,t} \mathrm{d}i$.

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:

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

[9] 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 E.3 discusses this and provides an example for further illustration. Moreover, note that we set a non-rival flow of working data φ_t here, while the accumulated data $\Phi_{i,t}$ remains rival. Obviously, $\Phi_{i,t}$ is identical among different varieties i due to homogeneity.

[10] 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 often modeled as a special type of capital, while Abis and Veldkamp (2024) discuss an alternative way of combining data with capital. Although working data in our framework also accumulate and provide productive services over time, we interpret them as an intangible intermediate input rather than a final investment good. The resources used to create working data are embedded in firms' production costs, so data investment does not appear as a separate component of final expenditure such as C_t or I_t. Consistent with this treatment, working data do not enter the household budget constraint directly but instead affect households through factor payments and equilibrium prices. This reduced-form approach allows us to capture the productive role of data without imposing a specific national accounting classification.

where $E > 0$ 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—and in the open economy, the same pool of raw records can be processed into separate datasets for domestic and foreign producers at once without depleting the underlying source. 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. Note that the data intermediary allows us to focus on the non-rivalry of data at the international level: the intermediary produces working data for domestic usage and for exporting, which are non-perfect substitutes. Moreover, cross-border flows of data often concern processed data in reality, and it is captured by the working data in the model.

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.

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.

2.3 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 economy (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:

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

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 eq. (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 economy. 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 according to the following CES technology:

$$Y_{i,t} = \left(Y_{H,i,t}^{\frac{m-1}{m}} + Y_{F,i,t}^{\frac{m-1}{m}} \right)^{\frac{m}{m-1}}. \tag{6}$$

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.[11]

Moreover, a country may impose restrictions (e.g., tariffs) on imported goods to protect its domestic firms. Such restrictions can also be interpreted as the "iceberg costs" widely examined in international trade studies. Specifically, let $P_{F,i,t}$ denote the price of intermediate goods that foreign firms sell to the home country. Firms in the home country must pay a markup of d_P to purchase these goods.[12] This

[11] The wholesale producers are importers or assemblers who combine domestically produced and imported intermediate varieties into composite goods for downstream use, analogous to manufacturers integrating components sourced from multiple countries. Domestically produced and imported varieties of the same good generally differ, as they are manufactured in different factories under potentially different processes.

[12] Similarly, we denote d_P^* as the markup faced by foreign firms when purchasing intermediate goods produced in the

variable plays a crucial role in trade decisions, which are analyzed in detail alongside the restrictions on data flows in Section 3.3.

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:

$$Y_{H,i,t} + Y^*_{H,i,t} = A_t(\Phi^{\xi}_{i,t} k_{i,t})^{\alpha} n^{1-\alpha}_{i,t}, \tag{7}$$

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

$$\Phi_{i,t+1} = (1 - \delta_{\Phi})\Phi_{i,t} + \varphi_t. \tag{8}$$

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

$$\varphi_t = \left[\chi^{\frac{1}{\omega}} \varphi_{H,t}^{\frac{\omega-1}{\omega}} + (1-\chi)^{\frac{1}{\omega}} \varphi_{F,t}^{\frac{\omega-1}{\omega}}\right]^{\frac{\omega}{\omega-1}}, \tag{9}$$

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.

Similarly, 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 $d^*_D f_t P_{\varphi,F,t}$. Here, d^*_D is the cost multiplier of using imported working data, which captures various

home country. In the following analysis, we refer to d_P and d^*_P as the goods import barriers for the two countries.

frictions.[13] 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:

$$\varphi_{H,t} = ED_t^{\gamma} l_{H,t}^{1-\gamma} \tag{10}$$

and

$$\varphi_{H,t}^{*} = ED_t^{\gamma} (l_{H,t}^{*})^{1-\gamma}. \tag{11}$$

The corresponding functions in the foreign country can be defined similarly.[14] This modeling choice is both a simplification and empirically motivated: major data providers such as Nielsen and Circana (formerly IRI) routinely process the same pool of raw consumer data into separate datasets tailored for clients in different countries or regulatory jurisdictions, incurring additional labor costs for localization, compliance, and re-labeling—costs captured in our model by the separate labor inputs $l_{H,t}$ and $l_{H,t}^{*}$. What crosses borders is therefore working data, not raw records; and the welfare cost of data export—the additional privacy disutility governed by b in eq. (1)—is borne by domestic households at the point of raw data supply, regardless of where the processed output is ultimately used. This is

[13] Note that this cost multiplier is determined by the foreign country that exports data, so we use the superscript "*" to indicate which country sets this variable. Meanwhile, we denote the effective price paid by the foreign country to acquire data generated in the home country as $d_D P_{\varphi,H,t}^{*}$. In the following analysis, we refer to d_D and d_D^{*} as the data export barriers. This stands in contrast to goods trade barriers, which are conventionally imposed by the importing country to protect domestic producers; data barriers, by contrast, are more naturally placed on the exporting side, as countries typically welcome data inflows for their productivity-enhancing role while remaining reluctant to allow domestic data to flow abroad.

[14] 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.

why data export policy operates as a form of household welfare policy rather than purely as industrial policy.

In the open economy, we anchor the generation of raw data to the location of the data subject. Raw data arise as a byproduct of individual consumption or usage behavior, and the economically relevant location of data generation is therefore defined by the residence of the individual whose actions generate the data, rather than by the location of firms, servers, or digital infrastructure. For example, when a consumer accesses a foreign-provided digital service, the consumption activity occurs locally and the associated behavioral data are generated in the consumer's country, even if the data are subsequently processed or stored abroad. This assumption is a modeling abstraction that allows us to separate the origin of data generation from the cross-border use of working data.

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

- 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 eq. (2), and data accumulation follows eq. (8). 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 $B_{H,t} = 0$; $B^*_{F,t} = 0$, $B^*_{H,t} = B_{F,t}$, 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 $(C^*_t)^{-\sigma}/C_t^{-\sigma} = P^*_t$.

2.4 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 eq. (9) becomes $\varphi_t = \varphi_{H,t}$. The data intermediary's optimization in (B.5) becomes:

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

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 compare the welfare levels in the different models and extend the models to allow a data divide between two asymmetric countries. Finally, we offer initial discussions on optimal goods trade and data flow policies. We also analyze steady states of key variables under different forms of data combination in Appendix C. Throughout the analysis, we follow the literature to build a parsimonious theory to provide an initial investigation of the data economy in the international context (e.g., Clarida et al., 2002; Galí and Monacelli, 2005). Future studies when data 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).[15] 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.[16] 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 "iceberg cost" of imported intermediate goods d_P and d_P^*, the cost multipliers of imported working data d_D and d_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.

The parameter ξ, representing the importance of data, is central to our analysis. Although its precise value remains a topic of debate, initial estimates suggest that the revenue share attributable to data ranges from 10% to 40%.[17] In our model, this implies that the income share of data, calculated as $\alpha\xi/(1+\alpha\xi)$, falls within this range. Given that α typically takes the standard value of $1/3$, ξ is approximately between 0.3 and 2. Based on this, we use the foreign country's ξ (denoted as ξ^*) as a benchmark and explore how equilibrium outcomes change as ξ varies from 0 to nearly 2. This

[15] The values of ρ and m do not affect the results significantly, given that they are much larger than 1, which makes the substitution highly elastic. In Appendix D.4, we discuss the steady states with alternative parameter values for illustration.

[16] Some studies suggest that data depreciate much faster than physical capital (Farboodi and Veldkamp, 2026). To illustrate this case, Appendix D.1 reports results for a higher data depreciation rate, $\delta_\Phi = 0.3$, and shows that the main qualitative conclusions remain essentially unchanged; results under a higher physical capital depreciation rate, $\delta_k = 0.05$, are also robust and available from the authors upon request.

[17] See the report *Data Monetization in Capital Markets* by the Arab Federation of Capital Markets, and Tang (2021). While national accounts-based measures typically imply lower income shares for data, our objective is not to reconcile alternative accounting perspectives, but to study how equilibrium outcomes vary with the importance of data.

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	Christensen and Dib (2008)
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_π, ρ_f	Persistence of exogenous shocks	0.95	Alessandria et al. (2013)
δ_Φ	Data depreciation rate	0.025	Benchmark
χ	Share of domestic data in the data composite	0.5	Normalization
E	Efficiency term in working data generation	1	Normalization
γ	Contribution of raw data in working data generation	0.5	Normalization
b	Additional disutility caused by cross-border data flows	1	Normalization

approach allows us to capture a broad spectrum of policy environments in different countries. It is important to note that, due to limited data on the data economy, we select a few reasonable parameter values and check their robustness. Our primary goal here is to illustrate the main predictions of the model rather than to perform a full-scale calibration.

3.2 Welfare Analysis: the Role of Data and Data Flows

We examine the welfare in each country (the steady state of total utility) and its link to ξ in the closed economy introduced in Section 2.2, 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. We also provide further robustness checks by differing the efficiency term of working data generation in Appendix D.2. 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 1. We also fix Π, d_P, and d_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 $d_P^* = d_D^* = 1$, for illustration.

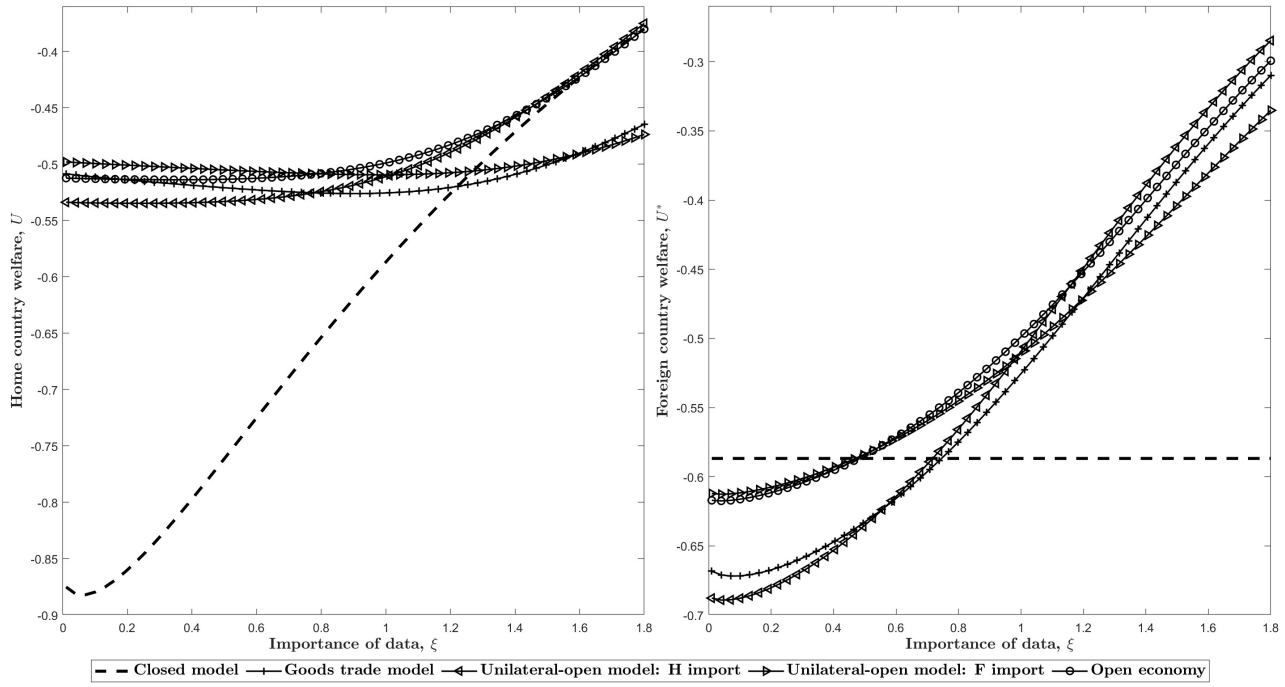


Figure 1: Welfare Levels with Different Importance of Data (ξ) in the Two Countries

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 economy, 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 economy 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-light country (i.e., with low data importance) is willing to trade with a data-intensive country (with high data importance), with or without data flows. In contrast, trade liberalization is not always desirable for the data-intensive country, especially when the divergence between ξ and ξ^* is large—a large pre-existing data divide. We show in the upper panels of Figure 2 the welfare improvements due to trade for regions of ξ where both countries are willing to trade, measured by the consumption equivalent variation (CEV): the permanent percentage increase in baseline consumption that would make households in the baseline economy equally well off as in the comparison economy, which is

denoted as λ. Formally, CEV solves

$$(1+\lambda)^{1-\sigma}\frac{C_b^{1-\sigma}}{1-\sigma}-\Omega\frac{N_b^{1+\eta}}{1+\eta}-\Pi(1+b)D_b^2=U_o,$$

where subscript "b" denotes the baseline economy and U_o denotes welfare in the comparison economy, yielding

$$\lambda=\left[\frac{U_o+\Omega\frac{N_b^{1+\eta}}{1+\eta}+\Pi(1+b)D_b^2}{\frac{C_b^{1-\sigma}}{1-\sigma}}\right]^{\frac{1}{1-\sigma}}-1.$$

The empty regions correspond to where trade breaks down. In the upper two panels of the figure, the baseline economy is the closed economy, while in the lower two panels, it is the goods trade model. We also show in Table 2 some selected values of ξ to present the numerical results in a more explicit way. 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-intensive country exports more goods but imports less from its trading partner, reducing its welfare gain from trade. This phenomenon can also be explained from the perspective of the real exchange rate. Since the model assumes the two countries do not use different currencies, their nominal exchange rate is set to 1. Higher prices in the low-ξ country and lower prices in the high-ξ country cause a real depreciation in the high-ξ country. As a result, the real purchasing power of consumers in the high-ξ country decreases, leading to welfare losses in this environment.

We should also notice that domestic and foreign goods do not perfectly substitute in the open economy, though we set the elasticity of substitution ρ to be very large. In the closed economy, households only demand and consume domestic goods. In contrast, in open economies, households must consume at least a small fraction of foreign goods to achieve a higher level of welfare. While the composition of wholesale goods changes as we move away from the closed economy, this adjustment

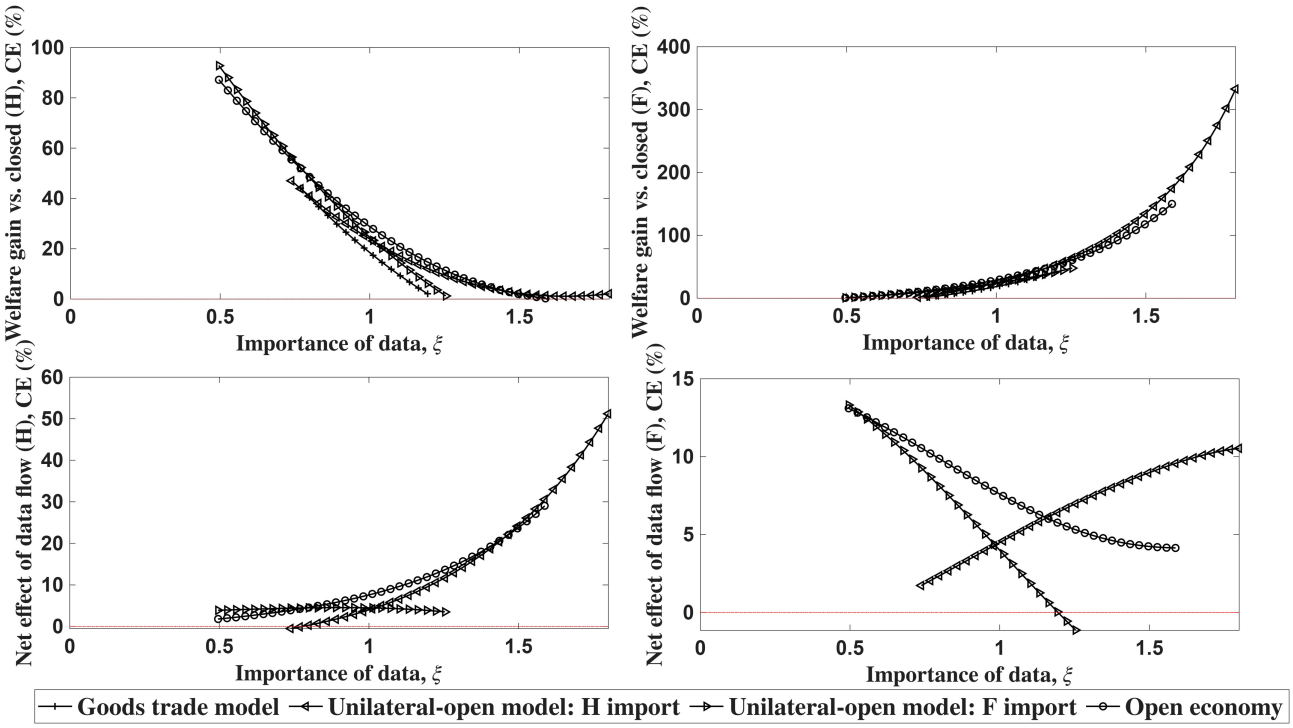


Figure 2: Welfare Improvements from Trade and Data Flows under Different Importance of Data (ξ)

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. Welfare improvements are measured by the consumption equivalent variation (CEV), reported in percentage terms. 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).

is necessary to effectively integrate models with varying degrees of openness. This is a friction arising from trade liberalization, and the introduction of data as a productivity-augmenting factor amplifies it and makes the results more complex.[18] 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-intensive 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.

[18] In fact, if we assume perfect substitution between intermediate goods from the two countries (i.e., a simple sum of goods rather than a CES aggregation), households in both countries consume only goods produced in the country with the higher ξ, as these goods are cheaper, making the problem trivial for further discussion. Thus, we apply imperfect substitution with a relatively large elasticity instead, as shown in eq. (6), to capture the more realistic scenario. For instance, semiconductor chips manufactured by firms in different countries—such as logic processors, memory modules, and AI accelerators—serve similar functions but differ in architecture, process technology, power efficiency, and software compatibility, making them imperfect substitutes for downstream producers even within the same product category.

Table 2: CEV Welfare Gains (%) at Selected Values of ξ

	$\xi = 0.3$	$\xi = 0.5$	$\xi = 0.8$	$\xi = 1.0$	$\xi = 1.2$	$\xi = 1.5$
Panel A: Welfare gain relative to closed economy (%)						
Home country (H)						
Goods trade model	–	–	40.26	18.48	–	–
Unilateral-open economy: H import	–	–	40.82	23.88	11.81	2.32
Unilateral-open economy: F import	–	92.13	48.19	24.55	5.62	–
Open economy	–	86.59	48.42	28.80	14.33	1.93
Foreign country (F)						
Goods trade model	–	–	3.15	18.48	–	–
Unilateral-open economy: H import	–	–	5.71	24.56	53.88	135.46
Unilateral-open economy: F import	–	0.80	11.87	23.88	41.77	–
Open economy	–	0.63	13.78	28.81	52.42	118.55
Panel B: Net effect of data flows relative to goods trade model (%)						
Home country (H)						
Unilateral-open economy: H import	–	–	0.32	4.00	9.73	24.34
Unilateral-open economy: F import	–	3.90	4.48	4.49	3.85	–
Open economy	–	1.82	4.61	7.60	12.10	23.81
Foreign country (F)						
Unilateral-open economy: H import	–	–	2.37	4.49	6.49	8.98
Unilateral-open economy: F import	–	13.26	8.09	4.00	-0.10	–
Open economy	–	13.06	9.85	7.60	5.71	4.24

Notes. Each cell reports the consumption equivalent variation (CEV, in %), defined as the permanent percentage increase in baseline consumption that equates baseline welfare to the comparison economy's welfare. Panel A uses the closed economy as the baseline; Panel B uses the goods trade model as the baseline. The elasticity of substitution of data is $\omega = 5$ and the foreign country's data importance is fixed at $\xi^* = 1$. Cells marked "–" correspond to parameter regions where at least one country prefers autarky.

As shown in the right panel of Figure 1, 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 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. This occurs because data flows reallocate productivity rather than output: they raise the effective capital composite in the data-light country without requiring goods movement or terms-of-trade adjustment, compressing the productivity gap before trade in goods takes place. Technology diffusion cannot play the same role over the short horizons we study because it requires firm-level adoption investment and does not scale automatically with the size of the consumer base supplying raw data.

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 2, 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 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 economy 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 3 shows that as the divide shrinks, welfare improvements from goods trade decrease while those from data flows increase in the data-light country. Conversely, in the data-intensive country, welfare improvements from goods trade increase while those from data flows decrease. Furthermore, goods trade can lead to welfare loss in the data-intensive 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-light country becomes less important, which reduces the demand for labor in the data intermediary in that country. Moreover, the data-light 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-intensive country's importation of working data from the data-light country, and subsequently increases welfare in the data-intensive country. However, as data divide becomes very large, the welfare loss from goods trade becomes dominant, and the data-intensive country may refuse to trade goods or data.

Feasible intervals of international trade and data flows. We present in Table 3 intervals of ξ in the home country for different levels of ξ^* in the foreign country so that trade (and data flows) is feasible.

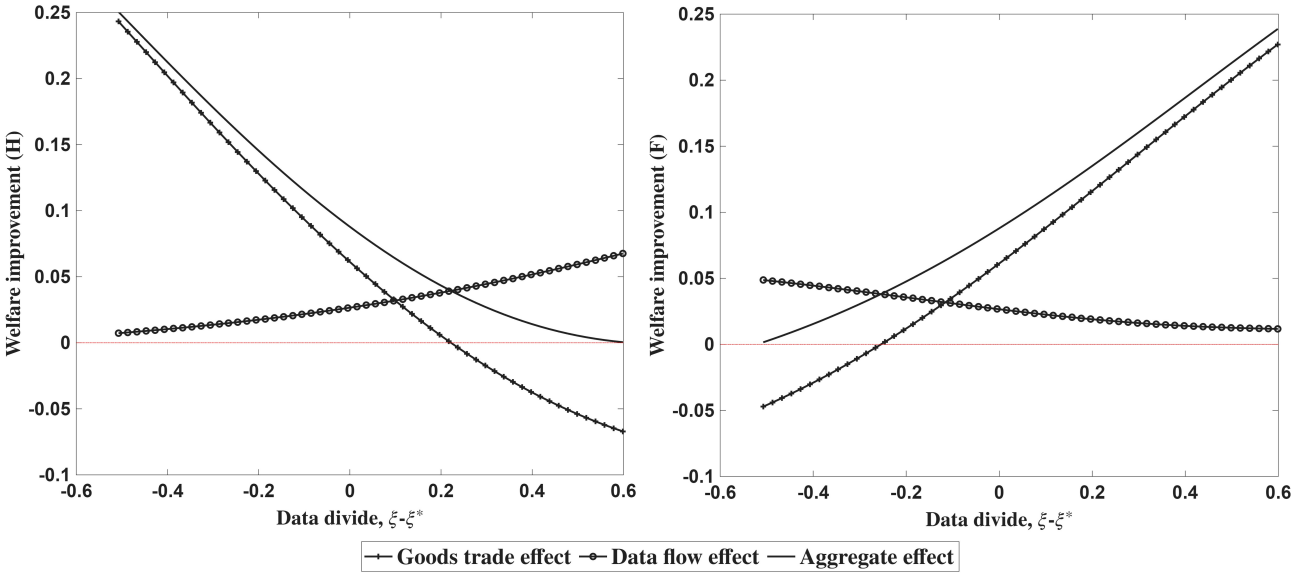


Figure 3: Decomposition of Welfare Improvements under Different Levels of Data Divide

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 $\xi^* = 1$.

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_D^*) to demonstrate the variation of the intervals. Appendix D.3 presents the welfare improvements due to trade and data flows with different levels of d_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 4 how trade and data flows alter the gaps in these outcomes with different levels of data divide. In the figure, we consider ξ^* at three levels, namely 0.5, 1.0, and 1.5, and adjust ξ continuously.

In general, we observe a negative relationship between the usage of raw data and working data. The data-light country provides more raw data for the generation of working data, while the data-intensive country uses more working data for production. The country with a larger ξ also ends up producing more final goods. The data-light 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

Table 3: 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: $d_D^* = 1$	$[0, 1.02]$	$[0.48, 1.61]$	$[0.93, 1.75]$
	Open Economy: $d_D^* = 1.5$	$[0, 0.91]$	$[0.49, 1.39]$	$[0.94, 1.75]$
	Open Economy: $d_D^* = 2$	$[0, 0.87]$	$[0.50, 1.33]$	$[0.96, 1.75]$
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: $d_D^* = 1$	$[0.13, 1.02]$	$[0.48, 1.61]$	$[0.93, 1.75]$
	Open Economy: $d_D^* = 1.5$	$[0.15, 0.91]$	$[0.49, 1.39]$	$[0.94, 1.75]$
	Open Economy: $d_D^* = 2$	$[0.17, 0.87]$	$[0.50, 1.33]$	$[0.96, 1.75]$

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 $d_D = 1$. The values of other parameters are shown in Table 1.

of data is high enough for the data-intensive country to agree to trade.

3.3 Policy Implications: Trade VS. Data Flow

The results shown above align with the prevailing trends in the global digital economy. Zheng (2021) documents and compares the distinct regulatory paradigms governing cross-border data flows among three major economies. The United States of America, with its advanced digital economy, attracts a substantial volume of global data due to its relatively permissive data flow policies until recently. In contrast, the EU, with a less developed digital economy, receives less data from other countries as a result of its stricter regulations. China falls somewhere in between, possessing a relatively more developed digital economy and supportive policies that enable it to accumulate a sizable amount of working data. In this subsection, we further examine the policy implications for trade and data flows.

Policy interactions between symmetric countries. Considering the increasingly intense competition between the U.S. and China in AI and the digital economy, it is also important to examine the

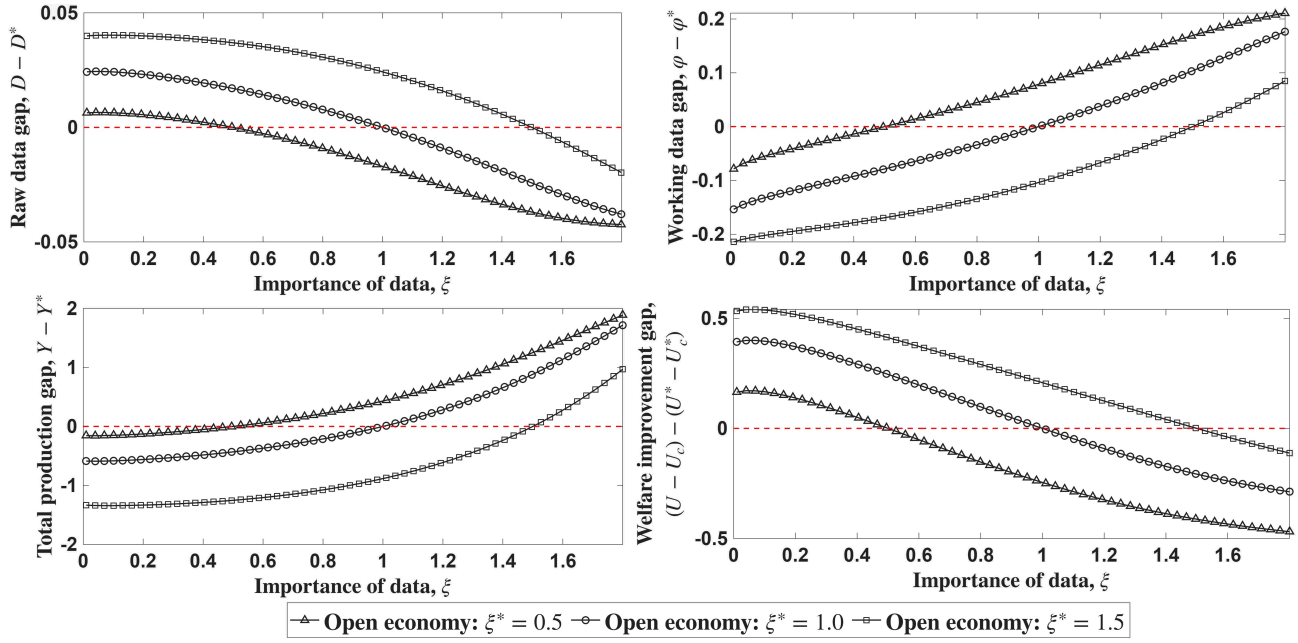


Figure 4: Gaps between Asymmetric Countries with Different Importance of Data (ξ)

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.

strategic interactions between the two countries. In Figure 5, we analyze two symmetric countries with varying levels of goods and data trade policies, illustrating how their optimal policies adjust in response to the actions of their counterparts. Specifically, we present a contour map depicting the optimal goods and data trade policies of the home country that maximize its welfare, given different combinations of policies implemented in the foreign country. Furthermore, in Figure 6, we illustrate the welfare changes relative to the scenario with no trade barriers when a country adopts different trade policies, with potential policy implications.

From Figure 5, we observe that the optimal goods import barrier d_P should always be minimized ($d_P = 1$), regardless of the foreign country's policy choices. However, this does not imply that the foreign country can set its barrier arbitrarily. Figure 6 shows that a country incurs welfare losses if it imposes high barriers, while in some scenarios, its counterpart may even benefit. This finding aligns with the prevalent view that trade barriers are generally undesirable for both countries.[19]

Meanwhile, further insights emerge when considering the optimal data trade policy. Figure 5(a)

[19]One might argue that a country could choose to raise trade barriers to protect its domestic firms and promote production. This would indeed be the case if production were explicitly considered in the objective functions. In this case, we find that the optimal decisions on goods trade policy increase to a high level but remain unchanged when varying the foreign country's policies, similar to the case presented here. Thus, we focus on the results in which governments in both countries aim solely to maximize consumer welfare within their respective economies.

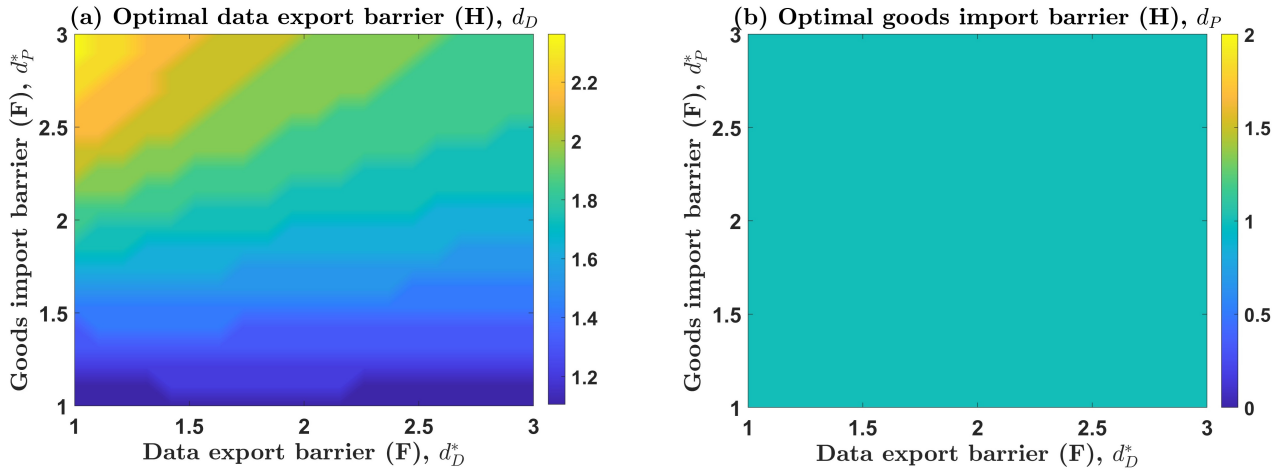


Figure 5: Optimal Policies of the Home Country

Notes. These figures illustrate the optimal policies (welfare maximization) of the home country under different combinations of trade policies implemented by the foreign country. The light and yellow areas indicate scenarios where a high-level barrier is chosen, while the dark and blue areas correspond to a low-level barrier. Subfigure (b) demonstrates that the optimal goods import barrier remains constant at $d_P = 1$, regardless of changes in the foreign country's policies.

shows that when the home country faces a high goods import barrier from the foreign country, it chooses a higher data export barrier than its counterpart when data barriers in the foreign country are relatively low. However, when the foreign country already imposes a high data barrier, the home country cannot sustain an equally stringent data export policy and instead concedes by reducing its own data export barrier. Conversely, when the goods import barrier in the foreign country is low, it is optimal for the home country to maintain a low data export barrier, regardless of the data trade policy chosen by its counterpart. This result can also be seen from Figure 6(c) and (d). When both the home and foreign countries set relatively low goods trade barriers ($d_P = d_P^* = 1$), welfare maximization requires that at least one country maintain a low data barrier. Further, Figure 6 indicates that the impact of data trade policy is relatively insignificant compared to that of goods trade policy. The choice of goods trade policy primarily determines economic outcomes, while the total sum of data trade policies follows them. This finding aligns with previous conclusions that countries can still achieve welfare improvements through unilateral data flows, even when they only export data. The underlying asymmetry is that importing working data raises domestic productivity without imposing privacy costs on domestic households, whereas exporting requires domestic households to supply additional raw records and bear disutility even when the productivity gain accrues abroad. This household-level cost—absent in conventional goods trade or technology licensing—explains why optimal data export barriers are positive across a range of parameter configurations and why a

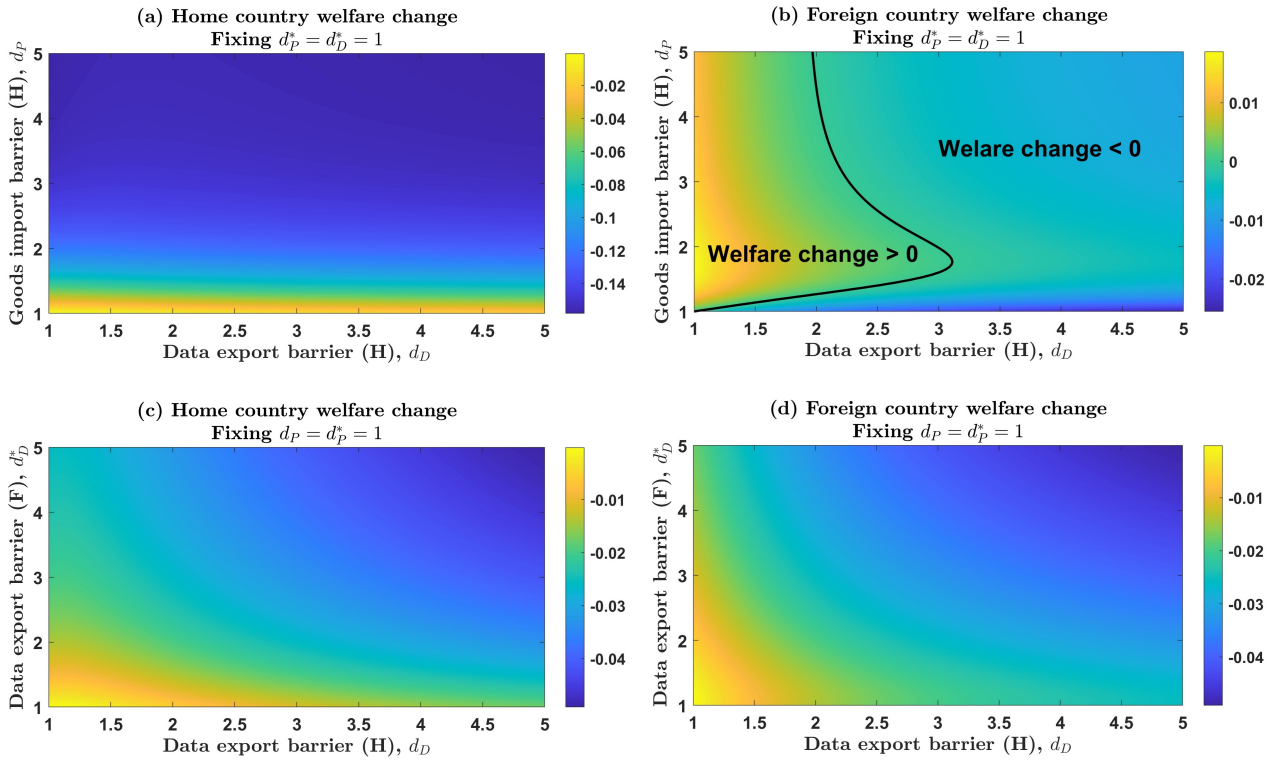


Figure 6: Welfare Changes With Different Combinations of Policies

Notes. These figures illustrate the welfare changes in the home and foreign countries relative to the corresponding cases without trade barriers. The light and yellow areas indicate scenarios with higher welfare levels, while the dark and blue areas correspond to lower welfare levels. In subfigure (b), a solid line represents the frontier between positive and negative welfare changes. The upper two subfigures depict the interactions between goods and data policies, while the lower two subfigures illustrate the interactions of data export barriers between the two countries.

country's willingness to export data is more sensitive to its partner's policies than its willingness to import.

In our framework, banning data flows prevents a country from achieving welfare improvements by limiting its access to a larger pool of data. Although our results indicate that partially open economies experience greater welfare gains when a country imports data rather than exports it, such decisions are typically reciprocal: once a country's data inflows are restricted by another country, it is likely to impose restrictions on its data outflows as well. Geopolitical tensions often trigger changes in data policies, which, in turn, mitigate their effects on promoting trade. Moreover, such policy shifts often extend beyond the scope of the welfare analysis presented in this paper and further analyses are warranted.

The long-term perspective. Finally, although data-light countries experience greater welfare improvements in our framework—often referred to as the latecomer's advantage—this does not nec-

essarily translate into sustained growth. Our analysis focuses on comparative statics across steady states, whereas in the long run, as data-intensive countries accumulate more working data, this will inevitably fuel their economic growth. Meanwhile, if data-light countries aim to achieve sustained growth, they must invest more in improving their efficiency in utilizing data. However, this is often not a feasible policy for most underdeveloped countries in reality.

Further discussions. From the perspective of U.S.–China competition in the digital economy, the analysis above should be understood as offering a stylized interpretation of strategic policy interactions, rather than a direct mapping to observed regulatory arrangements. The model highlights that goods trade barriers play a dominant role in shaping welfare outcomes, while data trade policies tend to adjust in response and often reinforce existing frictions, especially when bilateral tensions intensify. At the same time, differences in data intensity and productivity imply that similar policy choices can generate very different consequences across countries. As a result, the framework does not aim to explain specific policy episodes one-to-one, but to clarify how underlying structural conditions and relative positions in the global data economy jointly shape incentives, constraints, and long-run trade-offs in policy competition.

4. Transition Dynamics

In our framework, the adjustment of working data differs from that of conventional inputs, because the effects of data depend jointly on household behavior, policy frictions, and cross-border access. As a result, exogenous shocks to productivity, privacy costs, or data policies can generate transition dynamics in data stocks and data flows that are qualitatively distinct from those of traditional goods and capital. Studying these transition paths is therefore informative not only for understanding short-run fluctuations, but also for clarifying how data openness affects volatility, persistence, and cross-country transmission of economic shocks. To compute these transition dynamics, we log-linearize the equilibrium conditions around the relevant steady state and obtain impulse responses using a standard perturbation-based solution method. Specifically, we consider productivity shocks A_t (shown in eq. (7)), disutility shocks to the usage of raw data π_t (shown in (eq. 1)), and cost shocks to importing data for production f_t (shown in eq. B.4) in the Appendix). 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 three 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. Finally, we study the transition path associated with a regime switch from the steady state of the closed economy to that of the open economy, characterizing the adjustment path. 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 11 data-related variables after the shocks. In each of the analyses, we provide the results of six models: the closed economy, 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). For completeness, we show the changes of other important variables in Appendix E.1.

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.[20] 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.

The variations of data cyclicality observed in the closed economy and in the open economies are caused by different sources of the data and goods traded. In the closed economy and the goods trade model, data only come from domestic households and are crowded out by other factors, such

[20] In the closed economy, we only have impulse response functions regarding the following variables: φ, D, and Φ. In the following figures, we also present the results in a similar way.

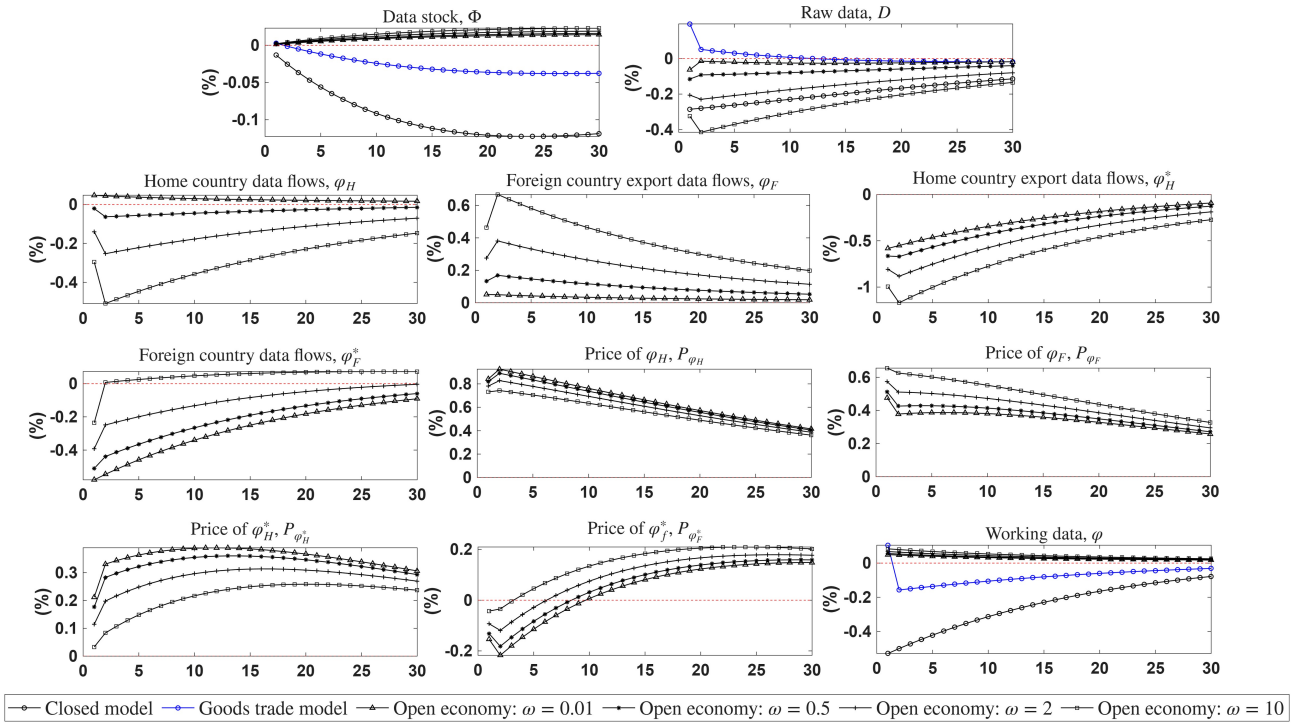


Figure 7: Impulse Responses of the Positive Productivity Shock A_t in Symmetric Countries

Notes. This figure illustrates the impulse responses of 11 data-related 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 economy, 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.

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.[21] Notably, the goods trade model exhibits a smoother variation than that in the closed economy due to the buffering effect of goods trade on data flow variations during shocks. These differences in data cyclicality highlight why different policies may be needed for cross-border data flows to mitigate aggregate shocks.

We gain further insights by distinguishing wholesale goods, denoted as Y, and consumption, referred to as C, within the contexts of closed and open economies, which are shown in Appendix E.1. In the absence of international trade, the total production and consumption levels exceed those

[21] 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-cyclicality 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.

observed when trade becomes open. The rationale behind this phenomenon lies in 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. 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."

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 eq. (B.4). In addition, we present the transitions in the closed economy and the goods trade model using the same notation π_t to represent the disutility shock in the corresponding utility function (1).

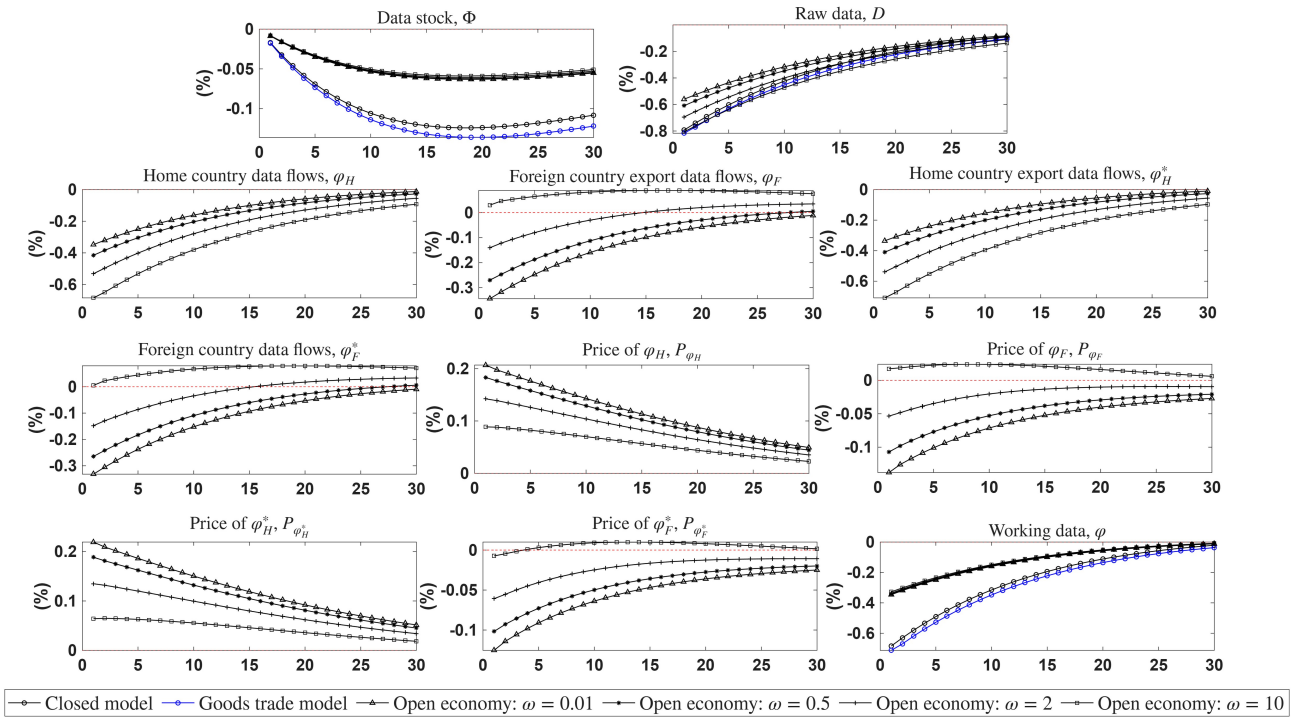


Figure 8: Impulse Responses of the Positive Disutility Shock to Raw Data π_t in Symmetric Countries

Notes. This figure shows the impulse responses of 11 data-related 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 economy, 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 Figure 8, it is evident that the closed economy and goods trade model exhibit distinct behaviors compared to the open economies across most variables. The two models without data flows exhibit

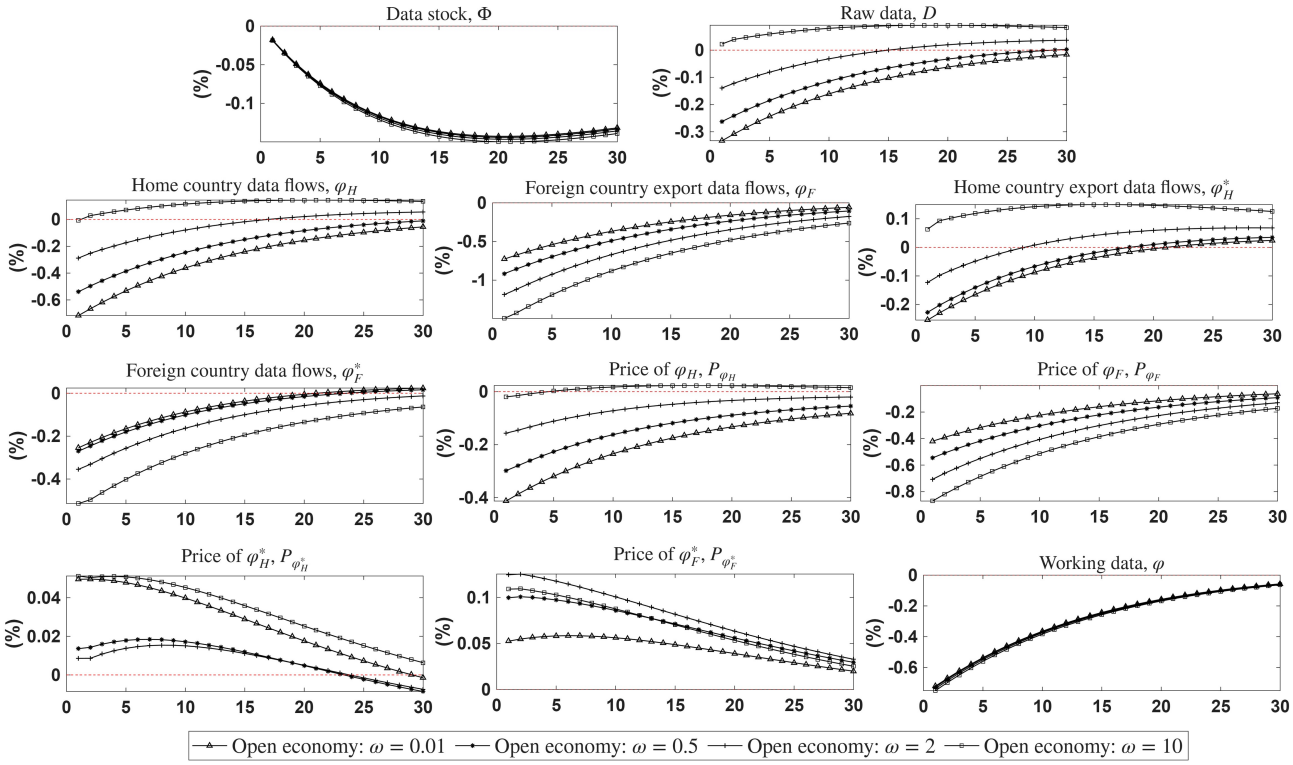


Figure 9: Impulse Responses of the Positive Cost Shock to Imported Data Flows f_t in Symmetric Countries

Notes. This figure depicts the impulse responses of 11 data-related 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 $\omega = 10$).

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 Figure 9, we observe that the shock to the price of imported data flows $P_{\varphi,F}$ results 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 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, policymakers should also carefully consider the fluctuation from these two types of shocks. The former shock usually relates to changes in privacy concerns and data regulation, whereas the latter shock 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 hinder the data economy development in the long run.

Summary. Taken together, the transition dynamics highlight that cross-border data flows respond systematically to macroeconomic and regulatory shocks, with patterns that depend on data intensity and policy frictions. Although comprehensive empirical evidence on these dynamics is still limited, recent work on international bandwidth usage and digital services trade suggests that data flows are highly sensitive to both economic conditions and regulatory changes (Ferencz, 2019; Sun and Trefler, 2023). Our results provide a set of qualitative benchmarks that can guide future empirical work in identifying key structural parameters, such as the elasticity of substitution between domestic and foreign data, the depreciation rate of data stocks, and the magnitude of data trade frictions.

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 E.2.1.

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 (see Appendix E.1). 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-intensive 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.

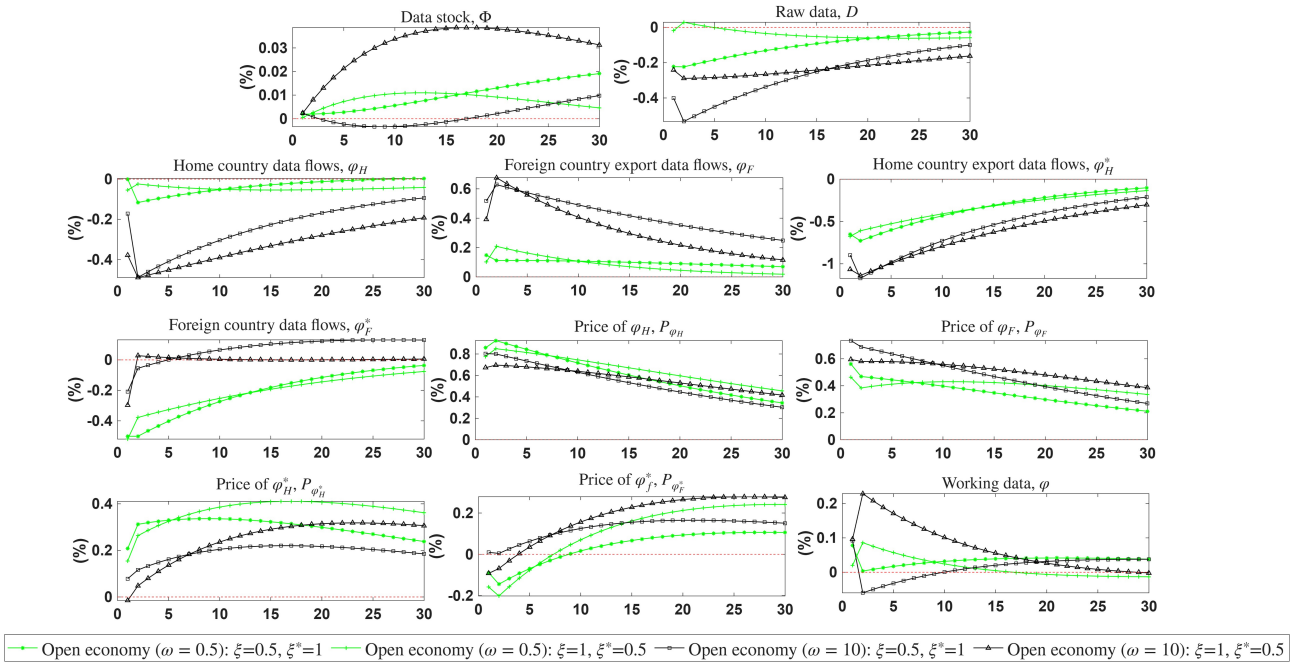


Figure 10: Impulse Responses of the Positive Productivity Shock A_t in Asymmetric Countries

Notes. This figure illustrates the impulse responses of 11 data-related 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 (only in the home country) 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 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-light country in a disadvantaged position, even if that country experiences a positive productivity shock. As a result, the data-intensive country experiences a production expansion and absorbs more working data from the data-light country, leading to a counter-cyclical pattern of data usage in the latter country. We further support this argument in Appendix E.2.2, where we demonstrate that as the data divide between the two countries diminishes, working data become pro-cyclical when either country 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-intensive 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-intensive country, it becomes less willing to trade with the other country and the gap in production widens even further.

4.3 Welfare Dynamics of Trade Liberalization

We now consider a complementary and policy-relevant question: what are the welfare consequences of the regime change itself? Specifically, we compute the transition path when an economy permanently switches from the closed economy to the open economy at $t = 1$, and track how cumulative welfare evolves relative to the counterfactual of remaining closed. This exercise connects the steady-state welfare comparisons in Section 3 with the dynamic adjustment process studied above, and adds a temporal dimension to welfare gains from trade liberalization and cross-border data flows.

Setup. At $t = 0$ and prior, both countries reside in the closed economy steady state. At $t = 1$, the regime permanently switches to the open economy with bilateral goods trade and data flows. The predetermined state variables—capital K_t and working data stock Φ_t—are inherited from the closed economy steady state, while all jump variables (consumption, output, prices, data flows, etc.) adjust immediately to satisfy the open economy equilibrium conditions. The economy then converges to the open economy steady state. We compute the cumulative welfare difference as

$$\Delta W_t = \sum_{\tau=1}^{t} \beta^{\tau-1} \left(U_\tau^{\text{trans}} - U_{\text{ss}}^{\text{closed}} \right),$$

which measures the cumulative discounted welfare gain (or loss) from opening the economy up to period t, relative to the counterfactual of remaining in the closed steady state forever. We fix $\xi^* = 1$ for

the foreign country and vary the home country's data importance $\xi \in \{0.5, 0.8, 1.0, 1.3\}$ to examine how the pre-existing data divide affects the welfare dynamics of liberalization. The results are shown in Figure 11.

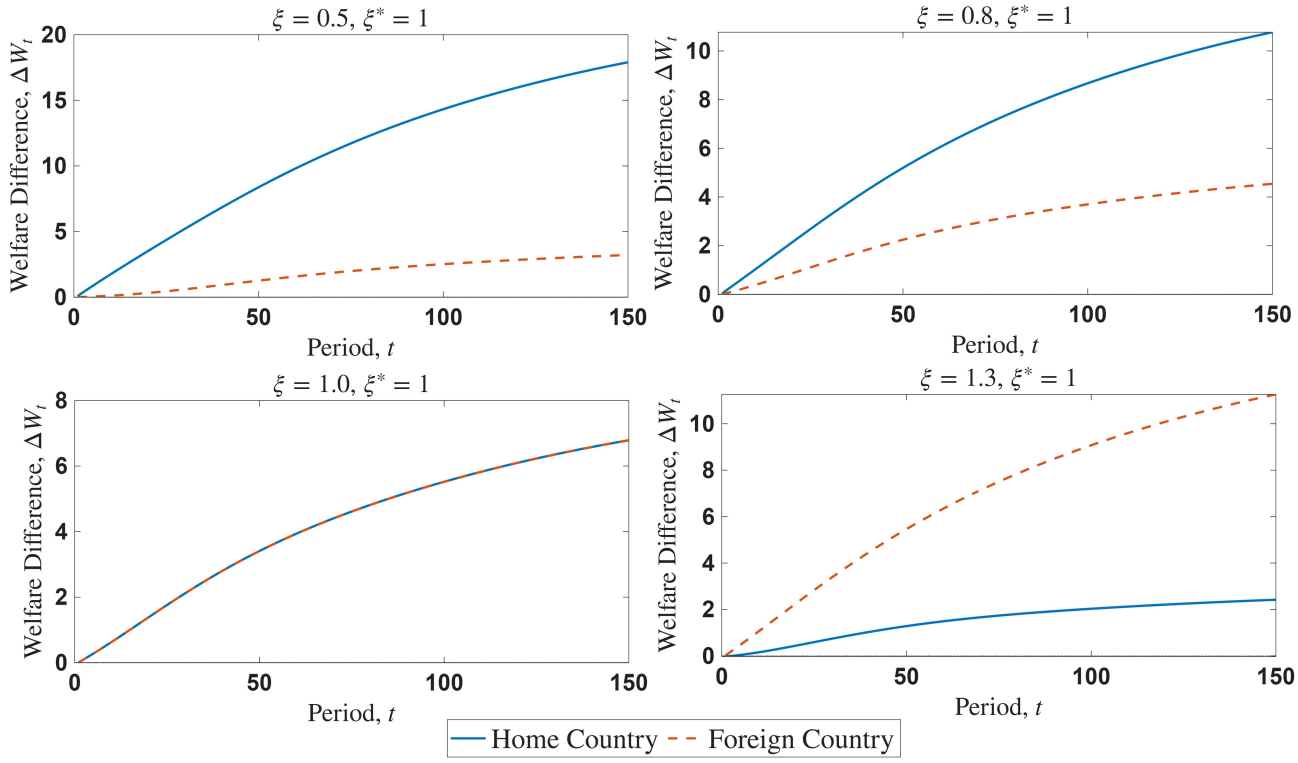


Figure 11: Cumulative Welfare Difference from Trade Liberalization

Notes. This figure depicts the cumulative discounted welfare difference ΔW_t between the transition path (from the closed economy to the open economy) and the counterfactual of remaining in the closed economy steady state, for both the home and foreign countries. The importance of data in the foreign country is fixed at $\xi^* = 1$. The four panels correspond to $\xi = 0.5$, $\xi = 0.8$, $\xi = 1.0$, and $\xi = 1.3$ for the home country.

Symmetric case. When $\xi = \xi^* = 1$, both countries gain from liberalization, and the cumulative welfare differences are virtually identical, as expected from the symmetry of the model. The welfare gains accrue gradually and monotonically, reflecting the slow adjustment of data stocks and capital toward their higher open economy steady-state values (see Appendix E.4.2 for the full transition paths). During the transition, both consumption and total production overshoot their open economy steady-state levels before settling, driven by the initial surge in data flows and working data accumulation when the economy opens.

Asymmetric cases and the latecomer's advantage. The most striking pattern emerges when $\xi \neq \xi^*$. When the home country is data-light ($\xi < \xi^*$), it enjoys a larger cumulative welfare gain than the foreign country. For instance, when $\xi = 0.5$, the home country's welfare gain at $t = 150$ is roughly

four times that of the foreign country. This is consistent with the steady-state "latecomer's advantage" documented in Section 3: a data-light country benefits disproportionately from trade liberalization because it gains access to working data generated from the data-intensive country's richer data environment. The transition dynamics (Appendix E.4.1) confirm the mechanism: upon opening, both consumption C_t and total production Y_t in the home country rise and converge toward the open economy steady state, while the working data stock Φ_t initially overshoots before settling, reflecting the rapid absorption of newly available cross-border data.

Conversely, when the home country is data-intensive (i.e., $\xi = 1.3 > \xi^* = 1$), the pattern reverses: the foreign (data-light) country captures a larger share of the welfare gains. Notably, the data-intensive country's welfare gain is substantially smaller but remains positive, suggesting that even the more advanced digital economy benefits from trade liberalization—though its gains stem primarily from access to a larger goods market rather than from data flows per se. This complements the steady-state finding that bilateral data flows can improve welfare for both countries even in the presence of a pre-existing data divide, provided the divide is not too large.

Transition costs and the role of the data divide. A key dynamic feature is visible in the $\xi = 0.8$ and $\xi = 1.3$ panels: in the early periods of liberalization, the foreign country's cumulative welfare difference briefly dips slightly below zero or stays near zero before eventually turning positive. This reflects a short-run adjustment cost: upon opening, the reallocation of data flows and labor across sectors can temporarily depress welfare in one country before the long-run gains from specialization and data sharing materialize. This phenomenon mirrors the well-known result in trade theory that liberalization can generate short-run losers even when it improves long-run welfare—here, the mechanism operates through the data channel rather than through traditional factor reallocation.

Discussion. These findings have several implications. First, the temporal profile of welfare gains reinforces the policy insight from Section 3 that data-light countries have strong incentives to embrace trade liberalization and cross-border data flows. Second, the transitory adjustment costs for the more data-intensive country suggest that phased liberalization or compensatory transfers could help facilitate the transition. Third, the monotonic and substantial welfare gains in the symmetric case indicate that, absent significant pre-existing data divides, simultaneous liberalization is unambiguously welfare-improving for both countries from the first period onward. These results complement and dynamically validate the steady-state analyses, and highlight that the welfare consequences of data-flow policies are not merely long-run phenomena but unfold meaningfully along the transition

path.

5. Conclusion

We develop a tractable general equilibrium framework that jointly analyzes 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-intensive country, while raw data primarily come from the data-light country. We investigate how policies concerning trade and data flows interact, offering insights for future policy-making. 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.

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 (2024). The Changing Economics of Knowledge Production. *Review of Financial Studies 37*(1), 89–118.

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.

Bontadini, F., C. Corrado, M. Iommi, and C. Jona-Lasinio (2022). Data, Intangible Capital, and Productivity: Literature review, Theoretical framework, and Empirical evidence on the UK. Working Paper.

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.

Corrado, C., J. Haskel, M. Iommi, C. Jona-Lasinio, and F. Bontadini (2025). Data, Intangible Capital, and Productivity. In S. Basu, L. Eldridge, J. Haltiwanger, and E. Strassner (Eds.), *Technology, Productivity, and Economic Growth*, pp. 193–243. University of Chicago Press.

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.

Ding, X. (2023). Industry Linkages from Joint Production. STEG Working Paper WP056, Strutural Transformation and Economic Growth.

Eeckhout, J. and L. Veldkamp (2024). Data and Markups: A Macro Information Frictions Perspective. NBER Working Papers 30022, National Bureau of Economic Research.

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 (2024). Valuing Financial Data. *Review of Financial Studies 38*(3), 938–980.

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 (2026). A Model of the Data Economy. *Review of Economic Studies, forthcoming*.

Ferencz, J. (2019). The OECD Digital Services Trade Restrictiveness Index. OECD Trade Policy Papers 221, OECD Publishing.

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 (2025). The Limits to Growth in the AI-Driven Economy. *China Economic Review 94*, 102510.

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 (2024). Neoclassical Growth in an Interdependent World. NBER Working Papers 31951, National Bureau of Economic Research.

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. NBER Working Papers 31250, National Bureau of Economic Research.

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

Ridder, M. D. (2024). Market Power and Innovation in the Intangible Economy. *American Economic Review 114*(1), 199–251.

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

Sun, R. and D. Trefler (2023). The Impact of AI and Cross-Border Data Regulation on International Trade in Digital Services: A Large Language Model. NBER Working Papers 31925, National Bureau of Economic Research.

Sun, T., Z. Yuan, C. Li, K. Zhang, and J. Xu (2024). The Value of Personal Data in Internet Commerce: A High-stake Field Experiment on Data Regulation Policy. *Management Science 70*(4), 2645–2660.

Tang, C. (2021). *Data Capital: How Data Is Reinventing Capital for Globalization*. Switzerland: Springer.

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 27*(5), 1545–1562.

Veldkamp, L. and C. Chung (2024). Data and the Aggregate Economy. *Journal of Economic Literature 62*(2), 458–484.

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. The Data Lifecycle

This appendix traces the lifecycle of data as it evolves through the economy, providing an interpretive map for the modeling choices and results in the main text. We show the data lifecycle in Figure A.1.



Figure A.1: Data lifecycle in the model

Notes. The diagram traces the four-stage lifecycle of data in the model. Raw data D_t arise as byproducts of household consumption and are sold at price $P_{D,t}$. The data intermediary transforms them into working data φ_t (structured information) using labor l_t. Working data are non-rival and sold simultaneously to all intermediate producers; in the open economy they can be exported. Producers accumulate firm-specific information stocks $\Phi_{i,t}$ that augment effective capital in production and depreciate at rate δ_Φ. The transition from non-rival flows (φ_t) to rival stocks ($\Phi_{i,t}$) reflects the firm-specific resources required to embed information into operational systems.

Stage (i): Generation of raw data. Raw digital records arise as byproducts of household consumption activity—for example, browsing histories, purchase records, and platform usage logs generated at the interface between households and digital services. Concretely, such raw data are produced

at the point of interaction between the household and the platforms or devices through which consumption takes place. Consistent with modern data governance frameworks such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States, we treat households as the effective owners of these records and allow them to sell raw data D_t to a data intermediary at a competitive price $P_{D,t}$. Because data sharing exposes households to privacy risks such as price discrimination, unauthorized access, or identity theft, households incur a convex disutility cost when contributing data, which is captured by the third term in the utility function (1).

Stage (ii): Processing into working data. The data intermediary acquires raw records from households and transforms them into usable working data—structured information—through costly processing activities including filtering, classification, labeling, feature extraction, and normalization. This transformation requires both raw data and labor as inputs, reflecting the real resource costs of making unorganized digital traces economically productive. The resulting working data φ_t constitute organized datasets that can be systematically interpreted and used as productive inputs by intermediate-good producers. In the open economy, the same pool of raw data can be processed into separate datasets for domestic and foreign use, capturing the non-rival nature of raw data at the intermediary stage.

Stage (iii): Use as a non-rival intermediate input. Working data are sold as a non-rival intermediate input to intermediate-good producers, who can simultaneously purchase the same processed dataset without depleting it. This non-rivalry is the defining economic property of data that distinguishes it from physical capital and labor: the same information can be used by all producers in the same period at no additional marginal cost of reproduction. In the open economy, working data can also be exported, so that foreign producers benefit from information generated by domestic households. It is this stage that gives rise to the paper's central welfare results: cross-border data flows expand the effective pool of working data available to producers in both countries, generating gains from trade that are governed by the relative importance of data in each country's production technology.

Stage (iv): Accumulation as a firm-specific information stock. Intermediate-good producers accumulate firm-specific stocks of processed information $\Phi_{i,t}$ that augment their effective capital and depreciate over time as datasets become obsolete and technologies evolve. This accumulation reflects the gradual embedding of working data into firm-level operational systems, algorithms, and organizational routines. Unlike the flow of working data φ_t—which is non-rival and purchased simultaneously by all producers—the accumulated stock $\Phi_{i,t}$ is rival across firms, because building and maintaining it requires firm-specific resources and effort. These stocks enter production as a capital-augmenting factor, as in eq. (4), and their depreciation rate δ_Φ governs the speed at which firms must refresh their information inputs, a margin that shapes both the steady-state demand for

working data and the transition dynamics analyzed in Section 4.

B. Model Solutions

B.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:

$$Y_{i,t} = P_{i,t}^{-\rho}Y_t. \tag{B.1}$$

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

$$\left(\int_0^1 P_{i,t}^{1-\rho}\mathrm{d}i\right)^{\frac{1}{1-\rho}} = 1. \tag{B.2}$$

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 eq. (3), (4), and (B.1), 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:

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

and

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

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

B.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 eq. (B.1) and (B.2). Meanwhile, the profit maximization problem for the wholesale producers maximizes $Y_{i,t} - P_{H,i,t}Y_{H,i,t} - d_P P_{F,i,t} Y_{F,i,t}$ by choosing $Y_{H,i,t}$ and $Y_{F,i,t}$. We can derive the demands for these two types of intermediate goods as:

$$Y_{H,i,t} = P_{H,i,t}^{-m} Y_{i,t}, \quad \text{and} \quad Y_{F,i,t} = (d_P P_{F,i,t})^{-m} Y_{i,t}. \tag{B.3}$$

Here, $P_{H,i,t}$ and $d_P P_{F,i,t}$ are the prices of domestic and imported intermediate goods, respectively. The price for the wholesale goods satisfies $\left[P_{H,i,t}^{1-m} + (d_P P_{F,i,t})^{1-m} \right]^{\frac{1}{1-m}} = 1$, 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:

$$\min_{\varphi_{H,t}, \varphi_{F,t}} P_{\varphi,H,t}\varphi_{H,t} + d_D^* f_t P_{\varphi,F,t}\varphi_{F,t}, \tag{B.4}$$

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.[22] Then, we derive the demands from the two sources of data as follows:

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

where the price index of the data composite is defined as

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

With $Q_{t,t+k}$ denoting the discount factor similarly as that in the closed economy, intermediate 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 eq. (7), (8), (B.3), 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:

$$\max_{P_{i,\varphi,H,t}, P_{i,\varphi,H,t}^*, D_t, l_{H,t}, l_{H,t}^*} \int_0^1 P_{i,\varphi,H,t} \varphi_{H,t} \mathrm{d}i + \int_0^1 P_{i,\varphi,H,t}^* \varphi_{H,t}^* \mathrm{d}i - P_{D,t} D_t - w_t (l_{H,t} + l_{H,t}^*), \tag{B.5}$$

subject to eq. (10), (11), and the zero-profit condition:

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

[22] 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.

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.

C. Steady-States of Key Variables

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

From the figures, we see that the steady states of the variables do not change as ω increases when $\Pi = \Pi^* = 2$ or $d_D = d^*_D = 1$ 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^*_D in the other models. In Figure C.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

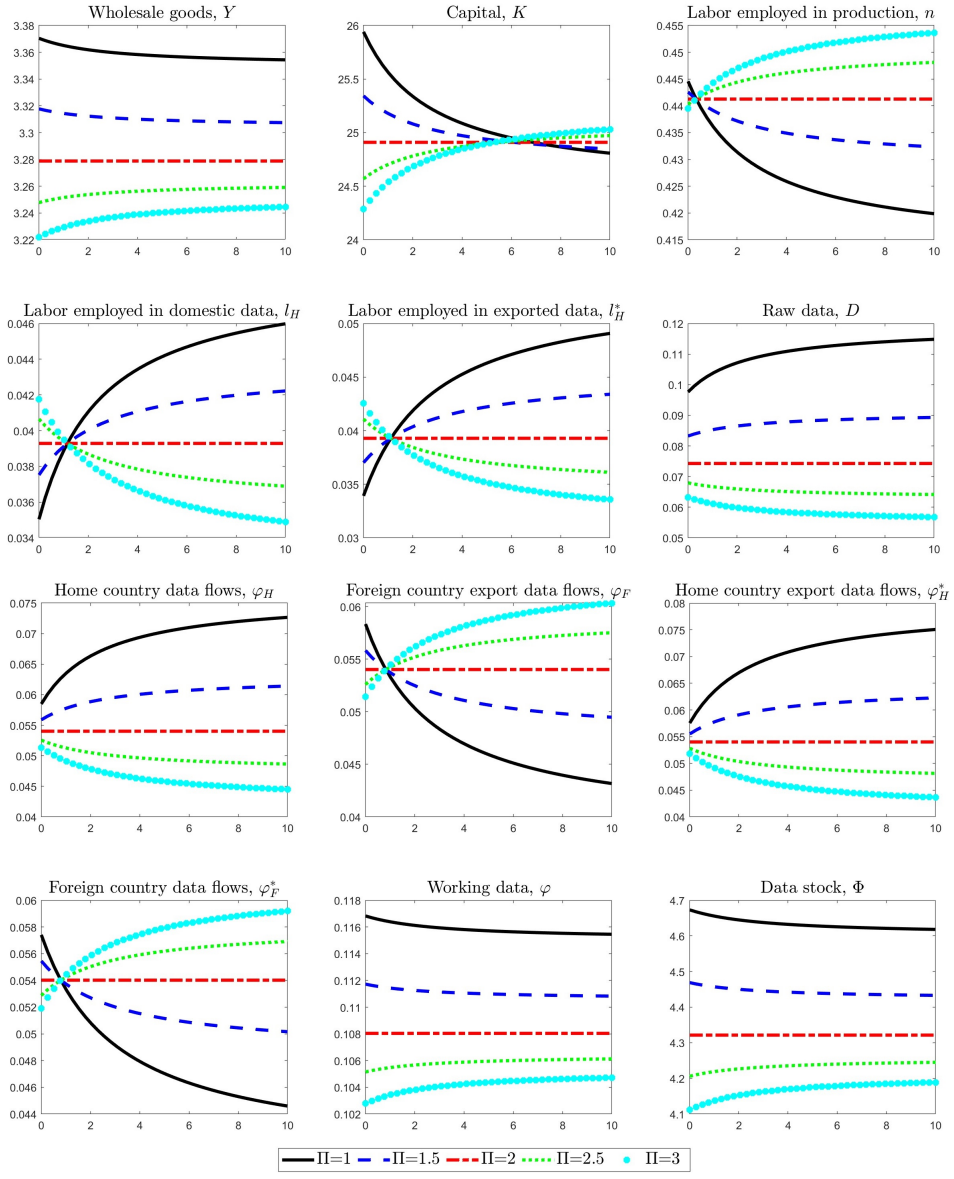


Figure C.1: Steady-States of the Main Variables with Different Values of Disutility Parameter Π

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: Π = 1 (full black line), Π = 1.5 (blue dashed line), Π = 2 (red dashed dotted line), Π = 2.5 (green dotted line), and Π = 3 (cyan star line). In the foreign country, we always have $\Pi^* = 2$.

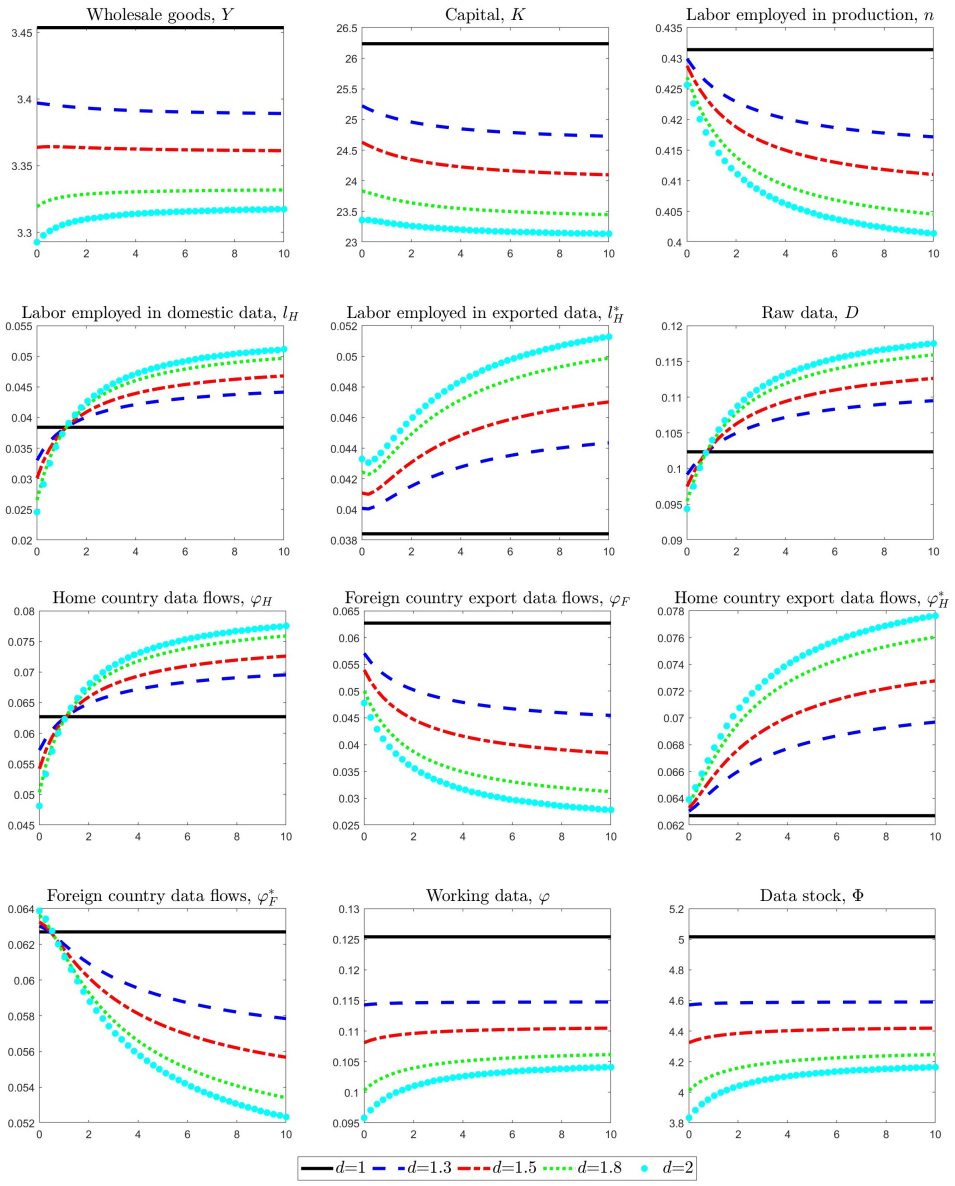


Figure C.2: Steady-States of the Main Variables with Different Values of Cost Multiplier d_D^*

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: $d_D^* = 1$ (full black line), $d_D^* = 1.3$ (blue dashed line), $d_D^* = 1.5$ (red dashed dotted line), $d_D^* = 1.8$ (green dotted line), and $d_D^* = 2$ (cyan star line). In the foreign country, we always have $d_D = 1$.

demand for raw data in foreign country D^* and further pushes φ_F^* to a higher level.[23] 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 C.2 displays further results when we alter the import cost multiplier d_D^*. First, production Y becomes lower as d_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_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.

D. Welfare Analyses

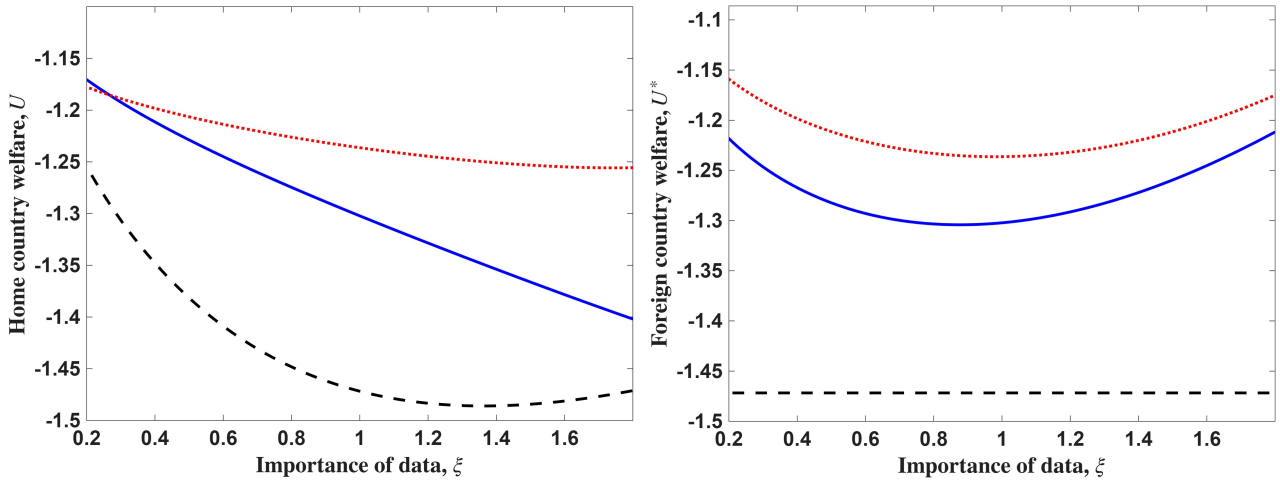
D.1 Alternative Data Depreciation Rate δ_Φ and Data Importance ξ

In this section, we conduct a robustness exercise by considering a relatively high data depreciation rate. Figure D.1 illustrates the welfare levels of the home and foreign countries when the data depreciation rate is set to $\delta_\Phi = 0.3$, compared to the benchmark value of $\delta_\Phi = 0.025$. To summarize the implications for the main conclusions of the paper, we vary the foreign country's data importance parameter ξ^* and examine how welfare changes as the home country's data importance parameter varies. To keep the analysis transparent, we report results for three canonical cases: the closed economy, the goods-trade-only economy, and the fully open economy.

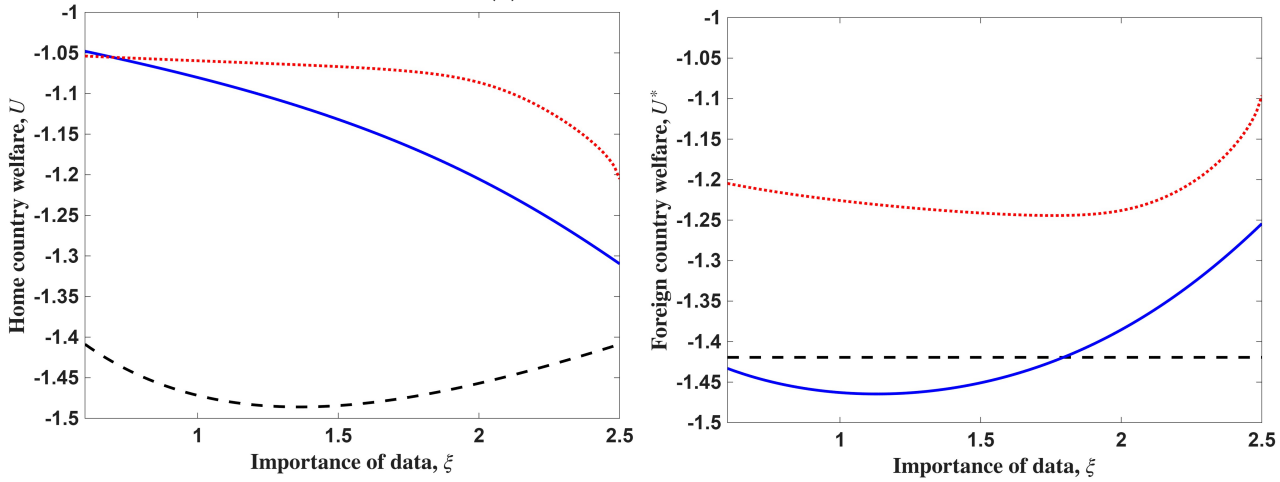
The figure illustrates welfare outcomes when the data depreciation rate is set to $\delta_\Phi = 0.3$, for three representative values of the foreign country's data importance parameter, $\xi^* = 1.0$, 2.4, and 3.7. These cases are chosen to highlight three qualitatively distinct patterns in how welfare responds to cross-country differences in data importance across model environments. Notably, the depreciation rate of capital here takes the value of 0.05 instead of 0.025 in the baseline analysis. The change in this depreciation rate does not change the results of the model much, but it can make the illustrations more explicit.

We begin with the case $\xi^* = 1.0$, which corresponds to the benchmark value used in the main text, with the only difference being the higher data depreciation rate. In this case, the welfare levels of the foreign country under both the goods trade model and the open economy rise substantially, so that they lie above the closed economy for the entire range of ξ. As a result, the foreign country strictly prefers trade in all scenarios. The underlying mechanism is that a higher data depreciation

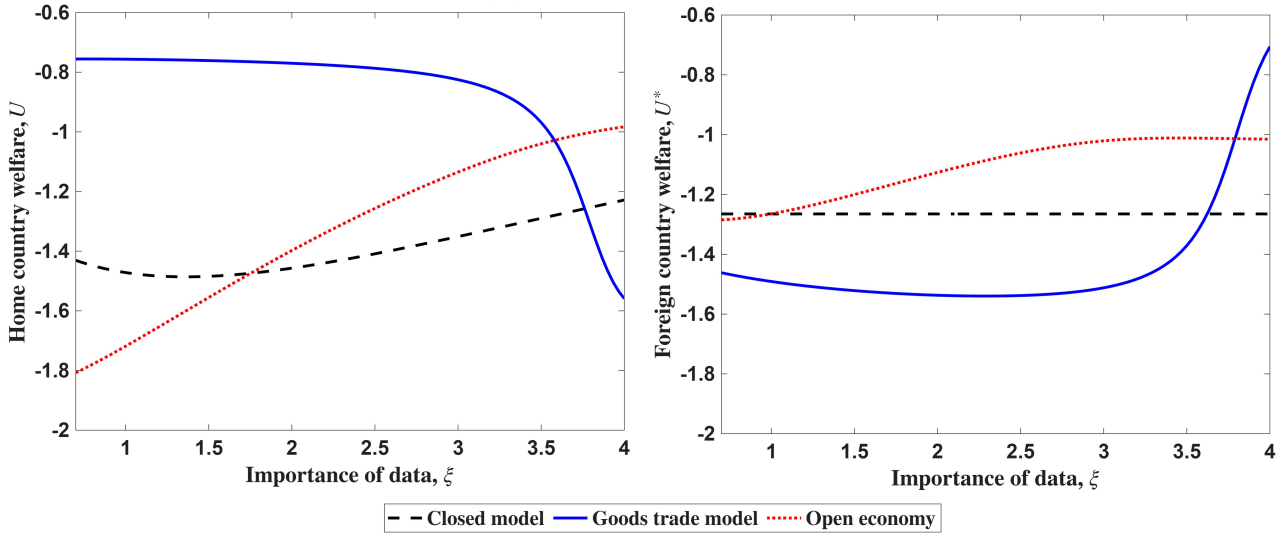
[23] The raw data within a country are non-rival, which means they can be used in generating working data used in both countries simultaneously. As a result, the increasing demand for one type of data pushes the usage of raw data higher, and then further pushes the supply of the other type of data higher.



(a) $\delta_\Phi = 0.3$ and $\xi^* = 1.0$



(b) $\delta_\Phi = 0.3$ and $\xi^* = 2.4$



(c) $\delta_\Phi = 0.3$ and $\xi^* = 3.7$

Figure D.1: Welfare Levels when $\delta_\Phi = 0.3$ and ξ^* takes different values

rate increases the need for continual data updates, making both countries more reliant on newly updated data. Consequently, situations in the benchmark calibration where the foreign country would lose from trade no longer arise once data depreciate sufficiently fast.

This naturally raises the question of whether, even with a high data depreciation rate, there remain circumstances under which one country would choose not to trade. The answer is affirmative once data importance is also sufficiently high. This is illustrated by the case $\xi^* = 2.4$. Here, when the cross-country gap in data importance becomes large, welfare under the goods trade model for the foreign country falls below the closed economy level, although welfare under the open economy remains above it. Intuitively, a higher data importance parameter implies more efficient use of data, which reduces the marginal need for additional data inputs. As a result, the gains from data flows decline, making trade less attractive in environments with large asymmetries.

Finally, when $\xi^* = 3.7$, these effects are further amplified. In this case, for sufficiently large differences in data importance across countries, the foreign country again experiences welfare losses from trade liberalization and therefore prefers to remain closed. Moreover, in this configuration, we observe that when the home country's data importance is far below that of the foreign country, opening data flows can reduce welfare relative to the goods trade model, that is, welfare under the open economy falls below that under goods trade alone. This occurs because, when both countries use data very efficiently, the additional privacy disutility associated with cross-border data flows can outweigh the benefits from increased data availability.

In intuitive terms, a higher data depreciation rate makes data more like a flow input that must be continuously renewed, which generally strengthens incentives to trade. However, the basic trade-off identified in the main text remains unchanged: the desirability of trade continues to hinge on how closely countries' data technologies are aligned. When data importance parameters are sufficiently similar, opening to trade remains welfare improving, even under rapid depreciation. When they differ substantially, the same tension between data-related gains and privacy costs re-emerges. Thus, higher data depreciation shifts the range over which trade is attractive, but does not overturn the core economic logic emphasized in the paper.

D.2 Heterogeneous Efficiency Term of Working Data Generation

In addition to cross-country differences in the importance of data in production, heterogeneity in the efficiency of working data generation may also play an important role in shaping trade incentives and welfare outcomes. In practice, countries differ not only in how intensively data are used in production, but also in their ability to transform raw data into usable working data, reflecting differences in digital infrastructure, human capital, algorithmic capabilities, and regulatory environments. Such heterogeneity can potentially affect relative productivity, data flows, and the desirability of openness in a manner similar to, but distinct from, differences in data importance.

To examine this channel, we extend our baseline analysis by allowing the efficiency parameter

in working data generation to differ across countries. Specifically, we fix the data-importance parameters at $\xi = \xi^* = 0.5$ and normalize the foreign country's efficiency parameter to $E^* = 1$, while varying the home country's efficiency parameter E.[24] Importantly, we adopt a destination-based perspective: all working data used in the home country, regardless of whether it is generated domestically or imported from abroad, are processed with the same efficiency parameter E, while all working data used in the foreign country are processed with E^*. This choice is motivated by two considerations. First, as the baseline analysis shows, data prices in equilibrium are primarily determined by conditions in the country where the data are utilized, since it is the destination-country producers' demand that pins down the marginal value of working data. Second, in practice, working data destined for use in a given jurisdiction must comply with that jurisdiction's regulatory standards, quality requirements, and technical specifications—for example, data formatting, labeling protocols, and privacy-compliance procedures—so that all working data employed within the same country are effectively produced under a common set of standards, justifying a uniform efficiency parameter at the destination level. We then compare welfare outcomes under the closed economy, the goods trade model, and the open economy with bilateral data flows, which is shown in Figure D.2. This exercise allows us to isolate the role of efficiency heterogeneity in shaping the gains from trade and data flows, independently of differences in data importance.

The figure shows that countries' incentives to engage in trade continue to depend primarily on the cross-country gap in data importance. In this exercise, we set $\xi = \xi^* = 0.5$, which corresponds to a symmetric environment in terms of data importance. According to the baseline analysis, such symmetry implies that both countries are willing to trade, and this prediction is confirmed by the results reported in the figure.

The left panel further indicates that variations in data-processing efficiency have the largest impact on welfare in the closed economy, where neither goods nor data can be sourced from abroad. In this environment, domestic productivity and welfare depend exclusively on internal production and data-processing capabilities, making efficiency heterogeneity particularly consequential. From the perspective of the home country, welfare under the goods trade model is lower than under the open economy, but the difference remains relatively small. This pattern suggests that even in the absence of cross-border data flows, international goods trade alone can generate substantial welfare gains for both countries. At the same time, allowing data to flow internationally provides additional welfare improvements on top of these gains, consistent with the baseline results.

An important additional finding is that when the cross-country gap in data-processing efficiency becomes sufficiently large, in particular when the home country's efficiency parameter E is very low, the foreign country may no longer find trade desirable. In this case, welfare under openness for the foreign country can fall below that under autarky or goods trade, leading to a breakdown

[24] In the baseline model and throughout the main text, we do not differentiate the efficiency parameter across countries, i.e., $E = E^* = 1$. The heterogeneous treatment here is introduced solely for the purpose of this robustness exercise, in order to examine how cross-country differences in data-processing efficiency affect trade incentives and welfare outcomes.

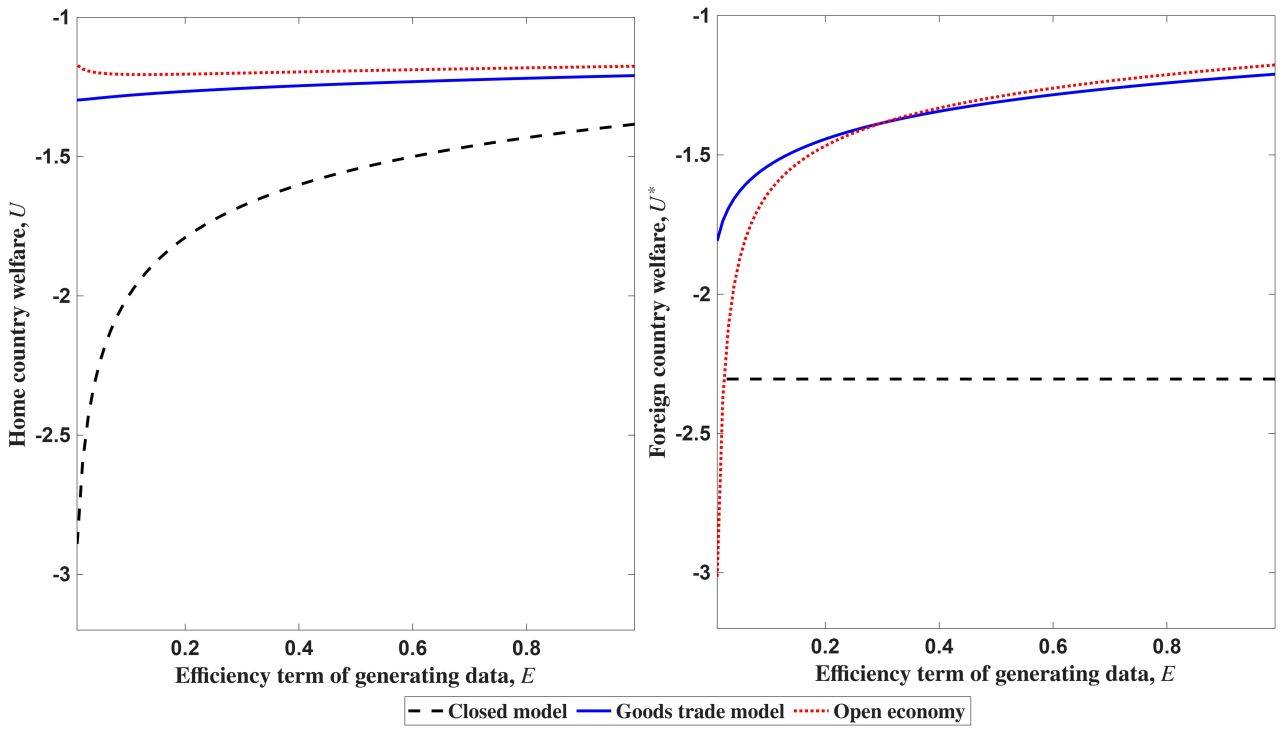


Figure D.2: Welfare Effects of Heterogeneous Data-Processing Efficiency

Notes. This figure shows the welfare levels under the closed economy, the goods trade model, and the open economy as functions of the home country's data-processing efficiency parameter E. We fix the data importance parameters at $\xi = \xi^* = 0.5$ and normalize the foreign country's efficiency parameter to $E^* = 1$.

of mutually beneficial trade. Moreover, only when efficiency differences remain moderate does the open economy dominate the goods trade model from the foreign country's perspective. This pattern highlights that, similar to differences in data importance, large disparities in data-processing efficiency can also undermine the sustainability of international trade and data flows.

D.3 Welfare and Data Flows Under Various Import Frictions

In this section, we show in Figure D.3 the welfare improvements and the net effects of data flows under various import frictions. From the figure, we see that the welfare improvements and data flow effects both decrease as the data import friction, d_D^*, increases. The feasible interval becomes narrower as d_D^* increases.

D.4 Welfare Levels Under Alternative Elasticities ρ and m

In this section, we show the welfare level when the elasticities of substitution ρ and m become different from the calibrated values in Table 1, especially when $\rho < m$. We show in Figure D.4 that our conclusions are robust, given that these two parameters are sufficiently larger than 1.

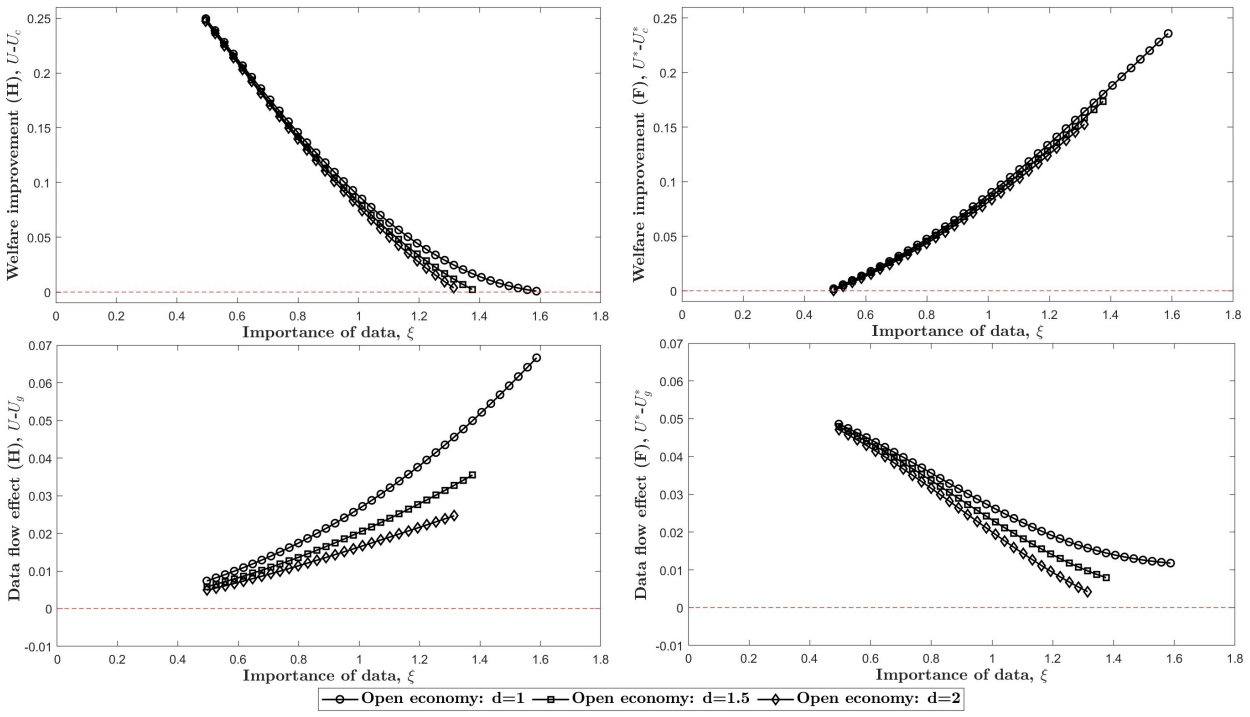


Figure D.3: Welfare Improvements and Data Flows with Different Importance of Data in the Two Countries When d_D^* Varies

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 in the composite factor ξ in the home country in open economies with different levels of d_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).

E. Transition Dynamics

E.1 Symmetric Countries in Transitions: Additional Variables

In this section, we show all the 20 main variables in the economy when the two countries are symmetric with respect to the following three shocks: productivity shock A_t, disutility shock π_t, and imported data shock f_t. The results are shown in Figures E.1, E.2, and E.3, respectively.

E.2 Asymmetric Countries in Transitions: Other Shocks

E.2.1 Disutility Shock and Cost Shock Under Extant Data Divide

In Figure E.4, we show all 16 main variables in response to a positive productivity shock A_t in the home country for completeness.

From Figure E.5 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.

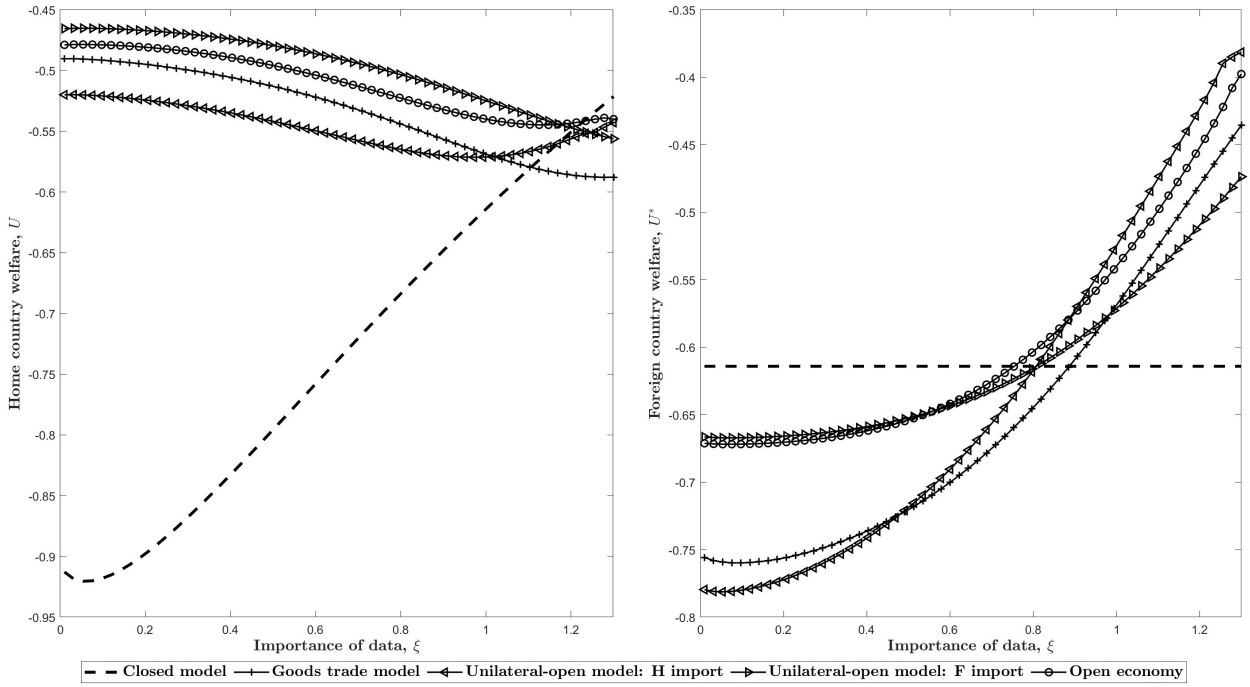


Figure D.4: Welfare Levels with Different Importance of Data ($\rho = 5$ and $m = 21$)

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 economy, 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 $\xi^* = 1$. In this figure, we have $\rho = 5$ and $m = 21$.

From Figure E.6 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.

E.2.2 Productivity Shocks Under Small Data Divide

From Figure E.7, 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 supports the results discussed in Section 4.2.

E.3 The Effect of Data Accumulation and Depreciation in Transitions

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 cyclicality 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 E.8 plots the transition dynamics after a positive productivity shock when

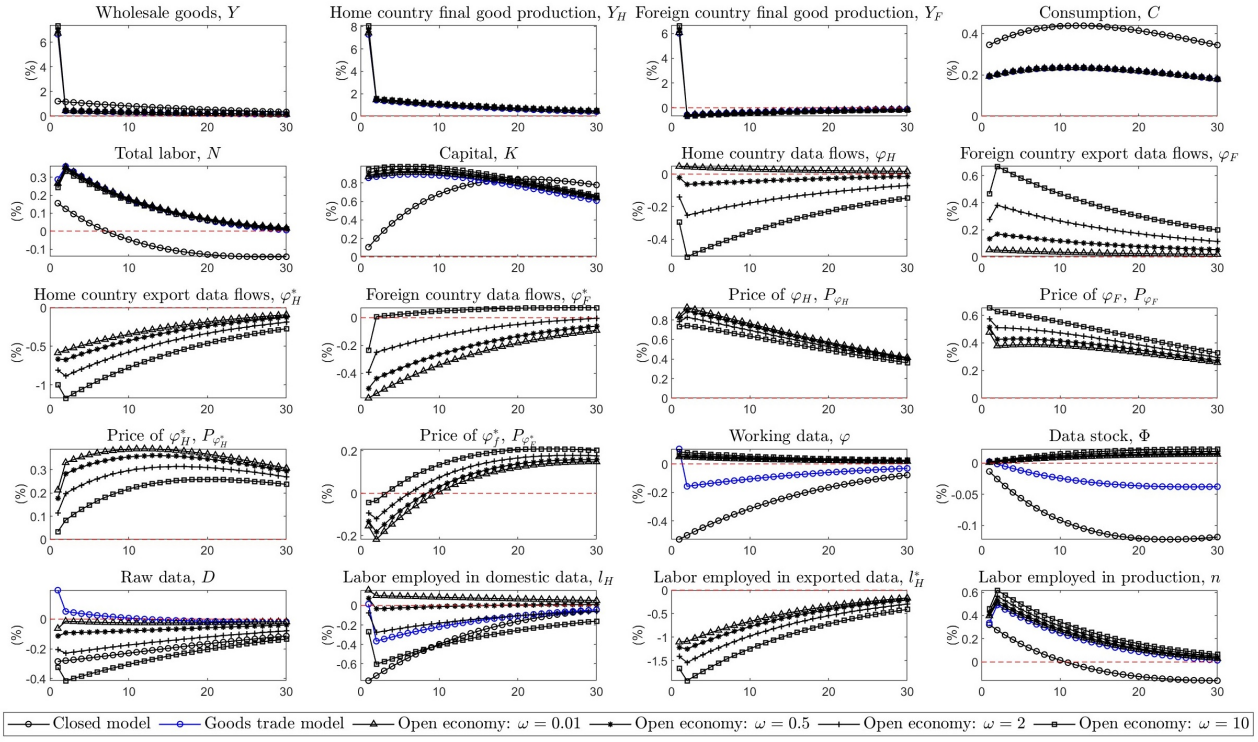


Figure E.1: Impulse Responses of the Positive Productivity Shock A_t in Symmetric Countries

Notes. 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 economy, 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.

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.

Juxtaposing 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 is thus not reported here.

E.4 Transition Dynamics from Closed Economy to Open Economy

This appendix presents the full transition paths of key macroeconomic variables when the economy permanently switches from the closed economy to the open economy at $t = 1$, complementing the cumulative welfare analysis in Section 4.3. In each figure, the solid blue line traces the transition path, the red dashed line marks the open economy steady state, and the gray dotted line indicates the closed economy steady state.

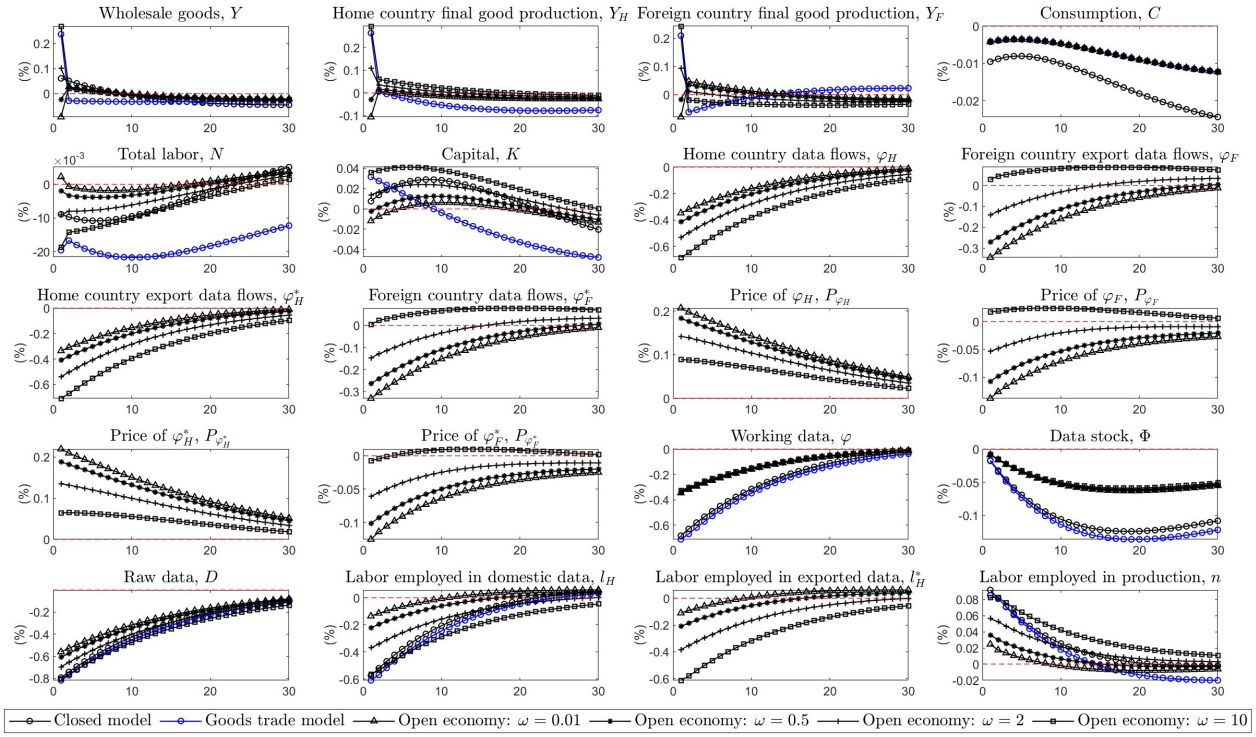


Figure E.2: Impulse Responses of the Positive Disutility Shock to Raw Data π_t in Symmetric Countries

Notes. 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 economy, 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 $\omega = 10$).

E.4.1 Asymmetric Countries ($\xi = 0.8$, $\xi^* = 1$)

Figure E.9 illustrates the transition dynamics when the two countries differ in their data importance. Several patterns are noteworthy. First, consumption in both countries jumps upward at $t = 1$ before gradually converging to the open steady state, consistent with the immediate welfare gains from accessing a larger market. Second, total production Y_t exhibits a notable jump at $t = 1$ followed by a hump-shaped path, reflecting the interplay between the immediate reallocation of factors and the slow build-up of the data stock. Third, raw data supply D_t initially spikes as households respond to new data demand from both countries, then declines toward the open steady state, which is below the closed economy steady-state level due to the higher privacy cost ($\Pi(1+b)D_t^2$ vs. ΠD_t^2) in the open economy. Fourth, the working data stock Φ_t exhibits a pronounced overshooting, rising above the open steady state before settling, as the initial surge in cross-border data flows temporarily boosts data accumulation beyond its long-run level. In the foreign country, the data stock Φ^* rises monotonically from a lower initial level, reflecting the foreign country's higher ξ^* and thus stronger data accumulation capacity. These asymmetric adjustment speeds are the dynamic counterpart of

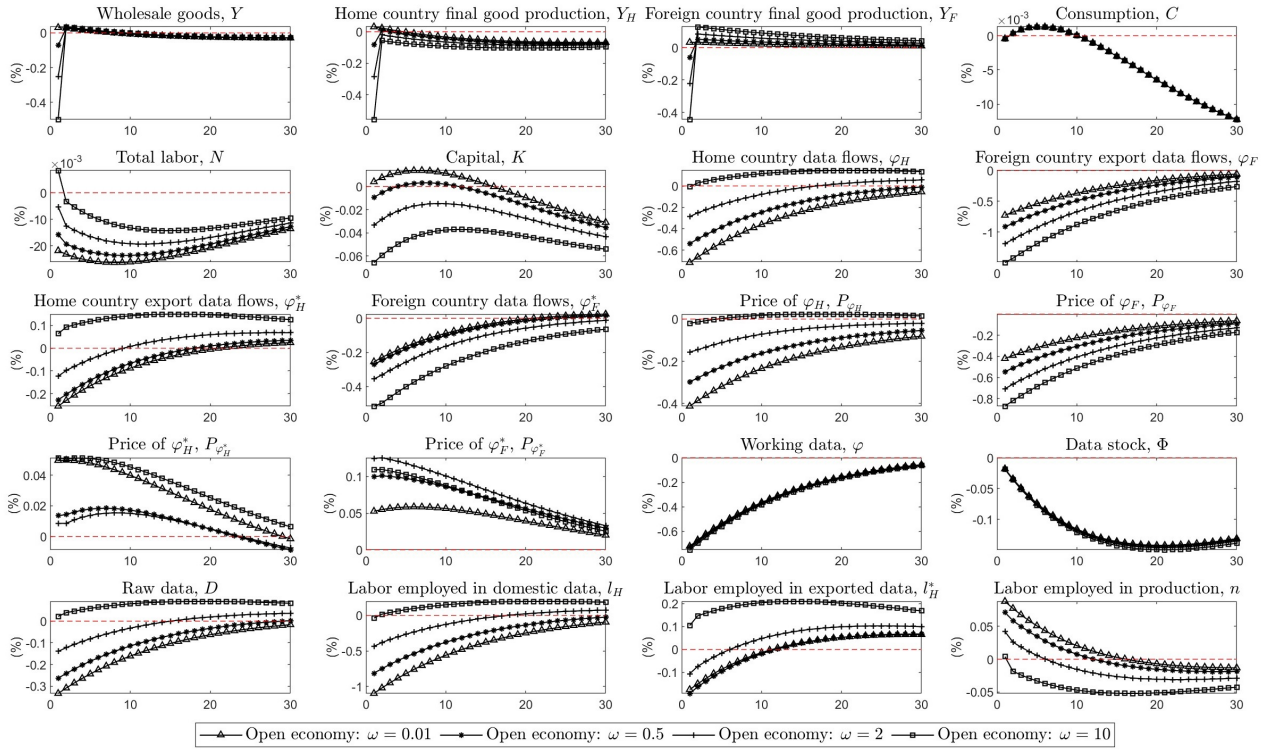


Figure E.3: Impulse Responses of the Positive Cost Shock to Imported Data Flows f_t in Symmetric Countries

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 $\omega = 10$).

the latecomer's advantage discussed in Section 4.3.

E.4.2 Symmetric Countries ($\xi = 1$, $\xi^* = 1$)

Figure E.10 presents the symmetric benchmark. As expected, the transition paths are identical across countries. Consumption overshoots the open steady state in the short run as households anticipate higher lifetime income from the open economy, then converges from above. Total production similarly exhibits a hump-shaped overshooting, driven by the rapid initial accumulation of working data. The working data stock Φ_t displays the most pronounced overshooting: it rises sharply as both countries simultaneously begin exchanging data, temporarily exceeding the long-run level before depreciation and the settling of data flows bring it back toward the steady state. Raw data D_t declines monotonically from its higher closed-economy level, reflecting the increased privacy costs under the open regime. The symmetric paths confirm that, absent any pre-existing data divide, the transition to an open economy is smooth and generates welfare gains from the first period, with no country experiencing even transitory losses.

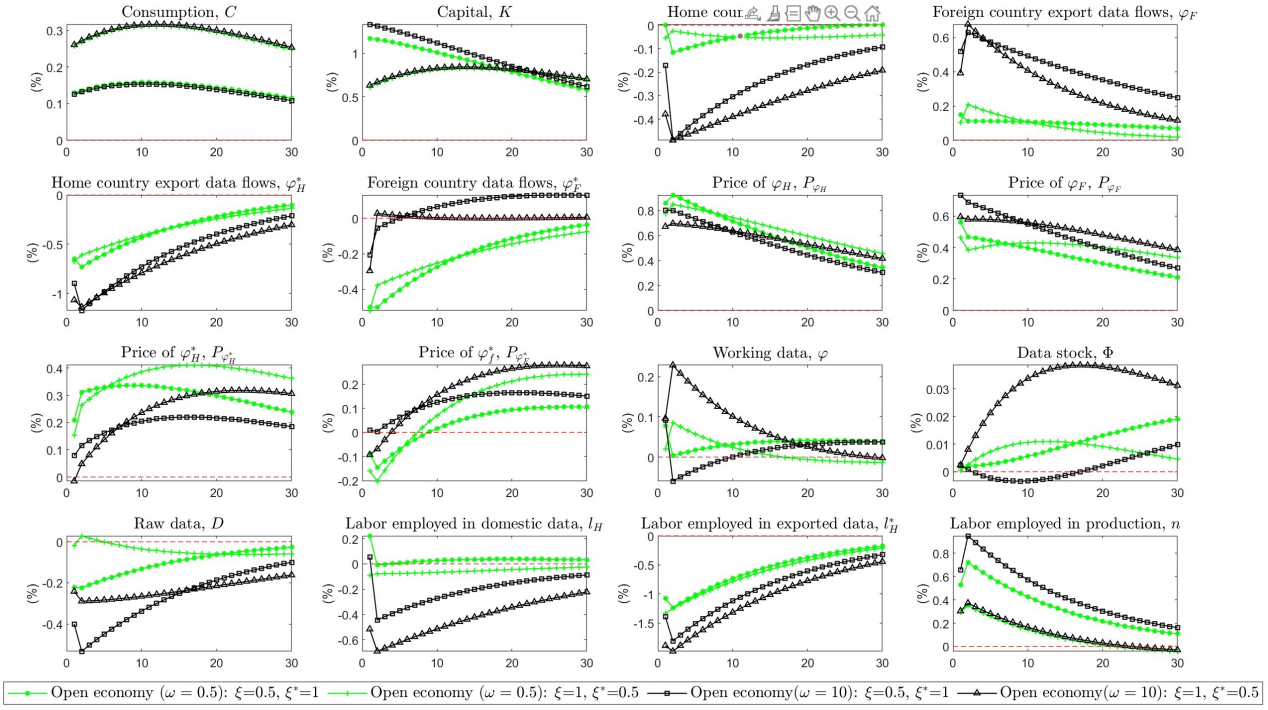


Figure E.4: Impulse Responses of the Positive Productivity Shock A_t in Asymmetric Countries

Notes. 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 (only in the home country) 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 $\xi^* = 0.5$).

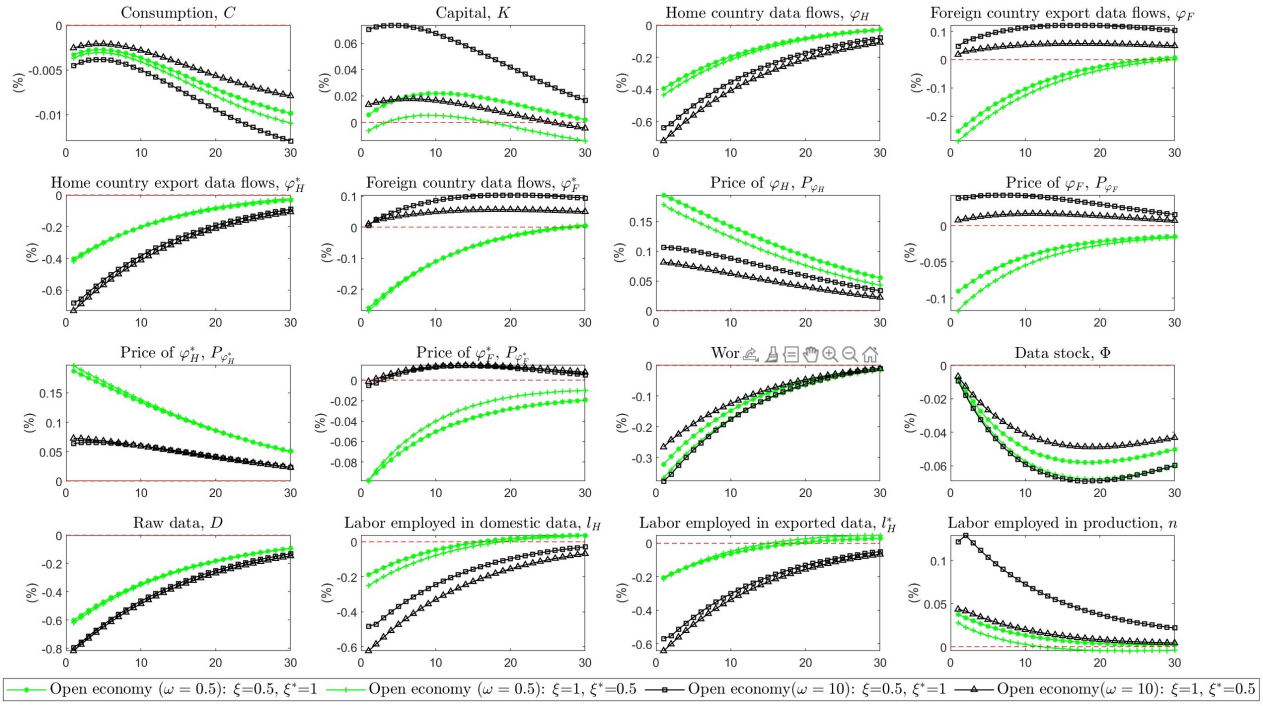


Figure E.5: Impulse Responses of the Positive Disutility Shock π_t in Asymmetric Countries

Notes. 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.

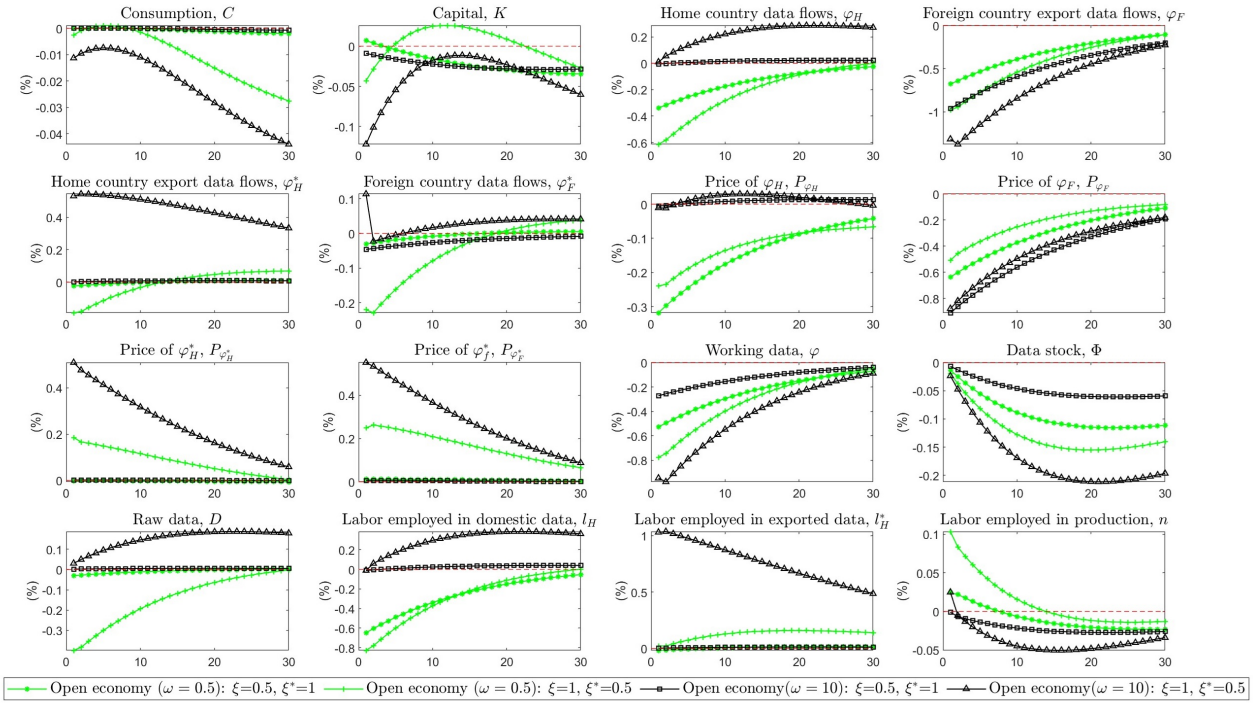


Figure E.6: Impulse Responses of the Positive Cost Shock f_t in Asymmetric Countries

Notes. 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 ($\omega = 0.5$ 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 cost shock only happens in the home country.

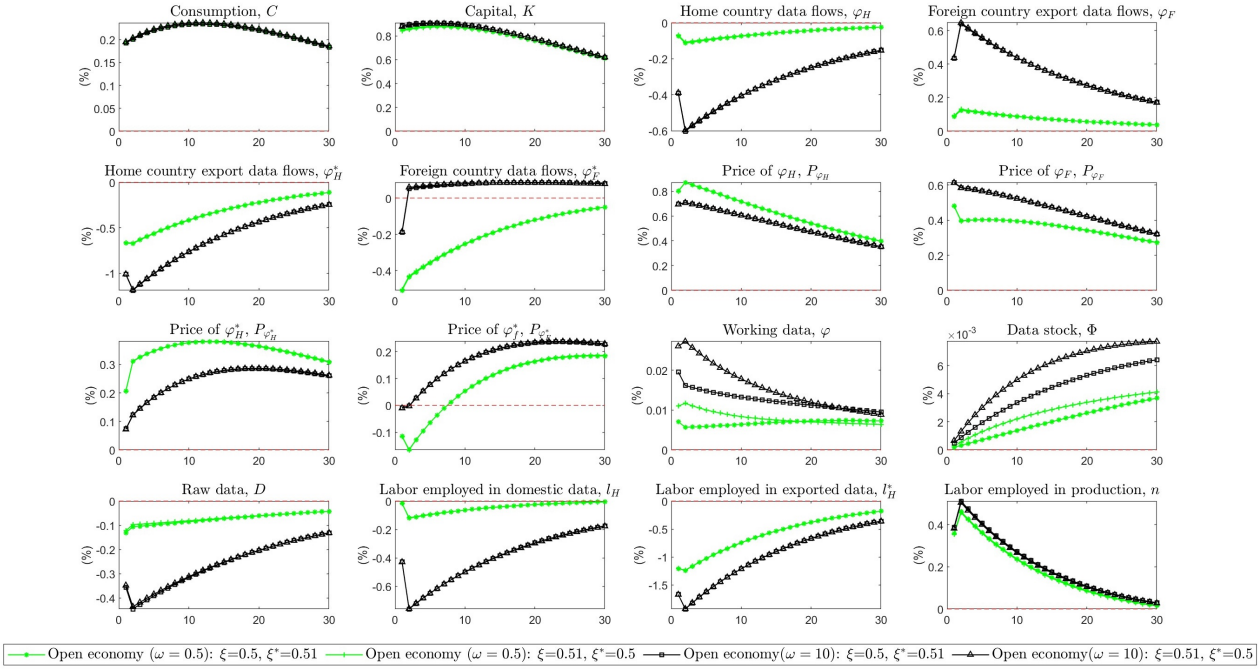


Figure E.7: Impulse Responses of the Positive Productivity Shock A_t in Asymmetric Countries When the Data Divide Is Small

Notes. 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.

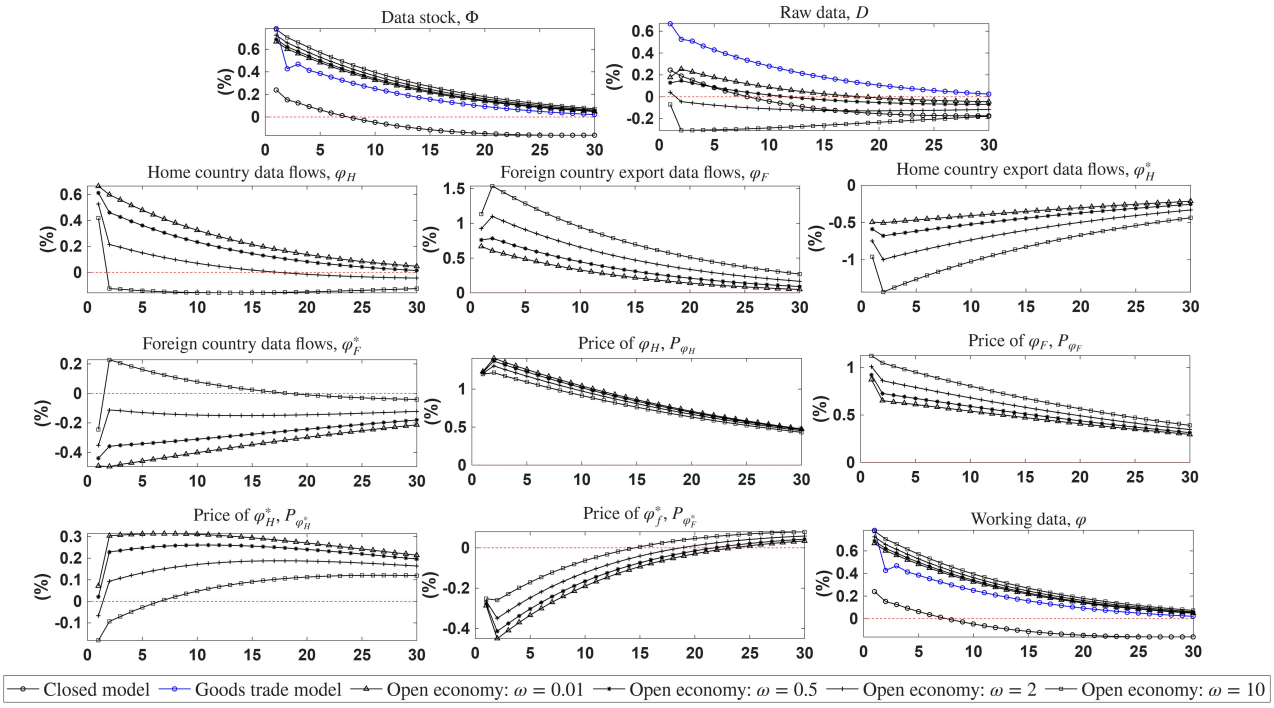


Figure E.8: Impulse Responses of the Positive Productivity Shock A_t in Symmetric Countries When Data are Fully Depreciated

Notes. This figure illustrates the impulse responses of 11 data-related 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 economy, 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.

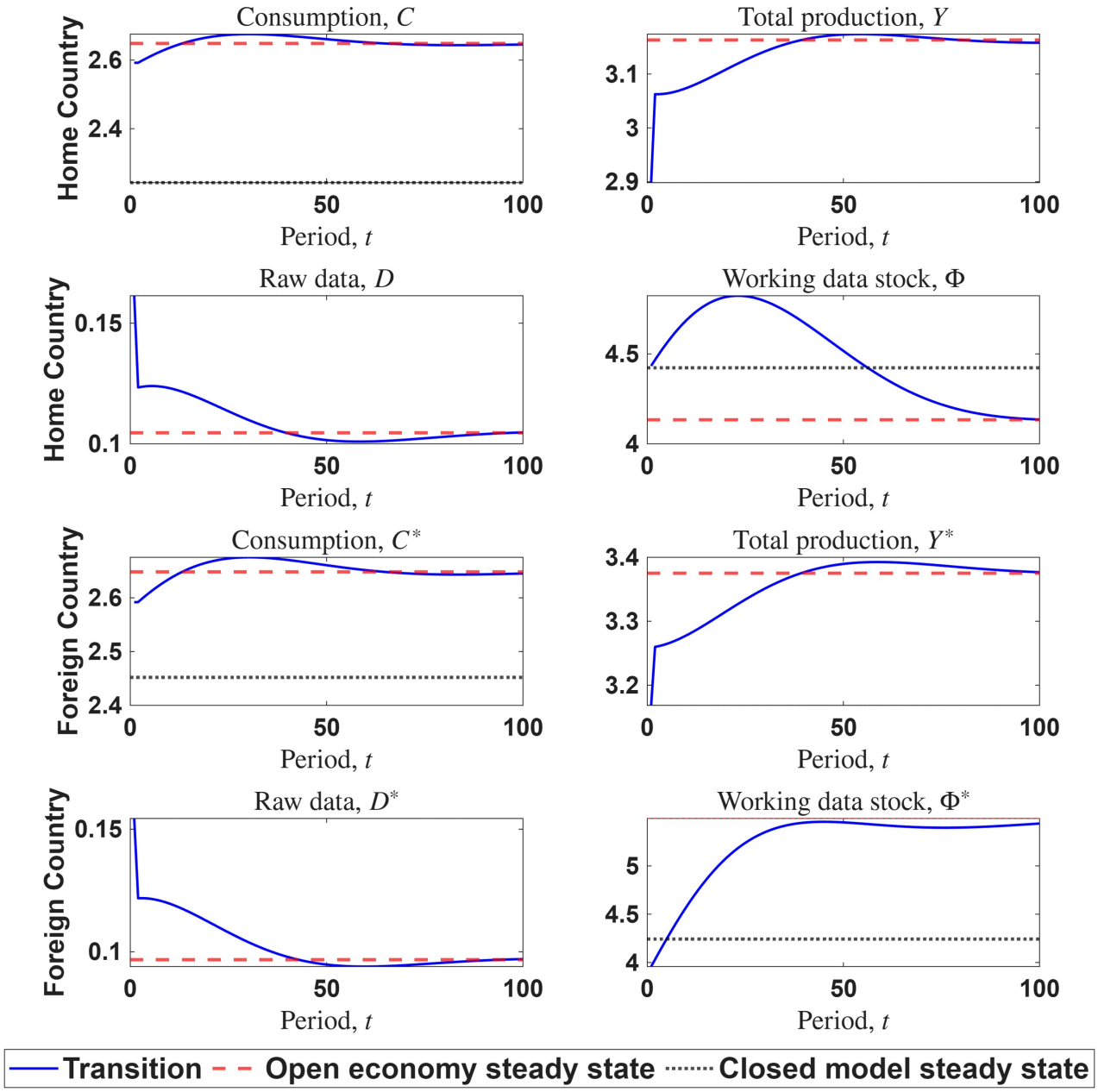


Figure E.9: Transition Dynamics from Closed Economy to Open Economy ($\xi = 0.8$, $\xi^* = 1$)

Notes. This figure depicts the transition paths of consumption C_t, total production Y_t, raw data D_t, and working data stock Φ_t for both the home and foreign countries when the economy permanently switches from the closed economy to the open economy at $t = 1$. The home country has lower data importance ($\xi = 0.8$) relative to the foreign country ($\xi^* = 1$). The solid blue line is the transition path, the red dashed line is the open economy steady state, and the gray dotted line is the closed economy steady state.

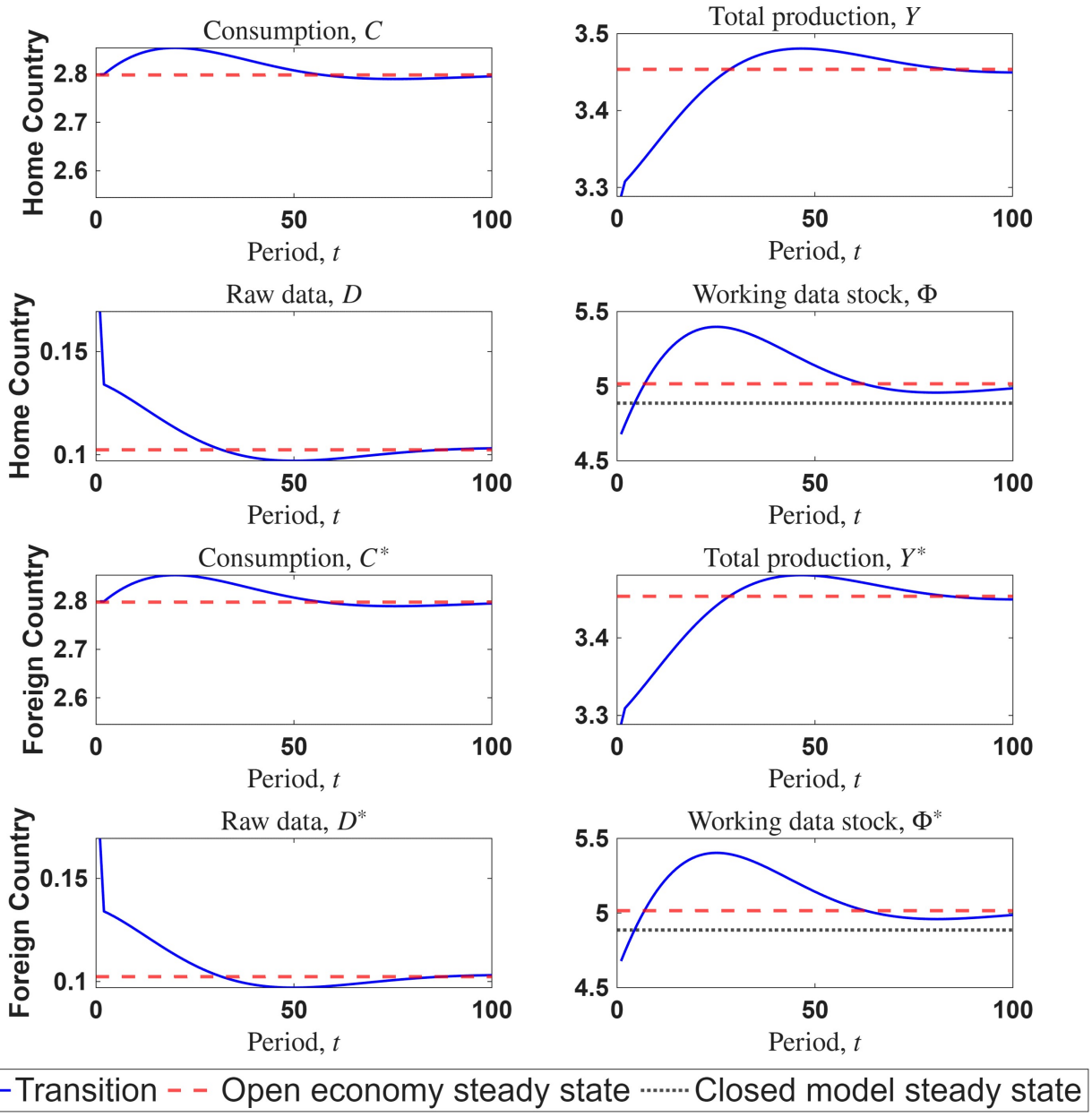


Figure E.10: Transition Dynamics from Closed to Open Economy ($\xi = 1$, $\xi^* = 1$)

Notes. This figure depicts the transition paths of consumption C_t, total production Y_t, raw data D_t, and working data stock Φ_t for both the home and foreign countries when the economy permanently switches from the closed economy to the open economy at $t = 1$. Both countries have the same data importance ($\xi = \xi^* = 1$). The solid blue line is the transition path, the red dashed line is the open economy steady state, and the gray dotted line is the closed economy steady state.