The Impact of Data Transfer Restrictions on Data Centre Geography

Otto Kässi*[1,2] and Vili Lehdonvirta[1]

[1]Aalto University
[2]Etla Economic Research

April 30, 2026

Abstract

We study whether national laws limiting cross-border personal-data transfers reshape where computation occurs. We combine a panel of data-centre openings in 158 countries with classifications of transfer regimes. Using event studies, we find no entry response in most countries. Effects concentrate in downstream economies: GDPR-level tightenings are followed by one-third more openings, while high-GDP economies show weaker patterns. We find flat pre-trends, no spillovers, and null effects for crypto data centres, which are unaffected by personal-data transfer laws. Motivated by these findings, we develop a compliance-quota model showing how tighter rules can reduce demand for compute while increasing domestic data-centre investment.

JEL F14, L51, L86, O33, D22

Keywords: Cross-border data flows; data localisation; data transfer restrictions; data centres; digital infrastructure; regulation; investment; compliance constraints; cloud computing; geography of computation; GDPR; weaponised interdependence; digital trade

*We thank Linh Ngo for her excellent work in collecting and analysing the law dataset, and Martina Ferracane, Markus Holmgren, Stefka Schmidt, Francesco Venturini, and participants at the PRIN Workshop at LUISS University for helpful comments. This work is supported by an ERC grant (GEOCLOUD, 101141671). Author names are listed in alphabetical order. Emails: Kässi: otto.kassi@aalto.fi; Lehdonvirta: vili.lehdonvirta@aalto.fi.

1 Introduction

Many countries now regulate cross-border flows of personal data. A large empirical literature studies how these rules shape intangible outcomes such as innovation, productivity, and digital trade. Yet digital production still relies on fixed, place-based capital. Regulation of intangible cross-border flows reallocates this complementary physical capital across space. Data centres are the factories and warehouses of the digital age: they bundle servers, cooling, and specialized real estate to deliver data storage and processing services. Little is known about how laws regulating data transfers shape these material infrastructures, even though they determine where digital production can take place and at what cost. Their geography also matters for a host of other reasons: they generate employment and tax revenues, consume electricity, water, and land, anchor digital activities within particular jurisdictions, and shape countries' exposure to digital supply chain risks.

This study links data transfer regulation to the geography of data centre investment. Prior work finds that regulations like the GDPR suppress compute demand: compliance costs raise the effective price of data processing, and firms use less of it (Demirer et al., 2024; Chang et al., 2023; Janßen et al., 2022). The focus here is different: not how much firms compute, but where. When cross-border transfers are restricted, firms might shift workloads to domestic locations; total compute may fall even as local demand for data centre capacity rises. We formalise this geographic margin in a compliance quota model that can reproduce both patterns. We ask: for which countries do tighter data transfer laws increase domestic data centre entry?

To answer this, we combine two new pieces of measurement. First, we construct a country–year panel of global data centre openings from S&P Global Capital IQ Pro Real Estate, which tracks property-level data centre facilities across 158 countries using a globally standardised methodology. Second, we build a harmonised panel measure of national laws regulating data transfers by classifying each country's legal framework on a restrictiveness scale through LLM-based analysis of the full legal text, with targeted manual validation. In the empirical analysis, we focus on discrete changes, in particular the first adoption of laws at least as restrictive as the GDPR.

Whether domestic entry actually rises depends on whether a country can support local data centre capacity. We find that it does in countries with more downstream production structures (consumer-facing sectors where personal data is central to production) and secondarily in richer countries. Upstream and poorer countries face the same compliance obligations but show no measurable domestic entry response, which is why the full-sample average, though positive, is imprecisely estimated.

As of 2024, 172 countries and autonomous regions had enacted laws or rules that restrict or complicate the transfer of personal data abroad (Greenleaf, 2025). These regimes are heterogeneous, ranging from rules that largely enable free flow to systems that require case-by-case authorisation, and they vary considerably in legal scope and restrictiveness across jurisdictions and over time.[1] These policies are often justified in terms of privacy, sovereignty, or security, yet they may also alter the geography of computation. Tighter data transfer regimes can change

[1] These rules are catalogued in Greenleaf (2023), with full legal texts hosted at WorldLII (World Legal Information Institute, 2024); see also Ferracane (2017).

where firms carry out digital production by shifting data processing toward domestic locations. Despite the centrality of this issue for digital trade and production, systematic evidence on how data transfer restrictions affect data centre entry across countries remains limited.

Motivations for tightening cross-border data transfer restrictions are not limited to privacy. Governments increasingly value strategic autonomy and reduced exposure of data and computation to foreign jurisdictions — what Farrell and Newman (2019) call "weaponised interdependence". The policy family is also defended in industrial policy terms: by restricting cross-border digital activity, policymakers seek to tilt demand toward domestic providers and expand local data accumulation, which could strengthen incentives for domestic innovation in data-intensive markets (Zhou, 2024). At the same time, restrictions that reduce access to global information inputs can lower exposure to frontier knowledge and generate measurable innovation losses (Sun and Trefler, 2023; Zheng and Wang, 2020). Taken together, these channels imply that data transfer restrictions can simultaneously reduce cross-border data flows and reallocate processing to local data centres. This makes it important to study the physical side of the digital economy: where computation takes place and where data centres are built.

We implement a two-way fixed effects event-study design around regulatory tightenings. The effects are heterogeneous. The sharpest response is in countries whose industries are closer to final demand (i.e., more *downstream*; Antràs et al., 2012; Mancini et al., 2024); higher-GDP economies show a positive but noisier secondary pattern. The full-sample average is positive but imprecisely estimated.

Policy adoption is unlikely to be strictly exogenous, and we cannot rule out every confound. We do, however, provide evidence against the most plausible alternatives. Pre-treatment coefficients are flat, arguing against gradual pre-existing divergence in data centre investment and against tightening that responds to anticipated investment changes. State enforcement capacity is unlikely to drive the result: the effect is similar after controlling for measures of state legal and regulatory capacity. General data-centre industrial policy is unlikely to drive it either: a falsification using crypto mining facilities, which sit outside the scope of personal-data law but are exposed to the same industrial-policy instruments, shows no response. The estimated effect also scales with the restrictiveness of the legal regime in a dose-response fashion. Together, these patterns are more consistent with a reaction to compliance to personal data transfer restrictions rather than with a general digital economic policy taking place concurrently.

Our model implies that the investment response depends on the relative costs of local and remote compute. In the model, tightening increases data centre entry when the price gap between remote and local compute is small enough that reallocation does not produce a large increase in the unit price of compute. Consistent with this, we find that the post-tightening increase in domestic data centre entry is higher in countries with cheaper local electricity. To quantify the aggregate scale of these effects, we construct a back-of-the-envelope counterfactual. Taking the estimated post-treatment effect to be the causal impact of the tightening, the cumulative number of data centre sites in countries with more downstream production structures would be roughly 22% lower in its absence.

Our paper connects to several literatures. First, it relates to work on how trade policy affects domestic investment (e.g., Pierce and Schott, 2018; Amiti et al., 2020; Handley and Limao, 2015;

Handley and Limão, 2017; Caldara et al., 2020). A common finding in this literature is that raising trade barriers can reduce investment by increasing input costs and policy uncertainty, particularly for firms exposed through global value chains. At the same time, when policy restricts access to a foreign good, it can induce substitution toward domestic substitutes. We document entry patterns consistent with substitution toward domestic compute.

Second, we contribute to a growing empirical literature on the economic effects of the GDPR and related privacy regulation (e.g., Peukert et al. (2022); Frey and Presidente (2024); Demirer et al. (2024); Goldberg et al. (2024); Aridor et al. (2020); Johnson et al. (2023); Janßen et al. (2022); Koski and Valmari (2020); Gupta et al. (2022); Sisto and Van der Marel (2025); Zhang et al. (2025); Jia et al. (2025); Sun and Trefler (2023)). We view the GDPR as one prominent instance within a broader set of cross-border data transfer regimes and study domestic entry responses across countries using a globally comparable panel.

Third, we contribute to a stream of research that studies data as an intangible factor of production, including Farboodi and Veldkamp, 2021; Corrado et al., 2022; Agrawal et al., 2019; Goldfarb and Tucker, 2019. Our model formalises the complementarity between intangible data and physical compute and adds a location choice between local and remote compute that is distorted by regulation.

Fourth, our paper links to a small but growing literature on the 'geography of cloud computing', which studies the economic and policy determinants of data centre locations (Pan Fang and Greenstein, 2025; Lehdonvirta et al., 2025, 2024; Bonfiglioli et al., 2025; Tian et al., 2025).

Finally, our results have bearing on recent debates on "weaponised interdependence" (Farrell and Newman, 2019) and "digital sovereignty" (Adler-Nissen and Eggeling, 2024), which are concerned with how dependence on foreign digital infrastructures may expose countries to espionage and coercion. Our contribution is to show, using cross-country evidence, that data transfer restrictions are followed by a shift in data centre entry toward domestic locations, consistent with the idea that laws governing data flows can be used to limit data-intensive activities' exposure to foreign territorial jurisdictions.

2 Data

2.1 Laws regulating the cross-border flow of data

Laws that regulate cross-border data transfers typically specify the legal and procedural requirements for sending data abroad, such as adequacy decisions, contractual safeguards, consent-based derogations, or government approvals (Casalini and López González, 2019).

For the purposes of this study, we focus on such cross-border transfer restrictions on personal data, and treat them as a measure of the practical difficulty of processing personal data abroad. Our focus is on the legal scope and structure of these restrictions, which may vary across jurisdictions, and over time.

We interpret stricter regimes as expanding the set of data and use-cases for which cross-border processing is illegal, uncertain, or so operationally costly as to force firms to process data within the country's borders. Accordingly, ranking countries' laws by overall restrictiveness provides a consistent measure of how binding the domestic-processing requirement is in practice.

The classification is de jure: it captures what the law requires, not how vigorously regulators enforce it. Data centres are long-lived assets; when firms commit capital, they price in the full range of future enforcement scenarios, not just past enforcement rates. A law on the books raises expected compliance costs even where regulatory action has so far been limited. We assess the sensitivity of our results to cross-country heterogeneity in states' enforcement capacity separately.

Our dataset on national data transfer frameworks extends the cross-border data regulation dataset compiled by Greenleaf (2023); the underlying law texts are catalogued by World Legal Information Institute (2024). It documents the primary national data privacy laws that govern cross-border data transfers and the domestic storage of personal data up to the end of 2022. For each country, we identify the primary comprehensive privacy legislation, recording the introduction of new frameworks, replacements of existing ones, and their enactment dates. We have also manually updated the law catalogue for years 2023–2024. Minor amendments that do not alter the overall structure of the legal framework are excluded. This allows us to construct a balanced country–year panel of data transfer regimes. We focus on economy-wide ("comprehensive") personal-data frameworks rather than sectoral rules (e.g., health, finance) because the former are the relevant margin for broad data centre investment decisions and for cross-industry comparability.

While the data documented in Greenleaf (2023) provides a comprehensive inventory of relevant laws, it does not assess their restrictiveness. We extend the dataset by classifying each national framework using the taxonomy of cross-border data transfer regulation developed by Casalini and López González (2019). This taxonomy distinguishes between seven levels (0–6) of regulatory restrictiveness, ranging from the absence of specific cross-border provisions (0) to frameworks requiring case-by-case authorisation for each transfer (6). Intermediate levels correspond to conditional regimes—such as accountability-based systems (1), firm-assessed adequacy regimes (2), and public authority-based adequacy regimes with contractual or procedural safeguards (3–4). The taxonomy does not compare legal regimes by their objectives or institutional design. It ranks them by the stringency of their operative legal requirements for transferring personal data abroad. Our estimates therefore recover an average effect of adopting a more restrictive regime, primarily through new compliance obligations.

To assign each country–year observation to one of these levels, we employ a large language model classifier that analyses the full legal text of the relevant data transfer regime. The model is prompted with structured criteria derived from the taxonomy definitions in Casalini and López González (2019). The resulting classification provides a panel measure of data flow restrictiveness that is conceptually aligned with the OECD framework and enables consistent cross-country and temporal comparisons. We present the classification of laws with examples in Table 1.

Classifying legal texts involves a degree of subjectivity. Our procedure has three steps: we embed anchor classifications from Casalini and López González (2019) in the classification prompt as reference points. We then apply the prompt to the full dataset using an LLM (Claude Sonnet), and validate the output with an inter-coder check, in which a human reviewer independently classified a random sample of 20 laws; the LLM and human classifications agree in all 20 cases.

We present more details in the appendices. Appendix A presents representative examples with the key statutory provisions that determine each level; Online Appendix B reproduces the full classification prompt; robustness of the classifications to prompt perturbation is assessed in Online Appendix C.

Together, the human validation and prompt robustness checks follow the practices Ludwig et al. (2025) recommend for using LLM-based classifications in empirical research: checking for classification error and verifying that the classification is robust to reasonable variations in the prompt used.

In our empirical analysis, we use a binary indicator that equals one from the first year a country enacts a regime at least as strict as the GDPR (level 4 or higher) onwards. This dichotomisation reduces sensitivity to small ranking errors between intermediate levels. Remaining (classical) measurement error would bias estimated effects toward zero, making our results conservative.[2] The cost of this simplification is that countries at levels 4, 5, and 6 all receive the same treatment status, so the estimate reflects an average effect across regimes of differing restrictiveness rather than variation in restrictiveness within the treated group.

Choosing level 4 as our definition of treatment is motivated by both conceptual and practical reasons. Conceptually, it corresponds to a material tightening of transfer conditions. Although mechanisms such as Standard Contractual Clauses or Binding Corporate Rules remain legally available, their use typically involves added compliance cost, documentation, and regulatory scrutiny (European Parliament and Council of the European Union, 2016). Level 4 essentially marks the point at which transfers become conditional rather than routine. From a more practical perspective, a substantial share of the countries in our data—including all EU member states—are at level 4, so setting the threshold higher would sharply reduce the number of treated cohorts and thus statistical power.

Our "treated" units consist of 71 countries that implemented a level 4 or stricter regime by the end of the study period. The staggered adoption of the laws over time is visualised in Figure 1.

[2] We report a prompt-robustness exercise in Online Appendix C. The binary ≥ 4 classification agrees in 86% of cases between the two most divergent prompt versions; the implied 14% misclassification rate is consistent with the attenuation-bias interpretation.

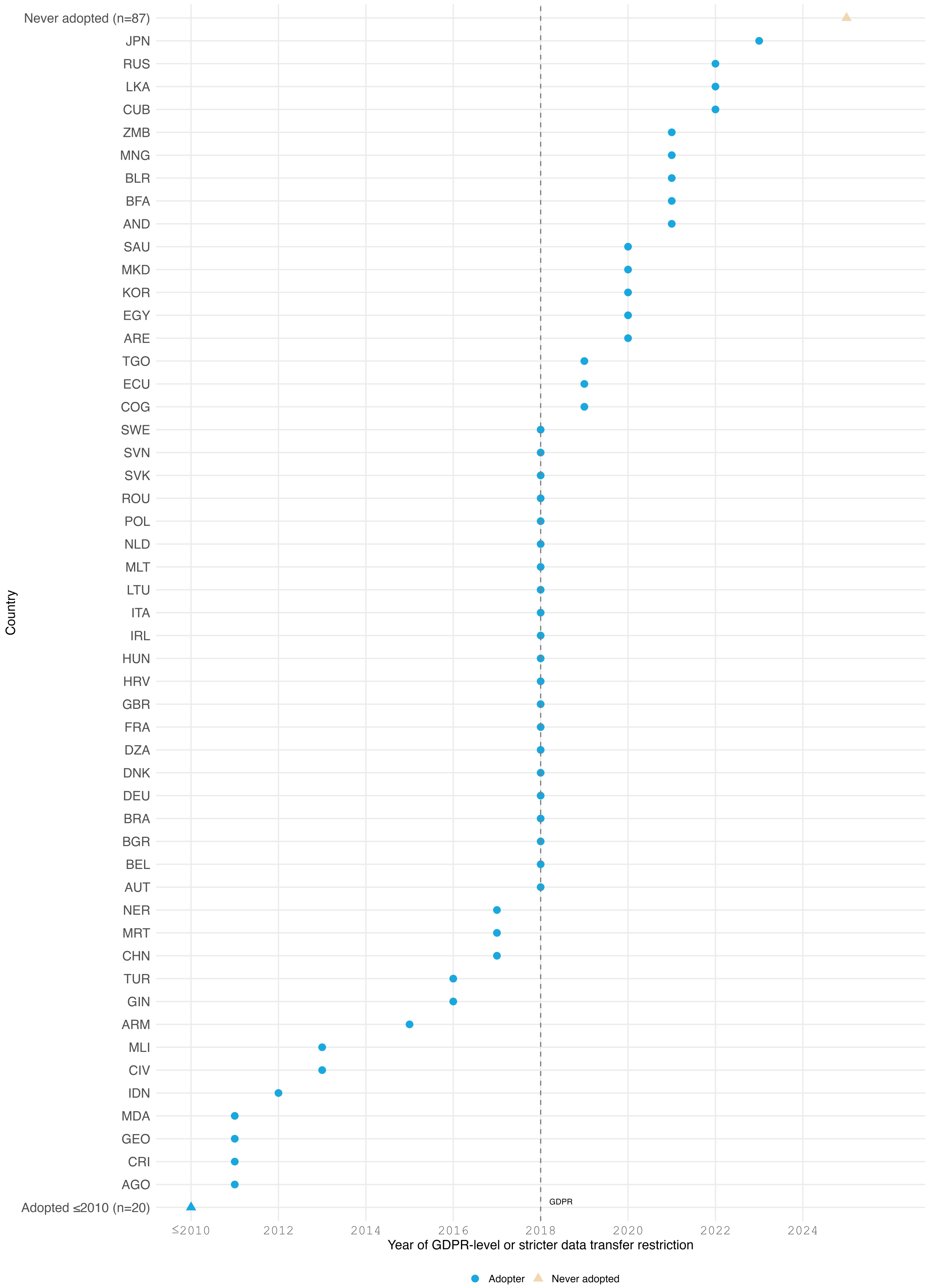


Figure 1. Adoption of GDPR-level or stricter laws regulating data transfers, 2010–2024.

Note: Each row represents a country. Shading indicates the first year of adopting a data transfer regime rated level 4 or above (GDPR-equivalent or stricter). Law classifications are based on Greenleaf (2023), extended through 2024, and assigned using the taxonomy of Casalini and López González (2019). 71 of the 158 countries in the sample adopted a level 4 or stricter regime by end of 2024.

Table 1. OECD Cross-Border Data Transfer Restrictiveness

OECD High-level Category	Sub-Level (0–6)	Description	Example Countries
No Regulation	0. No Regulation	No domestic rules governing cross-border flows.	United States, Afghanistan
Ex-Post Accountability	1. Free-Flow with Ex-Post Accountability	Transfers free; exporter liable if data misused abroad.	Canada, Hong Kong, India
Conditional on Safeguards	2. Self-Assessment of Adequacy + Fallbacks	Exporter judges adequacy; otherwise relies on consent, contracts, or necessity.	Australia, South Africa, Mexico
	3. Public-Assessment of Adequacy + Options	Public authority certifies "adequate" destinations; others need SCCs/BCRs/consent.	Japan, Nigeria
	4. Public-Assessment + Options + Processing Conditions	Same as 3 plus requirement that foreign processing apply sender's rules; exporter liable downstream.	European Union, United Kingdom
Conditional on Ad-Hoc Authorization	5. Case-by-Case Authorization with Adequacy Fallback	Transfers to "adequate" jurisdictions flow freely; others require regulator approval.	Russia, Saudi Arabia
	6. Case-by-Case Universal Authorization	All transfers require prior regulator approval; often tied to localization.	China, Egypt

Notes: The seven-level cross-border data transfer restrictiveness taxonomy is adapted from Casalini and López González (2019). Country-level assignments reflect the primary comprehensive personal-data privacy legislation in force and are produced by an LLM classifier applied to the full legal text. Some countries have shifted across levels over the 2010–2024 sample window (see Figure 1).

2.2 Data centres

The dependent variable measures the net change in data centre count in a given country–year, capturing both new openings and closures. These come from the S&P Global Capital IQ Pro Real Estate database (S&P Global Market Intelligence, 2025b), a commercial database of real-estate-related data. According to S&P, their dataset covers 95,000 properties across 101 countries; the 158 countries in our panel include jurisdictions with zero recorded data centre openings throughout the sample period, which are not counted in the S&P figure. Capital IQ analysts draw on public company disclosure data such as annual and quarterly financial reports, regulatory filings, investor presentations and other market documents. The main strength of the dataset for our analysis is that it is standardised to be globally consistent. Its methodology allows us to map data centre construction consistently across jurisdictions and time.[3]

The coverage of the dataset is extensive but not exhaustive. It is strongest for listed firms and large private operators that disclose property-level information through regulatory filings or

[3]This data has been extensively used in real estate economics and finance research (see, e.g., Ling et al. (2021,?)).

investor communications. Coverage is therefore biased toward jurisdictions with robust financial disclosure requirements and toward larger commercial or institutional properties, including hyperscale data centres owned by public or major private firms. The dataset also records facility type, distinguishing general-purpose data centres from, for example, cryptocurrency mining facilities. Facilities owned by smaller domestic operators or entities not subject to public reporting, such as municipal or sovereign facilities, small colocation providers, or data centres embedded in enterprise campuses, may be under-represented or missing.

A related concern is that GDPR adoption might itself improve a country's financial disclosure environment, inflating S&P counts. Crypto facilities sit in the same S&P dataset and face identical reporting requirements but fall outside the scope of personal-data law. They show no response to tightening (Section 4.4).

We complement the S&P data with the 451 Research Datacenter KnowledgeBase (S&P Global Market Intelligence, 2025a), a commercial database also available from S&P Global. It records installed UPS capacity in kilowatts for each facility, which we aggregate to megawatts at the country–year level. One downside of the 451 Research data is that it records each facility's most recently observed configuration rather than its capacity at the time of opening. Facilities expanded after entry therefore carry their current capacity in the entry-year observation, biasing the measure upward for older installations.

Both databases are thus oriented towards more formal, capital-intensive segments of the market, while omitting smaller-scale and informal capacity. If investments in the latter respond similarly, or more strongly, to restrictions on cross-border data flows, our estimates would likely understate the total impact on domestic data infrastructure investment.

2.3 Additional country-level data

We use additional country-level variables to explore heterogeneity in the effects of laws regulating data transfers. These variables capture differences in market size, economic development, industry composition, and enforcement capacity that may have an impact on the investment response to the regulatory shock.

Population and GDP are obtained from the World Bank World Development Indicators (World Bank, 2024a), using series SP.POP.TOTL (total population) and NY.GDP.MKTP.KD (real GDP). To assess sensitivity to cross-country differences in enforcement capacity, we also draw on two indicators from the World Bank Worldwide Governance Indicators (World Bank, 2024b; Kaufmann et al., 2011), a separate dataset from the same institution: Rule of Law (RL.EST) and Regulatory Quality (RQ.EST). Rule of Law captures legal institutions: contract enforcement, property rights, quality of courts and police, likelihood of crime, and general strength of legal institutions. Regulatory Quality captures the government's perceived ability to formulate and implement sound policies.

To proxy differences in industry composition between countries, we use the "downstreamness" indicator by Mancini et al. (2024). The indicator builds on a theoretical framework from Antràs et al. (2012), in which downstreamness measures how close to final demand an industry is. Intuitively, the measure captures an average of how many production stages lay between an industry's product and end-user demand. A higher value indicates that an industry's production is closer to end users, while lower values correspond to upstream, intermediate-goods production. The dataset provides industry-level scores, which we aggregate to the country level by summing across all sectors. We assume that industries closer to end users produce and handle more personal data — including granular customer transaction data — compared to upstream industries.

In addition, we assemble two country-level infrastructure measures that will be used later to test cost- and connectivity-related predictions from our theoretical framework. First, we use total electricity generation (TWh) from the U.S. Energy Information Administration's international electricity statistics (U.S. Energy Information Administration, 2024). Second, we use international bandwidth from the International Telecommunication Union (International Telecommunication Union, 2024).

Population, GDP, downstreamness, and the two infrastructure measures are incorporated as dichotomous high/low indicators, split at the global median. We use binary splits rather than continuous measures because of the relatively small sample of countries. All values are taken from 2010, the starting year of our sample, so these characteristics can be treated as exogenous with respect to subsequent law changes and data centre entry, reducing reverse causality concerns. Rule of Law and Regulatory Quality are entered as continuous, time-varying controls.

2.4 Descriptive Statistics

Descriptive statistics for the dependent variable — new data-centre sites per country–year — are shown in Table 2. The distribution is heavily right-skewed with many zeros: 68% of country–years record no new openings, while the top 1% reaches 56 and the maximum is 189. Treated country–years, those in which a GDPR-level or stricter regime is in force, have fewer zeros (59% vs. 71%) and higher mean entry (2.78 vs. 1.93), with a heavier upper tail.

Table 3 reports 2010 baseline characteristics separately for ever-treated and never-treated countries. Ever-treated countries are larger and more connected: higher GDPs, larger populations, and substantially higher international bandwidth, consistent with regulation being adopted earlier in bigger and more globally integrated economies. Downstreamness is similar across groups. The cumulative site counts in Figure 2 show that ever-treated countries also accumulate more sites in absolute terms over the sample period. Table 4 examines whether this reflects selection on observables, regressing an indicator for eventual treatment on the 2010 baseline characteristics. Coefficients are small and largely insignificant, and the model accounts for little of the cross-country variation in adoption (Adj. $R^2 = 0.034$), suggesting limited selection on observables.

Table 5 reports correlations among the binary baseline split indicators used for the heterogeneity analysis. Correlations are mostly low to moderate, so the splits capture different underlying dimensions rather than the same one. The clearest exception is GDP and the infrastructure measures: richer economies tend to have both greater electricity supply and higher connectivity.

Table 2. Distribution of new data-centre site openings by treatment status

Treated	N	Share zero	Mean	Std. dev.	Mean | count > 0	P50	P75	P90	P95	P99	Max
Not yet or never treated (Treated = 0)	1846	0.71	1.93	8.30	6.61	0	1	4	8	45	141
Treated country-years (Treated = 1)	682	0.59	2.78	13.46	6.86	0	2	5	9	83	189
All country-years	2528	0.68	2.16	9.96	6.69	0	1	4	8	56	189

Notes: The table reports the distribution of the annual count of newly opened data-centre sites by country-year. The sample is a country-year panel covering 158 countries over 2010–2024. Source: S&P Global Capital IQ Pro Real Estate. Treated = 0 denotes country-years that are not yet treated or never treated. The column "Mean | count > 0" reports the mean conditional on at least one new site opening in the country-year.

Table 3. Baseline (2010) country characteristics by ever-treated status. Cells report mean (SD).

Variable	N (Never)	Mean (SD) Never	N (Ever)	Mean (SD) Ever	Diff. means
Downstreamness (2010)	83	2.08 (0.32)	71	2.13 (0.44)	0.05
Population (2010, mn)	87	41.06 (139.00)	71	43.67 (161.87)	2.60
GDP (2010 USD, bn)	86	320.85 (1761.51)	70	515.68 (1178.87)	194.83
Electricity generation (2010, TWh)	86	98.81 (457.96)	68	175.64 (539.65)	76.83
International bandwidth (2010, Gbit/s)	86	254.13 (1092.33)	69	482.37 (1049.61)	228.24

Notes: The table reports baseline country characteristics measured in 2010. Population and GDP are from the World Bank World Development Indicators. Electricity generation is from the U.S. Energy Information Administration. International bandwidth is from the ITU. Downstreamness is from Mancini et al. (2024). See the main text for further details on variable construction and sample coverage.

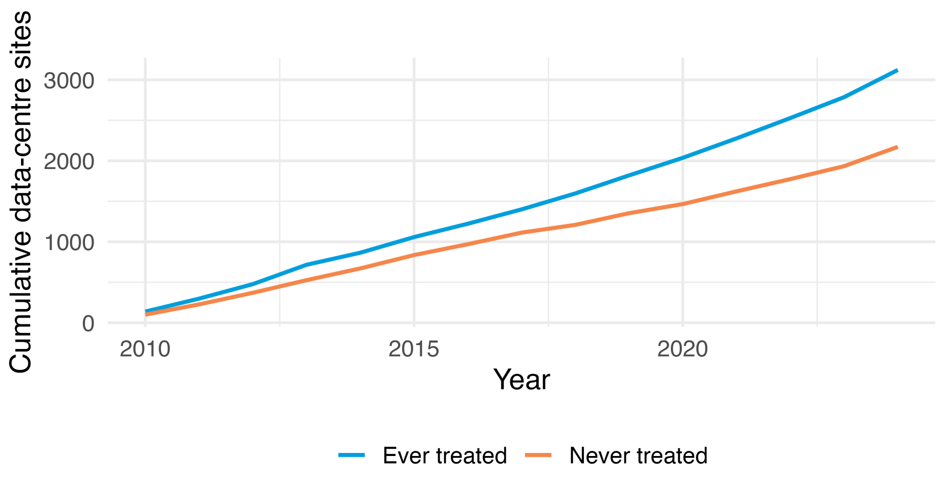


Figure 2. Cumulative number of data centre sites in ever-treated and never-treated countries.

Note: Cumulative sum of new data centre site openings, separately for ever-treated and never-treated countries. Ever-treated countries are those that adopt a data transfer regime rated level 4 or above at any point during 2010–2024; all other countries are classified as never-treated. Data from S&P Global Capital IQ Pro Real Estate. Sample: 158 countries, 2010–2024.

Table 4. Selection on observables: predictors of treatment adoption

	Coefficient
Intercept	-1.073*
	(0.513)
Log GDP (2010)	0.052.
	(0.027)
Log population (2010)	0.002
	(0.027)
Downstreamness (2010)	0.039
	(0.110)
N	152
Adj. R^2	0.034
F-statistic	2.943 (df = 3, 148; p = 0.035)

Notes: OLS regression. Dependent variable is an indicator equal to one if the country adopted a data-transfer restriction of category 4 or above for the first time in 2011 or later. Regressors are measured in 2010 from World Bank WDI (GDP, population) and Mancini et al. (2024) (downstreamness). Countries with *Lawcat* ≥ 4 prior to 2011 or never treated are coded zero. Heteroskedasticity-robust standard errors in parentheses. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.10$.

3 Estimation method and the empirical model

We estimate the standard two-way fixed effects model

$$\Delta Y_{c,t} = \alpha_c + \eta_t + \beta D_{c,t}^{\text{Tighten}} + \varepsilon_{c,t}, \tag{1}$$

where countries are indexed by c and years by t. $\Delta Y_{c,t}$ is the net change in the number of data centres in country c between years t and $t-1$, capturing both new openings and closures. $D_{c,t}^{\text{Tighten}}$ equals 1 for all years $t \geq G_c$, where G_c denotes the first year country c's law equals or exceeds level 4, and 0 otherwise (see Section 2 for the rationale behind this threshold). We refer to G_c as the *tightening year* and to country–years with $t \geq G_c$ as *treated*.

The global, staggered introduction of data transfer laws provides a useful source of quasi-experimental variation for studying their effects on local data centre investment. From the point of view of individual firms, the timing of a country's data transfer law can be treated as plausibly exogenous, since firms do not determine regulatory timing. At the same time, country-level adoption may still correlate with broader trends in digitalisation or economic development. A two-way fixed effects event-study framework helps account for time-invariant country differences and common global shocks, allowing us to isolate the within-country changes associated with data transfer laws.

Identification relies on several standard assumptions. The key requirement is that, absent the adoption of data transfer laws, treated and not-yet-treated or never-treated countries would have followed parallel trends in data centre entry, conditional on time and country fixed effects. In other words, the timing of law adoption should be unrelated to unobserved country-specific shocks that also affect investment. In addition, the Stable Unit Treatment Value Assumption

Table 5. Correlation matrix between the heterogeneity splits

	Downst. < med.	Pop. < med.	GDP < med.	Electricity < med.	Bandwidth < med.
Downst. < med.	1.00	−0.21	−0.21	−0.05	0.05
Pop. < med.		1.00	0.42	0.30	0.10
GDP < med.			1.00	0.73	0.63
Electricity < med.				1.00	0.60
Bandwidth < med.					1.00

Notes: The table reports pairwise correlations between binary indicators for whether a country is below the sample median of each baseline characteristic. The sample contains 158 countries, with baseline values measured in 2010.

(SUTVA) should hold. This requires that a data transfer law in one country does not itself deter or induce data centre entry in other countries.

While both of these assumptions are ultimately untestable in an observational setting, we report pre-trend and placebo estimates that support them.

Another key challenge with staggered adoption designs is that the standard two-way fixed effects models compare units with different treatment timings, sometimes treating already-treated units as controls for those treated later. To address this, we follow Sun and Abraham (2021), who propose an event-study estimator that reweights group-time average treatment effects to avoid such "forbidden" comparisons.[4] In this framework, countries that have not yet adopted a data transfer law serve as valid controls for those that have, while already-treated countries are excluded from the control group once they enter treatment.

The Sun–Abraham estimator decomposes the event-study coefficients into cohort-time average treatment effects. Let G_c denote country c's first year of treatment (the "cohort"), and let the outcome be the first difference ΔY_{ct}:

$$ATT_{g,t} = \mathbb{E}[\, \Delta Y_{gt}(1) - \Delta Y_{gt}(0) \,|\, G_c = g] \,. \tag{2}$$

Each $ATT_{g,t}$ is identified from never-treated and not-yet-treated countries; already-treated countries do not enter the control group. Event-time coefficients β_k are cohort-size-weighted averages of $ATT_{g,\,g+k}$, and the post-treatment summary coefficient β is a cohort-size-weighted mean of the β_k. This eliminates the contaminated comparisons to earlier-treated units in standard TWFE.

4 Results

4.1 Main results

We start with average post-treatment effects of Equation (1) estimated with the Sun–Abraham weights described above, summarised in Table 6. In the full sample, the tightening of the law regulating transfers of personal data is followed by roughly two-thirds of an additional data centre per year relative to the pre-period. Relative to the pre-treatment mean, this implies an increase of about 21%.

In the second and third columns, we split the sample by industries' position in the value chain. Upstream industries show no measurable effect. For downstream industries (those closer to end-users), the estimate is about 1.3 additional data centres per year, roughly 33% above the pre-treatment mean, significant at the 5% level. The pattern is consistent with more customer-facing sectors being more exposed to regulatory frictions in cross-border data flows.

[4] An alternative is to pool all observations and interact event-time indicators with the moderator, which would give a formal test of heterogeneity. The cost is that year fixed effects must be common across groups — a non-trivial assumption, since GDP, population, and downstreamness all predict the level and cyclicality of digital investment, so the groups likely follow different aggregate trends. Pooled year effects would absorb these jointly, biasing the interaction estimates in an unpredictable direction. Sample splitting lets year fixed effects vary by group, which avoids this at the cost of a formal heterogeneity test. Sun and Abraham's estimator also does not support binary moderator interactions directly; accommodating them would require reverting to standard TWFE and losing the heterogeneity-robust weighting.

In columns 4 and 5, we repeat the analysis using data split into two groups by population size. We find that the effect is statistically indistinguishable from zero in the low-population group of countries. For the high-population group, the point estimate is positive at about 1.1 additional data centres per year, amounting to a 24% increase over the group-specific pre-treatment mean, but it is not statistically significant at conventional levels.

Finally, in columns 6 and 7, we split the data by country GDP. Countries below the median show no measurable effect. For countries above the median, the estimate is roughly one additional data centre annually (around 22% relative to their pre-treatment mean), significant at the 10% level.

Of the three heterogeneity splits, downstream yields the sharpest and most precisely estimated effect. The GDP split is consistently positive but only clears the 10% threshold; the population split points in the same direction but is statistically indistinguishable from zero. That GDP matters and population does not is itself informative: it is economic market size (spending power, not headcount) that appears to drive the secondary pattern. The differences between high and low subgroup ATTs are themselves statistically significant: the gap between high- and low-downstreamness countries has a p-value of 0.06, and the gap between high- and low-GDP countries has a p-value of 0.09, both significant at the 10% level.

The subgroup comparisons also serve an identification role: the subgroups are arguably more homogeneous in unobservables (e.g. only high-GDP countries), and therefore treated and control units differ less on unobservable characteristics than they would do in the full sample, which makes the parallel trends assumption more defensible.

Table 6. Effect of cross-border data transfer restriction on new data centre buildup

	Full sample	Upstream	Downstream	Low population	High population	Low GDP	High GDP
Estimate	0.680	0.075	1.325*	0.143	1.123	0.005	1.045.
Std. error	(0.965)	(0.292)	(0.664)	(0.159)	(1.494)	(0.044)	(0.578)
N countries	158	77	77	79	79	78	78
N treated	71	33	38	33	38	26	44
N observations	2528	1232	1232	1264	1264	1248	1248
Treated pre mean	3.185	2.179	3.982	0.704	4.702	0.039	4.805

Notes: This table presents average post-treatment effects estimated using the Sun and Abraham (2021) staggered difference-in-differences estimator. The outcome is the annual count of new data centre sites opening in a country. Treatment is defined as the first year a country adopts a data transfer regime rated level 4 or above (at least as restrictive as the GDPR). All specifications include country and year fixed effects. Subsample splits use the cross-sectional median of downstreamness (Mancini et al., 2024), 2010 population, and 2010 GDP. "Treated pre mean" is the mean of the outcome in treated countries over their pre-treatment years. Sample sizes vary across splits due to missing data. Standard errors are clustered at the country level and reported in parentheses. . $p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$.

4.2 Event study analysis

One potential concern for our event-study analysis is that it might be driven by pre-tightening trends. These could be driven by policy anticipation: if governments announce laws regulating data transfers well in advance of when they come into force, operators may adjust their investment in anticipation of the policy, producing an increase in data centre entry prior to the tightening. Moreover, the policy timing itself could be endogenous: countries experiencing a data centre boom for reasons unrelated to policy might be more likely to tighten laws regulating data transfers, which would generate a correlation between pre-treatment growth and policy choices. Both of these factors would threaten the identification strategy, which relies on measuring a weighted difference between not-yet treated and treated units' time trajectories.

The event study plots (Figure 3) speak against both concerns. Pre-tightening trends in data centre openings track closely between treated and control countries: across all seven specifications, pre-treatment coefficients sit near zero, with only a handful differing from zero at the 5% level.

Flat pre-treatment coefficients, however, do not rule out all forms of endogenous policy timing, including cases in which governments tighten rules in anticipation of a discrete investment increase or cases of very short-run anticipation. They do, however, argue against the empirically more visible alternatives: a gradual pre-existing divergence in data-centre investment or systematic investment responses several years before the tightening date.

The post-period coefficients are noisier, as one would expect given the high variance of the dependent variable. One pattern holds across specifications: the downstream split shows a positive post-treatment effect throughout. The GDP split is less consistent but is suggestively positive in most specifications.

4.3 Evidence against spillovers across countries

Another threat to identification is spillovers. If a policy in country c reduces investment in other countries, this would constitute a spillover from country c to those countries, which would in turn violate the Stable Unit Treatment Value Assumption (SUTVA) and invalidate the research design.

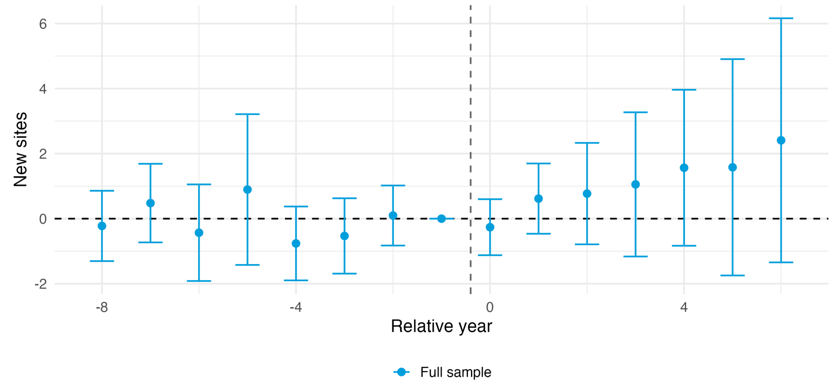
A simple way to assess whether such spillovers are present is to ask, whether the neighbouring countries' outcomes systematically move when country c tightens its data transfer laws. The logic is the following: if the SUTVA holds, the potential outcomes of country $j \neq c$ must be unaffected by the treatment status of country c. This test is intentionally local: it is designed to detect displacement or complementarities to geographically proximate countries.[5] It does not test for reallocation to distant locations outside the 'nearby' set (such as reallocation between the EU and the US).

Unlike the standard approaches that capture spillovers by aggregating outcomes across neighbouring units (e.g. Di Tella and Schargrodsky (2004); Jardim et al. (2024)), we isolate average spillovers by constructing a leave-one-out outcome. Specifically, we define:

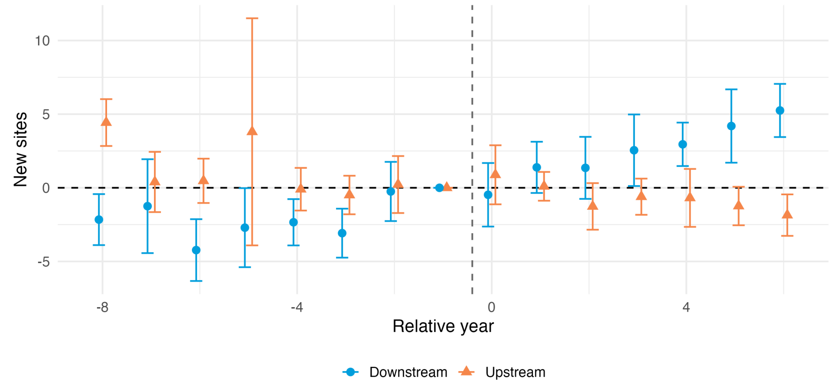
$$\Delta Y_{c,t}^{\text{near}} = \sum_{j \in r(c), j \neq c} \Delta Y_{j,t}, \tag{3}$$

where $r(c)$ denotes the region containing country c. If law tightening in c diverts investments away from nearby countries (or triggers complementary investments in them), the sum of neighbours' outcomes will exhibit a systematic pattern around country c's treatment date. If SUTVA holds, $\Delta Y_{c,t}^{\text{near}}$ should remain approximately flat apart from noise.

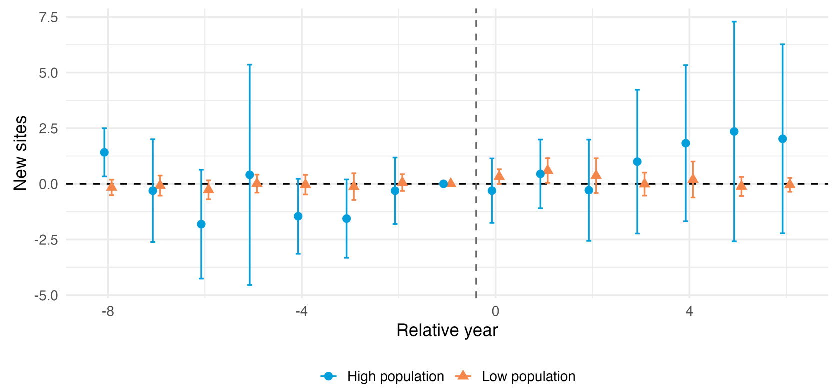
[5] Intra-regional connectivity between neighbouring countries is generally not a binding constraint on investment relocation (Dang et al., 2021).



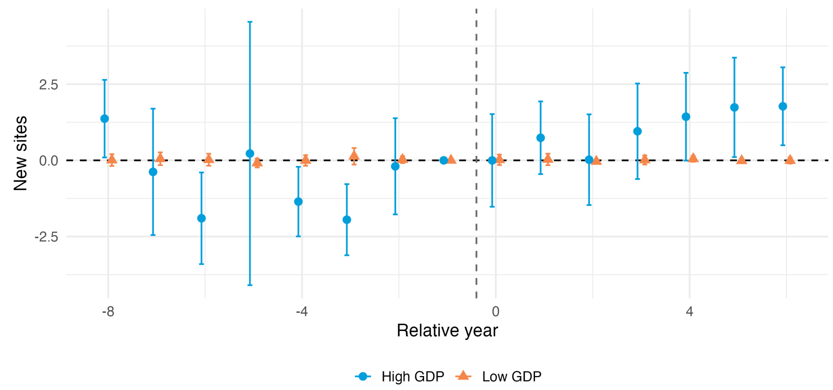
(a) Event study: full sample



(b) Event study: Downstreamness split



(c) Event study: population split



(d) Event study: GDP split

Figure 3. Event-study estimates of the effect of regulatory tightening on domestic data centre entry.

Note: Each panel shows Sun and Abraham (2021) event-study estimates using not-yet-treated and never-treated countries as controls. The outcome is the annual count of new data centre site openings per country. Event time 0 is the first year a country adopts a data transfer regime rated level 4 or above (GDPR-equivalent or stricter). Error bars are 95% confidence intervals based on standard errors clustered at the country level. High/low splits are defined by the sample median of the 2010 value of the splitting variable: downstreamness (Mancini et al., 2024), total population, and GDP (World Bank WDI). Sample: 158 countries, 2010–2024.

We then estimate the Sun–Abraham event study model outlined in Equation (2) using $\Delta Y_{c,t}^{\text{near}}$ as the dependent variable. This produces an average contamination profile: if the coefficients are approximately zero pre- and post-treatment, this is evidence against contamination.

We define "nearby" countries as countries that are in the same region $r(c)$ according to the 22-region classification used by the World Bank and others.[6]

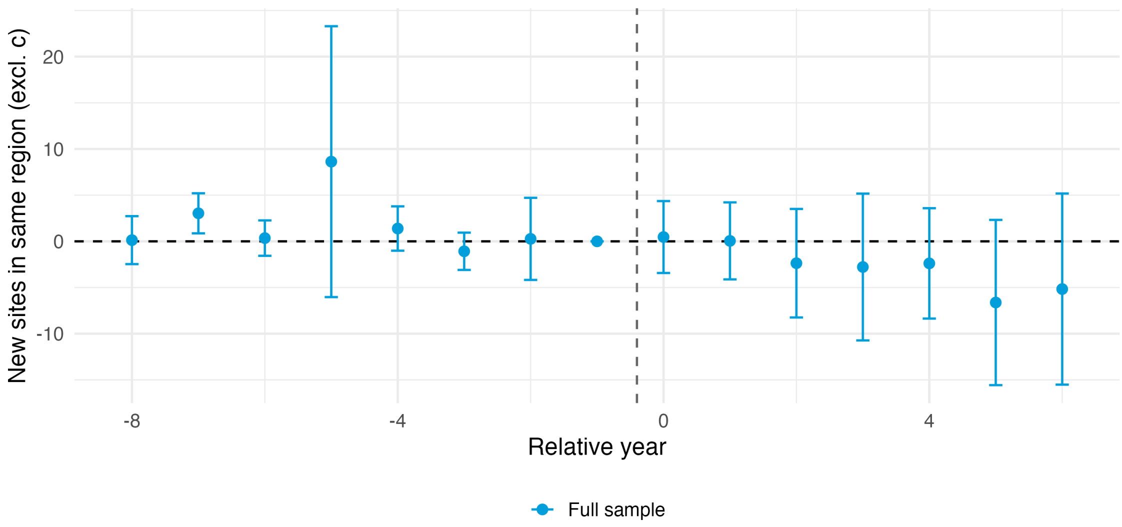
By construction, a flat $\Delta Y_{c,t}^{\text{near}}$ rules out systematic within-region spillovers. It remains silent on possible reallocation toward countries outside of $r(c)$. However, such distant displacement is unlikely in this context. Data centre placement is strongly constrained by latency and physical proximity to end users. Long-distance reallocation (e.g., from Europe to North America) would violate practical latency limitations and is not supported by observed cloud infrastructure practices.[7]

We run two versions of the spillover test: one using the full sample and one with data winsorised at the 95th percentile. Results are shown in Figure 4.

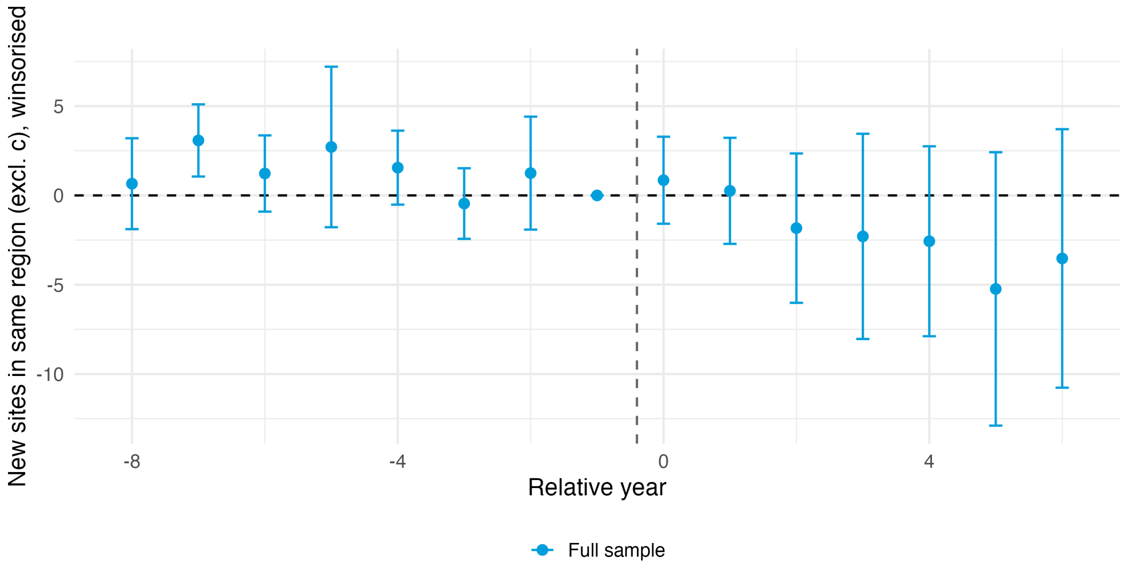
The estimated coefficients are flat and centered around zero both before and after the law change, with no evidence of systematic within-region spillovers. This supports SUTVA against local contamination. The regional design is also directly informative about intra-EU displacement. If tightening in Germany diverted investment to Ireland or Finland, those countries are in the same World Bank region, so their additional entry would appear as a positive post-treatment coefficient in the leave-one-out sum. The flat regional estimates provide no evidence of systematic within-region displacement. Given latency constraints, they also reduce concern over spillovers to more distant regions.

[6] We use the `countrycode` R package (Arel-Bundock et al., 2018); regions follow the UN M.49 sub-region scheme (e.g., Southern Europe, Eastern Asia).

[7] Cloud services are typically designed in a way that latency sensitive applications are served from infrastructure that is physically close to the user. This matters empirically: median end-to-end latencies across continents exceed 80–100 ms, while intra-regional traffic stays below 20 ms, which is critical for real-time performance (Corneo et al., 2021; Dang et al., 2021).



(a) Test for regional treatment spillover.



(b) Test for regional treatment spillover (winsorised sample).

Figure 4. Regional spillover tests.

Note: Each panel shows Sun and Abraham (2021) event-study estimates where the outcome is the leave-one-out sum of data centre openings across all countries in the same World Bank region as the treated country, excluding the treated country itself. Flat coefficients indicate no systematic within-region spillovers, supporting SUTVA. The lower panel uses the outcome winsorised at the 95th percentile at the country-year level. Standard errors are clustered at the country level; error bars are 95% confidence intervals. Sample: 158 countries, 2010–2024.

4.4 Falsification: crypto data centres

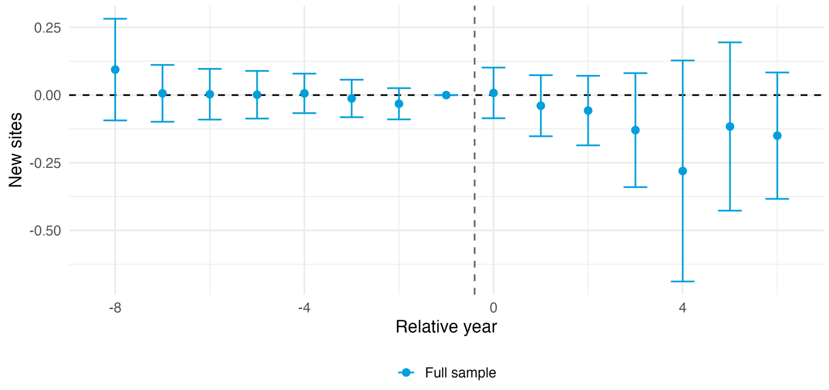
As an additional falsification check, we show that laws regulating data transfers have no statistically discernible effect on crypto data centre entry, identified using the facility-type field in the same S&P Global dataset. Cryptocurrency mining is structurally different from most other data processing as it does not rely on personal data. Consequently, crypto data centres do not face compliance burdens under the legal regimes we study. However, crypto data centre investment might still respond to investment incentives, internet infrastructure spending, and other policies aimed at facilitating digitalisation and attracting data centre investments. If the main estimates were primarily driven by such broader policy trends rather than by compliance to data transfer laws, we would expect to see a similar pattern of increased crypto data centre entry following the legal tightening. This makes the crypto data centre entry a useful placebo test for the mechanism we propose.

We report the results of this exercise in Figure 5. Cryptocurrency data centres' location decisions primarily depend on energy and cooling costs rather than data transfer legislation. If our estimates primarily reflected a broader trend in data centre construction rather than a compliance response to law tightening, we would expect similar patterns for crypto facilities. Finding no effect suggests that our main estimates reflect compliance with regulation rather than picking up a general data centre entry trend. The null result also bears on the financial disclosure concern raised in Section 2: crypto facilities are drawn from the same S&P dataset and are subject to the same financial reporting requirements as general-purpose data centres. If GDPR-level regulation improved financial disclosure in treated countries and thereby inflated site counts, crypto facilities would show the same pattern. They do not.

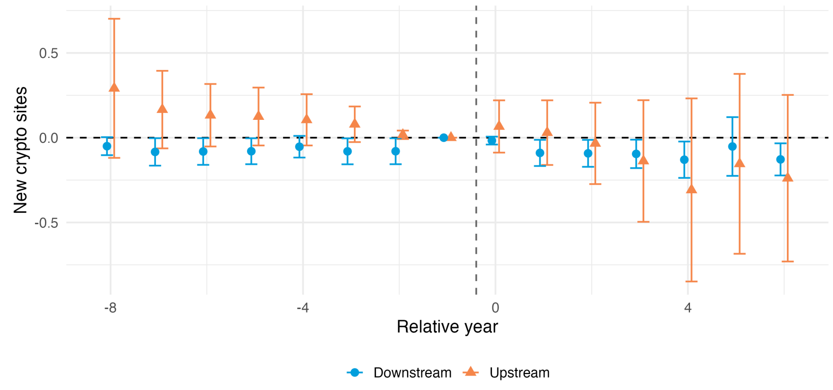
4.5 Sensitivity of the results to enforcement heterogeneity

Our treatment variable captures what laws require, not how vigorously states enforce them. Countries with weaker institutional capacity may adopt strict transfer rules without consistent application, attenuating our estimates toward zero. We re-estimate the main specifications adding the World Bank Rule of Law and Regulatory Quality as time-varying controls, separately and jointly.

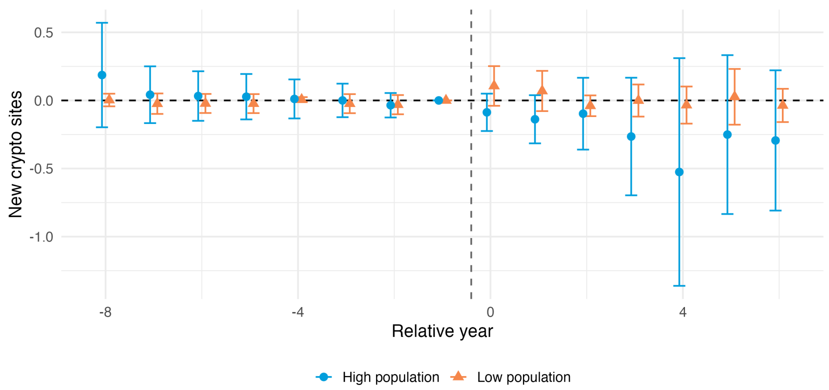
Figure 6 shows four sets of estimates: baseline with no governance controls (blue), adding Regulatory Quality (orange), adding Rule of Law (light purple), and adding both (sage green). The estimates are virtually identical across all four specifications. Cross-country differences in institutional quality do not account for the main results.



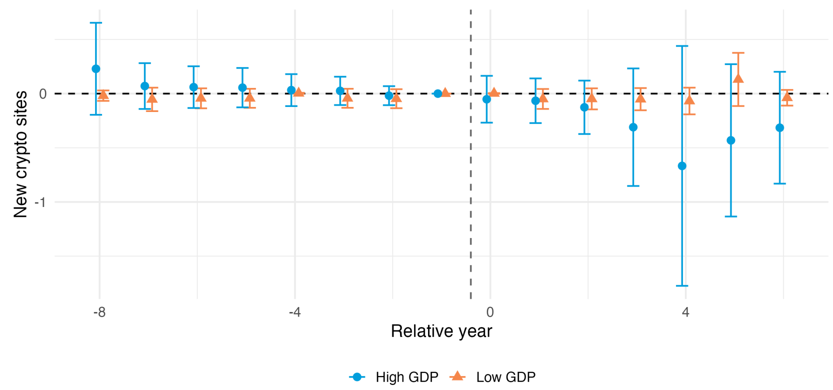
(a) Event study: full sample



(b) Event study: downstreamness split



(c) Event study: population split



(d) Event study: GDP split

Figure 5. Falsification test: impact of data transfer restrictions on cryptocurrency data centre sites.

Notes: Replication of the main event-study specification using cryptocurrency mining data centre openings as the outcome. Cryptocurrency mining facilities are identified via the facility-type field in S&P Global Capital IQ Pro Real Estate (see Section 2). Cryptocurrency facilities are not subject to personal data compliance requirements and should not respond to data transfer regulation if the compliance mechanism drives the main results. The estimator is Sun and Abraham (2021); standard errors are clustered at the country level; error bars are 95% confidence intervals. High/low splits defined as in Figure 3. Sample: 158 countries, 2010–2024.

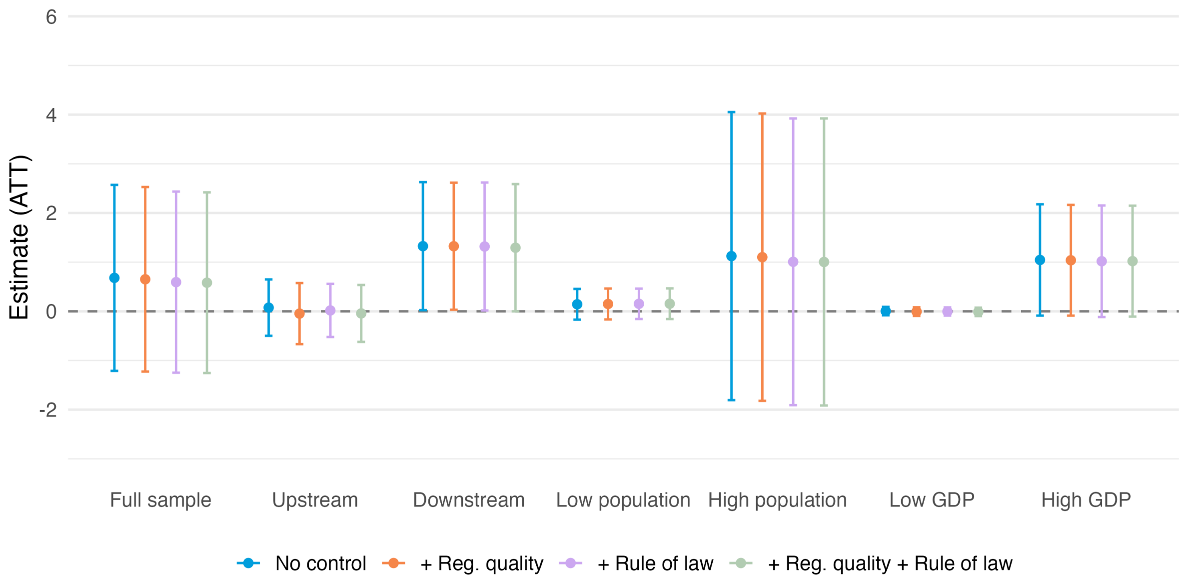


Figure 6. Sensitivity of results to controlling for institutional quality. Each panel shows the cohort-averaged ATT from the Sun and Abraham (2021) estimator under four specifications: baseline with no governance controls (blue), adding Regulatory Quality (orange), adding Rule of Law (light purple), and adding both (sage green). All indicators are from the World Bank Worldwide Governance Indicators (Kaufmann et al., 2011) and enter as time-varying controls. Error bars are 95% confidence intervals based on standard errors clustered at the country level. Subgroup splits defined as in Table 6. Sample: 158 countries, 2010–2024.

4.6 Sensitivity of the results to the treatment threshold

Our baseline treatment indicator is defined at level 4 or above. The rationale is discussed in Section 2, but the estimates could be sensitive to this choice. We therefore re-estimate the main specifications at two alternative thresholds: level 3 or above, which captures weaker conditional regimes, and level 5 or above, which covers only the most restrictive frameworks.

Figure 7 shows the results. Two features stand out. First, the sign of the estimated effects is consistent across all three thresholds in every subgroup: groups that respond positively at the baseline (downstream industries, high-GDP, high-population countries) do so at the alternative thresholds as well, while non-responding groups (upstream, low-GDP, low-population) remain near zero. Second, point estimates rise with regulatory restrictiveness—from ≥ 3 to ≥ 4 to ≥ 5—in the subgroups where an effect is present. Stricter regimes expand the set of data and use-cases for which cross-border processing is infeasible, forcing a larger share of compute onshore.

The ≥ 5 estimates should be interpreted with caution, as few cohorts are treated at this threshold (a ≥ 6 cutoff would not be identified) and the Sun–Abraham specification is costly in terms of degrees of freedom. Nevertheless, comparisons across thresholds are consistent with a dose-response gradient, with tighter regulation associated with stronger effects.

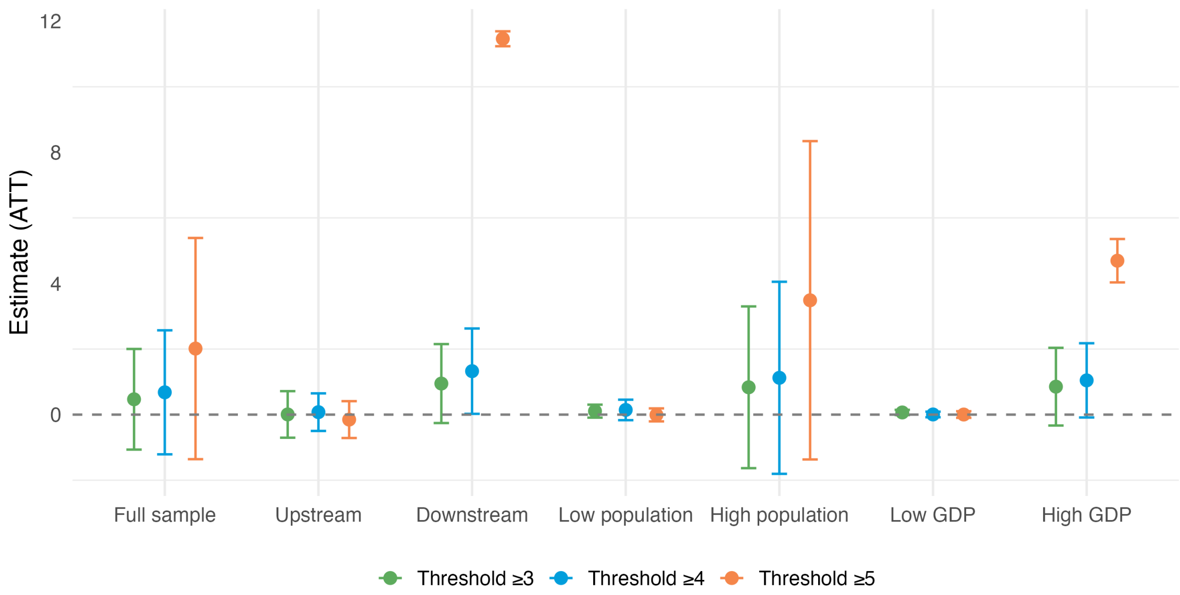


Figure 7. Sensitivity of results to the treatment threshold. Each group of dots shows the cohort-averaged ATT under three alternative treatment definitions: level ≥ 3 (green), level ≥ 4 (blue, baseline), and level ≥ 5 (orange). Estimates use the Sun and Abraham (2021) estimator; error bars are 95% confidence intervals based on standard errors clustered at the country level. The ≥ 5 confidence intervals are likely understated due to the small number of treated cohorts at that threshold. Subgroup splits defined as in Table 6. Sample: 158 countries, 2010–2024.

4.7 Capacity-based outcomes

The main outcome counts facility openings but says nothing about facility size. To check whether the estimated effects are driven by large or small data centres, we replace the site-count outcome with annual new UPS capacity in megawatts, drawn from the 451 Research Datacenter KnowledgeBase (see Section 2).

Table 7 reports the results. The full-sample estimate is 21.5 MW and imprecisely estimated. The downstream estimate is 78.7 MW, significant at the 0.1% level, against a pre-treatment mean of 17.3 MW. This mirrors the site-count pattern: the effect is concentrated in sectors close to final demand, where personal-data obligations are most binding, and it is large relative to the baseline. The estimate for the high-GDP group is positive but not statistically significant at conventional levels.

Pre-treatment event-study coefficients for the MW outcome are reported in Online Appendix D; they are flat and close to zero across all specifications, supporting the parallel trends assumption for this measure.

4.8 Sensitivity of the results to outliers

One potential concern is that the results are based on conditional means of right-skewed count outcomes. Given the distribution of the outcome variables, it is at least theoretically possible that the estimation results are driven by a small number of extreme observations. To demon-

Table 7. Effect of cross-border data transfer restriction on new data-centre capacity (MW)

	Full sample	Upstream	Downstream	Low population	High population	Low GDP	High GDP
Estimate	21.508	-27.698	78.731***	6.160.	22.719	-0.101	22.176
Std. error	(32.321)	(32.382)	(6.315)	(3.245)	(56.375)	(0.082)	(39.226)
N countries	158	77	77	79	79	78	78
N treated	71	33	38	33	38	26	44
N observations	2528	1232	1232	1264	1264	1248	1248
Treated pre mean	13.872	9.554	17.295	2.780	20.657	0.300	20.862

Notes: Sun–Abraham Sun and Abraham (2021) event-study estimator. Outcome is annual new UPS power capacity in megawatts (Total UPS Power / 1,000), summed over all facilities whose entry year equals t. Source: 451 Research Datacenter KnowledgeBase. Sample period 2010–2024. All specifications include country and year fixed effects. Standard errors (in parentheses) are clustered by country. "Treated pre mean" is the mean of the outcome for treated countries in the pre-treatment window. Columns split the sample at the cross-sectional median of: downstreamness Mancini et al. (2024); and population and GDP (2010 values). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.10$.

strate that this is not the case, we rerun our estimates with the country-level outcome winsorised at the 95% level.

The results are presented in Fig. 8. A comparison between the baseline estimates (in blue) and the winsorised estimates (in orange) reveals that the two sets of estimates are both quantitatively and qualitatively similar. This indicates that the main results are not driven by outliers. The only exception is the full-sample estimate, where winsorising reduces the standard error while leaving the point estimate largely unchanged, suggesting that a small number of extreme observations mainly affect precision rather than identification.[8]

4.9 Sensitivity of the results to functional form

Since our outcome variable is the annual count of newly opened data centres, the linear event-study specification implicitly assumes parallel trends in the conditional mean of the outcome. For non-negative count outcomes with many zeros, this assumption may be restrictive. As a robustness check, we therefore estimate a Poisson quasi-maximum likelihood model. Under the exponential mean specification, the identifying restriction is that parallel trends hold in the log of the conditional mean rather than in levels Wooldridge (2023).

We follow the staggered-adoption estimator proposed by Wooldridge (2023), which implements an extended two-way fixed effects (ETWFE) specification with cohort-by-relative time interactions in the Poisson conditional mean. This recovers cohort–time average treatment effects without imposing a linear conditional mean for the outcome.

We implement the estimator using the etwfe R package (McDermott, 2026). We report the results in Fig. 8.[9] Across subgroups, the ETWFE–Poisson estimates are close to the corresponding Sun–Abraham estimates, with somewhat tighter standard errors in the downstream and high-GDP subgroups. The qualitative conclusions remain unchanged, however.

[8]Our six subsample comparisons also raise a multiple-testing concern. Online Appendix E reports Romano–Wolf (Romano and Wolf, 2005) stepdown-adjusted p-values. In the raw specification, the downstream and high-GDP splits no longer reach conventional significance after correction ($p_{\mathrm{RW}} = 0.18$ and 0.32). In the winsorised specification, both survive the correction ($p_{\mathrm{RW}} = 0.001$ and 0.007), consistent with the interpretation that outliers inflate noise rather than obscure a substantive effect.

[9]The ETWFE specification does not recover pre-period coefficients (McDermott, 2026), so we cannot inspect pre-trends for this estimator; pre-trend evidence comes from the Sun–Abraham event-study estimates in Figure 3.

4.10 Sensitivity of the results to estimator choice

To demonstrate that the results are not driven by the choice of estimator, we re-estimate the model using the estimator proposed by Callaway and Sant'Anna (2021). This estimator provides an alternative approach to Sun and Abraham (2021) for estimating treatment effects in TWFE settings with staggered treatment adoption.

The resulting estimates are also reported in Fig. 8. Across all subsamples, the Callaway–Sant'Anna estimates (in sage green) are close to the corresponding Sun–Abraham estimates in terms of point estimates, but have systematically larger confidence intervals in several cases.

This pattern is consistent with the small-sample properties of the two estimators. In our setting, sample sizes and treated cohorts are mostly small, which reduces the effective sample size available. The Sun–Abraham estimator often achieves higher efficiency, while delivering estimates that are quite similar to those obtained from the more flexible Callaway–Sant'Anna approach, but with smaller standard errors. This builds confidence that our main results are not driven by the estimator choice.

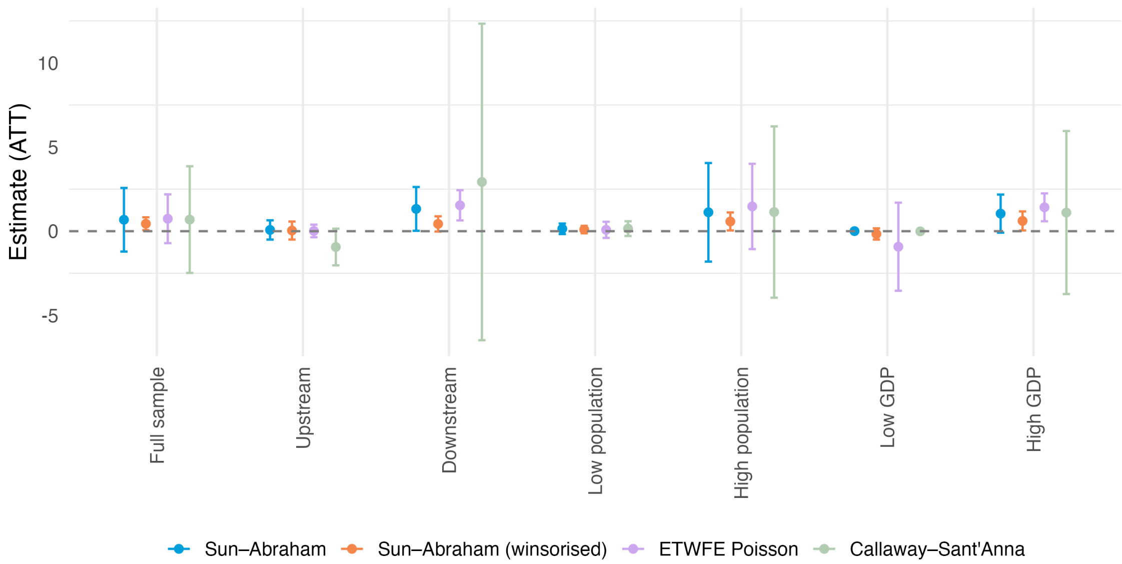


Figure 8. **Sensitivity to outliers and estimator choice.** The outcome variable is the annual count of newly opened data-centre sites in a country–year. The sample is a country–year panel covering 158 countries over 2010–2024. Treatment is defined as the adoption of a cross-border personal data transfer regime of level ≥ 4. Blue: Sun–Abraham baseline; orange: Sun–Abraham with the outcome winsorised at the 95% level; light purple: ETWFE Poisson; sage green: Callaway–Sant'Anna. Standard errors are clustered at the country level. Horizontal bars show 95% confidence intervals.

5 Theoretical Framework

The patterns, stronger effects in downstream and high-GDP economies, documented above motivate a simple model that generates these comparative statics.

The model combines a standard production technology with compute and data as complements and a compliance requirement that forces a minimum domestic share of compute. This setup can reproduce the main empirical patterns from the previous section. It also accounts for the near-zero full-sample average: the model predicts that the investment response scales with market size and data intensity, so the effect is larger in downstream, data-intensive economies and smaller or absent elsewhere. Averaging across a heterogeneous sample of countries with widely varying data intensity and market size naturally produces an attenuated mean effect. The subgroup analyses are therefore directly tied to the theory.

A common approach in macroeconomics and international trade is to represent policy and regulatory frictions as "wedges" that distort firms' choices (Chari et al., 2007; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). In the data-economy context in particular, recent papers such as Demirer et al. (2024) and Chang et al. (2023) model regulation as a wedge that lowers the marginal productivity of computational resources and data. A general implication of this assumption is that such wedges reduce the use of the regulated input and, through complementarity, also reduce the use of its complements. Our estimates point in the opposite direction: in some countries, tighter regulation is followed by more data centre investment.

We take an alternative route. Following Kwon and Chun (2009), we model regulation as a compliance quota. Rather than imposing an explicit cost on cross-border data transfers, we require that a minimum share of compute be performed domestically. Because regulated data must remain within the country, compliance can be interpreted as forcing the associated compute to take place within the country. Under this interpretation, the policy index τ is a reduced-form measure of regulatory scope. Higher τ corresponds to a broader set of data and use-cases for which cross-border processing is infeasible, uncertain, or administratively burdensome enough that firms treat it as effectively non-transferable.

Formally, we assume that a firm's workload consists of many task types, each requiring compute $c(j)$. Regulation identifies a subset of tasks for which the associated data cannot be transferred. The minimum local compute share then equals the compute-weighted share of restricted tasks, which we denote $\omega(\tau)$. In this sense, $\omega(\tau)$ is the share of total compute associated with tasks tied to regulated (effectively non-transferable) data. Next, we demonstrate that this model can theoretically generate the empirical patterns from the previous section.

5.1 Model

A representative firm in country c uses a strictly positive amount of compute C and data D_c to produce a digital good for the local market:

$$Y_c = C^{\alpha} \log(1 + D_c^{\gamma}), \tag{4}$$

where compute can be performed locally (C_L) or abroad in a remote data centre (C_R),

$$C = C_L + C_R. \tag{5}$$

We choose an asymmetric functional form because compute and data play different economic roles. We model compute as a standard rival input with diminishing returns, captured by C^{α}.

We model data differently: the economic value of additional data is primarily informational and is subject to diminishing marginal returns due to redundancy and diminishing marginal benefit in prediction. Accordingly, data enters the production function through the transform $\log(1 + D_c^\gamma)$, which behaves approximately like a power law for small D_c (since $\log(1 + D_c^\gamma) \approx D_c^\gamma$ when D_c^γ is small) but delivers sharply diminishing marginal returns as D_c grows. The multiplicative structure $C^\alpha \log(1 + D_c^\gamma)$ preserves complementarity between compute and data while keeping the model's comparative statics transparent.

We assume $C > 0$, $D_c > 0$, $0 < \alpha < 1$, and $0 < \gamma \leq 1$. Under these restrictions, compute and data are complements (the cross-partial $\frac{\partial^2 Y_c}{\partial C\, \partial D_c}$ is strictly positive), while both inputs exhibit diminishing marginal returns.[10] This specification is motivated by empirical evidence on the economic value of data (Bajari et al., 2019) and related theoretical work (Farboodi and Veldkamp, 2021), and is consistent with the strong diminishing returns suggested by empirical learning curves in machine learning (e.g. Hestness et al., 2017).

The restriction on cross-border data transfers enters the model through a compliance constraint:

$$C_L \geq \omega(\tau)C, \tag{6}$$

which simply states that at least a share $\omega(\tau)$ of compute needs to happen locally. $\dfrac{\partial \omega(\tau)}{\partial \tau} > 0$, and $0 < \omega(\tau) < 1$. We interpret tighter regulation as an increase in τ, which raises the required local compute share.[11]

We assume the following cost structure. Regarding data, we abstract from the costs of generating and maintaining data-generating activity (such as product development and user acquisition). These are treated as fixed costs of operating on the market, and are not affected by cross-border data transfer regulation.[12] At the same time, the firm faces costs for using compute. Remote compute is rented at a constant marginal cost p_R from a cloud service abroad.

$$\text{Cost}_R(C_R) = p_R C_R \tag{7}$$

Local compute however, has to be built (with a positive fixed cost F), and thus the average costs per unit of local compute decrease as C_L grows. Declining average costs of local compute arise exclusively from a fixed cost related to local compute. For simplicity, we abstract from technical economies of scale.

$$\text{Cost}_L(C_L) = F + p_L C_L. \tag{8}$$

[10] Agrawal et al. (2022) note that data can exhibit increasing returns if it helps generate market power. Since firms are price-taking in our model, we abstract from that channel. We also note that the main comparative statics below do not hinge on a particular assumption about the curvature of Y_c with respect to D beyond complementarity and diminishing marginal returns.

[11] The model we outline here is observationally equivalent to a cost-wedge model with high but finite substitutability between local and remote compute. In the latter, tighter regulation raises the effective price of remote compute and firms respond along a smooth interior margin rather than at a corner. The qualitative comparative statics are the same under either specification. We prefer the quota formulation because it has one fewer free parameter (σ) that our country-level data cannot identify.

[12] This simplified treatment matches the standard view that, once collected, data are largely non-rival and cheap to reuse and the economically relevant margin is acquisition rather than per-use pricing Farboodi and Veldkamp (2021).

Combining the production and costs functions yields the profit function:

$$\pi_c = N_c \left(C^\alpha \log(1 + D_c^\gamma) \right) - \mathrm{Cost}_L(C_L) - \mathrm{Cost}_R(C_R), \tag{9}$$

where N_c is the market size in country c.

Proposition 1: Effective unit price under regulatory compliance. The impact of regulation depends on the relative prices of local and remote compute. There are two options:

1. If $p_L \leq p_R$, firms' cost minimising allocation is always $C_L = C$, $C_R = 0$, and changes in τ do not affect firms' choices.

2. $p_L > p_R$, the constraint binds, and the effective unit price of compute can be expressed as

$$p(\omega) = p_R + (p_L - p_R)\omega(\tau), \tag{10}$$

and local and remote compute can be expressed as shares of total compute: $C_L = \omega(\tau)C$, $C_R = (1 - \omega(\tau))C$.[13]

Proof: In Online Appendix F.

According to Proposition 1, regulation only has an effect when $p_L > p_R$, i.e., when remote computing is cheaper than local. This is the empirically relevant case for this paper, so throughout the rest of this section, we only concentrate on Case 2. In this case, regulation forces a share of the remote compute to be local, and, thus increases the effective unit price.

When the regulation bites, the optimal level of total compute $C^*(\tau)$ solves for the following first-order condition, where the marginal revenue equals marginal cost.

$$N_c \alpha (C^*)^{\alpha - 1} \log(1 + D_c^\gamma) = p_R + (p_L - p_R)\omega(\tau), \tag{11}$$

and solving for $C^*(\tau)$ gives the optimal level of compute.

$$C^*(\tau) = \left(\frac{N_c \, \alpha \, \log(1 + D_c^\gamma)}{p_R + (p_L - p_R)\, \omega(\tau)} \right)^{\frac{1}{1-\alpha}}. \tag{12}$$

Proposition 2. Impact of regulation on demand for total compute.
When $p_L > p_R$, tightening regulation decreases the total compute used, or $\dfrac{dC^*}{d\tau} < 0$.

Proof: Optimal C^* is given by Equation 12. Since $\omega'(\tau) > 0$, and $p_L > p_R$, then the denominator rises with τ, hence $C^*(\tau)$ falls.

Proposition 2 indicates that, if regulation increases the price of compute, it also mechanically decreases its demand.

The effect of τ on the demand for local compute C_L, is ambiguous. Proposition 2 indicates that regulation decreases the equilibrium amount of total compute. But at the same time, the

[13] For simplicity, we assume that all costs from regulation accrue to firms via their choice of local and remote compute. It would be possible to include compliance costs such as legal services and audits as an extra per-unit compliance cost term in the cost function. In that case, the effective unit costs could be expressed as $p(\omega(\tau)) = p_R + (p_L - p_R)\omega(\tau) + G(\tau)$, where $G'(\tau) > 0$. This shifts the threshold condition for when tighter regulation increases local compute, but it does not change the basic trade-off in the model. For conceptual clarity, we omit this channel.

regulation pushes the share of local compute up. Which of these effects dominates is given by Proposition 3.

When the regulation binds, the compliance constraint implies that optimal local compute is given by

$$C_L^*(\tau) = \omega(\tau)\, C^*(\tau). \tag{13}$$

Proposition 3. Impact of regulation on local compute.

When the compliance constraint binds, or $p_L > p_R$, $\dfrac{dC_L^*}{d\tau} > 0$ if and only if $p_R\,(1 - \alpha + \alpha\omega(\tau)) > \alpha\omega(\tau)p_L$.

Proof: In Online Appendix G.

Proposition 3 shows that when the price advantage of remote over local compute is sufficiently small, the increase in the required local share dominates the contraction in total compute demand, so tighter regulation raises local compute. Intuitively, regulation can shift the composition of demand toward local compute, which raises demand for domestic capacity even if it reduces the profit of compute users.[14]

Moreover, conditional on regulation increasing local compute, the magnitude of the response scales with market size and data intensity. This is formalised in Proposition 4. Proposition 4 demonstrates that our theoretical framework can match the empirical patterns we document.

Link to empirical measures. The model contains two country-level demand shifters, N_c and $\log(1 + D_c^\gamma)$ that capture the heterogeneity patterns we observed empirically in Section 4.

First, N_c measures the market size. In Section 4, we proxy N_c using either population (market size in persons) or aggregate GDP (market size in USD, i.e. the average income per person multiplied by population size). Under either proxy, we found that the tightening of regulation has a larger impact on C_L for larger market size.

Second, $\log(1 + D_c^\gamma)$ captures the importance of data in the firm's revenues. In the empirical section, we proxied data intensity using the industry-level downstreamness measure introduced in Section 2. In the model, higher downstreamness corresponds to a higher effective data term $\log(1 + D_c^\gamma)$, and thus larger level changes in local compute following a regulation shock.

Proposition 4. Level heterogeneity When the compliance constraint $p_L > p_R$ binds and the price level difference threshold $p_R\,(1 - \alpha + \alpha\omega(\tau)) > \alpha\omega(\tau)p_L$ holds, the effect of regulation on local compute can be expressed as

$$\frac{dC_L^*}{d\tau} = \omega'(\tau)C^*(\tau)\left[1 - \frac{\omega(\tau)}{1-\alpha}\,\frac{p_L - p_R}{p(\omega(\tau))}\right]. \tag{14}$$

The bracketed term only depends on prices, α and $\omega(\tau)$. Therefore, the magnitude of the response in C_L is increasing in $C^*(\tau)$, and increasing in market size N_c and in the data revenue shifter $\log(1 + D_c^\gamma)$.

In other words, for a common policy change $\Delta\tau$, countries with larger N_c, and larger $\log(1 + D_c^\gamma)$ experience larger level increases in local compute.

[14]The model is partial equilibrium and isolates the intensive-margin response of local compute C_L holding the mass of firms fixed. When the compliance constraint binds, tighter regulation can reduce maximised profits, but we abstract from entry/exit decisions. We map the theory to the data by assuming that domestic data centre openings are monotonically increasing in aggregate local compute demand $\sum C_L = M\,C_L$ (with M fixed and tied to market size). Hence, when Proposition 3 implies higher C_L, it implies higher domestic capacity investment/openings.

Proof: In Online Appendix H.

We have demonstrated that the compliance constraint model can replicate the empirical patterns we observed in the data.

The model also delivers additional implications about *where* the compliance constraint bites the most. These implications arise from cross-country heterogeneity in the relative costs of local and remote compute. In the model, local energy and construction costs map into p_L, while international connectivity and access to cloud services map into p_R. We derive comparative statics with respect to (p_L, p_R) next and discuss how they translate into testable heterogeneity.

5.2 Additional Testable Implications

We formalise the testable implications in Proposition 5.

Proposition 5: Cost and connectivity heterogeneity

Assume $p_L > p_R$, so that the compliance constraint binds, and consider an interior solution.

$$B(\tau) \equiv 1 - \frac{\omega(\tau)}{1-\alpha}\frac{p_L - p_R}{p(\omega(\tau))} > 0, \; p(\omega(\tau)) \equiv p_R + (p_L - p_R)\omega(\tau).$$

In this case, the marginal effect of tightening on local compute is expressed as:

$$\frac{dC_L^*(\tau)}{d\tau} = \omega'(\tau)\, C^*(\tau)\, B(\tau).$$

1. Local cost monotonicity: Holding $(N_c, D_c, \alpha, \omega(\tau), \omega'(\tau))$ fixed, the marginal effect is strictly decreasing in p_L:
$$\frac{\partial}{\partial p_L}\left(\frac{dC_L^*(\tau)}{d\tau}\right) < 0. \tag{15}$$
As a result, higher marginal cost of local compute (e.g. by higher electricity prices) produces a smaller reaction in C_L to tightened regulation.

2. Ambiguous impact of remote cost: Holding $(N_c, D_c, \alpha, \omega(\tau), \omega'(\tau))$ fixed, the effect is generally non-monotone in p_R because
$$\frac{\partial}{\partial p_R}\left(\frac{dC_L^*(\tau)}{d\tau}\right) = \omega'(\tau)\left[\underbrace{\frac{\partial C^*(\tau)}{\partial p_R}B(\tau)}_{<0} + \underbrace{C^*(\tau)\frac{\partial B(\tau)}{\partial p_R}}_{>0}\right]. \tag{16}$$
An increase in p_R raises the effective unit price $p(\omega)$ and reduces C^* (the first term in brackets). At the same time, it also narrows the relative cost gap between remote and local compute (the second term). The relative magnitude of the two terms is unclear.

Proof: In Online Appendix I.

We explore these predictions empirically in the next section.

6 Empirical Test: Cost and Connectivity Heterogeneity

The theory model outlined above gives specific predictions about the heterogeneity related to cost-environment faced by firms in different countries. The model predicts that the marginal response of local compute is smaller when local compute is more expensive (higher p_L) and that the cost of remote compute (p_R) has an a priori ambiguous effect on local compute.

In this section, we take these predictions to the data. We proxy p_L with country-level electricity generation measured in 2010. The rationale is that data centres are electricity-intensive, so locations with greater generation capacity plausibly face fewer supply constraints and, on average, lower wholesale electricity costs for large industrial users. While generation is an imperfect proxy for prices (since it ignores fuel mix, regulation, and cross-border power trade), it captures a basic constraint on the scale of energy supply that is relevant for hosting electricity-intensive facilities. At the same time, electricity generation is highly correlated with GDP and population (Table 5), so the split may partially capture market-size heterogeneity already documented earlier.

We measure the cost of remote compute, p_R, using international internet connectivity. For this, we use the international bandwidth usage measured by the International Telecommunication Union (International Telecommunication Union, 2024). The rationale is that using compute abroad requires reliable, high-capacity international links: greater bandwidth reduces effective frictions in accessing remote services (e.g., congestion and quality constraints) and is correlated with better integration into global digital networks. As with electricity generation, bandwidth is an imperfect proxy (it does not directly measure cloud prices or latency), but it captures an observable component of the infrastructure needed to rely on remote compute at scale. International bandwidth is likewise strongly related to income and market size (Table 5), implying that the "high bandwidth" group overlaps substantially with richer economies.

As before, to alleviate concerns related to reverse causality, we fix these measures at their 2010 values. Moreover, due to small sample sizes, we dichotomise the variables using a simple below/above median split. Because these splits are correlated with GDP and population, the event studies below should be interpreted as comparisons of higher- vs lower-electricity (or bandwidth) environments that may also differ in market size; they provide a check of whether the model's comparative statics appear in the data rather than a decomposition of mechanisms.

The resulting event studies are given in Figure 9. In panel (a), we report that tightening of cross-border data transfer law has, on average, a larger effect on new data centre entry in countries with high electricity production (and lower energy prices). This is consistent with theory outlined above.

For the remote-connectivity split (Figure 9 (b)), we find that data centre entry after tightening is concentrated in countries with high international bandwidth. While we cannot directly measure firms' sourcing choices, this is consistent with a domestic substitution mechanism: when cross-border connectivity is good, remote compute is effectively cheaper and firms rely more on it; tightening then requires a larger share of activity to be executed domestically, raising C_L in the tightening countries.

In terms of Proposition 5, higher connectivity lowers the effective marginal cost of remote compute p_R (where p_R is a combination of cloud rental price plus a connectivity or latency component). In the expression, $\frac{dC_L^*(\tau)}{d\tau} = \omega'(\tau)\, C^*(\tau)\, B(\tau)$, a lower p_R raises the term C^* by lowering the effective unit price, but also widens the local-remote price gap $(p_L - p_R)$ which pushes down the final term $B(\tau)$.

The observation that high-connectivity countries react more, implies that, in the region relevant in our data remote and local compute costs are sufficiently close, so tightening mainly shifts compute from abroad to domestic sites without a large contraction in total compute demand within the tightening country.

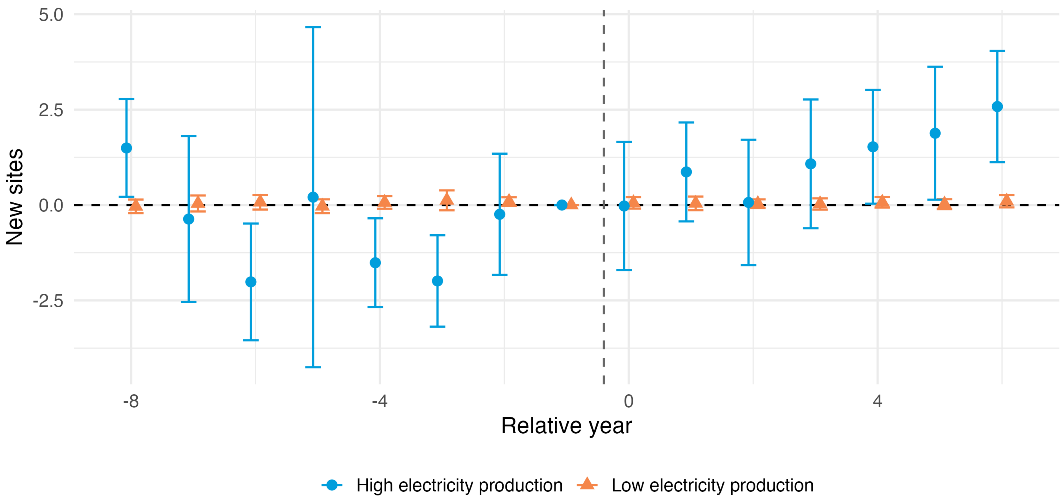
In Figure 9(b), the low-connectivity group shows post-treatment coefficients at or slightly below zero. If taken literally, this would suggest that stricter regulation neither raises domestic entry nor triggers substitution away from remote compute in weakly connected countries — possibly even the reverse. The estimates for this group are imprecise, however. In the language of the theory model, this pattern fits the region where $p_L > p_R$, and remote compute is not a meaningful margin to reallocate from. In that region, tightening would theoretically predict a zero effect on domestic investment. A systematically negative response might be seen as a result of an additional friction that shrinks the production of data-intensive goods rather than triggering a substitution into domestic compute.

7 Discussion

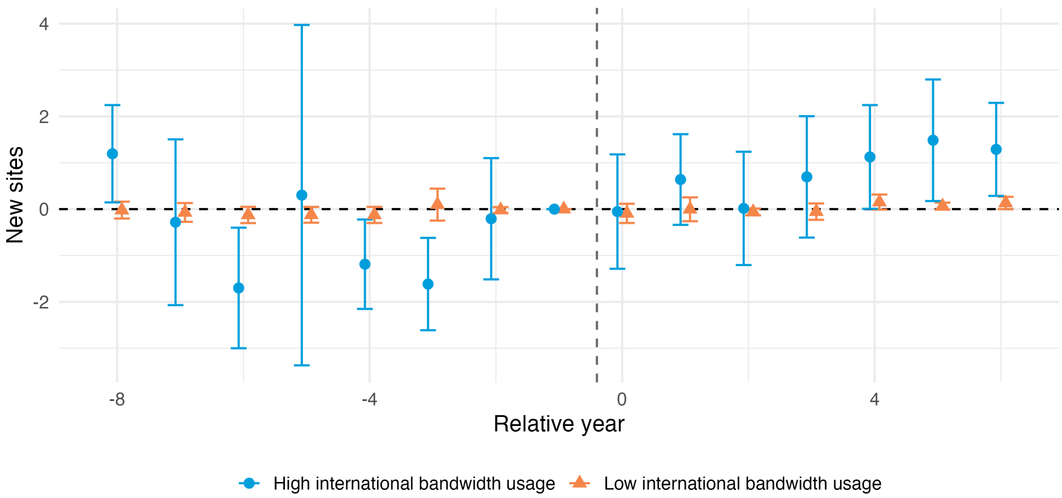
7.1 The Scale of Regulatory Fragmentation

To quantify the implications, we build a back-of-the-envelope counterfactual: for each treated country-year, we subtract the estimated treatment effect from observed new sites and cumulate through 2024. For countries with more downstream production structures (Figure 10, Panel (b)), the counterfactual gap is around 22%, based on the most precisely estimated ATT. The exercise treats the estimated post-treatment difference as the causal impact of the tightening — an assumption supported but not proven by the identification evidence in Section 4. This is a mechanical accounting exercise based on site counts rather than capacity or welfare, and it ignores general-equilibrium relocation. The regional spillover tests support the downstream result: there is no evidence of reallocation from neighbouring countries, consistent with the gap reflecting genuine new entry.

Taken together, the theoretical compliance constraint model and our empirical estimates suggest that the tightening of data transfer laws leads to fragmentation of compute, with more data centre sites built in countries that have tightened regulation. This also implies a loss of real resources. The theoretical model predicts fragmentation, resulting in parallel "stacks" across jurisdictions: investments, staffing, network buildout, and compliance processes are all duplicated. The theory model implies that even if total compute use decreases as a result of stricter regulation, average costs increase because fixed costs are spread over a smaller scale.



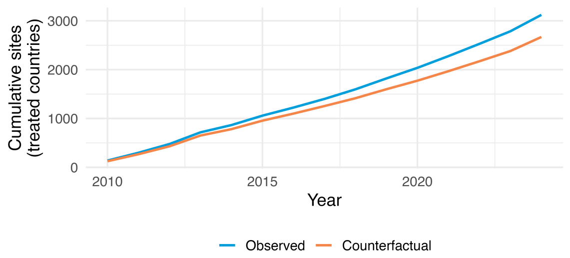
(a) Event study: electricity generation split



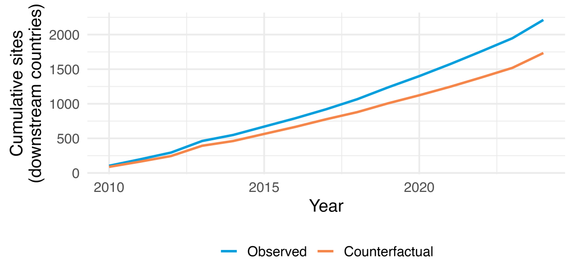
(b) Event study: international bandwidth split

Figure 9. Event-study estimates of the effect of regulatory tightening on domestic data centre entry, by electricity generation and international bandwidth.

Note: Each panel shows Sun and Abraham (2021) event-study estimates. The outcome is the annual count of new data centre site openings per country. Event time 0 is the first year a country adopts a data transfer regime rated level 4 or above. Error bars are 95% confidence intervals based on standard errors clustered at the country level. Panel (a): high/low split by total electricity generation in 2010 (TWh; U.S. Energy Information Administration). Panel (b): high/low split by international internet bandwidth in 2010 (Gbit/s; `https://datahub.itu.int/`). Both splits use the sample median. Sample: 158 countries, 2010–2024.



(a) All treated countries



(b) Countries with more downstream production structures

Figure 10. Cumulative data-centre sites: observed versus counterfactual without tightening of cross-border data-transfer restrictions.

Notes: The counterfactual subtracts the cohort-averaged post-treatment ATT estimate (Table 6, columns 1 and 3 respectively) from observed new site counts for each treated country–year and cumulates the result. No confidence band is plotted for visual clarity; the underlying uncertainty scales with the standard errors on the ATT estimates in Table 6. Panel (a) covers all 71 treated countries; panel (b) covers the 38 treated countries with above-median downstreamness. Sample period 2010–2024.

7.2 Implications and Limitations

The prior literature models cross-border data transfer law as a cost wedge that reduces firms' data use and, via complementarity, compute demand (Demirer et al., 2024; Chang et al., 2023). Our model can reproduce this effect, but adds a geographic margin: regulation can also reallocate compute by shifting *where* it is hosted. The wedge channel operates primarily through the level of activity, while the compliance channel operates through the composition of compute across locations; these forces can move domestic infrastructure investment in opposite directions.

The two heterogeneity patterns have different roots. The downstream effect comes from data intensity: compliance obligations bite hardest in consumer-facing sectors, where personal data flows are central to production, driving the strongest demand for onshore capacity. The market-size effect is different: it comes from fixed-cost scale economies. Building local data centre infrastructure carries overhead that is easier to amortise over a large domestic customer base, so the compliance shift is cheaper to execute in bigger economies. The two dimensions are correlated in practice, since large consumer-facing markets tend to be both downstream and wealthy, but, as Table 5 shows, they are not collinear.

Another direct implication of the theory model is that using laws regulating data transfers as industrial policy to attract domestic data centre investment is costly for local firms. In the model, there are two cases: if local compute is cheaper than remote compute ($p_L \leq p_R$), the compliance constraint is slack and regulation has little effect on firms' allocation choices or investment outcomes, because firms would already choose to process locally. On the other hand, if remote compute is cheaper ($p_L > p_R$), the constraint binds and firms are forced to shift part of their workload onshore. The compliance constraint raises the average cost of compute for local firms. They must cover the fixed cost of maintaining compliant local capacity (or contracting for it) — which can trigger additional local site entry — and pay the higher per-unit price $p_L > p_R$ for each unit processed domestically.

With fixed costs, the choice to host locally is easier to accommodate when local demand is large. The compliance constraint generates a size gradient with regard to local market size and local energy prices. Larger markets can host local hyperscale facilities, while smaller markets have to rely on smaller, higher-cost data centres. China, treated in 2017 under the Cybersecurity Law, provides an illustrative example: a large, high-GDP economy with an extensive downstream digital sector — all characteristics that, in the model, are associated with a stronger investment response. At the same time, regulation can effectively exclude some small or high-cost countries. As a result of regulation, a small country with high fixed costs might end up in a “corner solution” with no meaningful data centre capacity.

These predictions are consistent with the empirical patterns, but causal attribution is not clear: data transfer laws may be enacted alongside broader digital-economy reforms, any of which could independently drive data centre investment. Our results are more consistent with data transfer restrictions being the operative channel, however.

Pre-treatment trends are flat across all subgroups, which argues against a slow-moving digitalisation push predating the law. If data transfer restrictions are bundled with other digital policies, we would expect broad positive effects. Instead, the estimated effects are concentrated in countries with more downstream production structures, where data transfer restrictions are expected to bite the most. Crypto mining facilities do not process personal data, and we find no effect on crypto site openings. This is consistent with a compliance-based mechanism and less consistent with a generic pro-investment environment. In addition, the estimated effects become stronger as the treatment definition is tightened from ≥ 3 to ≥ 4 to ≥ 5, which is what one would expect if the operative margin is the restrictiveness of the transfer regime itself. That pattern has no obvious analogue under a broad bundled-policy interpretation.

None of these tests is individually conclusive, and we cannot fully rule out that other coincident policies are driving the results. But the pattern as a whole is hard to square with a reading in which the data transfer law is incidental to broader digitalisation policies: effects are concentrated in countries with more downstream sector exposure, crypto site openings are unaffected, and estimated effects strengthen with regulatory restrictiveness.

Setting aside identification concerns, we face the usual limitations of country-level data. The number of countries in the world is naturally limited, and the subset that has changed its data transfer legislation even more so. The small sample sizes combined with a highly right-skewed dependent variable and potential measurement error in the classification of the data transfer legislation all work against precision when trying to find statistically significant effects. All of these factors reduce precision and make it less likely for us to observe a statistically significant effect. The fact that we still find effects concentrated where the theory predicts is reassuring.

Our empirical and theoretical results rely on a set of simplifying assumptions that stem from limitations related to underlying data. First, our outcome is the number of sites, which does not map directly to aggregate data centre capacity. More sites need not mean proportionately more capacity, and an increase in sites is best read as evidence of new entry and duplicative fixed costs rather than a commensurate expansion of productive infrastructure. Section 4.7 replicates the main results using UPS capacity in megawatts as the outcome. The downstream effect remains large and significant under that specification, which makes it harder to argue the site-count results are simply picking up small-scale entries. That said, the MW measure records the most recently observed capacity per facility rather than capacity at opening, so it likely overstates the initial size of facilities expanded after entry.

Second, we do not differentiate between the types of compute available locally and remotely. For instance, frontier AI training requires specialised hardware such as advanced GPU chips, access to which can be geographically concentrated and, in some jurisdictions, constrained by export controls (U.S. Department of Commerce, Bureau of Industry and Security, 2022). When these non-modelled constraints bind, a compliance quota like the one modelled in this paper can increase domestic sites and shift compliance-sensitive workloads onshore, while leaving reliance on foreign frontier compute largely unchanged. While relevant in practice, we abstract away from these considerations.

Overlapping heterogeneity splits create a problem analogous to multicollinearity: when the same countries appear in multiple groups, it is hard to attribute effects to any one dimension. Table 5 shows that correlations among the splits are mostly low to moderate, so the groupings capture largely different underlying factors.

Our analysis captures de jure effects of legislation, not enforcement intensity. Some jurisdictions may have adopted stringent transfer rules without the regulatory capacity to apply them consistently. However, our results are robust to controlling for proxies for cross-country differences in legal enforcement capacity.

8 Conclusions

In this paper, we examine how restrictions on cross-border data transfers affect the geography of data centre entry. Using a Sun–Abraham two-way fixed effects event-study design (Sun and Abraham, 2021) on a global country–year panel of data centre openings and law changes, we find that tightenings of data transfer rules are followed by higher domestic data centre entry in countries with more downstream production structures and, secondarily, in wealthier economies. The dynamic estimates show stable pre-trends, and a within-region leave-one-out placebo test provides no evidence of systematic changes in nearby countries' entry around the tightening date, supporting the interpretation of the estimates as a domestic entry response.

To rationalise these findings, we present a simple micro-foundation model where regulation affects firms' choices through a compliance quota, which states that at least a certain share of compute must take place locally. We demonstrate that this model can replicate a wide range of comparative statics we estimate from the data.

Treating our ATT estimates as the causal impact of regulation, we find that, in the absence of regulation, there would be 15% fewer data centres; for downstream economies the corresponding counterfactual estimate is 22%. However, both are back-of-the-envelope estimates with large uncertainty. The benefits of personal data transfer regulation are not evenly spread: larger, wealthier, and more connected economies capture most of the infrastructure response, while smaller and less connected countries bear compliance costs with little domestic investment to show for it. To that extent, data transfer regulation concentrates digital economy activity in countries already well-placed to host it.

Despite these costs, tighter data transfer laws can still be justified as insurance against exposure to foreign-controlled infrastructure — the "weaponised interdependence" argument (Farrell and Newman, 2019). Our evidence is consistent with this premise. In countries with enough domestic market and data-processing activity to turn compliance pressure into genuine investment, such rules are followed by domestic entry. For smaller or more upstream economies, tighter laws tend to generate compliance costs without much infrastructure to show for it, which limits how much traction digital sovereignty arguments can get in practice.

One open question our data cannot answer is who actually invests. We cannot distinguish between entry by global hyperscalers and entry by smaller domestic operators.

The competitive implications of the two types of entrants differ: if the response is driven mainly by hyperscalers, regulation may entrench market concentration in cloud services even while meeting localisation requirements. Whether data flow policies weaken or reinforce hyperscaler dominance remains open, and is of direct relevance to competition policy and digital sovereignty debates.

References

Adler-Nissen, R. and K. A. Eggeling (2024). The discursive struggle for digital sovereignty: Security, economy, rights and the cloud project gaia-x. *JCMS: Journal of Common Market Studies 62*(4), 993–1011.

Agrawal, A., J. Gans, and A. Goldfarb (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Agrawal, A., J. Gans, and A. Goldfarb (2022). *Prediction machines, updated and expanded: The simple economics of artificial intelligence*. Harvard Business Press.

Amiti, M., S. H. Kong, and D. Weinstein (2020). The effect of the us-china trade war on us investment. Technical report, National Bureau of Economic Research.

Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012). Measuring the upstreamness of production and trade flows. *American Economic Review 102*(3), 412–416.

Arel-Bundock, V., N. Enevoldsen, and C. Yetman (2018). countrycode: An r package to convert country names and country codes. *Journal of Open Source Software 3*(28), 848.

Aridor, G., Y.-K. Che, and T. Salz (2020). *The economic consequences of data privacy regulation: Empirical evidence from GDPR*. National Bureau of Economic Research Cambridge, MA, USA.

Bajari, P., V. Chernozhukov, A. Hortaçsu, and J. Suzuki (2019). The impact of big data on firm performance: An empirical investigation. In *AEA papers and proceedings*, Volume 109, pp. 33–37. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bonfiglioli, A., R. Crinò, M. Filomena, and G. Gancia (2025). Data, power and emissions: The environmental cost of ai. Technical report, CESifo Working Paper.

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.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230.

Casalini, F. and J. López González (2019, January). Trade and cross-border data flows. OECD Trade Policy Papers 220, OECD Publishing, Paris.

Chang, Q., L. W. Cong, L. Wang, and L. Zhang (2023). Production, trade, and cross-border data flows. *Available at SSRN 4672185*.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007). Business cycle accounting. *Econometrica 75*(3), 781–836.

Corneo, L., M. Eder, N. Mohan, A. Zavodovski, S. Bayhan, W. Wong, P. Gunningberg, J. Kangasharju, and J. Ott (2021). Surrounded by the clouds: A comprehensive cloud reachability study. In *Proceedings of the Web Conference 2021*, WWW '21, New York, NY, USA, pp. 295–304. Association for Computing Machinery.

Corrado, C., J. Haskel, C. Jona-Lasinio, and M. Iommi (2022). Intangible capital and modern economies. *Journal of Economic Perspectives 36*(3), 3–28.

Dang, T. K., N. Mohan, L. Corneo, A. Zavodovski, J. Ott, and J. Kangasharju (2021). Cloudy with a chance of short rtts: analyzing cloud connectivity in the internet. In *Proceedings of the 21st ACM Internet Measurement Conference*, IMC '21, New York, NY, USA, pp. 62–79. Association for Computing Machinery.

Demirer, M., D. J. J. Hernández, D. Li, and S. Peng (2024). Data, privacy laws and firm production: Evidence from the gdpr. Technical report, National Bureau of Economic Research.

Di Tella, R. and E. Schargrodsky (2004). Do police reduce crime? estimates using the allocation of police forces after a terrorist attack. *American Economic Review 94*(1), 115–133.

European Parliament and Council of the European Union (2016). Regulation (eu) 2016/679 (general data protection regulation). Official Journal of the European Union, L 119, 4.5.2016, pp. 1–88.

Farboodi, M. and L. Veldkamp (2021). A model of the data economy. Technical report, National Bureau of Economic Research Cambridge, MA, USA.

Farrell, H. and A. L. Newman (2019). Weaponized interdependence: How global economic networks shape state coercion. *International security 44*(1), 42–79.

Ferracane, M. F. (2017). Restrictions on cross-border data flows. Technical report, ECIPE working paper.

Frey, C. B. and G. Presidente (2024). Privacy regulation and firm performance: Estimating the gdpr effect globally. *Economic Inquiry 62*(3), 1074–1089.

Goldberg, S. G., G. A. Johnson, and S. K. Shriver (2024). Regulating privacy online: An economic evaluation of the gdpr. *American Economic Journal: Economic Policy 16*(1), 325–358.

Goldfarb, A. and C. Tucker (2019). Digital economics. *Journal of economic literature 57*(1), 3–43.

Greenleaf, G. (2023). Global tables of data privacy laws and bills (8th ed.). Technical report, University of New South Wales, Faculty of Law.

Greenleaf, G. (2025). Global data privacy laws 2025: 172 countries, twelve new in 2023/24. Technical report, University of New South Wales, Faculty of Law. Published April 2, 2025.

Gupta, S., P. Ghosh, and V. Sridhar (2022). Impact of data trade restrictions on it services export: A cross-country analysis. *Telecommunications Policy 46*(9), 102403.

Handley, K. and N. Limao (2015). Trade and investment under policy uncertainty: theory and firm evidence. *American Economic Journal: Economic Policy 7*(4), 189–222.

Handley, K. and N. Limão (2017). Policy uncertainty, trade, and welfare: Theory and evidence for china and the united states. *American economic review 107*(9), 2731–2783.

Hestness, J., S. Narang, N. Ardalani, G. Diamos, H. Jun, H. Kianinejad, M. M. A. Patwary, Y. Yang, and Y. Zhou (2017). Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

International Telecommunication Union (2024). ITU DataHub: International internet bandwidth. Online database.

Janßen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022). Gdpr and the lost generation of innovative apps. Technical report, National Bureau of Economic Research.

Jardim, E., M. C. Long, R. Plotnick, J. Vigdor, and E. Wiles (2024). Local minimum wage laws, boundary discontinuity methods, and policy spillovers. *Journal of Public Economics 234*, 105131.

Jia, J., G. Z. Jin, M. Leccese, and L. Wagman (2025). How does privacy regulation affect transatlantic venture investment? evidence from gdpr. Technical report, National Bureau of Economic Research.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2023). Privacy and market concentration: Intended and unintended consequences of the gdpr. *Management Science 69*(10), 5695–5721.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2011). The worldwide governance indicators: Methodology and analytical issues1. *Hague journal on the rule of law 3*(2), 220–246.

Koski, H. and N. Valmari (2020). Short-term impacts of the gdpr on firm performance. Technical report, ETLA Working Papers.

Kwon, C.-W. and B. G. Chun (2009). Local content requirement under vertical technology diffusion. *Review of Development Economics 13*(1), 111–124.

Lehdonvirta, V., B. Wú, and Z. Hawkins (2024). Compute north vs. compute south: The uneven possibilities of compute-based ai governance around the globe. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Volume 7, pp. 828–838.

Lehdonvirta, V., B. Wu, and Z. Hawkins (2025). Weaponised interdependence in a bipolar world: how economic forces and security interests shape the global reach of us and chinese cloud data centres. *Review of International Political Economy*, 1–26.

Ling, D. C., A. Naranjo, and B. Scheick (2021). There is no place like home: Information asymmetries, local asset concentration, and portfolio returns. *Real Estate Economics 49*(1), 36–74.

Ling, D. C., C. Wang, and T. Zhou (2021). The geography of real property information and investment: Firm location, asset location and institutional ownership. *Real Estate Economics 49*(1), 287–331.

Ludwig, J., S. Mullainathan, and A. Rambachan (2025). Large language models: An applied econometric framework. Technical report, National Bureau of Economic Research.

Mancini, M., P. Montalbano, S. Nenci, and D. Vurchio (2024). Positioning in global value chains: World map and indicators, a new dataset available for gvc analyses. *The World Bank Economic Review 38*(4), 669–690.

McDermott, G. (2026). *etwfe: Extended Two-Way Fixed Effects*. R package version 0.6.1.

Pan Fang, T. and S. Greenstein (2025). Where the cloud rests: The economic geography of data centers. *Strategy Science 10*(4), 404–420.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science 41*(4), 746–768.

Pierce, J. R. and P. K. Schott (2018). Investment responses to trade liberalization: Evidence from us industries and establishments. *Journal of International Economics 115*, 203–222.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics 11*(4), 707–720.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Sisto, E. and E. Van der Marel (2025). Privacy at a price? an empirical analysis of gdpr's impact on eu trade flows. Technical report, ECIPE Occasional Paper.

S&P Global Market Intelligence (2025a). 451 research datacenter Knowledgebase. Online database.

S&P Global Market Intelligence (2025b). Capital IQ Pro real estate database. Online database. Property-level data on data centre facilities; accessed via S&P Capital IQ Pro platform.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics 225*(2), 175–199.

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. Technical report, National Bureau of Economic Research.

Tian, C., R. K. Chellappa, and J. G. Martinez (2025). Supply, demand or policy driven? an empirical examination of data center location strategies in the united states. In *Proceedings of the International Conference on Information Systems (ICIS 2025)*. Association for Information Systems (AIS).

U.S. Department of Commerce, Bureau of Industry and Security (2022, October 28). Bis updated public information page on export controls imposed on advanced computing and semiconductor manufacturing items to the people's republic of china (prc). Accessed January 7, 2026.

U.S. Energy Information Administration (2024). International electricity generation statistics. Online database.

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal 26*(2), C31–C66.

World Bank (2024a). World development indicators. Online database. Series SP.POP.TOTL (population) and NY.GDP.MKTP.KD (real GDP).

World Bank (2024b). Worldwide governance indicators. Online database. Series RL.EST (Rule of Law) and RQ.EST (Regulatory Quality).

World Legal Information Institute (2024). National data privacy legislation collection. Online repository.

Zhang, H., H. Kim, and S. Moon (2025). The effects of data restriction policies and institutions on digital financial service and trade. *Emerging Markets Finance and Trade 61*(9), 2745–2762.

Zheng, Y. and Q. R. Wang (2020, December). Shadow of the great firewall: The impact of Google blockade on innovation in China. *Strategic Management Journal 41*(12), 2234–2260.

Zhou, J. (2024, December). Firewall for innovation. Job Market Paper.

A Appendix: Qualitative Evaluation of Law Classification

Table A1 illustrates how the classification rubric applies to seven representative laws drawn from our dataset. For each law, we report the key statutory provision that determines its level assignment. The examples are chosen for global prominence and geographic spread.

Table A1. Qualitative evaluation of law classification: selected laws and key statutory provisions by restrictiveness level.

Level	**Country (Law, Year)**	**Key statutory provision**	**Why this level**
0	United States (Privacy Act, 1974)	Covers only federal government agencies. No provision addresses private-sector cross-border data transfers; no adequacy mechanism, no DPA, no mention of international data flows in operative provisions.	No regulation of private-sector transfers.

Table A1 continued

Level	Country (Law, Year)	Key statutory provision	Why this level
1	India (DPDPA, 2023)	S. 16(1): "The Central Government *may*, by notification, restrict the transfer of personal data ... to such country or territory outside India as may be so notified." Transfers are free by default; the government holds a blacklisting power only. No adequacy list, SCC, or authorisation gate anywhere in the Act.	Free flow by default; only ex-post accountability.
2	South Africa (POPIA, 2013)	S. 72(a): A responsible party may not transfer unless "the third party who is the recipient of the information is subject to a law, binding corporate rules or binding agreement which provide an adequate level of protection" — adequacy is assessed by the data exporter, not by any government body or the Information Regulator. No government adequacy list exists.	Exporter self-assesses adequacy; no public authority involved.
3	Japan (APPI, 2015)	Art. 24: The Personal Information Protection Commission (PPC) determines adequate countries by regulation; for non-adequate countries operators must obtain prior individual consent. No authorisation, license, or permit from the PPC is required for individual transfers. No ongoing obligation on the exporter to monitor the recipient's post-transfer data processing.	Public adequacy mechanism with self-executing fallbacks; exporter's obligation ends at point of transfer.
4	Germany (BDSG/GDPR, 2018)	BDSG § 1(5): "The provisions of this Act shall not apply where the law of the European Union, in particular Regulation (EU) 2016/679... directly applies" — cross-border transfers governed entirely by GDPR Chapter V. GDPR Art. 46(1): Transfers require "appropriate safeguards" ensuring "essentially equivalent" protection (Schrems II standard); the controller remains liable for the recipient's compliance; enforceable data subject rights must be available at the destination. The European Commission maintains the operative adequacy list (Art. 45).	Public adequacy mechanism; exporter has ongoing liability for recipient's actual processing under Schrems II.

Table A1 continued

Level	Country (Law, Year)	Key statutory provision	Why this level
5	Turkey (KVKK, 2016)	Art. 9(3): The Board determines adequate countries; transfers to adequate destinations bypass authorisation. Art. 9(2)(b): for non-adequate countries, transfers require Board authorisation *plus* written sufficiency guarantees from both the Turkish and foreign controllers. Consent is the only self-executing bypass of Board authorisation; no SCC or BCR mechanism avoids Board approval.	Genuine authorisation gate for non-adequate transfers; consent alone insufficient for routine commercial flows.
6	China (PIPL, 2021)	Art. 40: Critical information infrastructure operators "shall store the personal information collected and generated within the territory of the People's Republic of China" — mandatory data localisation. Art. 38: all processors must satisfy one of four government-controlled conditions (CAC security assessment; CAC-designated certification; CAC-formulated standard contract; or other conditions prescribed by CAC). Art. 39: individual consent is required *in addition to* satisfying an Art. 38 condition, not instead of it.	Universal authorisation and mandatory localisation; no self-executing transfer pathway exists.

Notes: Illustrative country–law pairings across the seven levels of the Casalini and López González (2019) taxonomy. The level assigned to each entry is the LLM classifier output, manually verified against the legal text cited in column three.

Online Appendix

B LLM Classification Prompt

We classified laws using Claude Sonnet (claude-sonnet-4-5-20251001) with the following structured prompt. The prompt instructs the model to focus on the operative mechanism of each law rather than surface features such as the phrase "adequate protection." It defines the seven levels, lists eight common misclassification traps, and specifies a ten-step decision protocol.

The prompt instructs the model to act as a legal classification assistant and classify a given law text into a single OECD restrictiveness level from 0 to 6. The opening instruction emphasises that the phrase "adequate protection" alone does not determine the level; what matters is the *operative mechanism* — who assesses adequacy, what happens when adequacy is absent, and what obligations fall on the data exporter regarding the recipient's behaviour.

Classification framework

Level 0 — No regulation. The country has no law regulating personal data, or the law does not address cross-border transfers.

Level 1 — Free transfer with ex-post accountability. Transfers are permitted without pre-conditions. The exporter has accountability obligations after the fact (records, liability for misuse), but no gate or approval is needed before transferring.

Level 2 — Self-assessment of adequacy. The data exporter (not a government body) decides whether the recipient provides adequate protection. Fallback mechanisms (consent, contractual clauses) are available. The law does not impose substantive requirements on the safeguards themselves. *Key test*: if the law says "the controller shall ensure adequate protection" without designating a government authority and without imposing substantive obligations on the safeguard mechanisms, this is Level 2.

Level 3 — Public adequacy determination with fallback options. A government body or data protection authority determines which countries are adequate. Fallback mechanisms (consent, SCCs, BCRs) are available for non-adequate destinations. The exporter uses a transfer mechanism and is done — no further obligations regarding the recipient's actual data handling. *Key test*: the law establishes a public adequacy mechanism and provides fallbacks, but does not impose ongoing obligations on the exporter to monitor or guarantee protection at the recipient's end.

Level 4 — Public adequacy + enforceable safeguard requirements (GDPR-type). A public adequacy mechanism exists, and the law imposes *substantive ongoing obligations* on the data exporter regarding the recipient's behaviour. All of the following must be present: (a) the exporter must ensure or guarantee that the recipient provides protection "essentially equivalent" to domestic standards; (b) the exporter remains liable for the recipient's data processing after transfer; (c) specific safeguard mechanisms (SCCs, BCRs) must themselves ensure an adequate level of protection. EU GDPR Chapter V (Articles 44–49) is the paradigmatic Level 4 regime. *Key test*: would a compliance officer need to actively monitor how the foreign recipient processes the data, or merely check a box and move on? If the former, Level 4; if the latter, Level 3.

Level 5 — Case-by-case authorisation with adequacy fallback. Prior authorisation, license, or permit from a regulator is required as a genuine operative gate. The authorisation requirement is not bypassed by standard derogations (consent, contract, etc.) — those derogations may exempt the exporter from the adequacy condition, but the authorisation itself remains mandatory. Transfers to adequate countries may bypass the authorisation requirement. *Key test*: does the law impose an authorisation requirement *separate* from the adequacy condition? If consent bypasses adequacy but a license/permit is still required, this is Level 5.

Level 6 — Universal authorisation or data localisation. All transfers require government approval, with no alternative pathway — even adequate countries or consent do not bypass the authorisation requirement. Or the law mandates local storage/processing of personal data. No adequacy list, consent exception, or SCC pathway removes the need for regulatory approval.

Common misclassification traps

Trap 1. "Adequate protection" $\neq$ Level 4. At Level 3, adequacy is a gateway condition. At Level 4, it is an ongoing obligation on the exporter regarding the recipient's actual processing.

Trap 2. Derogations from *adequacy* $\neq$ derogations from *authorisation*. When a law has two separate requirements (adequate destination; regulator authorisation), check which requirement the derogation bypasses. If derogations bypass authorisation $\rightarrow$ Level 3. If derogations bypass only adequacy while the license requirement persists $\rightarrow$ Level 5 or 6.

Trap 3. Do not classify based on which country the law is from. Classify based on what the text says.

Trap 4. "Minister *may* specify adequate countries" $\neq$ public adequacy already in effect. If the mechanism has not been activated, the actual regime may be self-assessment (Level 2).

Trap 5. Level 4 without a formal adequacy list is possible but rare — only if the law explicitly requires safeguard mechanisms to ensure an adequate level of protection *and* a regulatory body has oversight over the safeguards.

Trap 6. Qualified consent (e.g., "punctual" or "non-massive" transfers only) is not a broad fallback. Check whether a company doing regular, ongoing cross-border transfers could rely on the consent derogation.

Trap 7. Mandatory DPA notification/verification before transfer $\approx$ an authorisation gate, not mere notification.

Trap 8 (Level 3 vs. 5). Consent-only bypass of authorisation $\neq$ Level 3 when no mechanism-based pathway (SCCs, BCRs) independently bypasses authorisation. If individual consent is the only DPA-free transfer path and all mechanism-based transfers require Board authorisation, the authorisation gate is operative for routine commercial flows $\rightarrow$ Level 5.

Decision protocol

1. Read the entire law text end-to-end before classifying.
2. Identify the operative transfer mechanism: how does a company actually transfer data abroad under this law?
3. Ask: who determines adequacy — the exporter or a public authority?
4. Ask: what happens for non-adequate destinations — is there a practical fallback, or must the exporter obtain DPA approval?
5. Ask: if the law requires authorisation/license/permit, do the derogations (consent, contract, etc.) bypass the *authorisation itself*, or only the adequacy condition? If authorisation persists even with consent, classify as Level 5 or 6.

6. Ask: are derogations (especially consent) qualified or limited (e.g., "punctual only," "non-massive," "emergency")? If so, they do not cover routine business transfers.
7. Ask: does the DPA have a mandatory verification role *before* transfers? If yes, this functions as a de facto authorisation gate (Level 5).
8. Ask: does the law impose ongoing obligations on the exporter regarding the recipient's actual data processing (Level 4), or is it "use a mechanism and you're done" (Level 3)?
9. Ask: if consent bypasses authorisation, does the law also provide a mechanism-based pathway (SCCs, BCRs) that bypasses authorisation? If consent is the only bypass, classify as Level 5 (Trap 8).
10. Classify based on the answers above. When ambiguous, prefer the level whose criteria are most explicitly stated in the text.

The full prompt, including eight worked examples covering Levels 2–6, is available in the replication package.

C Robustness to Prompt Perturbation

To assess whether the classification results depend on the specific framing of the prompt, we reclassified the laws passed later than 2022 using the main prompt and two perturbations. The three prompts are described below:

Prompt A (the main specification used in the text). Legal assistant framing. Level definitions presented in ascending order (0–6). Includes ten-step decision protocol and eight misclassification traps. Used for the full dataset.

Prompt B. Compliance-analyst framing. Level definitions presented in *descending* order (6–0) with "compliance officer" language replacing "legal assistant." Eight pitfalls replace the misclassification traps.

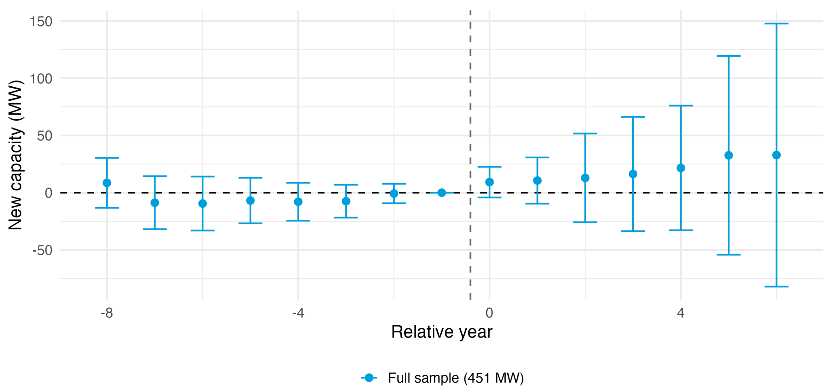
Prompt C. Decision-tree framing. Six binary yes/no questions replace the level definitions. No traps; instead, each decision node includes a brief caution.

Table C1 reports agreement statistics across the three prompts. Exact agreement is high between Prompts A and B (84.9%), and binary agreement at the economically relevant ≥ 4 threshold exceeds 86% in all pairwise comparisons. Krippendorff's $\alpha = 0.794$ across all three prompts indicates substantial ordinal agreement.

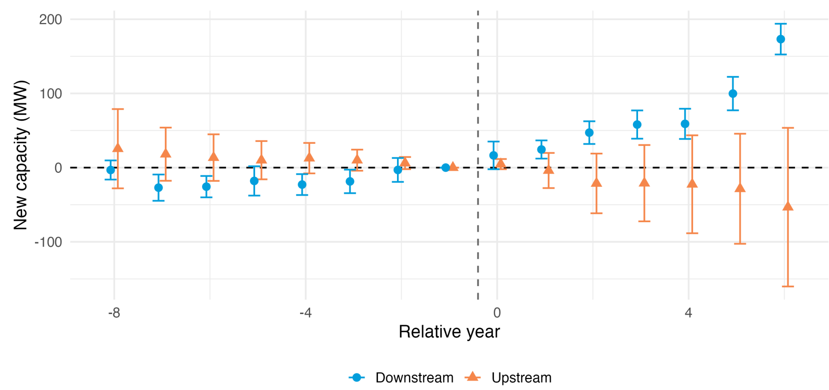
The replication package, law texts, and the full pipeline are available at `https://version.aalto.fi/gitlab/diesl/GEOCLOUD/-/tree/main/cross_border_transfer_laws`.

D Pre-treatment trends in UPS capacity

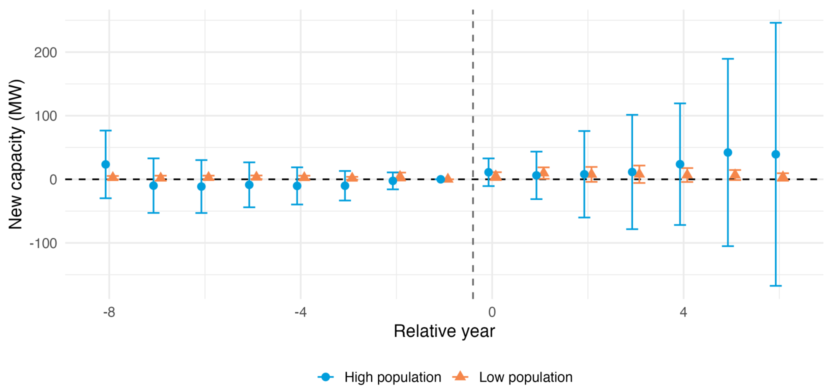
Figure D1 replicates the event-study design from Section 4 of the main paper using annual new UPS capacity in megawatts as the outcome. Pre-treatment coefficients are close to zero across all four specifications, with confidence intervals that overlap zero throughout the pre-period. The flat pre-trends support the parallel trends assumption for this outcome and are inconsistent with a scenario in which the measurement properties of the 451 Research data create differential pre-treatment trends between treated and control countries.



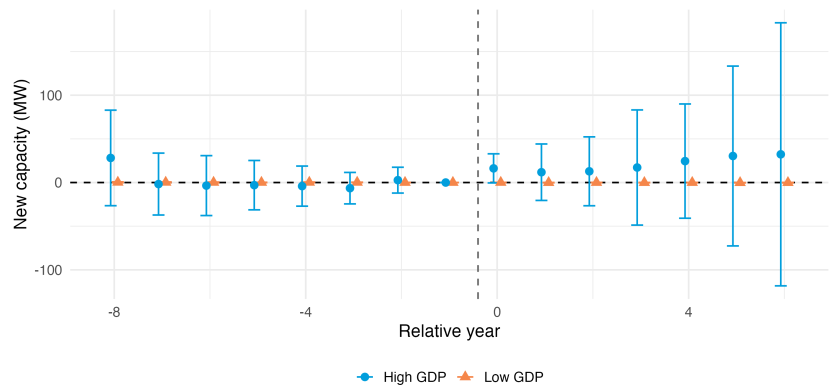
(a) Event study: full sample



(b) Event study: Downstreamness split



(c) Event study: population split



(d) Event study: GDP split

Figure D1. Event-study estimates of the effect of regulatory tightening on domestic data centre capacity (UPS megawatts).

Note: Each panel shows Sun and Abraham (2021) event-study estimates using not-yet-treated and never-treated countries as controls. The outcome is annual new UPS capacity in megawatts per country, drawn from the 451 Research Datacenter KnowledgeBase. Event time 0 is the first year a country adopts a data transfer regime rated level 4 or above (GDPR-equivalent or stricter). Error bars are 95% confidence intervals based on standard errors clustered at the country level. High/low splits are defined by the sample median of the 2010 value of the splitting variable. Sample: 158 countries, 2010–2024.

Table C1. Robustness of law classification to prompt perturbation ($N = 186$ laws reclassified under each prompt variant).

Comparison	N	Exact agreement	Pearson r	Binary ≥ 4 agreement
A vs. B	186	84.9%	0.928	92.5%
A vs. C	186	71.0%	0.739	86.0%
Krippendorff's α (ordinal)	186		0.794	

Notes: $N = 186$ laws enacted after 2022, classified independently under each of the three prompts described in the surrounding text. "Exact agreement" is the share of laws receiving the identical level assignment (0–6) across two prompts; "Binary ≥ 4 agreement" is the share classified identically as ≥ 4 or < 4. Krippendorff's α is computed treating the seven-level scale as ordinal.

E Multiple-testing adjustment

Table E1 reports Romano–Wolf (Romano and Wolf, 2005) stepdown-adjusted p-values for the six subsample splits in Table 6 of the main paper, alongside the uncorrected p-values, for the raw and winsorised specifications. In the raw specification, the correction raises the downstream and high-GDP p-values above 0.1. In the winsorised specification, the adjusted p-values are close to the uncorrected ones, and the downstream and high-GDP estimates remain significant at the 1% level. This is consistent with Figure 8 and Section 4.8 of the main paper. Outliers in the raw outcome inflate standard errors rather than drive point estimates.

Table E1. Post-treatment ATT across six subsample splits: raw vs. winsorised specification, with Romano–Wolf multiple-testing adjustment.

	Raw outcome				Winsorised (95th pct, by country)			
Split	Est.	S.E.	p	p_{RW}	Est.	S.E.	p	p_{RW}
Upstream	0.002	(0.297)	0.995	0.996	0.144	(0.235)	0.540	0.790
Downstream	1.325	(0.665)	0.046	0.180	1.476	(0.403)	<0.001	0.001
Low population	0.170	(0.209)	0.416	0.873	0.130	(0.152)	0.394	0.773
High population	1.085	(1.678)	0.518	0.890	1.330	(1.321)	0.314	0.773
Low GDP	0.022	(0.049)	0.658	0.890	0.009	(0.042)	0.823	0.825
High GDP	1.154	(0.681)	0.090	0.315	1.241	(0.398)	0.002	0.007

Notes: Sun–Abraham post-treatment ATT from subsample splits (cf. Table 6). Standard errors (in parentheses) are clustered by country. p is the two-sided pointwise p-value from the asymptotic normal approximation; p_{RW} is the Romano–Wolf (Romano and Wolf, 2005) stepdown-adjusted p-value, controlling the family-wise error rate across the six subsample tests. The joint null distribution is simulated from $N(0, \hat{R})$, where $\hat{R}$ is the correlation matrix of the six ATT estimates obtained by leave-one-cluster-out jackknife on each subsample. The winsorised specification caps *new_sites* at each country's own 95th percentile before estimation. Based on $B = 10{,}000$ simulation draws.

F Proof of Proposition 1

Proof. Fix total compute $C > 0$ and consider the cost-minimisation problem over the allocation (C_L, C_R):

$$\min_{C_L, C_R \geq 0} \ \text{Cost}_L(C_L) + \text{Cost}_R(C_R) \quad \text{s.t.} \quad C_L + C_R = C, \quad C_L \geq \omega(\tau)C,$$

where $\text{Cost}_L(C_L) = F + p_L C_L$ and $\text{Cost}_R(C_R) = p_R C_R$.

Using the accounting identity $C_R = C - C_L$, the objective can be written as

$$F + p_L C_L + p_R(C - C_L) = F + p_R C + (p_L - p_R)C_L.$$

Since $F + p_R C$ is constant for fixed C, the problem reduces to choosing C_L to minimise $(p_L - p_R)C_L$ subject to the feasibility interval $C_L \in [\omega(\tau)C,\ C]$.

Case 1: $p_L \le p_R$. Then $p_L - p_R \le 0$, so the objective is weakly decreasing in C_L. The cost-minimising choice is $C_L = C$ and $C_R = 0$. The compliance constraint is slack, so changes in τ (and thus $\omega(\tau)$) do not affect the allocation.

Case 2: $p_L > p_R$. Then $p_L - p_R > 0$, so the objective is strictly increasing in C_L. The cost-minimising choice is the smallest feasible C_L, i.e. the compliance constraint binds:

$$C_L = \omega(\tau)C, \qquad C_R = (1 - \omega(\tau))C.$$

Substituting this allocation into variable cost gives

$$p_L C_L + p_R C_R = p_L\omega(\tau)C + p_R(1 - \omega(\tau))C = [p_R + (p_L - p_R)\omega(\tau)]\,C.$$

Defining $p(\omega) = p_R + (p_L - p_R)\omega(\tau)$ gives the effective unit price $p(\omega)C$. □

G Proof of Proposition 3

Proof. Assume that the regulation binds, i.e. $p_L > p_R$, so that $C_L = \omega(\tau)C$ and $C_R = (1 - \omega(\tau))C$.

Step 1: Reduced-form profit and optimal C^*.

Since $C = C_L + C_R$, profits can be expressed as

$$\begin{aligned}\pi(C;\tau) &= N_c\, C^\alpha \log(1 + D_c^\gamma) - F - p_L\,\omega(\tau)\,C - p_R\,(1 - \omega(\tau))\,C \\ &= N_c\, C^\alpha \log(1 + D_c^\gamma) - F - p(\omega(\tau))\,C,\end{aligned}$$

where $p(\omega) = p_R + (p_L - p_R)\omega$.

The F.O.C for an interior optimum C^* is

$$N_c\,\alpha\,(C^*)^{\alpha-1}\log(1 + D_c^\gamma) = p(\omega). \tag{G.1}$$

This implies that

$$C^*(\omega) = \left(\frac{N_c\,\alpha\,\log(1 + D_c^\gamma)}{p(\omega)}\right)^{\frac{1}{1-\alpha}}. \tag{G.2}$$

Step 2: Local compute and its derivative with respect to ω. In the binding regime,

$$C_L^*(\omega) = \omega\,C^*(\omega). \tag{G.3}$$

Differentiate with respect to ω:

$$\frac{dC_L^*}{d\omega} = C^*(\omega) + \omega\,\frac{dC^*}{d\omega}. \tag{G.4}$$

From (G.2), using $p'(\omega) = p_L - p_R$,

$$\frac{dC^*}{d\omega} = -\frac{1}{1-\alpha}\,C^*(\omega)\,\frac{p'(\omega)}{p(\omega)} = -\frac{1}{1-\alpha}\,C^*(\omega)\,\frac{p_L - p_R}{p(\omega)}.$$

Substitute into (G.4):

$$\frac{dC_L^*}{d\omega} = C^*(\omega)\left[1 - \frac{\omega}{1-\alpha}\frac{p_L - p_R}{p(\omega)}\right]. \tag{G.5}$$

Step 3: Obtain the threshold for the partial derivative.

Since $C^* > 0$, and $\omega'(\tau) > 0$, the sign of $\dfrac{dC_L^*}{d\omega}$ determined by the sign of the bracketed term. C_L increases with ω (and also with τ), if:

$$1 - \frac{\omega}{1-\alpha}\frac{p_L - p_R}{p(\omega)} > 0. \tag{G.6}$$

substituting $p(\omega) = p_R + (p_L - p_R)\omega$ and rearranging gives:

$$p_R\left(1 - \alpha + \alpha\omega(\tau)\right) > \alpha\omega(\tau)p_L, \tag{G.7}$$

which is the desired expression. □

H Proof of Proposition 4

Proof. From the first-order condition (G.1), the interior optimum satisfies

$$C^*(\tau) = \left(\frac{N_c\,\alpha\,\log(1 + D_c^\gamma)}{p(\omega(\tau))}\right)^{\frac{1}{1-\alpha}},$$

which is increasing in N_c and $\log(1 + D_c^\gamma)$. From the proof of Proposition 3, the derivative of local compute is

$$\frac{dC_L^*}{d\tau} = \omega'(\tau)\,C^*(\tau)\left[1 - \frac{\omega(\tau)}{1-\alpha}\frac{p_L - p_R}{p(\omega(\tau))}\right].$$

The bracketed term depends only on prices, α, and $\omega(\tau)$, and is common across countries. Therefore the magnitude of $\frac{dC_L^*}{d\tau}$ is proportional to $C^*(\tau)$, and hence increasing in N_c and $\log(1 + D_c^\gamma)$. □

I Proof of Proposition 5

Proof. Fix τ so that $\omega = \omega(\tau) \in (0,1)$ and $\omega' = \omega'(\tau) > 0$ are constants, and assume $p_L > p_R$ so the compliance constraint binds. Then $C_L = \omega C$, $C_R = (1-\omega)C$, and the effective unit price of compute is

$$p(\omega) = p_R + (p_L - p_R)\omega = \omega p_L + (1-\omega)p_R.$$

Let $A = \frac{\omega}{1-\alpha} > 0$ and define

$$B(\tau) = 1 - A\frac{p_L - p_R}{p(\omega)}.$$

From Proposition 4 for an interior solution,

$$\frac{dC_L^*}{d\tau} = \omega'\,C^*(\tau)\,B(\tau).$$

(i) Local cost p_L monotonicity. Since

$$C^*(\tau) = \left(\frac{\alpha N_c\log(1 + D_c^\gamma)}{p(\omega)}\right)^{\frac{1}{1-\alpha}}, \qquad \frac{\partial C^*}{\partial p(\omega)} = -\frac{1}{1-\alpha}\frac{C^*}{p(\omega)} < 0,$$

and $\frac{\partial p(\omega)}{\partial p_L} = \omega$, we have

$$\frac{\partial C^*}{\partial p_L} = -\frac{\omega}{1-\alpha}\frac{C^*}{p(\omega)} < 0.$$

Moreover,

$$\frac{\partial B}{\partial p_L} = -A\frac{p(\omega) - (p_L - p_R)\,\partial p(\omega)/\partial p_L}{p(\omega)^2} = -A\frac{p(\omega) - (p_L - p_R)\omega}{p(\omega)^2} = -\frac{\omega}{1-\alpha}\frac{p_R}{p(\omega)^2} < 0.$$

Therefore,

$$\frac{\partial}{\partial p_L}\left(\frac{dC_L^*}{d\tau}\right) = \omega'\left(\frac{\partial C^*}{\partial p_L}B + C^*\frac{\partial B}{\partial p_L}\right) = -\omega' C^*\frac{\omega}{1-\alpha}\left(\frac{B}{p(\omega)} + \frac{p_R}{p(\omega)^2}\right).$$

Under interior condition $B(\tau) > 0$ (equivalently $p_R(1-\alpha+\alpha\omega) > \alpha\omega p_L$), the bracket is strictly positive, hence

$$\frac{\partial}{\partial p_L}\left(\frac{dC_L^*}{d\tau}\right) < 0.$$

(ii) Remote cost p_R is ambiguous. We have $\frac{\partial p(\omega)}{\partial p_R} = 1 - \omega > 0$, so

$$\frac{\partial C^*}{\partial p_R} = \frac{\partial C^*}{\partial p(\omega)}\frac{\partial p(\omega)}{\partial p_R} = -\frac{1-\omega}{1-\alpha}\frac{C^*}{p(\omega)} < 0.$$

Also,

$$\frac{\partial}{\partial p_R}\left(\frac{p_L - p_R}{p(\omega)}\right) = \frac{-p(\omega) - (p_L - p_R)(1-\omega)}{p(\omega)^2} = -\frac{p_L}{p(\omega)^2} < 0,$$

so

$$\frac{\partial B}{\partial p_R} = -A\frac{\partial}{\partial p_R}\left(\frac{p_L - p_R}{p(\omega)}\right) = \frac{\omega}{1-\alpha}\frac{p_L}{p(\omega)^2} > 0.$$

Thus

$$\frac{\partial}{\partial p_R}\left(\frac{dC_L^*}{d\tau}\right) = \omega'\left[\underbrace{\frac{\partial C^*}{\partial p_R}B}_{<0} + \underbrace{C^*\frac{\partial B}{\partial p_R}}_{>0}\right],$$

which is of ambiguous sign. □