Chokepoint Compute: Export Controls, Substitution, and the Welfare Cost of Fragmenting the AI Hardware Stack

Joel N. Christoph
FernUniversität in Hagen
joel.christoph@proton.me

September 2026

Question

Since October 2022 the United States has restricted exports of advanced AI accelerators and the equipment used to make them, with successive revisions through 2025. Existing work treats these measures largely as trade policy. This paper models them as a restriction on access to a productive input, which puts them closer to a capital control than to a tariff, and asks how the resulting distortion falls on economies that depend on foreign compute for production.

Model

A downstream economy produces using compute services drawn from two sources. A dominant stack is governed by a gatekeeper jurisdiction that can restrict access. An alternative stack is available at a lower quality-adjusted level. The two combine through a constant-elasticity aggregator with elasticity σ. The gatekeeper chooses a restriction that acts as a wedge on the price of dominant-stack compute.

Two comparative statics follow directly. The wedge the gatekeeper can impose before the downstream economy switches away rises with the dominant stack's expenditure share s, and falls with σ. Restricting access therefore has a dynamic counterpart: it raises the return to investment in the alternative stack, lowering s in later periods and reducing the wedge available in the future.

Results

First, restriction is self-limiting. Higher current restriction raises alternative-stack investment, which lowers the gatekeeper's future position. The gatekeeper faces a trade-off between present restriction and future access to the chokepoint.

Second, third-country re-export caps the binding wedge at the smuggling premium. Once diversion through intermediary jurisdictions is available at a premium τ, the effective restriction cannot exceed τ regardless of how broad the listed items are. Enforcement intensity, not the breadth of the control list, determines the effective restriction.

Third, separation into distinct compute blocs lowers aggregate welfare relative to an integrated market, because each bloc forgoes gains from specialisation in hardware production. An importer able to source from both blocs collects a neutrality dividend, since it faces the lower of the two quality-adjusted prices while contributing to neither bloc's fixed investment.

Calibration

China's sourcing of AI accelerators shifted from a dominant-stack expenditure share near 0.95 in 2022 to roughly 0.55 in 2025, following the successive control revisions. Matching that shift within the model implies σ between 3 and 5. At those values, full fragmentation produces an output loss on the order of 10 percent for an economy fully dependent on the dominant stack, falling toward 3 percent as the alternative stack approaches quality parity. The calibration is illustrative and treats the observed sourcing shift as driven by the controls.

Identification in the full draft

The full draft separates two components of the observed price movement. Quality improvement in the alternative stack lowers its effective price, while the control wedge raises the effective price of the dominant stack. Measured grey-market premia for restricted accelerators identify the second component directly. Estimation uses an event study around the control revisions of October 2022, October 2023, December 2024 and the 2025 amendments, combined with mirror statistics comparing reported exports from supplier jurisdictions against reported imports in destination and intermediary jurisdictions over 2022 to 2026.

Relevance to emerging and frontier markets

Most economies outside the supplier jurisdictions are in the position modelled here: compute-intensive production depends on hardware governed by another state. The framework gives the conditions under which such an economy faces a binding restriction, the conditions under which diversion limits it, and the size of the output loss under separated blocs. It also identifies which economies gain from remaining outside either bloc.

Disclosure

The author is employed by FernUniversität in Hagen. Related research on compute governance is supported by a grant from Coefficient Giving for a compute governance training programme, and separate research on multilingual AI is supported by a grant from Imminent (Translated Research Center). The author holds no financial relationship with any semiconductor manufacturer, AI developer, or firm subject to the export controls analysed here, and receives no support from any government. The paper is unpublished and not under consideration elsewhere.