Identification and Estimation of Semi-nonparametric Gravity Equation

Ryo Kimoto

October 28, 2026

Abstract

The Homothetic with Single Aggregator (HSA) demand system developed by Matsuyama and Ushchev (2017) is a nonparametric generalization of the Constant Elasticity of Substitution (CES) demand system, a core building block of many models of international trade. We study the identification and estimation of the HSA demand system in a simple endowment economy commonly called an Armington model. We specify an explicit data-generating process based on repeated equilibrium outcomes. The restrictions imposed on the entire matrix of bilateral trade flows by utility maximization and market clearing identify the demand system under the stated conditions. We use this result to construct a sieve estimator and establish its weak consistency. We also develop a statistical test of the CES restriction and establish the large-sample properties of counterfactual estimators commonly used by trade economists. A simulation study shows favorable finite-sample performance. An application to the World Input–Output Tables (Timmer et al., 2012) illustrates estimation of the demand system and comparison with CES using annual trade data.

Introduction

Quantitative gravity models use observed bilateral trade flows to evaluate the effects of trade policy on trade and welfare (Costinot and Rodríguez-Clare, 2014; Yotov, 2024). These exercises require solving a system of market-clearing conditions for a counterfactual equilibrium. Their conclusions depend on how consumers respond when access to foreign goods changes. The Constant Elasticity of Substitution (CES) demand system makes this calculation tractable, but imposes strong restrictions on substitution patterns. In particular, its Independence of Irrelevant Alternatives property restricts how expenditure shifts between competing suppliers. In the gross-substitutes case commonly used in quantitative trade, consumers can substitute toward domestic goods as imports become more expensive, and no imported good is essential. These restrictions can therefore matter for the welfare consequences of trade barriers (Matsuyama, 2023; Ossa, 2015).

In this paper, we study the identification and estimation of the Homothetic with Single Aggregator (HSA) demand system proposed by Matsuyama and Ushchev (2017) in one of the simplest quantitative general-equilibrium models of international trade, commonly referred to

in the literature as the Armington model (Armington, 1969). HSA is a nonparametric generalization of CES that retains homotheticity while allowing more flexible substitution patterns. Our contribution is to show how restrictions on the entire bilateral trade-flow matrix identify the unknown demand system and provide a basis for constructing an estimator. We develop an explicit sampling framework, construct a sieve estimator, and establish large-sample properties that give estimation and counterfactual analysis a statistical foundation. The simple endowment economy allows us to isolate the role of general equilibrium restrictions without introducing additional supply-side mechanisms.

The key source of information is the joint determination of bilateral trade flows in each realization of the general equilibrium. Utility maximization requires expenditure shares to sum to one, while market clearing requires each country's total sales to equal its income. The same equilibrium price vector links these restrictions across the entire matrix. A candidate demand system must therefore account for all bilateral outcomes jointly, rather than fit each bilateral relationship in isolation. Under suitable support and rank conditions, variation in equilibrium prices across observations, together with the restrictions imposed by utility maximization and market clearing, identifies the HSA demand system.

To make this argument statistical, we specify what generates each observation and what is repeated in the sampling experiment. We observe independent equilibrium outcomes for a fixed set of N countries and let the number of realizations R increase. This differs from conventional gravity asymptotics based on a growing number of countries. Holding N fixed serves two purposes, corresponding to utility maximization and market clearing. First, adding countries in an Armington model also adds goods, changing the dimensionality of the consumption set over which the utility function is defined. Thus, the demand system obtained by solving the utility maximization problem need not remain stable along the sequence $N \to \infty$. Second, the standard fixed-point argument establishes equilibrium existence for each finite N, but does not by itself establish existence in a limiting economy. Our asymptotic experiment keeps the underlying economic environment fixed, thereby fixing the object of interest—the demand system—while generating repeated equilibrium observations. Section 2 develops this distinction.

Based on the identification result, we construct a sieve least-squares estimator of the unknown demand system. Although we approximate the unknown share function by a linear combination of basis functions, the HSA structure complicates estimation because the single price aggregator—the SA in HSA—makes the criterion nonlinear in the sieve coefficients. The aggregator is implicitly determined by the adding-up restriction and must be recomputed at each candidate demand system. Establishing large-sample properties therefore requires controlling this dependence uniformly over the sieve parameter space. We derive entropy bounds and conditions on the tuning parameters that establish weak consistency and a convergence rate for the estimator. We do not claim that this rate is optimal.

The framework also supports inference and counterfactual analysis. We use the sharp characterization of CES in Proposition 1, Assertion (iii), of Matsuyama and Ushchev (2017) to de-

velop a statistical test of the CES restriction. To study the test under the CES null, we establish point identification of the elasticity of substitution, which conventional quantitative applications often calibrate using estimates from the literature (Costinot and Rodríguez-Clare, 2014; Head and Mayer, 2014). We then establish the large-sample properties of counterfactual estimators commonly used by trade economists. The explicit data-generating process makes it possible to interpret these calculated outcomes as estimators of the unobserved counterfactual outcomes, rather than solely as solutions to a calibrated model.

A simulation study evaluates the estimator, the test, and the counterfactual procedure. In the designs considered, the HSA estimator recovers the trade-cost parameters and the distribution of the unobserved counterfactual outcomes under both CES and HSA demand. When CES is misspecified, its estimates instead target pseudo-true values and its counterfactual distribution can differ from the truth. The test has favorable power against the alternatives considered, although its size can be conservative. These exercises illustrate both the potential gains from estimating demand flexibly and the finite-sample qualifications relevant to inference.

We illustrate estimation using annual observations from the World Input–Output Tables (Timmer et al., 2012), a dataset frequently used by trade economists. The application compares CES with HSA across sieve dimensions and examines the sensitivity of the estimated trade-cost parameters and demand system. The main empirical specification restricts the inverse-power coefficients to be nonnegative, producing a globally admissible demand system. We report marginal and general-equilibrium welfare effects under this specification and CES, with signed-coefficient estimates as robustness comparisons.

After a brief review of the related literature, the paper proceeds as follows. Section 1 presents the HSA gravity model. Section 2 specifies the data-generating process and sampling framework. Section 3 establishes identification and develops the sieve estimator. Section 4 studies the Matsuyama–Ushchev test. Section 5 studies counterfactual estimators. Section 6 presents the simulation evidence. Section 7 presents the empirical application. Section 8 concludes.

Related Literature

Our paper relates to work on HSA demand, flexible gravity, counterfactual analysis, and semiparametric identification and estimation.

Following their development of HSA in Matsuyama and Ushchev (2017), Matsuyama and Ushchev use the demand system to study constant pass-through, innovation dynamics, and firm selection and sorting in Matsuyama and Ushchev (2020), Matsuyama and Ushchev (2022a), and Matsuyama and Ushchev (2022b). HSA has since begun to be adopted by other researchers for theoretical analysis. Fujiwara and Matsuyama (2026) use HSA to study the relationship between competition and the Phillips curve. Grossman, Helpman, and Lhuillier (2023) use it to study optimal policy in the presence of supply-chain disruptions. These applications show that replacing CES with HSA can generate qualitatively different implications while preserving

theoretical tractability. See Matsuyama (2023) for a review.

Our paper also relates to efforts to relax the substitution restrictions imposed by CES, particularly independence of irrelevant alternatives. Relaxing these restrictions need not entail a nonparametric demand system. On the demand side of quantitative gravity, existing work has introduced both parametric and shape-based alternatives. Novy (2013) develops and estimates translog gravity using the parametrization of translog demand proposed by Feenstra (2003). Building on this translog structure, Chen and Novy (2022) use a finite-dimensional specification in which a trade-cost variable is interacted with a predicted bilateral import share. They also estimate the translog gravity specification directly. As Matsuyama (2023) shows, the resulting translog demand system, under suitable restrictions, is a parametric subclass of the broader HSA class and is therefore covered by our setup. Fajgelbaum et al. (2024) use translog demand to study trade reallocation following the US–China trade war. Carrère, Mrázová, and Neary (2020) theoretically study subconvex demand rather than estimate a particular demand system. Subconvexity roughly corresponds to a property of demand systems called Marshall's Second Law of Demand (see Matsuyama, 2023). The intersection of the HSA class and the subconvex demand class is nonempty, as illustrated by Grossman, Helpman, and Lhuillier (2023). When the true demand system belongs to this intersection, our framework can recover it nonparametrically.

A different approach introduces heterogeneity in consumer preferences. Piveteau and Smagghue (2024) aggregate CES demands across consumers with heterogeneous substitution elasticities and develop a tractable estimation procedure for the resulting mixed-CES system. This relaxes IIA while retaining CES demand at the individual-consumer level. Head and Mayer (2026) examine the consequences of preference heterogeneity for counterfactual predictions. These approaches differ from HSA, which retains a single common price aggregator while allowing unknown share functions. Our framework does not encompass mixed CES in general.

Within the HSA class, Baqaee, Farhi, and Sangani (2024) nonparametrically calibrate an HSA demand system in a model with monopolistic competition, heterogeneous markups, and fixed costs to quantify how market size affects welfare through reallocations, firm exit, and changing markups. We differ from this work in two respects. First, we study a gravity equation implied by HSA demand in a general-equilibrium model of international trade. Second, we establish frequentist guarantees for estimators of the HSA demand system and for counterfactual estimators constructed using the estimated demand system.

Flexible gravity has also been developed through the supply side. Lind and Ramondo (2023) allow productivity to be correlated across countries in a Ricardian model, generating richer substitution patterns through technology rather than a relaxation of consumer demand. Adão, Arkolakis, and Ganapati (2026, *JPE* forthcoming) use semiparametric gravity equations to estimate nonparametric elasticity functions that summarize firm entry and export decisions. Their approach introduces nonparametric flexibility through the supply side while maintaining CES demand. Our nonparametric relaxation instead concerns the demand system itself.

At a more general level, Adão, Costinot, and Donaldson (2017) develop nonparametric

counterfactual analysis using a reduced factor demand system and implement their approach with mixed-CES factor demand. Their object is the demand for factor services implied jointly by preferences and technology. We instead identify and estimate the HSA consumer demand system within an endowment economy and study the large-sample properties of its estimator under repeated equilibrium observations.

Our counterfactual analysis relates to work on parametric structural gravity models. Yotov et al. (2016) describe the standard implementation, while Anderson, Larch, and Yotov (2018) develop a procedure that uses Poisson pseudo-maximum likelihood estimates to compute general-equilibrium counterfactuals. Allen, Arkolakis, and Takahashi (2020) establish existence, uniqueness, and robustness of counterfactual changes in a broad class of gravity models. Anderson and Yotov (2016) construct confidence intervals for estimated general-equilibrium effects, and Pfaffermayr (2020) derives the asymptotic distribution of counterfactual comparative statics using the delta method. We complement this literature by establishing the large-sample properties of counterfactuals constructed from a nonparametrically estimated HSA demand system.

Finally, our paper relates to the econometric literature on semiparametric identification and estimation. Chow, Kasahara, and Sugita (2026) identify production functions and consumer demand from firm revenue data under monopolistic competition without observing firm-level output quantities. Under the HSA restriction, they also identify consumer preferences. Their empirical implementation uses the Constant Pass Through (CoPaTh) subclass of HSA proposed by Matsuyama and Ushchev (2020), assesses the CES restriction through its curvature parameter, and quantifies the welfare costs of market power using short-run partial-equilibrium counterfactuals holding factor prices and aggregate expenditure fixed. We use variation across repeated equilibria in cross-country trade flows, together with general-equilibrium restrictions, to identify HSA demand. Our trade-policy counterfactuals allow equilibrium prices and incomes to adjust, and we establish consistency of the resulting counterfactual estimators under our sampling assumptions.

A second econometric connection is to classical single-index models. HSA has a single-index-like structure because the conditional mean of each bilateral expenditure share is an unknown function of a scalar effective relative price. This reduction to a scalar argument resembles the classical single-index models studied by Ichimura (1993), Powell, Stock, and Stoker (1989), and Klein and Spady (1993). For a bilateral outcome, let $\lambda_{ij,(r)}$ denote the expenditure share of the variety supplied by country (i) to country (j) in realization (r), and let $\mathbf{P}_{0,(r)} = (P_{01,(r)}, \ldots, P_{0N,(r)})$ denote the equilibrium price vector in that realization. In stripped-down notation, the HSA demand system implies

$$\mathbb{E}\left[\lambda_{ij,(r)} \mid \mathbf{P}_{0,(r)} = \mathbf{p}\right] = F_0\left(\frac{p_i g(W_{ij}'\theta_0)}{A_j(\mathbf{p}, \theta_0, F_0, g)}\right).$$

Here, F_0 and θ_0 are unknown, g is known, and A_j is a known scalar-valued functional.

The analogy is incomplete in three ways. The index is nonlinear in θ_0. The unknown function F_0 enters both as the link function and through A_j. Finally, A_j depends on the entire equilibrium price vector, so every bilateral share is linked to the other shares in the same equilibrium realization. We use these general-equilibrium restrictions to identify F_0 and θ_0. Although we approximate F_0 with a linear sieve, the criterion is nonlinear in the sieve coefficients because A_j must be recomputed at every candidate. We derive an entropy bound for the resulting loss class and provide conditions on the sieve dimension and coefficient restrictions under which the estimator is consistent and admits a global rate of convergence.

1 Gravity under the Homothetic with Single Aggregator (HSA) Demand System

In this section, we introduce the HSA demand system into a simple endowment economy of international trade commonly referred to as the Armington (1969) model. As the demand system is no longer parametrically specified, we arrive at a semi-nonparametric gravity equation[1]. Since HSA includes the popular CES demand model (c.f. Anderson and Van Wincoop (2003)) as a special case, our semi-nonparametric gravity equation includes the classical CES gravity equation as a special case.

The model has two layers. Conditional on prices, utility maximization determines bilateral trade flows and imposes an adding-up restriction for each destination. Market clearing then jointly determines the common equilibrium price vector and imposes a second set of restrictions across exporters and destinations. Consequently, the bilateral trade flows in an equilibrium realization are jointly determined rather than separate outcomes.

This section therefore constructs the equilibrium mapping that serves as the economic foundation of the data-generating process in Section 2.

We begin by describing the demand side of the economy. We assume that there are $N < \infty$ countries, with each country inhabited by a representative household indexed by $i = 1, \ldots, N$. The economy has N goods and each consumer i has a preference over the N goods. Let $U(q_1, \ldots, q_N)$ denote the utility function representing the consumer's taste over N varieties $\mathbf{q} = (q_1, \ldots, q_N)$. Each consumer i is endowed with $L_i > 0$ units of variety i and does not own any other variety $j \neq i$[2]. Under this setup, the country index $i = 1, \ldots, N$ represents both the country and the good/variety.

Consider a utility maximization problem (UMP) of a representative consumer in country $i = 1, \ldots, N$.

$$\max_{q_{1i}, q_{2i}, \cdots, q_{Ni}} U(q_{1i}, q_{2i}, \ldots, q_{Ni}),$$

[1] We follow Chen (2007) and refer to a model where both the finite-dimensional and infinite-dimensional parameters are objects of interest as semi-nonparametric, as opposed to a semiparametric model where the latter is treated as a nuisance parameter.

[2] Alternatively, we can assume that (1) each country is endowed with L_i units of labor, and (2) the production of the good is linear in L_i.

subject to the budget constraint $\sum_{j=1}^{N} p_{ji}q_{ji} \leq E_i$. Here, q_{ji} represents the quantity i imports from country j, p_{ji} is the price for the imported good q_{ji}, and $E_i > 0$ is the total expenditure of the representative agent in country i. Each consumer may sell its good to countries $j = 1, \ldots, N$ at prices p_{ij}. The value of country i's endowment at the factory-gate price is $p_i L_i$. In equilibrium, household expenditure therefore satisfies $E_i = p_i L_i$. We assume that the preference represented by the utility function $U(\cdot)$ is symmetric across countries. The consumers $i = 1, \ldots, N$ are assumed to take prices $\mathbf{p}_i = (p_{1i}, \ldots, p_{Ni}) \in \mathbb{R}^N_{++}$ as given.

We assume that trading across borders is costly. To this end, we employ the so-called "Iceberg" trade cost specification. We assume that we have $p_{ij} = p_i \tau_{ij}$ so instead of having N^2 prices $\{\mathbf{p}_i\}_{i=1}^N$, we have an N-dimensional price vector $\mathbf{p} \in \mathbb{R}^N_{++}$ and N^2 trade costs $\{\tau_{ji}\}_{j,i}$, which we treat as parameters. Formally, for country j to deliver quantity $q_{ji} > 0$ to country i, it must ship $\tau_{ji} q_{ji}$ units of the good. The difference between what is shipped to i and what i actually receives is assumed to be lost, as if it had melted away (hence the name "Iceberg"). We will assume that the importer bears the cost for this lost shipment.

The trade costs are assumed to satisfy the following:

Assumption 1. *The trade cost parameter $\{\tau_{ji}\}_{j,i}$ satisfies:*

1. *For all distinct i, j, we have $\tau_{ji} > 1$.*
2. *For all three distinct indices i, j, and ℓ, we have $\tau_{ij} < \tau_{i\ell}\tau_{\ell j}$.*
3. *For all i, $\tau_{ii} = 1$.*

Assumption 1 imposes natural restrictions on the trade costs. First, $\tau_{ji} > 1$ for all $j \neq i$ means that it is costly to export goods across a border. Second, the triangle inequality $\tau_{ij} < \tau_{i\ell}\tau_{\ell j}$ ensures that it is always cheaper to buy goods directly from the source. Lastly, the condition $\tau_{ii} = 1$ for all i is a normalization of the trade cost. Let us collect the trade costs in a (N, N) real matrix $\mathbf{T}$ where the (j, i) element of $\mathbf{T}$ is τ_{ji}.

For there to be a solution to the UMP, certain regularity conditions must be imposed on $U(\cdot)$. We will assume that $U : \mathbb{R}^N_+ \to \mathbb{R}$ is continuous, strongly increasing, and strictly quasiconcave. Then by a basic result from consumer theory (cf. Theorem 5.1 in Jehle and Reny (2011)), the UMP has a unique solution for each $\mathbf{p}_i \in \mathbb{R}^N_{++}$ and a continuous demand function. Since every element in $\mathbf{T}$ is greater than or equal to 1, it follows that the demand function exists for every $\mathbf{p} \in \mathbb{R}^N_{++}$.

Let us write $q_{ji}(\mathbf{p}, E_i \mid \mathbf{T})$ to represent the Marshallian demand function obtained as a solution to the UMP for a given matrix of trade costs $\mathbf{T}$. We will restrict the demand function $q_{ji}(\cdot, \cdot \mid \cdot)$ to a class of demand system named Homothetic with Single Aggregator (HSA) introduced by Matsuyama and Ushchev (2017).

The HSA demand system is characterized by

$$q_{ji}(\mathbf{p}, E_i \mid \mathbf{T}) = \frac{E_i}{p_j \tau_{ji}} S_j \left(\frac{p_j \tau_{ji}}{A_i} \right), \tag{1}$$

where $S_j(\cdot)$ denotes the expenditure share function of good j in country i which satisfies several regularity conditions described below in Assumption 2. The constant A_i is defined implicitly as a solution to

$$\sum_{j=1}^{N} S_j\left(\frac{p_j \tau_{ji}}{A_i}\right) = 1. \tag{2}$$

It is important to note that the implicitly defined constant A_i depends not only on the prices $\mathbf{p}$ and the trade costs $\mathbf{T}$, but also on the share functions $\mathbf{S} = (S_1, \ldots, S_N)$. Thus, we would sometimes write $A_i(\mathbf{p}, \mathbf{S}, \mathbf{T})$ to make its dependence on $(\mathbf{p}, \mathbf{T}, \mathbf{S})$ explicit when the notation does not become too cumbersome. We also use notation such as $g(\mathbf{p} \mid \mathbf{S}, \mathbf{T})$ to emphasize that $\mathbf{S}$ and $\mathbf{T}$ are held fixed.

The aggregator A_i is therefore not a free destination-specific term. It is jointly determined by all effective prices faced by destination i and by the entire demand system. A change in one effective price or one share function changes A_i and thereby changes every predicted bilateral share for that destination. This is the first source of cross-equation restrictions in the model.

Remark 1. Note that we are allowing the share functions S_i to be different for each good i, but maintaining the symmetry of the underlying preference over the goods across countries. That is, if the expenditure share for variety i is given by the share function $S_i(\cdot)$ for one consumer, then the same function $S_i(\cdot)$ should also give the share for all other consumers.

Remark 2. Homotheticity of the demand system allows us to write the expenditure share allocated to good sourced from j by the share function $S_j(\cdot)$. The name "Single Aggregator" comes from the fact that the share function $S_j(\mathbf{p} \mid \mathbf{T})$ takes a single argument $p_j \tau_{ji} / A_i(\mathbf{p} \mid \mathbf{S}, \mathbf{T})$.

As we can see from the argument of the share function $S_j(\cdot)$, the expenditure share of good j in country i is determined by the relative price between the effective price p_{ji} and the single aggregator $A_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})$. The single aggregator $A_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})$ serves as a summary measure of the cross-variety competitive effects.

A related but distinct object is the ideal price index $P_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})$. The ideal price index is related to the indirect utility $V_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})$ via the relationship $V_i = E_i P_i^{-1}$. Matsuyama and Ushchev (2017) showed that the two objects are related via the relationship:

$$\ln P_i = \ln A_i + \sum_j \int_c^{p_j \tau_{ji} / A_i} \frac{S_j(z)}{z} \mathrm{d}z, \ (c > 0), \tag{3}$$

and crucially, when the number of goods N exceeds 2, $P_i \propto A_i$ if and only if the underlying aggregator/demand system is CES.

We restrict the space in which the share functions $S_1, S_2, \ldots, S_N$ lie.

Assumption 2. *For each $i = 1, \ldots, N$, the share function S_i is an element of the set $\mathcal{S}_i$ defined by:*

$$\mathcal{S}_i = \left\{ g : \mathbb{R}_{++} \to I_i \subset \mathbb{R}_+ \;\middle|\; \begin{array}{l} g \in C^2(\mathbb{R}_{++}) \text{ and } g'(z) < 0 \text{ for every } z \in \mathbb{R}_{++}, \\ \text{and } \lim_{z \to 0^+} g(z) > 1/N > \lim_{z \to \infty} g(z) \end{array} \right\}.$$

Let us denote the infinite-dimensional parameter space by $\mathcal{S} \equiv \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$.

Assumption 2 is a sufficient condition for the demand system to be well-defined. Proposition 1 in Matsuyama and Ushchev (2017) states that when (1) $zS_i'(z) < S_i(z)$, (2) $S_i(z_i)S_j(z_j) \geq 0$, the demand system described by equations (1) and (2) is well defined. This is clearly satisfied for any $\mathbf{S} = (S_1, \ldots, S_N) \in \mathcal{S}$ by the strict monotonicity requirement.

The condition $\lim_{z\to 0^+} \sum_i S_i(z) > 1 > \lim_{z\to\infty} \sum_i S_i(z)$ is sufficient to ensure that the price aggregator A_j exists. If such a constant exists, the monotonicity of S_i implies that it is unique. However, we work with the stronger condition $\lim_{z\to 0^+} S_i(z) > N^{-1} > \lim_{z\to\infty} S_i(z)$ for all i for reasons that become apparent when we study the problem of identification in Section 3.

We also impose twice continuous differentiability of the share functions to ensure that the parameters in the CES model, which is a special case of the HSA model, are asymptotically normal (Subsection 3.1).

The demand system described by the system (1) and (2) covers the CES demand system as a special case. The following example is taken from Matsuyama (2023):

Example 1. Let $\varepsilon \in [0, 1)$, and $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}_+^N$ be a vector of Cobb–Douglas weights satisfying $\sum_i \alpha_i = 1$. Let $\sigma > 1$. Then, consider a utility function parametrized by a CES–Cobb–Douglas mixture as in

$$U(q_{1i}, \ldots, q_{Ni}) = \left(\prod_{j=1}^{N} q_{ji}^{\alpha_j}\right)^{\varepsilon} \left(\sum_{j=1}^{N} q_{ji}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}(1-\varepsilon)}.$$

This produces a demand system given by

$$S_j(z) = \varepsilon\alpha_j + (1-\varepsilon)z^{1-\sigma}$$

with the price aggregator for destination i given by $A_i = \left(\sum_j (p_j\tau_{ji})^{1-\sigma}\right)^{1/(1-\sigma)}$.

Up to the fixed normalization constant, the ideal price index P_i is given by

$$P_i \propto A_i^{1-\varepsilon} \left(\prod_{j=1}^{N} (p_j\tau_{ji})^{\alpha_j}\right)^{\varepsilon}.$$

This example is interesting because the demand system allows the goods to be essential when $\alpha_j > 0$ while simultaneously being gross substitutes when $\sigma > 1$. Consequently, when every country transitions into autarky, countries that lose access to essential goods would suffer greatly, as is evident from the fact that $P_i \to \infty$, which implies that the indirect utility $V_i \to 0$.

Remark 3. Setting $\varepsilon = 0$ in Example 1, we obtain the CES model studied, for example, by Anderson and Van Wincoop (2003) and in Section 2.1 of the handbook chapter by Costinot and Rodríguez-Clare (2014).

Define the trade flow from j to i as $X_{ji} \equiv p_{ji}q_{ji}$. Then the demand function (1) can be written

as

$$X_{ji} = S_j \left(\frac{p_j \tau_{ji}}{A_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})} \right) E_i. \tag{4}$$

Equation (4) is our gravity equation.

It is worth noting that Equation (2) imposes a normalization such that summing the trade flows (4) over exporters j, the budget constraint is satisfied. This can be seen by

$$\sum_{j=1}^{N} X_{ji} = E_i \sum_{j=1}^{N} S_j \left(\frac{p_j \tau_{ji}}{A_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})} \right) = E_i. \tag{5}$$

Equation (5) is the destination-side adding-up restriction. Because A_i is common to every exporter serving destination i, the predicted bilateral trade flows in a row of the trade-flow matrix must be determined simultaneously.

Next, we describe the supply side of our economy. We have assumed that each representative consumer in country j is endowed with $L_j > 0$ units of good j which is assumed to be of a different variety to the good L_i, for all $i \neq j$. Iceberg trade costs specify that $\tau_{ij} q_{ij}$ units must be shipped for q_{ij} units to arrive at the destination. This is represented in our model as

$$L_i = \sum_{j=1}^{N} \tau_{ij} q_{ij}(\mathbf{p}, E_j \mid \mathbf{T}, \mathbf{S}) \tag{6}$$

for all i.

The revenue of consumer i is given by

$$\begin{aligned} \pi_i &= \sum_{j=1}^{N} X_{ij}, \\ &= \sum_{j=1}^{N} p_{ij} q_{ij}(\mathbf{p}, E_j \mid \mathbf{T}, \mathbf{S}). \end{aligned} \tag{7}$$

So combining (6), (7), we have

$$\begin{aligned} \pi_i &= p_i \sum_{j=1}^{N} \tau_{ij} q_{ij}(\mathbf{p}, E_j \mid \mathbf{T}, \mathbf{S}), \\ &= p_i L_i \ (= E_i). \end{aligned} \tag{8}$$

Equation (8) is a natural restriction which states that revenue and expenditure coincide country by country, so that $\pi_i = E_i = p_i L_i$ for every $i = 1, \ldots, N$. Equivalently, total exports of country i satisfy

$$\sum_{j=1}^{N} X_{ij} = p_i L_i.$$

This is the exporter-side market-clearing restriction. Together with Equation (5), it links the rows and columns of the entire bilateral trade-flow matrix.

Figure 1 arranges destinations in rows and exporters in columns. Each row satisfies the

adding-up restriction imposed by utility maximization through the destination-specific aggregator. Each column satisfies the market-clearing restriction that equates the exporter's revenue with its expenditure.

	Exporter 1	Exporter 2	$\cdots$	Exporter N	Utility maximization and adding up
Destination 1	X_{11}	X_{21}	$\cdots$	X_{N1}	$\sum_{j=1}^{N} X_{j1} = E_1$
Destination 2	X_{12}	X_{22}	$\cdots$	X_{N2}	$\sum_{j=1}^{N} X_{j2} = E_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Destination N	X_{1N}	X_{2N}	$\cdots$	X_{NN}	$\sum_{j=1}^{N} X_{jN} = E_N$
Market clearing	$\sum_{i=1}^{N} X_{1i} = p_1 L_1$	$\sum_{i=1}^{N} X_{2i} = p_2 L_2$	$\cdots$	$\sum_{i=1}^{N} X_{Ni} = p_N L_N$	Revenue equals expenditure

Figure 1: General-equilibrium restrictions on the bilateral trade-flow matrix

Every bilateral trade flow X_{ji} lies at the intersection of one destination row and one exporter column. It must therefore satisfy both sets of restrictions simultaneously. This intersection is precisely the general-equilibrium information that we use for identification.

We introduce additional notation to define the equilibrium of the model. Let us collect the endowments into a vector $\mathbf{L} \equiv (L_1, \ldots, L_N) \in \mathbb{R}^N_{++}$, and define the excess demand function for good $j = 1, \ldots, N$ by

$$Z_j(\mathbf{p} \mid \mathbf{L}, \mathbf{T}, \mathbf{S}) := \sum_{i=1}^{N} \tau_{ji} q_{ji}(\mathbf{p}, E_i \mid \mathbf{L}, \mathbf{T}, \mathbf{S}) - L_j, \tag{9}$$

$$= \sum_{i=1}^{N} p_j^{-1} S_j \left(\frac{p_j \tau_{ji}}{A_i(\mathbf{p} \mid \mathbf{T}, \mathbf{S})} \right) p_i L_i - L_j, \tag{10}$$

where we used equations (1) and (7) in the second equality.

Definition 1. The equilibrium is a price vector $\mathbf{p}^*(\mathbf{L}, \mathbf{T} \mid \mathbf{S}) \in \mathbb{R}^N_{++}$ that satisfies the following conditions.

1. The excess demand function $Z_j(\mathbf{p}^* \mid \mathbf{L}, \mathbf{T}, \mathbf{S})$ given in (9) is zero for all j.
2. The scale normalization $\sum_{i=1}^{N} p_i^* = 1$.

We defer the argument for existence and uniqueness up to normalization until Section 2.

Conditional on the endowment vector, the market-clearing conditions jointly determine the equilibrium price vector. The same price vector enters every bilateral trade flow and every destination-specific aggregator. Each equilibrium realization therefore produces an entire trade-flow matrix satisfying both destination-side adding-up and exporter-side market-clearing restrictions. These general-equilibrium restrictions provide the identifying information studied in Section 3.

2 Data-Generating Process

Section 1 defines the equilibrium mapping for a fixed endowment vector, trade-cost matrix, and HSA demand system. This section turns that deterministic economic model into a statistical model. We parameterize the trade-cost matrix, draw the endowment vector from a fixed distribution, and use the equilibrium mapping to generate prices and the bilateral trade-flow matrix. Our asymptotic experiment is obtained by repeating this process independently.

We first specify how each equilibrium observation is generated and what the researcher observes. We then introduce the sampling framework and explain why our asymptotic argument holds the number of countries N fixed while increasing the number of equilibrium realizations. Next, we relate the general-equilibrium restrictions to the row and column restrictions on the trade-flow matrix introduced in Section 1. Finally, we compare our sampling framework with conventional CES gravity estimation.

2.1 Data-Generating Process

To turn the equilibrium model in Section 1 into a data-generating process, we first distinguish the fixed design, structural parameters, and random primitives. We begin by parameterizing the bilateral trade costs, which are unknown to the researcher. We use the parametric specification

$$\tau_{ij}(\theta) = \exp(W_{ij}'\theta)$$

for all $i \neq j$, where $W_{ij} \in \mathbb{R}^d$ is fixed and known. We set $W_{ii} = 0$, so that $\tau_{ii}(\theta) = 1$. Let

$$\mathbf{W}_j = (W_{1j}, \ldots, W_{j-1,j}, W_{j+1,j}, \ldots, W_{Nj})' \in \mathbb{R}^{(N-1)\times d},$$

and collect the fixed design points in $\mathcal{W} = \{\mathbf{W}_1, \ldots, \mathbf{W}_N\}$.

The feasible parameter space incorporates the restrictions on iceberg trade costs imposed in Assumption 1. Define

$$\Theta_1 = \left\{\theta \in \mathbb{R}^d \mid W_{ij}'\theta > 0 \text{ for all } i \neq j\right\}$$

and the parameter restriction

$$\Theta_2 = \left\{\theta \in \mathbb{R}^d \mid (W_{ij} - W_{i\ell} - W_{\ell j})'\theta < 0 \text{ for all distinct } i, j, \ell\right\}.$$

We set $\Theta = \Theta_1 \cap \Theta_2$. The first restriction gives $\tau_{ij}(\theta) > 1$ for $i \neq j$. The second gives the strict triangle inequality $\tau_{ij}(\theta) < \tau_{i\ell}(\theta)\tau_{\ell j}(\theta)$. The feasible set need not be bounded. A compact subset of this set is introduced in Section 3 for estimation.

We impose a sufficient condition ensuring that the feasible set is nonempty and full dimensional. For a vector Y, let $Y^{(k)}$ denote its k-th element.

Assumption 3. *The fixed design points satisfy the following conditions.*

1. *For every $i \neq j$, the first element $W_{ij}^{(1)}$ is strictly positive.*

2. *For every three distinct indices i, j, ℓ,*

$$W_{ij}^{(1)} < W_{i\ell}^{(1)} + W_{\ell j}^{(1)}.$$

Under Assumption 3, Θ is a nonempty, full-dimensional convex cone. Appendix A.1 provides the argument. Bilateral distance is a natural variable satisfying the two conditions. It is strictly positive across countries and satisfies the triangle inequality.

The random primitive of the equilibrium model is the endowment vector $\mathbf{L}$. Its distribution $\mu_{\mathbf{L}}$ is fixed and has compact support

$$\mathbb{S}(\mathbf{L}) \subset \mathbb{R}_{++}^{N}.$$

Write $\mathcal{I} = (\mathcal{W}, \mathbf{L})$ and

$$\mathbb{S}(\mathcal{I}) = \{\mathcal{W}\} \times \mathbb{S}(\mathbf{L}).$$

The structural parameter is $\beta = (\theta, \mathbf{S}) \in \mathcal{B}$, where $\mathcal{B} = \Theta \times \mathcal{S}$.

The next proposition ensures that the equilibrium mapping used in the DGP is well defined.

Proposition 1. *For the fixed and known $\mathcal{W}$, any $\ell \in \mathbb{S}(\mathbf{L})$, and any $\beta \in \mathcal{B}$, the equilibrium price vector exists uniquely up to normalization. Under the normalization in Definition 1, it lies in the interior of the unit price simplex $\Delta_{\mathbf{P}} \setminus \partial\Delta_{\mathbf{P}}$.*

Proof. See Appendix A.2. □

Fix the true parameter $\beta_0 = (\theta_0, \mathbf{S}_0)$, where $\mathbf{S}_0 = (S_{01}, \ldots, S_{0N})$. Let ν_ε denote the fixed distribution of a scalar measurement error. The moment and support restrictions on this distribution are stated in Assumption 4.

Definition 2. One observation is generated as follows.

1. Draw

$$\mathbf{L} \sim \mu_{\mathbf{L}}.$$

2. Given $\mathbf{L}$, β_0, and the fixed $\mathcal{W}$, compute the unique normalized equilibrium price vector

$$\mathbf{P}_0 = \mathbf{p}(\beta_0, \mathbf{L} \mid \mathcal{W}).$$

3. Construct the latent equilibrium expenditure shares

$$\lambda_{ij}^{*} = S_{0i} \left(\frac{P_{0i}\tau_{ij}(\theta_0)}{A_j(\mathbf{P}_0, \beta_0 \mid \mathcal{W})} \right)$$

for all $i, j = 1, \ldots, N$. Equivalently, the latent bilateral trade flows generated by the economic model are

$$X_{ij}^{*} = E_j \lambda_{ij}^{*}.$$

4. Draw the error matrix $\mathcal{E} = [\varepsilon_{ij}]_{i,j=1}^{N}$, where its entries have common distribution ν_ε and are independent of $\mathbf{L}$. The researcher observes bilateral trade flows and expenditures and constructs the expenditure shares

$$\lambda_{ij} = \lambda_{ij}^{*} + \varepsilon_{ij}.$$

Equivalently, the observed trade flow satisfies

$$X_{ij}^{\text{obs}} = E_j \lambda_{ij} = X_{ij}^{*} + E_j \varepsilon_{ij}.$$

5. The observed outcome is

$$D = (\Lambda, \mathbf{P}_0, \mathbf{L}).$$

The fixed design $\mathcal{W}$ is known to the researcher.

The primitive probability measure for one observation is

$$\mathbb{Q}_0 = \mu_{\mathbf{L}} \otimes \nu_\varepsilon^{\otimes N^2}.$$

The equilibrium prices and latent shares are measurable functions of the primitive draw and β_0. The distribution of the observed outcome D is therefore the probability measure induced by $\mathbb{Q}_0$ through the mapping in Definition 2. This makes explicit what is random, what is fixed, and what the researcher observes.

2.2 Statistical Setup

Definition 2 specifies one equilibrium observation. We obtain the sample by repeating that experiment independently. Let $Y_{(r)}$ denote the value of an object in observation r.

Definition 3. The primitive vectors

$$\left\{ (\mathbf{L}_{(r)}, \mathcal{E}_{(r)}) \right\}_{r \in \mathbb{N}_+}$$

are IID with distribution $\mathbb{Q}_0$. For each $r = 1, \ldots, R$, Definition 2 generates

$$D_{(r)} = (\Lambda_{(r)}, \mathbf{P}_{0,(r)}, \mathbf{L}_{(r)}).$$

Thus, under Definition 3, the random sample

$$\{D_{(r)}\}_{r=1}^{R}$$

is IID from the distribution induced by $\mathbb{Q}_0$.

The equilibrium price vector is observed in each observation. In the endowment economy, expenditure satisfies $E_{i,(r)} = P_{0i,(r)} L_{i,(r)}$. Hence, when expenditure and endowments are ob-

served, the equilibrium price is recovered as

$$P_{0i,(r)} = \frac{E_{i,(r)}}{L_{i,(r)}}.$$

Randomness in $\mathbf{P}_{0,(r)}$ is induced by the IID endowment draw through the equilibrium mapping. The design points $\mathcal{W}$, the distribution $\mu_{\mathbf{L}}$, and the true parameter β_0 remain fixed across observations.

Our asymptotic sequence holds N fixed and lets R tend to infinity. The number of realizations R is the sample size, while the number of countries N determines the dimension of the economy. Holding N fixed serves two purposes. First, it keeps the structural object of interest $\mathbf{S}_0$ fixed along the asymptotic sequence. Countries and varieties correspond one to one in the Armington economy. Increasing N therefore changes the number of goods in the representative household's utility-maximization problem and generally changes the demand system being estimated. For example, $S_{1|N}$ describes demand for good 1 relative to $N-1$ alternatives, while $S_{1|M}$ describes demand for the same good relative to $M-1$ alternatives. These are generally different functions.

Second, holding N fixed ensures that the DGP remains well defined along the asymptotic sequence. For each finite N, Proposition 1 establishes the existence and uniqueness of the normalized equilibrium price vector by applying a fixed point argument to a finite system of excess-demand functions. Letting N tend to infinity instead produces a sequence of different economies with an increasing-dimensional price system. The finite-dimensional existence result does not establish the existence or behavior of equilibrium in the limiting economy. Doing so would require a separate analysis of an infinite-dimensional commodity space, which is not the object of this paper.

Increasing R instead gives repeated observations of the same static economy under different endowment realizations. The preference represented by $\mathbf{S}_0$, the set of countries, and the fixed design remain unchanged. We refer to this sequence as repeated-outcome asymptotics rather than time-series asymptotics because the model does not specify a dynamic law of motion.

2.3 General-Equilibrium Information

Having defined the sampling unit as an entire equilibrium outcome, we now describe the within-observation restrictions that provide the identifying information in Section 3. The equilibrium model imposes restrictions across all bilateral shares within an observation. For every destination j, utility maximization and the definition of the single aggregator imply

$$\sum_{i=1}^{N} \lambda_{ij}^{*} = 1.$$

Thus, every column of the latent share matrix adds up to one. Market clearing gives the corresponding restriction across destinations. For every supplier i,

$$\sum_{j=1}^{N} \lambda_{ij}^{*} E_j = P_{0i} L_i = E_i.$$

Hence, the value of total exports by country i, including domestic sales, equals its total expenditure. The inclusion of domestic flows is also important in structural gravity estimation (Yotov, 2022). The first restriction links suppliers within a destination. The second links destinations within a supplier.

The additive measurement errors mean that the observed shares need not satisfy these equations exactly. Under Assumption 4, their conditional means satisfy the restrictions. Our criterion therefore uses all N^2 shares in an observation simultaneously. A separate regression for one bilateral flow would discard this cross-equation information.

2.4 Comparison to Conventional CES Gravity Estimation

The preceding subsection shows how our sampling framework retains the restrictions linking all bilateral shares within an equilibrium outcome. We now contrast this framework with conventional CES gravity estimation along two dimensions. The first is how general-equilibrium restrictions are used for identification. The second is the asymptotic experiment used to define the population objective.

Structural gravity models determine bilateral trade flows jointly through market clearing and multilateral resistance terms (Anderson and Van Wincoop, 2003). This structure arises from several models with different microeconomic foundations. See Head and Mayer (2014) and Yotov (2024) for reviews. We focus on how it is used in conventional CES gravity estimation.

We begin with the identification comparison. Suppressing time subscripts, a conventional empirical specification for one cross-section writes the conditional mean of the observed trade flow X_{ij}^{obs} from exporter i to destination j as

$$\mathbb{E}\left[X_{ij}^{\text{obs}} \mid W_{ij}, T_{ij}, \delta_j^I, \delta_i^E\right] = \exp\left\{W_{ij}'\theta_0 + \alpha_0 T_{ij} + \delta_j^I + \delta_i^E\right\}, \tag{11}$$

where W_{ij} contains observed bilateral controls such as distance and an indicator for sharing a border, while T_{ij} is a treatment or policy variable. The terms δ_j^I and δ_i^E are importer and exporter components that contain the inward and outward multilateral resistance terms.

Empirical work commonly controls for these general-equilibrium components with importer and exporter fixed effects (Hummels, 2001) and estimates the gravity equation by Poisson pseudo-maximum likelihood (Silva and Tenreyro, 2006). In panel applications, these controls are typically exporter-time and importer-time fixed effects. Directional country-pair fixed effects are also commonly used to absorb time-invariant bilateral trade costs (Yotov, 2024). These fixed effects absorb the multilateral resistance terms in each observed equilibrium. These

controls are economically necessary. Since the multilateral resistance terms depend on bilateral trade costs, omitting them generally produces correlation between the error term and the trade-cost regressors. Baldwin and Taglioni (2006) call the resulting omitted-variable bias the "gold medal mistake."

Conventional gravity estimation therefore recognizes general equilibrium in the conditional mean. Poisson pseudo-maximum likelihood uses the adding-up restrictions to match observed exporter and importer totals. As shown by Fally (2015), the resulting fixed effects can be mapped into the multilateral resistance terms of the fitted CES equilibrium. This mapping permits structural gravity to be implemented as an estimated general-equilibrium system (Anderson, Larch, and Yotov, 2018; Yotov et al., 2016). Our distinction is therefore not that conventional gravity estimation discards general equilibrium. It concerns which structural parameters the equilibrium restrictions identify and the sampling experiment under which identification is studied.

The role of general-equilibrium information in our approach is especially transparent under CES. The latent shares satisfy

$$\lambda_{ij}^{*} = \left(\frac{P_{0i}\tau_{ij}(\theta_0)}{A_j(\mathbf{P}_0, \theta_0, \sigma_0 \mid \mathcal{W})} \right)^{1-\sigma_0}.$$

Taking the log share ratio relative to the domestic supplier gives

$$\log \frac{\lambda_{ij}^{*}}{\lambda_{jj}^{*}} = (1-\sigma_0)\left[\log \frac{P_{0i}}{P_{0j}} + W_{ij}'\theta_0 \right].$$

The same elasticity appears in every bilateral equation. Variation in observed equilibrium relative prices, together with the cross-equation restrictions implied by adding up and market clearing, allows us to distinguish σ_0 from the trade-cost parameters. By contrast, when the price and aggregator components are absorbed by unrestricted fixed effects, the coefficient on the trade-cost variables identifies only the composite parameter

$$\gamma_W = (1-\sigma_0)\theta_0.$$

To see this, for every admissible value of σ, one can choose $\theta_\sigma = (1-\sigma_0)\theta_0/(1-\sigma)$ and obtain the same composite coefficient. The conventional fixed-effect equation therefore does not separately identify the two parameters. Accordingly, Anderson and Van Wincoop (2003, p. 180) state that "the substitution elasticity σ cannot be estimated." Proposition 3 shows that it is separately identified when we retain the model-implied relationship between equilibrium prices, aggregators, and all bilateral shares.

This argument relies on the equilibrium price vector. Conventional gravity estimation typically does not observe factory-gate prices and absorbs the exporter price components into fixed effects. Some contributions use price information for particular purposes. For example, Feenstra (1994) constructs an exact import price index and Eaton and Kortum (2002) combine trade

flows with retail price data. Our DGP does not treat the equilibrium price vector as a primitive price series observed directly by the researcher. The researcher observes expenditure and endowments. The equilibrium condition implies

$$E_{i,(r)} = P_{0i,(r)} L_{i,(r)}, \qquad P_{0i,(r)} = \frac{E_{i,(r)}}{L_{i,(r)}}.$$

This relationship ties the recovered price to observed expenditure and endowments. The way the entire price vector moves with endowments is then determined by the equilibrium mapping that solves the market-clearing system in Equation (9). In this sense, we impose the structural model directly rather than replace its equilibrium objects with unrestricted fixed effects. The resulting normalized equilibrium price vector is recorded in each observation, as described in Subsection 2.2. Recovering prices alone does not deliver separate identification. We combine their variation across realizations with the restrictions that utility maximization and market clearing impose jointly on the complete share matrix.

We next compare the asymptotic experiments. Conventional fixed-effects gravity relies on an asymptotic sequence in which the number of countries increases. This is true in a cross-section with exporter and importer fixed effects and in a panel with exporter-time, importer-time, and country-pair fixed effects (Fernández-Val and Weidner, 2016; Weidner and Zylkin, 2021). Thus, $N \to \infty$ is indispensable to the conventional asymptotic argument. Suppressing time and letting ϑ collect the parameters of a conventional gravity equation, write

$$\widehat{Q}_N^{\text{conv}}(\vartheta) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \ell_{ij,N}(\vartheta).$$

Here, $\ell_{ij,N}(\vartheta)$ is a loss function for the bilateral outcome (i, j) in the economy with N countries. An identification argument based on this asymptotic sequence would require a well-defined limiting population objective such as

$$\widehat{Q}_N^{\text{conv}}(\vartheta) \xrightarrow{p} Q^{\text{conv}}(\vartheta) = \mathbb{E}_{I,J}[\ell_{IJ}(\vartheta)],$$

uniformly over ϑ in the parameter space, provided that such a limit exists. This sequence is suitable when the target is a finite-dimensional coefficient that is assumed to remain fixed as N increases. The growing collection of fixed effects can then be treated as nuisance parameters.

Our target is the demand system itself. Because countries and varieties correspond one to one in the Armington economy, increasing N changes the choice set and the equilibrium mapping. It may therefore change the demand system being estimated. As discussed in Subsection 2.2, a large-N argument would require additional assumptions linking these different economies and establishing the uniform limit above. A law of large numbers across bilateral pairs alone does not provide either result.

Our IID structure is instead defined across repeated equilibrium outcomes. If $\ell_{(r)}(\beta)$ denotes the loss computed from the complete share matrix in observation r, our sample objective

satisfies

$$\widehat{Q}_R(\beta) = \frac{1}{R}\sum_{r=1}^{R} \ell_{(r)}(\beta) \xrightarrow{p} Q(\beta) = \mathbb{E}\Big[\ell_{(r)}(\beta)\Big]$$

as $R \to \infty$ with N fixed. The distribution of the equilibrium outcome and the demand system remain fixed along this sequence. Each loss $\ell_{(r)}(\beta)$ is constructed from one complete share matrix, and the IID sampling occurs across equilibrium realizations. This is the population objective used for the identification analysis in Section 3.

3 Identification

We study identification first under CES and then when the common share function is unknown. The CES case is useful for showing how the general-equilibrium restrictions identify the elasticity of substitution. Subsection 2.4 explains why this elasticity is not separately identified when the equilibrium components are absorbed by unrestricted fixed effects. Under our sampling framework, repeated equilibrium outcomes generate variation in relative prices, while utility maximization and market clearing link all bilateral shares within each observation. The CES case is also needed to characterize the behavior of the test statistic under the null hypothesis that the demand system is CES in Subsection 4.2. The nonparametric HSA case is our main objective. For completeness, Appendix B considers the simpler case in which the share functions belong to a finite-dimensional family whose parameters are known or calibrated by the researcher.

First, we restrict the distribution of $\mathcal{E}_{(r)}$. For each r, let $\varepsilon_{ij,(r)}$ denote the (i,j) entry of the random error matrix $\mathcal{E}_{(r)}$. We assume that the errors are independent and identically distributed across exporters, importers, and repeated outcomes. They are also independent of the endowment vectors and, therefore, of the equilibrium price vectors generated by those endowments.

Assumption 4. *The collection* $\{\varepsilon_{ij,(r)} : (i,j,r) \in \{1,\dots,N\}^2 \times \mathbb{N}_+\}$ *is i.i.d., is independent of* $\{\mathbf{L}_{(r)}\}_{r\in\mathbb{N}_+}$*, and satisfies:*

1. *The support of* $\varepsilon_{ij,(r)}$ *is a compact subset of* $\mathbb{R}$*,*
2. *The first moment is normalized so that* $\mathbf{E}(\varepsilon_{ij,(r)}) = 0$*.*
3. *The variance* $\mathbf{E}(\varepsilon^2_{ij,(r)}) = \sigma^2_\varepsilon$ *is finite and satisfies* $\sigma^2_\varepsilon > 0$*.*

Remark 4. As noted in Subsection 2.1, conventional assumptions on stochastic error terms often rely on cross-section asymptotics. Here, independence across repeated outcomes allows us to use standard laws of large numbers and central limit theorems under R-asymptotics. Compact support is imposed to bound the empirical process which appears later on, and one may replace this with either a tail condition (Section 9.1 and 9.2 in van de Geer (2006)) on $\varepsilon_{ij,(r)}$ or introduce a truncation of the estimator at the expense of a lengthier proof. The strictly positive variance rules out a degenerate error distribution.

It is convenient to define the following object:

$$g_{ij,(r)}(\theta, \mathbf{S}, \mathbf{p}_0) \equiv S_i \left(\frac{p_{0i,(r)} \tau_{ij}(\theta)}{A_j(\mathbf{p}_{0,(r)}, \theta, \mathbf{S}, \mathcal{W})} \right), \tag{12}$$

which is a conditional mean function $\mathbb{E}(\lambda_{ij,(r)} \mid \mathbf{P}_0 = \mathbf{p}_0)$ taken with respect to the distribution of $\varepsilon_{(r)}$ given $\mathbf{P}_0 = \mathbf{p}_0$. Note that under Assumption 4, we have $g_{ij,(r)}(\theta_0, \mathbf{S}_0, \mathbf{P}_0) = E(\lambda_{ij,(r)} \mid \mathbf{P}_0) = \lambda^*_{ij,(r)}$.

Now consider the population objective function

$$Q(\theta, \mathbf{S}) := \mathbb{E}\left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\lambda_{ij,(r)} - g_{ij,(r)} \left(\theta, \mathbf{S}, \mathbf{P}_{0,(r)} \right) \right)^2 \right]. \tag{13}$$

The expectation is taken with respect to the joint distribution of $\mathcal{E}_{(r)}$ and $\mathbf{P}_{0,(r)}$. Because the complete observations are identically distributed across r, the right-hand side does not depend on the particular index r. Thus, there is a single population least-squares objective $Q(\theta, \mathbf{S})$. The corresponding sample objective, introduced below when defining the estimator, is an average of realized loss functions over $r = 1, \ldots, R$.

For the CES case considered in Subsection 3.1, we replace the objective function with $Q(\theta, \sigma)$ and restrict it to the compact parameter space introduced there. For the semi-nonparametric model, we impose a scale normalization and establish identification over the resulting normalized parameter space. The known-share-function case is recorded in Appendix B.

Below we will provide sufficient conditions to ensure that for any $\mathbf{S} \in \mathcal{S}$, $\sup_{\theta \in \Theta^C} Q(\theta, \mathbf{S})$ is bounded.

We restrict the parameter spaces as follows.

Assumption 5. *There exists some compact set $\Theta^C \subset \Theta$ such that $\theta_0 \in int(\Theta^C)$.*

Assumption 6. *For all $i = 1, \ldots, N$, there exist constants $0 < \underline{z_i} < \bar{z}_i < \infty$ such that uniformly over $S_i \in \mathcal{S}$, we have $S_i(\underline{z_i}) > 1/N$ and $S_i(\bar{z}_i) < 1/N$*

Assumption 5 is a standard assumption in the study of semi-nonparametric models. See Chen (2007). Assumption 6 places an additional condition in addition to the conditions we have already imposed on the infinite-dimensional parameter space $\mathcal{S}$ in Assumption 2. We use $\mathcal{S}$ to denote the infinite-dimensional parameter space and define the compact estimation parameter space $\mathcal{B}^C = \Theta^C \times \mathcal{S}$. Assumption 6 is sufficient to force the single aggregator A_j to lie in a compact subset of $\mathbb{R}_{++}$ uniformly over $\beta \in \mathcal{B}^C$, for all $j = 1, \ldots, N$. The continuity results and the uniform aggregator bounds used here are stated and proved in Appendix A.3. The following proposition is crucial for establishing identification in a semi-nonparametric case.

Proposition 2. *Let $\mathbf{p}_0(\mathcal{I}) = \mathbf{p}(\beta_0, \mathcal{I})$. For each $j = 1, \ldots, N$, there exists a compact set $K_j \subset \mathbb{R}_{++}$ such that*

$$\frac{p_{0i}(\mathcal{I}) \tau_{ij}(\theta)}{A_j(\mathbf{p}_0(\mathcal{I}), \beta \mid \mathcal{W})} \in K_j$$

for every exporter i, uniformly over $(\beta, \mathcal{I}) \in \mathcal{B}^C \times \mathbb{S}(\mathcal{I})$.

Proof. See Appendix A.4. □

Proposition 2 says that the effective relative price

$$\frac{p_{0i}\tau_{ij}(\theta)}{A_j(\mathbf{p}_0, \beta \mid \mathcal{W})}$$

lies in a compact subset of $\mathbb{R}_{++}$. This result has both an econometric and an economic implication. Econometrically, compactness allows us to bound the share functions, their derivatives, and the single aggregator uniformly over the relevant parameter space. These uniform bounds are used in the polynomial approximation and empirical process arguments below. Economically, the result rules out effective relative prices that approach zero or infinity. Equilibrium prices, trade costs, and the destination-specific aggregator therefore cannot combine to produce arbitrarily extreme prices over the support of the observed equilibrium outcomes.

For the remainder of the paper, take K_j to be the compact support of the model-generated effective relative prices associated with destination j, and define

$$K \equiv \bigcup_{j=1}^{N} K_j \subset \mathbb{R}_{++}.$$

Because this is a finite union of compact sets, K is compact. All comparisons between share functions in the nonparametric model are understood on K. Two elements $f, g \in \mathcal{S}_i$ are identified whenever $f(z) = g(z)$ for every $z \in K$. Thus, with a slight abuse of notation, $\mathcal{S}_i$ below denotes the corresponding space of restrictions to K. This convention does not discard any model-relevant variation because every effective relative price $z_{ij}(\mathbf{p}, \beta)$ belongs to K by Proposition 2.

3.1 CES Case

We begin with CES because it illustrates how the general-equilibrium restrictions separately identify the elasticity of substitution and because the CES estimator is needed to analyze the Matsuyama–Ushchev test under the null. Consider the CES specification $S_i(z; \sigma) = z^{1-\sigma}$, where $\sigma > 1$. Define its conditional mean function by

$$g^{CES}_{ij,(r)}(\theta, \sigma) \equiv \frac{P_{0i,(r)}^{1-\sigma} \exp\left((1-\sigma)W'_{ij}\theta\right)}{\sum_{\ell=1}^{N} P_{0\ell,(r)}^{1-\sigma} \exp\left((1-\sigma)W'_{\ell j}\theta\right)}. \tag{14}$$

Equivalently,

$$\lambda^{*}_{ij,(r)}(\mathbf{P}_{0,(r)}, \theta_0, \sigma_0, \mathcal{I}) = \left(\frac{P_{0i,(r)} \exp(W'_{ij}\theta_0)}{A_j(\mathbf{P}_{0,(r)}, \theta_0, \sigma_0 \mid \mathcal{W})}\right)^{1-\sigma_0}$$

for all i, j. The population criterion is

$$Q^{CES}(\theta, \sigma) = \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left(\lambda_{ij,(r)} - g^{CES}_{ij,(r)}(\theta, \sigma)\right)^2\right],$$

where the CES aggregator has the explicit expression

$$A_j(\mathbf{p}, \theta, \sigma \mid \mathcal{W}) = \left[\sum_{\ell=1}^{N} p_\ell^{1-\sigma} \exp(W'_{\ell j}\theta(1-\sigma))\right]^{\frac{1}{1-\sigma}}.$$

First, we restrict the parameter space for σ.

Assumption 7. *The parameter space for the elasticity of substitution σ is $[\sigma_L, \sigma_U] \subset (1, \infty)$ with $\sigma_0 \in int([\sigma_L, \sigma_U])$.*

Combining this restriction with Assumption 5, we set the parameter space for the CES model to $\mathcal{B}^{CES} = \Theta^C \times [\sigma_L, \sigma_U]$ with the true parameter $\beta_0 = (\theta_0, \sigma_0) \in \text{int}(\mathcal{B}^{CES})$. The objective function is bounded uniformly over $\mathbb{S}(\mathcal{I}) \times \mathcal{B}^{CES}$.

Identification requires variation in relative prices and sufficient variation in the trade-cost covariates. We impose the following two conditions.

Assumption 8. *The equilibrium price vector satisfies*

$$\Pr\left(\exists\, i \neq \ell : P_{0i,(r)} \neq P_{0\ell,(r)}\right) = 1.$$

Assumption 9. *The linear span of the differences $\{W_{i\ell} - W_{ij} - W_{j\ell} : i, j, \ell \text{ distinct}\}$ is $\mathbb{R}^d$.*

For $d > 1$, this condition requires the collection of triplet differences to contain d linearly independent directions; a single nonzero triplet is sufficient only when $d = 1$.

Proposition 3. *Suppose Assumptions 8 and 9 hold. Then $Q^{CES}(\theta, \sigma)$ is uniquely minimized over $\mathcal{B}^{CES}$ at (θ_0, σ_0).*

Proof. See Appendix C.3. □

An important feature of Proposition 3 is that the criterion uses the entire vector of bilateral shares within each observation,

$$\mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left(\lambda_{ij,(r)} - g^{CES}_{ij,(r)}(\theta, \sigma)\right)^2\right],$$

rather than a scalar textbook criterion $\mathbb{E}[(\lambda_{ij,(r)} - g^{CES}_{ij,(r)}(\theta, \sigma))^2]$ for one fixed pair (i, j). This distinction is crucial for identification. The observations are IID across repeated equilibrium outcomes r, but the bilateral shares within an observation are determined simultaneously by the same general-equilibrium price vector and CES aggregators. Consequently, the conditional means obey cross-equation restrictions, including adding up and the log-share-odds relations

used in the proof in Appendix C.3. Observing all shares allows us to compare suppliers within destinations and to combine comparisons across destinations, thereby eliminating the common aggregators and relative prices at the appropriate stages of the proof. A criterion based on only one fixed bilateral share would discard precisely this general-equilibrium variation and would not, under the conditions stated here, deliver the same identification argument.

Define the CES estimator $\hat{\beta} \equiv (\hat{\theta}, \hat{\sigma})$ as the solution to

$$\hat{\beta} := \arg \min_{(\theta,\sigma)\in\mathcal{B}^{CES}} \frac{1}{RN^2} \sum_{r=1}^{R} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\lambda_{ij,(r)} - g^{CES}_{ij,(r)}(\theta,\sigma)\right)^2.$$

Under standard extremum-estimation conditions, such as those stated in Newey and McFadden (1994), consistency and asymptotic normality of $\hat{\beta}$ follow as $R \to \infty$. These conditions include the compactness and identification conditions stated above, uniform convergence of the sample criterion to its population counterpart, and the usual differentiability, moment, and nonsingularity conditions for asymptotic normality.

The sample analogue of $\mathbb{E}[(\lambda_{ij,(r)} - g^{CES}_{ij,(r)}(\theta,\sigma))^2]$ may look like a standard nonlinear least-squares objective. The important distinction lies in how the predicted shares are constructed. For every observation r and every candidate (θ,σ), we impose the CES adding-up restriction by computing

$$A_j(\mathbf{P}_{0,(r)}, \theta, \sigma \mid \mathcal{W}) = \left[\sum_{\ell=1}^{N} \left(P_{0\ell,(r)} \exp(W'_{\ell j}\theta)\right)^{1-\sigma}\right]^{1/(1-\sigma)}$$

for each destination j. We then use these model-implied aggregators to construct all N^2 predicted shares jointly. Thus, evaluating the criterion requires imposing the nonlinear equilibrium relationship between the candidate parameters, observed prices, aggregators, and bilateral shares in every observation. A conventional gravity regression instead absorbs the aggregator and price components with unrestricted fixed effects. It therefore does not use this nonlinear cross-equation restriction.

The restriction $\theta \in \Theta^C \subset \Theta$ is also crucial. As shown in Lemma 1, this parameter space ensures that the candidate trade costs satisfy all three iceberg-cost axioms in Assumption 1. Thus, together with the remaining assumptions of Proposition 1, every admissible candidate defines a well-defined general equilibrium. Without these trade-cost restrictions, the equilibrium existence guarantee in that proposition no longer applies.

Remark 5. The sample criterion is generally nonconvex. Consequently, a numerical optimizer may converge to different local minima depending on its initialization. In practice, the researcher should solve the problem from multiple feasible starting values and retain the solution with the smallest attained criterion value.

The numerical results show that this procedure is feasible with both simulated and actual data. The simulation study in Appendix G examines whether the separate identification of θ_0 and σ_0, the large-sample properties of the point estimates, and inference for the gains-from-trade counterfactual are reflected in finite samples. The estimator recovers both parameters

and the counterfactual well. The bootstrap also delivers coverage close to the nominal level. Appendix I.1 implements the same multistart procedure in the empirical application.

3.2 Nonparametric Share Functions

We now treat the common share function as unknown. We first establish identification over a normalized population parameter space. We then define the normalized sieve estimator and derive its rate of convergence.

3.2.1 Identification

We first establish identification of the unknown HSA demand system. The central object in the argument is the effective relative price faced by each bilateral pair. For $\beta = (\theta, \mathbf{f})$, define

$$z_{ij}(\mathbf{p}, \beta) = \frac{p_i \tau_{ij}(\theta)}{A_j(\mathbf{p}, \beta \mid \mathcal{W})}.$$

The model has a scale indeterminacy. As noted by Matsuyama and Ushchev (2017), a common positive rescaling of the share functions produces the same equilibrium allocation. Thus, we work with a normalization. To see the indeterminacy, suppose that every share function is replaced by $\widetilde{f}_i(z) = f_i(z/c)$ for some $c > 0$. The associated aggregator becomes

$$A_j(\mathbf{p}, \theta, \widetilde{\mathbf{f}}) = \frac{A_j(\mathbf{p}, \theta, \mathbf{f})}{c}.$$

The predicted shares are unchanged because

$$\widetilde{f}_i\left(\frac{p_i \tau_{ij}(\theta)}{A_j(\mathbf{p}, \theta, \widetilde{\mathbf{f}})}\right) = f_i\left(\frac{p_i \tau_{ij}(\theta)}{A_j(\mathbf{p}, \theta, \mathbf{f})}\right).$$

We remove this indeterminacy by fixing $\bar{\mathbf{p}} \in \mathbb{S}(\mathbf{P}_0)$ and imposing

$$A_N(\bar{\mathbf{p}}, \beta) = 1.$$

In the CES case, writing the share function as $S_i(z) = cz^{1-\sigma}$, the conventional specification $S_i(z) = z^{1-\sigma}$ effectively imposes the normalization $c = 1$.

We focus on the symmetric HSA model. The normalized population parameter space is

$$\mathcal{B}^* = \left\{(\theta, \mathbf{f}) \in \Theta^C \times \mathcal{S} \;\middle|\; f_1 = \cdots = f_N,\ A_N(\bar{\mathbf{p}}, \theta, \mathbf{f}) = 1\right\}.$$

The true parameter β_0 is assumed to belong to $\mathcal{B}^*$.

Assumption 10. *We have $N \geq 3$. The true share functions are symmetric, so that $S_{0i} = S_0$ for every i, and $\beta_0 \in \mathcal{B}^*$. The following conditions also hold.*

1. *The support $\mathbb{S}(\mathbf{P}_0)$ is a compact and convex subset of the interior of the normalized price simplex*

and has nonempty relative interior. The distribution of $\mathbf{P}_0$ *has a density, relative to the* $(N-1)$*-dimensional Lebesgue measure on the simplex, that is strictly positive on this support.*

2. *For every* $i \neq \ell$,

$$\left\{ (W_{\ell j} - W_{ij})'\vartheta \;\middle|\; \vartheta \in \Theta^C,\ j = 1, \ldots, N \right\} \subset \operatorname{int}(R_{i\ell}),$$

where

$$R_{i\ell} = \{\log(p_i/p_\ell) \mid \mathbf{p} \in \mathbb{S}(\mathbf{P}_0)\}.$$

In addition,

$$\operatorname{span}\left\{ W_{\ell j} - W_{ij} \;\middle|\; i \neq \ell,\ j = 1, \ldots, N \right\} = \mathbb{R}^d.$$

The preceding assumption supplies the price and covariate variation used to identify the finite-dimensional parameter. Identification of the unknown share function requires an additional joint support condition.

Assumption 11 (Joint support propagation)**.** *For every* $\beta \in \mathcal{B}^*$*, there are* $J \geq 2$ *nondegenerate intervals* $I_1, \ldots, I_J$ *satisfying*

$$K \subseteq \bigcup_{a=1}^{J} I_a.$$

For any two distinct indices $a, b \in \{1, \ldots, J\}$*, there are an integer* $m \geq 1$ *and indices* $a_0, \ldots, a_m \in \{1, \ldots, J\}$*, with* $a_0 = a$ *and* $a_m = b$*, having the following property. For each* $q = 1, \ldots, m$*, there are a destination* j_q*, exporters* $i_q \neq \ell_q$*, and a relatively open set* $U_q \subset \mathbb{S}(\mathbf{P}_0)$ *such that*

$$I_{a_{q-1}} \times I_{a_q} \subseteq \left\{ \left(z_{i_q j_q}(\mathbf{p}, \beta), z_{\ell_q j_q}(\mathbf{p}, \beta) \right) \;\middle|\; \mathbf{p} \in U_q \right\}.$$

Assumption 11 is a joint support condition, not merely a condition on the marginal support of each effective relative price. Take two consecutive intervals in the sequence and arbitrary points $x \in I_{a_{q-1}}$ and $y \in I_{a_q}$. The displayed set inclusion requires a single equilibrium price vector $\mathbf{p} \in U_q$ for which

$$x = z_{i_q j_q}(\mathbf{p}, \beta) \qquad \text{and} \qquad y = z_{\ell_q j_q}(\mathbf{p}, \beta).$$

The same price vector must therefore generate both values simultaneously. This is stronger than requiring some price vector to generate x for exporter i_q and a possibly different price vector to generate y for exporter ℓ_q. The condition requires the map

$$\mathbf{p} \longmapsto \left(z_{i_q j_q}(\mathbf{p}, \beta), z_{\ell_q j_q}(\mathbf{p}, \beta) \right)$$

to attain every pair (x, y) in $I_{a_{q-1}} \times I_{a_q}$.

Economically, identification relies on observing how relative prices and expenditure shares change jointly across observations. It is not enough to observe one exporter's relative price

and its corresponding share in isolation. The relative prices and shares of different exporters must occur simultaneously in the same equilibrium realization, where they satisfy a common adding-up restriction. The support condition ensures sufficiently rich variation in these jointly observed relative prices to distinguish candidate share functions throughout the common support K.

These conditions are stronger than their counterparts in the CES model. Assumption 8 requires only that some pair of equilibrium prices differ with probability one. Part 1 of Assumption 10 instead requires a strictly positive density on a compact and convex support with nonempty relative interior. This stronger support condition supplies the continuous price variation needed to distinguish candidate share functions over an interval. Assumption 9 requires the CES triplet differences in the trade-cost covariates to span $\mathbb{R}^d$. Part 2 of Assumption 10 strengthens this requirement by combining a span condition with support coverage for the relevant covariate indices. These conditions use the variation in prices and design points to identify the trade-cost parameter θ_0. Assumption 11 has no counterpart in Assumptions 8 and 9. It connects the effective relative price intervals and thereby propagates identification of the unknown share function across its common support. This additional condition is unnecessary under CES because the unknown demand component is the scalar elasticity σ, whereas in the HSA case we need to distinguish candidate share functions over the interval K. By Proposition 1, Assertion (ii), of Matsuyama and Ushchev (2017), HSA demand is equivalent to CES demand when $N = 2$. That case is therefore covered by the preceding subsection.

Define the population objective function by

$$Q(\beta) = \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left\{\lambda_{ij,(r)} - f_i\big(z_{ij}(\mathbf{P}_{0,(r)}, \beta)\big)\right\}^2\right].$$

Theorem 1 (Identification of the symmetric HSA model)**.** *Suppose that Assumptions 4 and 10 and 11 hold. Then $Q(\beta)$ is uniquely minimized over $\mathcal{B}^*$ at β_0.*

Proof. The proof has four main steps. Part 1 of Assumption 10 converts equality almost surely into equality at every equilibrium price vector in the support. Strict monotonicity then converts equality of shares into restrictions on effective relative prices. Part 2 identifies θ_0. Part 3 uses Assumption 11 to propagate equality of the candidate and true functions across the common support K. The normalization fixes the remaining scale. Appendix D.1 gives the complete argument. □

For $\beta = (\theta, \mathbf{f})$ and $\widetilde{\beta} = (\widetilde{\theta}, \widetilde{\mathbf{f}})$, define

$$M(\beta, \widetilde{\beta}) = \left[\int_{S(\mathbf{P}_0)} \frac{1}{N^2}\sum_{i,j}\left|f_i\big(z_{ij}(\mathbf{p}, \beta)\big) - \widetilde{f}_i\big(z_{ij}(\mathbf{p}, \widetilde{\beta})\big)\right|^2 \mu(\mathrm{d}\mathbf{p})\right]^{1/2}.$$

We use this metric to study the large-sample properties of the estimator. Under Theorem 1, two normalized parameters are observationally equivalent only when they coincide at the truth.

3.2.2 The Normalized Sieve Estimator

The identification result concerns the infinite-dimensional population parameter space. To construct a feasible estimator, we approximate the unknown share function using a finite-dimensional sieve that preserves the economic restrictions and the scale normalization imposed above.

For $\alpha^{(i)} = (\alpha_0^{(i)}, \ldots, \alpha_\kappa^{(i)})'$, write

$$g_{\alpha^{(i)}}(z) = \alpha_0^{(i)} + \sum_{q=1}^{\kappa} \alpha_q^{(i)} z^{-q}.$$

For each i, let the sieve space be

$$\mathcal{S}_{i,\kappa} = \left\{ g_{\alpha^{(i)}} : K \to \mathbb{R}_+ \;\middle|\; \begin{array}{cc} \|\alpha^{(i)}\|_1 \le \rho_R^U, & \inf_{z \in K} g_{\alpha^{(i)}}(z) \ge \rho_R^L, \\ \multicolumn{2}{c}{\inf_{z \in K} \{-g'_{\alpha^{(i)}}(z)\} \ge \rho_R^L} \end{array} \right\}.$$

Here, $\|\alpha^{(i)}\|_1$ denotes the L^1 norm of the coefficient vector $\alpha^{(i)}$. The sieve dimension κ and the bounds ρ_R^U and ρ_R^L are tuning parameters. The restrictions impose positivity and strict monotonicity directly on K. The coefficient bound makes each finite-dimensional sieve compact. Define

$$\mathcal{S}_\kappa = \prod_{i=1}^{N} \mathcal{S}_{i,\kappa}$$

and define the enlarged sieve parameter space

$$\overline{\mathcal{B}}_\kappa = \Theta^C \times \mathcal{S}_\kappa.$$

This compact space is used to bound the empirical process.

For simplicity, we assume that all goods share the same share function. The identification argument compares goods supplied by two countries that share this function. This suggests that full symmetry could be relaxed by dividing countries into groups, with a common share function within each group but different functions across groups. Such an extension would require sufficient variation within groups to identify the trade-cost parameters and sufficient joint price variation to identify the different share functions. We do not pursue this extension here. Define the symmetric sieve subspace by

$$\mathcal{S}_\kappa^{\text{sym}} = \{(f_1, \ldots, f_N) \in \mathcal{S}_\kappa \mid f_1 = \cdots = f_N\}.$$

The normalized sieve parameter space is

$$\mathcal{B}_\kappa^* = \left\{ (\theta, \mathbf{f}) \in \Theta^C \times \mathcal{S}_\kappa^{\text{sym}} \;\middle|\; A_N(\bar{\mathbf{p}}, \theta, \mathbf{f}) = 1 \right\}.$$

For the approximation argument, define the set of normalized sieve functions at the true finite-

dimensional parameter by

$$\mathcal{S}^*_{\kappa,0} = \left\{ \mathbf{f} \in \mathcal{S}^{\text{sym}}_\kappa \;\middle|\; (\theta_0, \mathbf{f}) \in \mathcal{B}^*_\kappa \right\}.$$

Assumption 12. *For each R, the bounds ρ^L_R and ρ^U_R are chosen so that $\mathcal{S}^*_{\kappa,0}$ is nonempty for the sieve dimension $\kappa = \kappa_R$. If S_0 satisfies the smoothness condition in Proposition 5, these bounds are also chosen so that the normalized Jackson approximant constructed in Appendix D.6 belongs to $\mathcal{S}^*_{\kappa,0}$ for all sufficiently large R.*

Assumption 12 rules out choices of ρ^L_R and ρ^U_R that are so restrictive that no normalized sieve function is feasible at θ_0. The second requirement ensures that the sieve contains an approximant with the asserted Jackson rate. It is satisfied when ρ^L_R tends to zero and ρ^U_R dominates the coefficient norm of the normalized approximant. Appendix D.6 verifies this condition.

A popular way to impose shape restrictions is to use B-splines with shape-preserving properties, as discussed in Section 2.3.5 of Chen (2007), or Bernstein polynomials with coefficient restrictions, as in Wang and Ghosh (2012). Both approaches generally require a normalization that maps the argument of the unknown function to a known interval such as $[0, 1]$. This is difficult to implement in our setting for two reasons. First, although Proposition 2 shows that the effective relative price lies in a compact set, we do not know its precise bounds. Second, as in the CES case, the aggregator $A_j(\mathbf{p}, \beta \mid \mathcal{W})$ must be computed for every candidate β. The resulting effective relative price and the normalization of its domain therefore change with the candidate parameter. Inverse polynomials allow us to proceed without this rescaling. Finally, the workhorse CES share function is itself a power function, which makes inverse polynomials a natural choice. They nest the CES share function when its exponent is an integer and approximate general smooth share functions as κ increases.

For $D_{(r)} = (\Lambda_{(r)}, \mathbf{P}_{0,(r)}, \mathbf{L}_{(r)})$, write

$$\ell_\beta(D_{(r)}) = \frac{1}{N^2} \sum_{i,j} \left\{ \lambda_{ij,(r)} - f_i\big(z_{ij}(\mathbf{P}_{0,(r)}, \beta)\big) \right\}^2.$$

Let $P_R h = R^{-1} \sum_{r=1}^R h(D_{(r)})$ and $Ph = \mathbb{E}[h(D_{(r)})]$. The sieve estimator is

$$\hat{\beta}_\kappa \in \arg\min_{\beta \in \mathcal{B}^*_\kappa} P_R \ell_\beta.$$

The argmin set is nonempty because the sample objective is continuous and $\mathcal{B}^*_\kappa$ is compact. The sample minimizer need not be unique.

To establish weak consistency of the sieve estimator, first define the empirical-process component

$$T_{1\kappa} = \sup_{\beta \in \mathcal{B}^*_\kappa} \left| (P_R - P)\ell_\beta \right|$$

and

$$T_{2\kappa} = \inf_{\mathbf{f} \in \mathcal{S}^*_{\kappa,0}} M^2\big((\theta_0, \mathbf{f}), \beta_0\big).$$

A simple adaptation of Lemma 10.2 in Györfi et al. (2002) yields the following bound.

Proposition 4. *For each $\kappa \geq 1$ for which $\mathcal{B}_\kappa^*$ is nonempty, the sieve least-squares estimator satisfies*

$$M^2(\hat{\beta}_\kappa, \beta_0) \leq 2T_{1\kappa} + T_{2\kappa}.$$

Proof. See Appendix D.2. □

Proposition 4 reduces the convergence problem to an empirical-process component and a deterministic approximation component. This is the same basic decomposition used in the general sieve-estimation framework of Chen (2007), where convergence is governed by the stochastic variation of the sample criterion and the approximation error from restricting the unknown function to a sieve space.

3.2.3 Consistency and Rate of Convergence

The general-equilibrium structure that provides identification also makes the large-sample analysis nontrivial. For every candidate share function and trade-cost parameter, the single aggregator must be solved again, and the resulting change affects all predicted bilateral shares within the observation. We therefore derive an entropy bound for this composed loss class and impose restrictions on the tuning sequences that make its empirical-process component negligible.

We impose additional conditions under which the empirical-process component $T_{1\kappa}$ and the deterministic approximation component $T_{2\kappa}$ can be suitably controlled. These controls yield the rate of convergence of $\hat{\beta}_\kappa$ under the metric M.

The deterministic component is controlled by a simultaneous Jackson approximation. When S_0 admits $\gamma \geq 2$ continuous derivatives on a compact neighborhood of K, there is a normalized symmetric sieve approximant satisfying

$$\sup_{z \in K} |f_\kappa(z) - S_0(z)| = O(\kappa^{-\gamma}).$$

The construction preserves positivity, monotonicity, and the aggregator normalization. Appendix D.6 gives the complete argument.

The empirical-process component can be made negligible under the following conditions on the tuning parameters.

Assumption 13. *The deterministic tuning sequences satisfy*

$$\kappa_R \to \infty, \qquad \rho_R^U \to \infty, \qquad \rho_R^L \to 0, \qquad R^{-1}\kappa_R^2 \log\left(\frac{\kappa_R \rho_R^U}{\rho_R^L}\right) \to 0.$$

The final restriction regulates the logarithms of ρ_R^U and ρ_R^L. The covering number of a κ-dimensional Euclidean ball depends logarithmically on its diameter, which is why ρ_R^U enters the entropy bound through a logarithm. The lower bound ρ_R^L keeps the slope of the share function

away from zero. It enters the Lipschitz constant of the implicitly defined price aggregator $A_j(\cdot)$ through $(\rho_R^L)^{-1}$, and therefore also appears inside the logarithm in Assumption 13.

Following the general sieve framework in Chen (2007), we obtain a convergence bound by controlling the empirical-process and approximation errors. In our setting, this requires accounting for the dependence of the single price aggregator on the candidate share function. The following proposition gives the resulting bound.

Proposition 5. *Suppose that Assumptions 12 and 13 hold and that S_0 admits $\gamma \geq 2$ continuous derivatives on a compact interval $K^+ \subset \mathbb{R}_{++}$ satisfying $K \subset \text{int}(K^+)$. Then*

$$M^2(\hat{\beta}_\kappa, \beta_0) = O_p\left(\kappa\sqrt{\frac{\log(\kappa\rho_R^U/\rho_R^L)}{R}}\right) + O(\kappa^{-2\gamma}).$$

Proof. See Appendix D.7. □

The $O_p(\cdot)$ and $O(\cdot)$ terms correspond to the empirical-process component $T_{1\kappa}$ and the approximation component $T_{2\kappa}$, respectively. The metric M integrates with respect to the distribution μ of the random equilibrium price vector $\mathbf{P}_{0,(r)}$. Measurement errors do not enter M and instead contribute to the randomness in $T_{1\kappa}$. Conditional on the equilibrium price vectors, the remaining randomness comes from $\mathcal{E}_{(r)}$.

The empirical-process component is obtained from a global bound. The resulting order is larger than the familiar $O_p(\kappa/R)$ rate for the squared metric in the least-squares series benchmark of Newey (1997), apart from logarithmic terms. These rates are not directly comparable. Our result controls the criterion uniformly over the nonlinear sieve parameter space and should not be interpreted as establishing that the local rate of the HSA estimator is intrinsically slower. The general rate theorem in Chen (2007) permits nonconcave criteria, but it does not directly supply a rate for our HSA criterion. Although our sieve is linear, the criterion is nonlinear in the sieve coefficients and need not be concave because the price aggregator depends on the approximating function. Appendix D.4 derives our empirical-process rate by first verifying that the loss class is Lipschitz, then bounding its entropy using Section 2.7 of Vaart and Wellner (1996), and finally applying Theorem 2.14.2 of that book. To see why the resulting bound is inflated, let

$$d_\kappa(\beta, \widetilde{\beta}) = \|\theta - \widetilde{\theta}\|_1 + \sum_{i=1}^{N} \|\alpha^{(i)} - \widetilde{\alpha}^{(i)}\|_1$$

denote the distance between the finite-dimensional coordinates of two sieve elements. The single aggregator is implicitly defined by the adding-up restriction in equation (2). The first Lipschitz argument gives

$$\left|A_j(\mathbf{p}, \beta \mid \mathcal{W}) - A_j(\mathbf{p}, \widetilde{\beta} \mid \mathcal{W})\right| \leq L_{A,\kappa} d_\kappa(\beta, \widetilde{\beta}),$$

where $L_{A,\kappa}$ depends on κ, ρ_R^U, and ρ_R^L. This bound enters the effective relative price through

$$\left|z_{ij}(\mathbf{p},\beta) - z_{ij}(\mathbf{p},\widetilde{\beta})\right| \leq C_z(1 + L_{A,\kappa})d_\kappa(\beta,\widetilde{\beta})$$

for a finite constant C_z. A second Lipschitz argument is then required for the sieve functions themselves. It gives

$$\begin{aligned}
&\left|f_i(z_{ij}(\mathbf{p},\beta)) - \widetilde{f}_i(z_{ij}(\mathbf{p},\widetilde{\beta}))\right| \\
&\quad\leq C_\kappa\|\alpha^{(i)} - \widetilde{\alpha}^{(i)}\|_1 + L_{f,\kappa}\left|z_{ij}(\mathbf{p},\beta) - z_{ij}(\mathbf{p},\widetilde{\beta})\right| \\
&\quad\leq \left[C_\kappa + L_{f,\kappa}C_z(1 + L_{A,\kappa})\right] d_\kappa(\beta,\widetilde{\beta}).
\end{aligned}$$

Thus, the Lipschitz constant for the composed share contains the product $L_{f,\kappa}L_{A,\kappa}$. This iterated Lipschitz argument produces the stated global empirical-process bound. Appendix D.3 gives the detailed bounds.

Valid tuning sequences exist. For example, one may take

$$\kappa_R = \lfloor R^{1/4} \rfloor, \qquad \rho_R^U = \exp(CR^{1/4}), \qquad \rho_R^L = R^{-1/4},$$

where $C > 0$ is large enough to contain the coefficients of the approximating polynomials. Then the empirical-process component is $O_p(R^{-1/8})$, while the approximation component is $O(R^{-\gamma/2})$. More generally, one may take $\kappa_R = \lfloor R^a \rfloor$ for $0 < a < 1/3$, let $\log \rho_R^U = O(\kappa_R)$, and let $(\rho_R^L)^{-1}$ grow at a polynomial rate.

Finally, the benchmark CES share function $S_0(z) = z^{1-\sigma_0}$ is infinitely differentiable on every compact subset of $\mathbb{R}_{++}$. It therefore admits arbitrarily large finite values of γ in the approximation bound.

4 Large-Sample Properties of the Matsuyama–Ushchev Test

This section proceeds in two steps. We first establish a linear representation of the sieve estimator $\hat{\beta}_\kappa$ centered at the pseudo-true value β_κ^*. We then use this representation to derive the large-sample properties of the test statistic based on the characterization of CES in Matsuyama and Ushchev (2017).

4.1 Linear Representation of the Sieve Estimator

The consistency result in Proposition 5 is not sufficient by itself to derive the distribution of the test statistic. We therefore obtain a linear representation of the sieve estimator around its pseudo-true value. We use the symmetric normalized sieve defined in Subsection 3.2. Let

$$\xi_\kappa = (\theta', \alpha_1, \ldots, \alpha_\kappa)'$$

denote its reduced coordinate vector. For any such vector, define

$$\alpha_0(\xi_\kappa) = \frac{1}{N}\left\{1 - \sum_{q=1}^{\kappa} \alpha_q \sum_{i=1}^{N} [\bar{p}_i \tau_{iN}(\theta)]^{-q}\right\}.$$

The common sieve function is

$$f_\kappa(z; \xi_\kappa) = \alpha_0(\xi_\kappa) + \sum_{q=1}^{\kappa} \alpha_q z^{-q}.$$

This definition imposes $A_N(\bar{\mathbf{p}}, \theta, (f_\kappa, \ldots, f_\kappa)) = 1$. To see this, at the normalization point the adding-up equation evaluated at $A_N = 1$ is

$$\sum_{i=1}^{N} f_\kappa(\bar{p}_i \tau_{iN}(\theta); \xi_\kappa) = N\alpha_0(\xi_\kappa) + \sum_{q=1}^{\kappa} \alpha_q \sum_{i=1}^{N} [\bar{p}_i \tau_{iN}(\theta)]^{-q} = 1.$$

The final equality follows from the definition of $\alpha_0(\xi_\kappa)$. Since the single aggregator solving the adding-up equation is unique, its value at $\bar{\mathbf{p}}$ is $A_N = 1$. This is the scale normalization imposed in Subsection 3.2. Let Ξ_κ be the set of reduced coordinates for which $(\theta, f_\kappa, \ldots, f_\kappa) \in \mathcal{B}_\kappa^*$. Its dimension is $d + \kappa$.

We define the pseudo true parameters by

$$\xi_\kappa^* \in \arg\min_{\xi \in \Xi_\kappa} Q(\xi),$$

which is assumed to satisfy $\xi_\kappa^* \in \text{int}(\Xi_\kappa)$ for each $\kappa \geq 1$. We write $f_i^*(z) = f_\kappa(z; \xi_\kappa^*)$ for every i. Let us define

$$H_\kappa \equiv \frac{1}{2}\mathbb{E}\left[\nabla_{\xi'} \nabla_\xi Q_{(r)}(\xi_\kappa^*)\right].$$

Let $d_{ij,(r)}^* = S_0(z_{ij,(r)}(\beta_0)) - f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*)$. Equivalently,

$$H_\kappa = \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left\{\nabla_\xi f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*) \nabla_{\xi'} f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*) - d_{ij,(r)}^* \nabla_{\xi'} \nabla_\xi f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*)\right\}\right].$$

Thus, H_κ is one half of the full population Hessian. The $1/2$ appears because the population objective is quadratic. The second term need not vanish because the sieve approximation error need not be zero at the pseudo truth. We assume the following regularity conditions to derive the linear representation of the sieve estimates $\hat{\xi}_\kappa$ around the pseudo truth ξ_κ^*. Let us impose a typical eigenvalue condition on H_κ.

Assumption 14. *Let $\lambda_{min,\kappa}$ denote the smallest eigenvalue of H_κ. We assume that, along the sequence of tuning parameters $\kappa = \kappa_R \to \infty$, $\lambda_{min,\kappa} \to c_\lambda > 0$.*

Assumption 13 controls the empirical process used to obtain the convergence rate in Section 3. It does not by itself imply uniform convergence of the Hessian in operator norm. For the linear representation below, we additionally require the Hessian tuning condition in Equation

(D.1) of Appendix E.1.2. This additional condition is not vacuous. For example, one may take

$$\kappa_R = \lfloor \log\log R \rfloor, \qquad \rho_R^U = \exp(C_U \kappa_R), \qquad \rho_R^L = \exp(-C_L \kappa_R),$$

for fixed positive constants C_U and C_L, with C_U chosen large enough to contain the coefficients of the sieve approximants. This sequence satisfies both Assumption 13 and the additional Hessian tuning condition. We record the linear representation as follows.

Proposition 6. *Assume that Assumptions 13 and 14 are met. Suppose also that the additional Hessian tuning condition in Equation (D.1) holds. The sieve estimator satisfies*

$$\sqrt{R}\left(\hat{\xi}_\kappa - \xi_\kappa^*\right) = H_\kappa^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \psi_{\kappa,(r)} + o_p(1),$$

where

$$\psi_{\kappa,(r)} \equiv \frac{1}{N^2} \sum_{i,j} \left\{ \lambda_{ij,(r)} - f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*) \right\} \nabla_\xi f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*).$$

The population first-order condition at the interior pseudo truth implies $\mathbb{E}[\psi_{\kappa,(r)}] = 0$.

Proof. See Appendix E.1. □

The linear representation around the pseudo true parameter ξ_κ^* facilitates the analysis of the test statistic which we study in 4.2.

Remark 6. As is evident from the linear representation, the smoothness of the true share functions $S_{01}, \ldots, S_{0N}$ over the compact sets $K_1, \ldots, K_N$ given in Proposition 2 is crucial. In the CES benchmark, $S_{0i}(z) = z^{1-\sigma_0}$. Because K_i is a compact subset of $(0, \infty)$, S_{0i} is analytic on an open neighborhood of K_i and therefore possesses derivatives of every order on K_i.

4.2 The Matsuyama–Ushchev Test of the CES Model

We now use the linear representation from the preceding subsection to study a scalar implication that distinguishes CES from general HSA demand. This implication follows from the relationship between the ideal price index and the single aggregator established by Matsuyama and Ushchev. Let us denote the ideal price index of consumer j by $P_j^I(\mathbf{p}, \beta \mid \mathcal{W})$. Then, by Proposition 1 (Assertion (ii)) in Matsuyama and Ushchev (2017), for some constant $c > 0$, we have

$$\ln P_j^I(\mathbf{p}, \beta \mid \mathcal{W}) = \ln A_j(\mathbf{p}, \beta \mid \mathcal{W}) + \sum_{\ell=1}^{N} \int_c^{(p_\ell \tau_{lj}(\theta))/A_j(\mathbf{p},\beta|\mathcal{W})} \frac{S_\ell(z)}{z} \mathrm{d}z, \tag{15}$$

uniformly over $(\mathbf{p}, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$. The constant $c > 0$ is fixed and known. It only determines the lower limit of integration. A different choice shifts both Φ_j and $\Gamma(c, \sigma)$ by the corresponding normalization constant and does not change Marshallian demand or equilibrium share allocations. We maintain that the convex hull of $K \cup \{c\}$ is contained in the interior of the compact interval K^+ introduced in Proposition 5.

Importantly, when $N \geq 3$, Assertion (iii) of Proposition 1 in Matsuyama and Ushchev (2017) claims that we have:

$$\ln P_j^I(\mathbf{p}, \beta \mid \mathcal{W}) = \ln A_j(\mathbf{p}, \beta \mid \mathcal{W}) + \Gamma(c, \sigma_0), \tag{16}$$

if and only if the model is CES (i.e., for all ℓ, we have $S_\ell(z) = z^{1-\sigma_0}$ for some $\sigma_0 > 1$), for a known function $\Gamma(c, \sigma_0)$ which does not depend on $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. A simple calculation reveals that $\sigma \mapsto \Gamma(c, \sigma) = (1 - Nc^{1-\sigma})/(1-\sigma)$.

Let us define

$$\Phi_j(\mathbf{p}, \beta \mid \mathcal{W}) \equiv \sum_{\ell=1}^{N} \int_c^{(p_\ell \tau_{lj}(\theta))/A_j(\mathbf{p}, \beta \mid \mathcal{W})} \frac{S_\ell(z)}{z} \mathrm{d}z, \tag{17}$$

for each $j = 1, \ldots, N$. Then, Equation (16) implies that we have, μ-almost surely,

$$\Phi_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W}) = \Gamma(c, \sigma_0),$$

when β is CES. We use this sharp characterization to construct the Matsuyama–Ushchev test. The name distinguishes the economic characterization provided by Matsuyama and Ushchev (2017) from its econometric formalization developed here.

For the large-sample analysis of this scalar functional, we impose the following continuity condition.

Assumption 15. *There is a constant $C_\Phi < \infty$ such that, locally around β_0 in $\mathcal{B}^*$,*

$$\left\{ \mathbb{E}\left[\left| \frac{1}{N} \sum_{j=1}^{N} \left\{ \Phi_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W}) - \Phi_j(\mathbf{P}_{0,(r)}, \beta_0 \mid \mathcal{W}) \right\} \right|^2 \right] \right\}^{1/2} \leq C_\Phi M(\beta, \beta_0).$$

Assumption 15 ensures that candidates close to the truth in M also produce similar values of the scalar functional used in the test. This allows us to control the effect of approximating the unknown demand system on the test statistic. Appendix E.2.2 gives sufficient conditions under which the assumption holds.

Let us define:

$$\mathcal{B} \ni \beta \mapsto T_R(\beta) \equiv \frac{1}{RN} \sum_{r=1}^{R} \sum_{j=1}^{N} \Phi_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W}). \tag{18}$$

The normalization by N makes $T_R(\beta)$ the average across destinations and observations. Under CES, each destination-specific term equals $\Gamma(c, \sigma_0)$, so $T_R(\beta_0) = \Gamma(c, \sigma_0)$ almost surely. Let us consider the following object

$$\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) = T_R(\hat{\beta}_\kappa) - \Gamma(c, \hat{\sigma}), \tag{19}$$

which by a simple decomposition yields:

$$\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) = I_{R,\kappa} + II_{R,\kappa} + III_R + IV_R,$$

with

$$\begin{aligned} I_{R,\kappa} &\equiv T_R(\hat{\beta}_\kappa) - T_R(\beta_\kappa^*), \\ II_{R,\kappa} &\equiv T_R(\beta_\kappa^*) - T_R(\beta_0), \\ III_R &\equiv T_R(\beta_0) - \Gamma(c, \sigma_0), \\ IV_R &\equiv \Gamma(c, \sigma_0) - \Gamma(c, \hat{\sigma}). \end{aligned}$$

Under the usual regularity conditions for NLS, the results in Newey and McFadden (1994) give an observation-level influence representation

$$\sqrt{R}(\hat{\sigma} - \sigma_0) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \varphi_{\sigma,(r)} + o_p(1), \qquad \mathbb{E}[\varphi_{\sigma,(r)}] = 0.$$

Define

$$G_\kappa \equiv \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{P}_{0,(r)}, \xi_\kappa^*)\right], \qquad \zeta_{\kappa,(r)} \equiv G_\kappa' H_\kappa^{-1} \psi_{\kappa,(r)}.$$

Let

$$\chi_{\kappa,(r)} \equiv \zeta_{\kappa,(r)} - \dot{\Gamma}(c, \sigma_0)\varphi_{\sigma,(r)}, \qquad V_{\mathcal{T},\kappa} \equiv \mathbb{E}[\chi_{\kappa,(r)}^2],$$

where $\dot{\Gamma}(c, \sigma)$ denotes the derivative of $\Gamma(c, \sigma)$ with respect to σ.

Proposition 7. *Suppose that the model is CES and the conditions of Proposition 6 and Assumption 15 hold. Suppose that the tuning sequence is the one stated before Proposition 6. Then,*

$$\sqrt{R}\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \chi_{\kappa,(r)} + \sqrt{R} II_{R,\kappa} + o_p(1).$$

Proof. The proof is given in Appendix D. □

We impose the following nondegeneracy condition on the combined influence function.

Assumption 16. *Along the sequence $\kappa = \kappa_R$, we have $V_{\mathcal{T},\kappa} > 0$. For some $\delta > 0$,*

$$\frac{\mathbb{E}|\chi_{\kappa,(r)}|^{2+\delta}}{R^{\delta/2} V_{\mathcal{T},\kappa}^{1+\delta/2}} \longrightarrow 0.$$

Assumption 16 ensures that the leading stochastic term has positive variance and satisfies the conditions for a central limit theorem after standardization. The variance may decrease as the sieve dimension grows, provided that the stated moment condition holds.

Corollary 1. *Suppose that the conditions of Proposition 7 and Assumption 16 hold. If*

$$\sqrt{R} II_{R,\kappa} = o_p(V_{\mathcal{T},\kappa}^{1/2}),$$

then

$$\frac{\sqrt{R}\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma})}{V_{\mathcal{T},\kappa}^{1/2}} \xrightarrow{d} N(0,1).$$

Proof. See Appendix E.6. □

We next record the behavior of the test under a fixed alternative. Let σ^*_{CES} denote the elasticity component of the pseudo-true CES parameter and define

$$\Delta_0 \equiv \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} \Phi_j(\mathbf{P}_{0,(r)}, \beta_0 \mid \mathcal{W})\right] - \Gamma(c, \sigma^*_{CES}).$$

Proposition 8. *Suppose that the data are generated by a fixed HSA alternative. Suppose that the HSA estimator is consistent for β_0 and that the CES estimator is consistent for its pseudo-true parameter. If $\Delta_0 \neq 0$, then*

$$\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) \xrightarrow{p} \Delta_0.$$

Consequently, any two-sided test based on the absolute value of the root-R test statistic is consistent provided that its estimated standard deviation is bounded in probability and bounded away from zero in probability.

Proof. See Appendix E.7. □

Remark 7. The separation condition $\Delta_0 \neq 0$ is substantive. The statistic compares a scalar average of the HSA price-index functional with the corresponding pseudo-true CES functional. A non-CES alternative may nevertheless satisfy $\Delta_0 = 0$ through an accidental equality of these two scalar averages. The test is therefore consistent against fixed alternatives separated by this functional, but it is not an omnibus test against every departure from CES.

Remark 8. Appendix E.2.3 establishes $II_{R,\kappa} = O_p(\kappa^{-\gamma})$. A sufficient condition for the bias restriction in the corollary is therefore

$$\frac{\sqrt{R}\kappa^{-\gamma}}{V_{\mathcal{T},\kappa}^{1/2}} \longrightarrow 0.$$

If $V_{\mathcal{T},\kappa}$ is bounded away from zero, this reduces to $\sqrt{R}\kappa^{-\gamma} \to 0$. Theorem 6 in Newey (1997) imposes the corresponding undersmoothing condition $\sqrt{R}\kappa^{-s/r} \to 0$ for the asymptotic normality of a scalar functional, where s is the number of derivatives and r is the dimension of the function argument. Here $s = \gamma$ and $r = 1$ because each share function has the scalar effective relative price as its argument. Thus, when the asymptotic variance is bounded away from zero, our bias restriction agrees with Newey's condition. If $V_{\mathcal{T},\kappa}$ converges to zero, our condition is stronger because the approximation bias must be negligible relative to the standard deviation of the combined influence function.

4.3 The Feasible Matsuyama–Ushchev Test

The preceding results depend on population influence functions and an unknown asymptotic variance. We now replace these objects with sample analogues to obtain a feasible test. Let $\hat{\eta} = (\hat{\theta}', \hat{\sigma})'$ denote the CES estimator and let e_σ select its elasticity component, so that $e_\sigma' \hat{\eta} = \hat{\sigma}$. Write

$$Q_{(r)}^{CES}(\eta) = \frac{1}{N^2} \sum_{i,j} \left\{ \lambda_{ij,(r)} - g_{ij,(r)}^{CES}(\eta) \right\}^2.$$

Define

$$\widehat{H}_\kappa = \frac{1}{2R} \sum_{r=1}^{R} \nabla_{\xi'} \nabla_\xi Q_{(r)}(\hat{\xi}_\kappa), \qquad \widehat{G}_\kappa = \frac{1}{RN} \sum_{r=1}^{R} \sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{P}_{0,(r)}, \hat{\xi}_\kappa),$$

and

$$\widehat{\psi}_{\kappa,(r)} = \frac{1}{N^2} \sum_{i,j} \left\{ \lambda_{ij,(r)} - f_\kappa(z_{ij,(r)}(\hat{\xi}_\kappa); \hat{\xi}_\kappa) \right\} \nabla_\xi f_\kappa(z_{ij,(r)}(\hat{\xi}_\kappa); \hat{\xi}_\kappa).$$

For the CES estimator, define

$$\widehat{H}_{CES} = \frac{1}{2R} \sum_{r=1}^{R} \nabla_{\eta'} \nabla_\eta Q_{(r)}^{CES}(\hat{\eta})$$

and

$$\widehat{\psi}_{CES,(r)} = \frac{1}{N^2} \sum_{i,j} \left\{ \lambda_{ij,(r)} - g_{ij,(r)}^{CES}(\hat{\eta}) \right\} \nabla_\eta g_{ij,(r)}^{CES}(\hat{\eta}).$$

The estimated observation-level influence functions are

$$\widehat{\zeta}_{\kappa,(r)} = \widehat{G}_\kappa' \widehat{H}_\kappa^{-1} \widehat{\psi}_{\kappa,(r)}, \qquad \widehat{\varphi}_{\sigma,(r)} = e_\sigma' \widehat{H}_{CES}^{-1} \widehat{\psi}_{CES,(r)},$$

and

$$\widehat{\chi}_{\kappa,(r)} = \widehat{\zeta}_{\kappa,(r)} - \dot{\Gamma}(c, \hat{\sigma}) \widehat{\varphi}_{\sigma,(r)}.$$

Let $\overline{\widehat{\chi}}_\kappa = R^{-1} \sum_{r=1}^{R} \widehat{\chi}_{\kappa,(r)}$ and define

$$\widehat{V}_{\mathcal{T},\kappa} = \frac{1}{R} \sum_{r=1}^{R} \left(\widehat{\chi}_{\kappa,(r)} - \overline{\widehat{\chi}}_\kappa \right)^2.$$

Assumption 17. *The plug-in influence function satisfies*

$$\frac{1}{R V_{\mathcal{T},\kappa}} \sum_{r=1}^{R} \left(\widehat{\chi}_{\kappa,(r)} - \chi_{\kappa,(r)} \right)^2 \xrightarrow{p} 0.$$

This condition requires estimation of the two scores, the two Hessians, and G_κ to be sufficiently accurate relative to the standard deviation of the combined influence function. Appendix E.5 verifies that it yields relative consistency of the variance estimator.

Corollary 2. *Suppose that the conditions for the limiting result in the preceding subsection and As-*

sumption 17 hold. Then

$$\frac{\widehat{V}_{\mathcal{T},\kappa}}{V_{\mathcal{T},\kappa}} \xrightarrow{p} 1$$

and

$$\frac{\sqrt{R}\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma})}{\widehat{V}_{\mathcal{T},\kappa}^{1/2}} \xrightarrow{d} N(0,1).$$

Proof. See Appendix E.8. □

5 Counterfactual Analysis

Structural gravity provides a standard framework for counterfactual policy analysis (Yotov et al., 2016). Yotov (2024) emphasizes its evolution into an estimated computable general-equilibrium framework used for counterfactual projections. Allen, Arkolakis, and Takahashi (2020) establish existence, uniqueness, and robustness results for counterfactuals in a broad class of universal gravity models. Statistical uncertainty in counterfactual general-equilibrium effects has also been studied within parametric structural gravity. Anderson and Yotov (2016) construct confidence intervals for estimated general-equilibrium effects. Pfaffermayr (2020) derives the asymptotic distribution of counterfactual comparative statics using the delta method. We extend this line of work to HSA and study the large-sample properties of counterfactuals constructed from a nonparametrically estimated demand system.

Counterfactual analysis replaces the unknown parameter β_0 with its estimate $\hat{\beta}_\kappa$. The resulting counterfactual is a random object because it depends on the estimation error in $\hat{\beta}_\kappa$. The DGP in Section 2 allows us to give this object a clear statistical interpretation and study its large-sample properties. We focus on two counterfactual scenarios that are commonly considered by trade economists. The first scenario holds equilibrium prices fixed and studies a marginal change in trade costs. The second allows equilibrium prices to adjust following a general change in trade costs.

5.1 Marginal Change in Trade Costs

We begin with the marginal case because it isolates the effect of the estimated demand system while holding equilibrium prices fixed. We then show that the held-out estimated counterfactual converges to its unobserved true counterpart. We consider a scenario in which the trade costs $\tau_{ij}(\theta_0)$ are shifted marginally to $\tau_{ij}^* = \tau_{ij}(\theta_0) + \mathrm{d}\tau_{ij}$ for all distinct i, j. This marginal-change exercise follows Section 2 of Costinot and Rodríguez-Clare (2014). The domestic trade cost is kept fixed at $\tau_{ii}(\theta_0) = 1$, and we abstract from general-equilibrium adjustments in the price vector due to the shift in trade costs. Let $\mathbf{T}_0(\theta_0) = [\tau_{ij}(\theta_0)]_{1\le i\le N, 1\le j\le N}$ denote the matrix containing the trade costs before the shift, and similarly let $\mathbf{T}^*(\theta_0)$ denote the trade costs after the shift. Let us fix the endowments at $\ell^* \in \mathrm{int}(\mathrm{S}(\mathbf{L}))$ which fixes the equilibrium price vector at $\mathbf{p}_0^* = \mathbf{p}\left(\beta_0, \ell^* \mid \mathbf{T}_0(\theta_0)\right)$.

Due to homotheticity of HSA aggregation, the indirect utility $V_i(\mathbf{p}, E_i \mid \beta_0)$ is given by the ratio of the expenditure level E_i over the ideal price index $P_i^I(\mathbf{p}, \beta_0 \mid \mathbf{T}_0)$. Hence, totally differentiating the indirect utility function, we have:

$$\begin{aligned} \mathrm{d}\ln V_i(\mathbf{p}^*, E_i \mid \beta_0) &= \mathrm{d}\ln E_i - \mathrm{d}\ln P_i^I(\mathbf{p}^*, \beta_0 \mid \mathbf{T}_0), \\ &= \mathrm{d}\ln p_{0i}^* + d\ln \ell_i^* - \mathrm{d}\ln P_i^I\left(\mathbf{p}^*, \beta_0 \mid \mathbf{T}_0(\theta_0)\right), \\ &= -\mathrm{d}\ln P_i^I\left(\mathbf{p}^*, \beta_0 \mid \mathbf{T}_0(\theta_0)\right) \end{aligned} \tag{20}$$

where we used the equilibrium condition that expenditure satisfies $E_i = p_{0i}^* \ell_i^*$ in the second line, and the third line follows from the fact that we are only considering a shift in $\mathbf{T}_0(\theta_0)$ to $\mathbf{T}^*(\theta_0)$ keeping the prices fixed ($\mathrm{d}p_{0i}^* = 0$ for all i).

Using Shephard's lemma, we obtain the usual formula:

$$-\mathrm{d}\ln P_i^I\left(\mathbf{p}^*, \beta_0 \mid \mathbf{T}_0(\theta_0)\right) = -\sum_{j\neq i}^{N} S_{0j}\left(\frac{p_{0j}^*\tau_{ji}(\theta_0)}{A_i\left(\mathbf{p}_0^*, \beta_0 \mid \mathbf{T}_0(\theta_0)\right)}\right)\mathrm{d}\ln\tau_{ji}. \tag{21}$$

Let us consider a proportional shift where we have $\tilde{\tau}_{ji} = (1 + c_{ji})\tau_{ji}(\theta_0)$ for all i, j distinct and $c_{ji} > -1$. The corresponding log change is $\delta_{ji} \equiv \ln(1 + c_{ji})$. For a small change in trade costs chosen by the researcher, we have $\delta_{ji} \approx c_{ji}$ and a first order approximation for the percentage change in welfare level/ real consumption for consumer's $i = 1, \ldots, N$ by

$$\mathrm{d}\ln V_i(\mathbf{p}_0^*, E_i \mid \beta_0) \approx -\sum_{j\neq i}^{N} S_{0j}\left(\frac{p_{0j}^*\tau_{ji}(\theta_0)}{A_i\left(\mathbf{p}_0^*, \beta_0 \mid \mathbf{T}_0(\theta_0)\right)}\right) \times c_{ji}, \tag{22}$$

where we used the approximation $\mathrm{d}\ln\tau_{ji} = \delta_{ji} = \ln(1 + c_{ji}) \approx c_{ji}$ for all j, i distinct in the approximate equality. Let us set the counterfactual functional $\Delta_i(\mathbf{p} \mid \beta, \mathcal{W})$ as the right hand side of Equation (22). Throughout this section, the effective relative price for exporter j and destination i is

$$z_{ji}(\mathbf{p}, \beta) \equiv \frac{p_j\tau_{ji}(\theta)}{A_i(\mathbf{p}, \beta \mid \mathcal{W})}.$$

In most applications, the researcher evaluates the marginal change at one of the equilibrium observations contained in the dataset. The object of interest is therefore the marginal counterfactual at one observation rather than its average across observations. We use observations $r = 2, \ldots, R$ to estimate the model and reserve observation $r = 1$ for the counterfactual exercise. Define

$$\hat{\beta}_{\kappa,-1} \in \arg\min_{\beta\in\mathcal{B}_\kappa^*} \frac{1}{(R-1)N^2}\sum_{r=2}^{R}\sum_{i,j=1}^{N}\left|\lambda_{ij,(r)} - f_i\left(z_{ij,(r)}(\beta)\right)\right|^2.$$

At the held-out observation, define

$$\Delta_{i,(1)}^* \equiv \Delta_i(\mathbf{P}_{0,(1)} \mid \beta_0, \mathcal{W})$$

and

$$\hat{\Delta}_{i,(1)} \equiv \Delta_i(\mathbf{P}_{0,(1)} \mid \hat{\beta}_{\kappa,-1}, \mathcal{W}) = -\sum_{j \neq i} \hat{f}_{\kappa,-1}\left(z_{ji}(\mathbf{P}_{0,(1)}, \hat{\beta}_{\kappa,-1})\right) c_{ji}.$$

Proposition 9. *Suppose that the conditions of Proposition 5 hold. Suppose also that* $-1 < c_{ji}$ *and* $|c_{ji}| \leq \bar{c} < \infty$ *for all* j*. Then, for each* $i = 1, \ldots, N$*,*

$$\left|\hat{\Delta}_{i,(1)} - \Delta^*_{i,(1)}\right| \xrightarrow{p} 0.$$

Proof. See Appendix E. □

Proposition 9 also implies that $\hat{\Delta}_{i,(1)}$ converges in distribution to $\Delta^*_{i,(1)}$. Let Δ^*_i denote a generic random variable with the distribution of $\Delta_i(\mathbf{P}_0 \mid \beta_0, \mathcal{W})$ for $\mathbf{P}_0 \sim \mu$. Since $\hat{\Delta}_{i,(1)} - \Delta^*_{i,(1)} \xrightarrow{p} 0$ and $\Delta^*_{i,(1)}$ has the same distribution as Δ^*_i for every R, Slutsky's theorem gives

$$\hat{\Delta}_{i,(1)} \xrightarrow{d} \Delta^*_i.$$

Economically, the estimated counterfactual behaves in large samples as the unobserved true counterfactual $\Delta^*_{i,(1)}$. It is as if the researcher had sampled directly from the distribution of $\Delta^*_{i,(r)}$.

Proposition 9 concerns the marginal change at the held-out equilibrium observation. It does not concern the average marginal change across observations.

5.2 Counterfactual with General Equilibrium Adjustments

We now examine a proportional change in international trade costs. Unlike the first order approximation in Subsection 5.1, we have a general equilibrium adjustment of the price vector as the trade costs shift from $\mathbf{T}_0(\theta_0) \equiv [\tau_{ij}(\theta_0)]_{i,j=1}^N$ to $\mathbf{T}^{CF}(\theta_0 \mid c) \equiv [\tau^{CF}_{ij}(\theta_0 \mid c)]_{i,j=1}^N$, where

$$\tau^{CF}_{ij}(\theta \mid c) \equiv \begin{cases} (1+c)\tau_{ij}(\theta), & i \neq j, \\ 1, & i = j. \end{cases}$$

The constant $c \geq 0$ is fixed, bounded, known, and chosen by the researcher. In particular, $\mathbf{T}^{CF}(\theta_0 \mid 0) = \mathbf{T}_0(\theta_0)$.

As in Subsection 5.1, we use observations $r = 2, \ldots, R$ to estimate $\hat{\beta}_{\kappa,-1}$ and evaluate the counterfactual at the held-out observation $r = 1$. We condition on the fixed $\mathcal{W}$. We first define the true counterfactual outcome by first computing the equilibrium prices $\mathbf{p}_{0,(1)}(c) \equiv \mathbf{p}_0\left(\beta_0, \mathbf{L}_{(1)} \mid \mathbf{T}^{CF}(\theta_0 \mid c)\right)$ as per Definition 1. Proposition 1 implies that this equilibrium exists uniquely whenever $\mathbf{L}_{(1)} \in \mathbb{S}(\mathbf{L})$, which holds almost surely. We will denote $\mathbf{p}_{0,(1)}(0) \equiv \mathbf{p}_0\left(\beta_0, \mathbf{L}_{(1)} \mid \mathbf{T}^{CF}(\theta_0 \mid 0)\right) = \mathbf{p}_0\left(\beta_0, \mathbf{L}_{(1)} \mid \mathbf{T}_0(\theta_0)\right)$ as the initial equilibrium prices.

Our object of interest is the change in indirect utility of consumer j. Using Proposition 1 in Matsuyama and Ushchev (2017) to derive the change in the ideal price index, the change of

indirect utility from a change in prices from $\mathbf{p}_{0,(1)}(0)$ to $\mathbf{p}_{0,(1)}(c)$ is given by

$$\begin{aligned}\Delta^{*}_{j,(1)}(c) \equiv & \ln\left[p_{0j,(1)}(c)L_{j,(1)}\right] - \ln\left[p_{0j,(1)}(0)L_{j,(1)}\right] \\ & - \ln P^{I}_{j}\left(\mathbf{p}_{0,(1)}(c), \beta_0 \mid \mathbf{T}^{CF}(\theta_0 \mid c)\right) + \ln P^{I}_{j}\left(\mathbf{p}_{0,(1)}(0), \beta_0 \mid \mathbf{T}_0(\theta_0)\right) \\ = & \ln p_{0j,(1)}(c) - \ln p_{0j,(1)}(0) \\ & - \ln A_j\left(\mathbf{p}_{0,(1)}(c), \beta_0 \mid \mathbf{T}^{CF}(\theta_0 \mid c)\right) + \ln A_j\left(\mathbf{p}_{0,(1)}(0), \beta_0 \mid \mathbf{T}_0(\theta_0)\right) \\ & - \Phi_j\left(\mathbf{p}_{0,(1)}(c), \beta_0 \mid \mathbf{T}^{CF}(\theta_0 \mid c)\right) + \Phi_j\left(\mathbf{p}_{0,(1)}(0), \beta_0 \mid \mathbf{T}_0(\theta_0)\right).\end{aligned}$$

The feasible counterfactual is obtained by replacing β_0 with $\hat{\beta}_{\kappa,-1}$. In particular, the feasible counterfactual equilibrium is computed under $\mathbf{T}^{CF}(\hat{\theta}_{\kappa,-1} \mid c)$. Let us denote the resulting counterfactual by $\hat{\Delta}_{j,(1)}(c)$. The analysis of the Matsuyama–Ushchev test already established the required continuity and large-sample properties of the single aggregator A_j and the residual price-index component Φ_j. The additional work for the general-equilibrium counterfactual is therefore to establish that the counterfactual equilibrium price vector computed from $\hat{\beta}_{\kappa,-1}$ converges to the price vector computed from β_0.

Assumption 18. *Let $\mathbb{Z}^{CF}_j$ denote the excess-demand function defined in Equation (9) evaluated under the counterfactual trade-cost matrix $\mathbf{T}^{CF}(\theta \mid c)$. For each fixed c under consideration and for every sequence $\{\beta_n\}$ and $\{\mathbf{p}_n\}$ satisfying $M(\beta_n, \beta_0) \to 0$ and $\mathbf{p}_n \to \mathbf{p}$ in the normalized price set given in Proposition 1,*

$$\max_{1 \le j \le N} \left| \mathbb{Z}^{CF}_j(\mathbf{p}_n, \beta_n \mid c, \mathbf{L}_{(1)}) - \mathbb{Z}^{CF}_j(\mathbf{p}, \beta_0 \mid c, \mathbf{L}_{(1)}) \right| \to 0$$

almost surely. The same conditions hold at $c = 0$.

Assumption 18 is a high-level smoothness condition on the excess-demand system. It says that small changes in the estimated trade-share system and in prices produce only small changes in excess demand. Economically, it rules out environments in which an arbitrarily small estimation error causes a discrete reallocation of expenditure across origins and therefore a discontinuous change in goods-market imbalances. The assumption does not impose equilibrium uniqueness, which follows from Proposition 1. It ensures instead that the unique counterfactual equilibrium is stable with respect to small estimation errors in the primitives.

Proposition 10. *Suppose that the conditions of Proposition 5, Assumption 15, and Assumption 18 hold. Suppose that c is chosen so that the counterfactual trade costs $\mathbf{T}^{CF}(\theta \mid c)$ satisfy the three axioms of iceberg trade costs for every $\theta \in \Theta$. Then, for each $j = 1, \ldots, N$,*

$$\left| \hat{\Delta}_{j,(1)}(c) - \Delta^{*}_{j,(1)}(c) \right| \xrightarrow{p} 0.$$

Consequently, $\hat{\Delta}_{j,(1)}(c)$ converges in distribution to the distribution of the true counterfactual random variable $\Delta^{*}_{j,(1)}(c)$.

Proof. See Appendix F.2. □

The interpretation is similar to the marginal-change counterfactual in the previous subsection. Economically, the feasible general-equilibrium counterfactual $\hat{\Delta}_{j,(1)}(c)$ behaves in large samples as the unobserved true counterfactual $\Delta^*_{j,(1)}(c)$. It is as if the researcher had sampled directly from the distribution of the true counterfactual outcome, including the adjustment of equilibrium prices induced by the change in trade costs.

6 Simulation Study for the HSA Model

This section has three goals. First, we study the finite-sample performance of the sieve estimator whose consistency and rate of convergence are established in Proposition 5. Second, we study the finite-sample size and power of the Matsuyama–Ushchev test developed in Subsection 4.2. Third, we study the performance of the counterfactual estimator established in Proposition 9. For computational simplicity, we focus on the marginal change in trade costs rather than the counterfactual with general equilibrium adjustments. Implementation details are provided in Appendix H.

6.1 CES–Cobb–Douglas Mixture Design

This is the Cobb–Douglas–CES mixture in Example 1, where the Cobb–Douglas weights are set to $1/N$. The parameter $\eta_0 \in [0, 1)$ controls the departure from CES. The expenditure shares are

$$\lambda^*_{ij,(r)} = \frac{\eta_0}{N} + (1 - \eta_0) \left(\frac{P_{0i,(r)} \tau_{ij}(\theta_0)}{A^{CES}_j(\mathbf{P}_{0,(r)}, \theta_0, \sigma_0)} \right)^{1-\sigma_0}, \tag{23}$$

where

$$A^{CES}_j(\mathbf{p}, \theta, \sigma) = \left\{ \sum_{i=1}^{N} [p_i \tau_{ij}(\theta)]^{1-\sigma} \right\}^{1/(1-\sigma)}.$$

We set $\sigma_0 = 2$, $\theta_0 = 5$, and $\tau_{ij}(\theta) = \exp(W_{ij}\theta)$ for $i \neq j$. We set $\tau_{ii}(\theta) = 1$. The design is CES when $\eta_0 = 0$ and is an HSA alternative when $\eta_0 > 0$.

For each replication, we draw R endowment vectors independently from the fixed distribution μ. Each component is generated as $L_i = 500 + 50B_i$, where B_i has a Beta$(2, 2)$ distribution. For each draw, we solve the excess-demand system and normalize the numerical equilibrium price vector by setting $P_{0N,(r)} = 1$. This price normalization is distinct from the scale normalization imposed on the HSA representation below. We then form the observed shares as

$$\lambda_{ij,(r)} = \lambda^*_{ij,(r)} + \varepsilon_{ij,(r)},$$

where the disturbances are IID across observations. The disturbance scale is fixed at the middle noise specification described below.

We set $N = 27$, corresponding to the number of member states in the European Union, and $R = 20$. The bilateral characteristic W_{ij} is the Euclidean distance between two points drawn independently from the uniform distribution on $[0, 0.2]^2$. We draw these locations once and hold $\mathcal{W}$ fixed across replications. We consider $\eta_0 \in \{0, 0.25, 0.5\}$ and the sieve dimensions $\kappa \in \{1, 2, 3\}$. When $\sigma_0 = 2$, the share function in the CES–Cobb–Douglas mixture takes the form $S_0(z) = \alpha_0 + \alpha_1 z^{-1}$. The specification with $\kappa = 1$ therefore contains the true share function. We use one moderately noisy design with a signal-to-noise ratio of approximately 2. Because the variance of the true shares changes with η_0, we calibrate the disturbance scale separately for

each value of η_0. The resulting scales are 0.02991, 0.02243, and 0.01504 for $\eta_0 = 0$, 0.25, and 0.5, respectively.

The parameter and counterfactual exercises use $R = 20$. To examine whether the power of the Matsuyama–Ushchev test improves with the number of observed equilibrium outcomes, we repeat the test-only exercise with $R = 40$ and $R = 60$. All other features of the design are held fixed.

The normalized estimation procedure and the construction of the feasible test are described in Appendix H. The variance calculation is a direct application of the parametric delta method to the reduced sieve coordinates and the CES parameters.

6.2 Marginal Counterfactual

The counterfactual exercise follows Subsection 5.1. In each replication, observation 1 is held out. We estimate the model using observations $2, \ldots, R$ and evaluate the marginal effect of a 5 percent increase in all foreign trade costs into destination N at the prices and endowments of observation 1. The true and estimated effects are

$$\Delta^*_{N,(1)} = -0.05 \sum_{j \neq N} S_0(z_{jN,(1)}(\beta_0))$$

and

$$\widehat{\Delta}_{N,(1)} = -0.05 \sum_{j \neq N} \widehat{f}_{\kappa,-1}(z_{jN,(1)}(\hat{\beta}_{\kappa,-1})).$$

This design mirrors Proposition 9 and does not average the counterfactual over the estimation sample. We compare the empirical distributions of the true HSA effect, the normalized sieve estimate, and the CES estimate.

All numerical results in this section are regenerated using the normalized program described above. The replication-level output records convergence diagnostics, the normalization error, the minimum and maximum relative prices, the estimated test variance, and the held-out counterfactual error.

6.3 Results

Table 1 reports the parameter estimates from 1,000 replications. The HSA estimator recovers the true trade-cost parameter across all sieve dimensions and values of η_0. Every mean estimate lies within 0.1 of $\theta_0 = 5$. The estimator with $\kappa = 1$ is essentially unbiased in every design, while the estimates with $\kappa = 2$ and $\kappa = 3$ display a small downward bias. The mean of every additional sieve coefficient is smaller than 0.001, as it should be because the true share function belongs to the sieve with $\kappa = 1$. Under the CES null, the CES estimator recovers both $\theta_0 = 5$ and $\sigma_0 = 2$. Under the two alternatives, the CES estimator instead converges toward pseudo-true values. Its mean estimate of θ falls to 4.521 when $\eta_0 = 0.25$ and to 4.039 when $\eta_0 = 0.5$.

Table 2 reports the rejection frequency of the feasible Matsuyama–Ushchev test. The null

Table 1: Parameter Estimates

η_0	κ	Mean $\hat{\theta}_\kappa$	SD $\hat{\theta}_\kappa$	Mean $\hat{\theta}_{CES}$	Mean $\hat{\sigma}$
0	1	5.005	0.169	5.002	2.001
0	2	4.943	0.181	5.002	2.001
0	3	4.935	0.178	5.002	2.001
0.25	1	5.005	0.165	4.521	1.845
0.25	2	4.938	0.184	4.521	1.845
0.25	3	4.936	0.181	4.521	1.845
0.50	1	5.004	0.172	4.039	1.639
0.50	2	4.907	0.207	4.039	1.639
0.50	3	4.919	0.200	4.039	1.639

Notes. The true parameters are $\theta_0 = 5$ and $\sigma_0 = 2$. The CES estimates under $\eta_0 > 0$ are estimates of pseudo-true parameters.

rejection frequency is close to the 5 percent nominal level for $\kappa = 1$ and $\kappa = 2$ when $R = 40$ and $R = 60$. The test is conservative for $\kappa = 3$. It rejects in every replication under the $\eta_0 = 0.25$ alternative at these larger sample sizes. Under the $\eta_0 = 0.5$ alternative, power increases from 61.7, 34.0, and 24.7 percent at $R = 20$ to 95.1, 69.6, and 60.9 percent at $R = 60$ for $\kappa = 1, 2, 3$, respectively.

Power need not be monotone in η_0. The test evaluates one scalar implication of CES at $c = 5$. At $R = 60$, the absolute mean discrepancy is approximately 0.025 under $\eta_0 = 0.25$ and approximately 0.017 to 0.020 under $\eta_0 = 0.5$. The estimated variance is also larger under $\eta_0 = 0.5$. The standardized separation from the null is therefore smaller in that design. This is a feature of the scalar moment being tested rather than a failure of the HSA estimator. A joint test based on several values of c may improve power against a broader set of alternatives.

Table 2: Rejection Frequencies of the Matsuyama–Ushchev Test

η_0	κ	$R = 20$	$R = 40$	$R = 60$
0	1	0.084	0.064	0.064
0	2	0.081	0.064	0.066
0	3	0.036	0.023	0.022
0.25	1	1.000	1.000	1.000
0.25	2	0.990	1.000	1.000
0.25	3	0.980	1.000	1.000
0.50	1	0.617	0.849	0.951
0.50	2	0.340	0.523	0.696
0.50	3	0.247	0.414	0.609

Notes. Entries are rejection frequencies at the 5 percent level across 1,000 replications. The test is evaluated at $c = 5$. The $R = 20$ results use the datasets employed in the parameter and counterfactual exercises. The $R = 40$ and $R = 60$ columns come from a separate test-only experiment that holds the remaining design features fixed.

Finally, Table 3 compares the estimated and true distributions of the held-out marginal counterfactual. The HSA counterfactual has negligible bias and a small root mean squared error in every design. Its empirical CDF is also close to the empirical CDF of the true coun-

terfactual. Under the two alternatives, the corresponding CES distribution is substantially distorted. The difference becomes larger as the Cobb–Douglas weight increases.

Table 3: Marginal Counterfactual Performance

		HSA			CES		
η_0	κ	Bias	RMSE	ECDF distance	Bias	RMSE	ECDF distance
0	1	2.63×10^{-7}	1.12×10^{-5}	0.026	9.20×10^{-8}	8.43×10^{-6}	0.020
0	2	1.84×10^{-6}	1.16×10^{-5}	0.042	9.20×10^{-8}	8.43×10^{-6}	0.020
0	3	1.67×10^{-6}	1.14×10^{-5}	0.045	9.20×10^{-8}	8.43×10^{-6}	0.020
0.25	1	2.24×10^{-7}	8.17×10^{-6}	0.028	-4.09×10^{-5}	4.14×10^{-5}	0.504
0.25	2	1.02×10^{-6}	8.41×10^{-6}	0.041	-4.09×10^{-5}	4.14×10^{-5}	0.504
0.25	3	7.96×10^{-7}	8.29×10^{-6}	0.043	-4.09×10^{-5}	4.14×10^{-5}	0.504
0.50	1	2.94×10^{-9}	5.82×10^{-6}	0.018	-5.81×10^{-5}	5.83×10^{-5}	0.790
0.50	2	4.86×10^{-7}	6.01×10^{-6}	0.025	-5.81×10^{-5}	5.83×10^{-5}	0.790
0.50	3	2.60×10^{-7}	5.96×10^{-6}	0.024	-5.81×10^{-5}	5.83×10^{-5}	0.790

Notes. The ECDF distance is the Kolmogorov distance between the empirical distributions of the estimated and true counterfactual effects.

Figure 2 displays the corresponding empirical CDFs for all three sieve dimensions. When $\eta_0 = 0$, both the HSA and CES estimators recover the empirical distribution of the true counterfactual, as suggested by the convergence in distribution in Proposition 9. When CES is misspecified, its empirical CDF is visibly far from the empirical CDF of the true counterfactual. The HSA estimator continues to recover the true counterfactual distribution under both alternatives. The empirical CDFs obtained from the three sieve dimensions are visually indistinguishable.

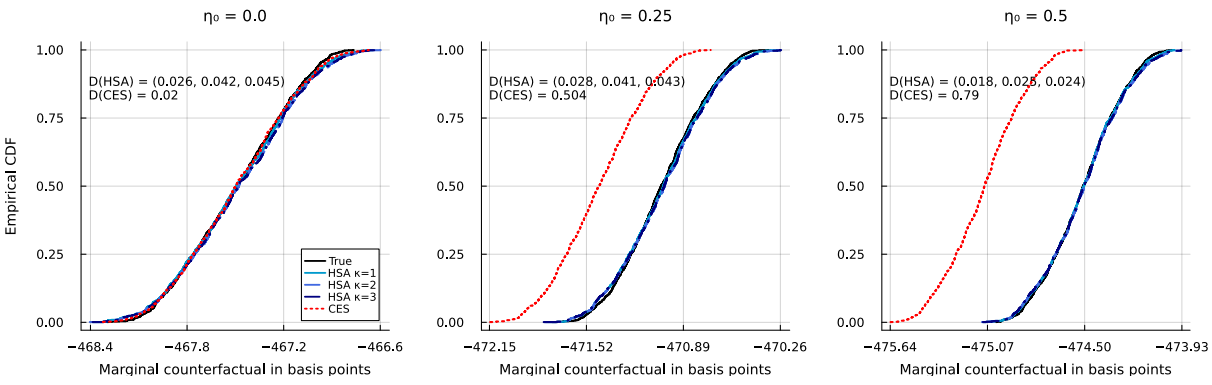


Figure 2: Empirical Distributions of the Marginal Counterfactual

Notes. Each panel contains 1,000 replications for each sieve dimension. The horizontal axis reports the marginal counterfactual in basis points. The reported distances are Kolmogorov distances from the empirical distribution of the true counterfactual. The three HSA distances are ordered by $\kappa = 1, 2, 3$.

All 9,000 HSA and CES estimation problems report convergence. The largest normalization error is 4.44×10^{-16}. Every estimated share is positive over the realized support, every estimated share function is decreasing, and every estimated share function is convex.

7 Empirical Application

We estimate CES and HSA demand using annual bilateral trade data for 1995–2014. The main HSA specification imposes nonnegative inverse-power coefficients and a nonnegative intercept, giving a share function that is positive and strictly decreasing for all positive relative prices. We use this specification to compare trade-cost estimates, local elasticities, and welfare effects with CES. Signed-coefficient estimates provide robustness comparisons for fit and local demand responses. We also evaluate predictive fit using 2014 as a validation sample. The empirical results are point estimates and descriptive comparisons.

We combine the 2013 and 2016 releases of the World Input–Output Tables (WIOT) (Timmer et al., 2012). We refer to the former vintage as WIOT2013 and the latter as WIOT2016. WIOT2013 spans from the year 1995 to 2010 and covers 40 countries, while WIOT2016 spans from the year 2000 to 2014 and covers 43 countries. We combine the two vintages to construct a dataset spanning 1995 to 2014, where the number of countries is 40 plus the rest of the world (RoW) category. The RoW is an aggregate of all countries that are not included among the 40 countries. We follow the online appendix of Costinot and Rodríguez-Clare (2014) to adjust the data so that it is compatible with a static economic model. This gives us 20 years of trade shares $\{\Lambda_{(r)}\}_{r=1995}^{2014}$.

For the trade cost variables $\{\mathcal{W}_{(r)}\}_{r=1995}^{2014}$, we use the CEPII dataset (Conte, Cotterlaz, and Mayer, 2022). The variables are the log distance between two capital cities (logdist), a dummy variable indicating whether the two countries share a common official language (comlang), and a dummy variable indicating whether they share a border (contig). For importer–exporter pairs that include the RoW, we set the dummy variables for common language and sharing a border to zero. The distance variable is a weighted average of the distances of all countries available in CEPII excluding the 40 countries in WIOT2013. When taking the weighted average, we use population weights.

The endowment and labor data $\{\mathbf{L}_{(r)}\}_{r=1995}^{2014}$ are taken from CREI's population data, and the expenditure levels $\{\mathbf{E}_{(r)}\}_{r=1995}^{2014}$ are constructed by summing all imports for each country and year using the trade flows contained in WIOT.

The main HSA specification uses $\kappa = 1$. Under the positivity restrictions, the higher-order coefficients for $\kappa = 2, 3$ reach their numerical lower bounds, yielding essentially the same fitted share function. We report signed $\kappa = 2, 3$ estimates as robustness comparisons. Appendix I.1 provides the CES estimation details, and Appendix I.3 describes the HSA implementation.

7.1 Demand Specification and Trade-Cost Estimates

We jointly estimate the trade-cost parameters and a symmetric share function. All specifications use the same annual observations and trade-cost covariates, so their comparison measures sensitivity to the demand restriction.

We approximate the common share function by

$$f_\kappa(z;\alpha) = \alpha_0 + \sum_{q=1}^{\kappa} \alpha_q z^{-q}, \qquad \kappa \in \{1,2,3\}. \tag{24}$$

The normalization determines α_0. For the main specification, we require $\alpha_0 \geq 0$, $\alpha_1 > 0$, and $N\alpha_0 < 1$. The resulting $\kappa = 1$ share function is globally positive and strictly decreasing, and the adding-up equation has a unique positive aggregator at every strictly positive delivered-price vector. For the signed $\kappa = 2,3$ robustness specifications, positivity and monotonicity are imposed on a fixed compact interval. Appendix I.3 describes both implementations.

For international pairs, trade costs satisfy

$$\tau_{ij}(\theta) = \exp\left(\theta_d \,\mathrm{logdist}_{ij} + \theta_\ell \,\mathrm{comlang}_{ij} + \theta_b \,\mathrm{contig}_{ij}\right), \qquad i \neq j, \tag{25}$$

with $\tau_{ii}(\theta) = 1$. We impose this same trade-cost specification under CES and each HSA sieve dimension. In every case, θ belongs to the cone Θ implied by the iceberg-cost restrictions, as established in Lemma 1. These restrictions do not impose individual coefficient signs.

Table 4 compares the point estimates using the same 20 annual observations. Distance raises fitted trade costs and a common language lowers them under all four specifications. The HSA distance and language coefficients are similar between the positive main specification and the signed robustness specifications, and substantially smaller in magnitude than their CES counterparts. The HSA border coefficient is close to zero and changes sign across these specifications. In contrast, CES produces a large positive border coefficient.

Table 4: Trade-Cost Parameter Estimates under CES and HSA

Demand specification	Distance	Common language	Shared border
CES	17.1161	-34.7247	21.8525
Positive HSA, $\kappa = 1$	1.0911	-2.5002	-0.0230
Signed HSA, $\kappa = 2$	1.1086	-2.5747	0.0115
Signed HSA, $\kappa = 3$	1.1092	-2.5540	-0.0108

Notes. All specifications use annual data for 1995–2014 and the same exponential iceberg-cost specification and cone restrictions. The CES elasticity estimate is 1.0393. The positive $\kappa = 1$ model is the main specification. Signed $\kappa = 2,3$ estimates impose positivity and monotonicity on a fixed compact interval. Entries are point estimates without standard errors.

These differences show that trade-cost parameter estimates are sensitive to the demand specification, even with the same covariates and sample. The simulation study in Section 6 illustrates one source of this sensitivity: under misspecified CES demand, the estimator targets pseudo-true parameters rather than the structural trade-cost parameters. The empirical comparison documents sensitivity to the demand restriction without identifying which specification is correctly specified.

7.2 Estimated Demand

Table 5 reports the estimated share-function coefficients and least-squares criteria. The main positivity-restricted estimate is

$$\widehat{S}_1(z) = 0.003891 + 0.828192z^{-1}.$$

This specification admits a relative-price-dependent elasticity while imposing global positivity and monotonicity. Under nonnegative inverse-power coefficients, the quadratic and cubic terms are approximately 10^{-10}, their numerical lower bound. Increasing the order within the restricted family therefore leaves the fit essentially unchanged.

The signed quadratic robustness estimate is

$$\widehat{S}_2(z) = 0.003895 + 0.914608z^{-1} - 0.086711z^{-2}.$$

Allowing signed coefficients reduces the criterion by 0.077 percent for $\kappa = 2$ and 0.117 percent for $\kappa = 3$, relative to their positive-coefficient counterparts. These modest gains in fit accompany similar distance and language coefficients. The signed specifications are admissible on their fixed estimation domain; they are used here to assess sensitivity of fitted demand and local elasticities rather than to extrapolate general-equilibrium welfare effects outside that domain.

Figure 3 reports the local trade-cost elasticity

$$\varepsilon_\tau(z) = -\left.\frac{\partial \log S(z_{ij})}{\partial \log \tau_{ij}}\right|_{p_i, A_j} = -\frac{zS'(z)}{S(z)}.$$

This elasticity measures the response of the bilateral expenditure share to trade costs, holding the producer price and destination aggregator fixed. Under CES it is constant at $\hat{\sigma} - 1 = 0.0393$. Under HSA it varies with relative prices, allowing the local response to depend on the initial price configuration. Figure 3 evaluates the estimates over their common fitted relative-price range within the admissible interval. The curves describe local demand responses; general-equilibrium effects also incorporate adjustments in producer prices and aggregators.

Table 6 reports prediction errors for foreign bilateral expenditure shares in 2014. Parameters and admissible intervals are estimated using 1995–2013, and predictions are evaluated on the same observations across specifications. Errors are expressed in percentage points of expenditure shares. Because 2014 informed the choice of sieve dimension, these results are interpreted as a validation exercise.

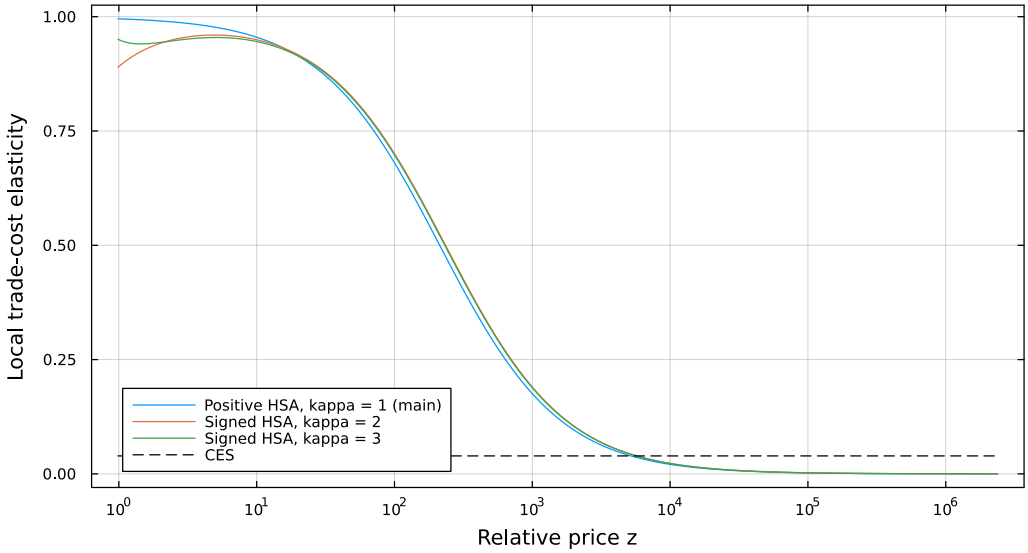
The main and robustness estimates produce similar trade-cost coefficients and local elasticities, despite their different shape restrictions. The positivity restriction sacrifices little fit while providing a globally admissible demand system for the policy exercises. The empirical Matsuyama–Ushchev test statistic is -4.0814. Its absolute value exceeds the conventional standard-normal cutoff of 1.96.

Table 5: HSA Share-Function Estimates

Specification	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	Criterion
Positive, $\kappa = 1$	0.003891	0.828192	–	–	0.000262986
Signed, $\kappa = 2$	0.003895	0.914608	-0.086711	–	0.000262784
Signed, $\kappa = 3$	0.003915	0.908932	-0.125521	0.043854	0.000262680

Notes. Estimates use annual observations for 1995–2014. The main specification is positive $\kappa = 1$; the signed specifications are robustness comparisons. The criterion is the mean squared share residual. The main share function is globally positive and strictly decreasing. The signed functions satisfy these restrictions on $y = 1/z \in [2.09949 \times 10^{-7}, 2.02904]$. All specifications cover the 820 destination-year columns. Entries are point estimates; standard errors are not reported. Appendix I.3 describes the numerical procedure.

Figure 3: Local Trade-Cost Elasticities



Notes. The curves use the full-sample positive HSA main estimate ($\kappa = 1$), signed HSA robustness estimates ($\kappa = 2, 3$), and the CES estimate. The horizontal axis uses a logarithmic scale. The plotted range is the intersection of the fitted relative-price ranges of the three HSA specifications, restricted to the admissible domain. The CES elasticity is $\hat{\sigma} - 1$.

Table 6: Foreign-Share Prediction Errors in 2014

	All destinations		United States	
Model	RMSE	MAE	RMSE	MAE
Positive HSA, $\kappa = 1$	1.2574	0.5969	0.4788	0.3951
Signed HSA, $\kappa = 2$	1.2578	0.5975	0.4808	0.3969
Signed HSA, $\kappa = 3$	1.2580	0.5974	0.4809	0.3966
CES	1.2681	0.5840	0.4623	0.3294

Notes. Parameters and the HSA intervals are estimated using 1995–2013. Errors exclude domestic shares and compare models on identical observations. All 41 destination columns are covered. The intervals for the signed models are not expanded during validation. The 2014 sample was used to compare specifications and is therefore treated as validation rather than an untouched test sample.

7.3 Welfare Effects of Trade-Cost Changes

We compare welfare effects under the positive HSA main specification and CES using the full-sample estimates. The experiments increase iceberg trade costs while holding endowments at their 2014 values. Domestic trade costs remain one, and the cost increases generate no tariff revenue. Each model uses its own estimated trade costs, so HSA–CES differences reflect the jointly estimated demand and trade-cost systems. All welfare effects are plug-in point estimates.

For a marginal 5 percent increase in foreign trade costs into the United States, holding observed 2014 producer prices fixed, Equation (22) gives

$$100\,\mathrm{d}\log V_{\mathrm{US}} = -5\sum_{j\neq \mathrm{US}} \widehat{S}\left(\frac{p_{j,2014}\widehat{\tau}_{j,\mathrm{US}}}{\widehat{A}_{\mathrm{US},2014}}\right).$$

The implied first-order U.S. welfare loss is 0.9143 percent under HSA and 0.6598 percent under CES. The signed quadratic and cubic estimates give 0.9150 and 0.9163 percent, respectively, close to the main HSA estimate. These local log-welfare approximations depend on fitted foreign expenditure shares at the baseline prices and do not require resolving the equilibrium.

Table 7: U.S. Welfare Effects of Iceberg Trade-Cost Increases

Experiment	Positive HSA, $\kappa = 1$	CES
Marginal 5%, observed producer prices	−0.9143	−0.6598
Unilateral 15%, general equilibrium	−2.1665	−1.5778
Bilateral 15%, general equilibrium	−2.6576	−1.6365

Notes. The marginal row reports $100\,\mathrm{d}\log V$ at observed 2014 producer prices. The general-equilibrium rows report $100(V^{\mathrm{cf}}/V^{0}-1)$ relative to each model's reconstructed equilibrium with 2014 endowments. They therefore use different baseline prices from the marginal row. The unilateral experiment raises foreign costs into the United States by 15 percent. The bilateral experiment also raises costs of U.S. exports by 15 percent. The HSA main specification is globally positive and monotone. No tariff revenue is collected.

For counterfactual analysis with general equilibrium adjustments, we first solve for equilibrium prices under each model using the estimated trade costs and 2014 endowments. We then increase trade costs and solve for the new equilibrium prices. Welfare changes are measured relative to the initial equilibrium of the same model.

When the cost of shipping foreign goods to the United States rises by 15 percent, U.S. welfare falls by 2.1665 percent under HSA and 1.5778 percent under CES. When the cost of shipping U.S. goods abroad also rises by 15 percent, the U.S. welfare losses increase to 2.6576 percent under HSA and 1.6365 percent under CES. Thus, both models predict larger losses when trade costs rise in both directions. HSA predicts a larger U.S. welfare loss than CES in both experiments. These differences reflect both the estimated demand functions and the estimated trade costs.

8 Conclusion

This paper develops a statistical framework for estimating the HSA demand system in a quantitative model of international trade. Our starting point is an explicit data-generating process in which the researcher observes repeated equilibrium outcomes for a fixed set of countries. Each complete bilateral share matrix is one observation. Within an observation, utility maximization imposes adding-up restrictions and market clearing links total sales to expenditure. Treating these outcomes jointly allows the statistical analysis to use the general-equilibrium restrictions that generated the data.

Under the stated identification conditions, these restrictions identify the demand system. In the CES special case, they separately identify the trade-cost parameters and the elasticity of substitution, which conventional gravity estimation recovers only through a composite parameter. For the nonparametric HSA model, variation in equilibrium prices identifies the trade-cost parameters and the common share function up to a scale normalization. We use this result to construct a normalized sieve least-squares estimator and establish its consistency and a global convergence bound as the number of equilibrium realizations increases.

The same sampling framework permits statistical analysis of two objects that are central to quantitative trade applications. First, we convert the sharp characterization of CES in Matsuyama and Ushchev (2017) into the Matsuyama–Ushchev test. We derive its large-sample behavior under the CES null and establish consistency against fixed alternatives with a nonzero population discrepancy. Second, we show that counterfactual estimators constructed from the estimated HSA demand system converge to their unobserved population counterparts. This gives counterfactual analysis a statistical interpretation under both marginal changes in trade costs and changes that allow equilibrium prices to adjust.

The simulation study indicates that these theoretical results are useful in finite samples. The HSA estimator accurately recovers the finite-dimensional parameters and the counterfactual distribution under the non-CES designs. The Matsuyama–Ushchev test has good power against the alternatives considered, although its size can be conservative for the larger sieve dimension. The results also illustrate the cost of CES misspecification. When the true demand system is not CES, the CES estimator converges toward pseudo-true parameters and produces a visibly different counterfactual distribution.

The empirical application uses annual World Input–Output Tables (Timmer et al., 2012) for 1995–2014. The main HSA specification restricts the inverse-power coefficients to be nonnegative. Allowing signed coefficients as a robustness comparison yields similar local demand estimates and reduces the least-squares criterion by less than 0.12 percent. The HSA trade-cost estimates differ substantially from CES, and the estimated HSA trade-cost elasticity varies with relative prices rather than remaining constant.

For a marginal 5 percent increase in the cost of shipping foreign goods to the United States, the first-order U.S. welfare loss is 0.9143 percent under HSA and 0.6598 percent under CES. This calculation holds observed producer prices fixed and does not require resolving the equi-

librium. With general equilibrium adjustments, a 15 percent increase in these costs reduces U.S. welfare by 2.1665 percent under HSA and 1.5778 percent under CES. When the cost of shipping U.S. goods abroad also rises by 15 percent, the losses increase to 2.6576 and 1.6365 percent, respectively. HSA therefore predicts larger U.S. welfare losses in both experiments. These differences reflect both estimated demand and estimated trade costs. The empirical Matsuyama–Ushchev test statistic is -4.0814. Its absolute value exceeds the conventional standard-normal cutoff of 1.96.

The analysis deliberately uses a simple Armington endowment economy to isolate the statistical role of general equilibrium. Extending the framework to environments with production, firm heterogeneity, and additional sources of equilibrium variation is a natural direction for future work. The main lesson extends beyond this particular model. When economic outcomes are jointly generated in general equilibrium, the restrictions linking bilateral outcomes within an equilibrium can provide identifying information that is lost when those outcomes are treated as separate sampling units. Using that information makes it possible to retain the theoretical flexibility of HSA while bringing the demand system into quantitative empirical work.

References

Adão, R., Arkolakis, C., and Ganapati, S. (2026). "From heterogeneous firms to heterogeneous trade elasticities: The aggregate implications of firm export decisions". *Journal of Political Economy*. Forthcoming.

Adão, R., Costinot, A., and Donaldson, D. (2017). "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade". *American Economic Review* 107.3, pp. 633–689.

Allen, T., Arkolakis, C., and Takahashi, Y. (2020). "Universal Gravity". *Journal of Political Economy* 128.2, pp. 393–433.

Anderson, J. E., Larch, M., and Yotov, Y. V. (2018). "GEPPML: General Equilibrium Analysis with PPML". *World Economy* 41.10, pp. 2750–2782.

Anderson, J. E. and Van Wincoop, E. (2003). "Gravity with gravitas: A solution to the border puzzle". *American economic review* 93.1, pp. 170–192.

Anderson, J. E. and Yotov, Y. V. (2016). "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990–2002". *Journal of International Economics* 99, pp. 279–298.

Armington, P. S. (1969). "A Theory of Demand for Products Distinguished by Place of Production". *Staff Papers-International Monetary Fund*, pp. 159–178.

Baldwin, R. and Taglioni, D. (2006). *Gravity for Dummies and Dummies for Gravity Equations*. Working Paper 12516. National Bureau of Economic Research.

Baqaee, D. R., Farhi, E., and Sangani, K. (2024). "The darwinian returns to scale". *Review of Economic Studies* 91.3, pp. 1373–1405.

Carrère, C., Mrázová, M., and Neary, J. P. (2020). "Gravity without Apology: The Science of Elasticities, Distance, and Trade". *Economic Journal* 130.628, pp. 880–910.

Chen, N. and Novy, D. (2022). "Gravity and Heterogeneous Trade Cost Elasticities". *Economic Journal* 132.644, pp. 1349–1377.

Chen, X. (2007). "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models". In: *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 5549–5632.

Chow, C. P., Kasahara, H., and Sugita, Y. (2026). *Identification and Estimation of Production Function and Consumer Demand Function under Monopolistic Competition from Revenue Data*.

Conte, M., Cotterlaz, P., and Mayer, T. (2022). *The CEPII gravity database*. Vol. 5. CEPII Paris.

Costinot, A. and Rodríguez-Clare, A. (2014). "Trade Theory with Numbers: Quantifying the Consequences of Globalization". In: *Handbook of International Economics*. Vol. 4. Elsevier, pp. 197–261.

Eaton, J. and Kortum, S. (2002). "Technology, geography, and trade". *Econometrica* 70.5, pp. 1741–1779.

Fajgelbaum, P., Goldberg, P., Kennedy, P., Khandelwal, A., and Taglioni, D. (2024). "The US–China Trade War and Global Reallocations". *American Economic Review: Insights* 6.2, pp. 295–312.

Fally, T. (2015). "Structural Gravity and Fixed Effects". *Journal of International Economics* 97.1, pp. 76–85.

Feenstra, R. C. (1994). "New Product Varieties and the Measurement of International Prices". *The American Economic Review* 84.1, pp. 157–177.

— (2003). "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity". *Economics Letters* 78.1, pp. 79–86.

Fernández-Val, I. and Weidner, M. (2016). "Individual and Time Effects in Nonlinear Panel Models with Large N, T". *Journal of Econometrics* 192.1, pp. 291–312.

Fujiwara, I. and Matsuyama, K. (2026). "Competition and the Phillips Curve". *Journal of Monetary Economics* 159, p. 103920.

Geer, S. A. van de (2006). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.

Grossman, G. M., Helpman, E., and Lhuillier, H. (2023). "Supply Chain Resilience: Should Policy Promote International Diversification or Reshoring?" *Journal of Political Economy* 131.12, pp. 3462–3496.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A distribution-free theory of nonparametric regression*. Springer.

Head, K. and Mayer, T. (2014). "Gravity Equations: Workhorse, Toolkit, and Cookbook". In: *Handbook of International Economics*. Vol. 4. Elsevier, pp. 131–195.

— (2026). "Poor substitutes? Counterfactual methods in IO and trade compared". *Review of Economics and Statistics* 108.1, pp. 241–256.

Hummels, D. (2001). "Toward a Geography of Trade Costs". Unpublished manuscript.

Ichimura, H. (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". *Journal of econometrics* 58.1-2, pp. 71–120.

Jehle, G. A. and Reny, P. J. (2011). *Advanced Microeconomic Theory*. Google-Books-ID: 1c2tbwAACAAJ. Financial Times/Prentice Hall.

Klein, R. W. and Spady, R. H. (1993). "An efficient semiparametric estimator for binary response models". *Econometrica*, pp. 387–421.

Lind, N. and Ramondo, N. (2023). "Trade with Correlation". *American Economic Review* 113.2, pp. 317–353.

Matsuyama, K. (2023). "Non-CES aggregators: a guided tour". *Annual Review of Economics* 15.1, pp. 235–265.

Matsuyama, K. and Ushchev, P. (2017). "Beyond CES: three alternative classes of flexible homothetic demand systems". *Global Poverty Research Lab Working Paper* 17-109.

— (2020). "Constant Pass-Through". *CEPR Discussion Paper No. DP15475*.

— (2022a). "Destabilizing Effects of Market Size in the Dynamics of Innovation". *Journal of Economic Theory* 200, p. 105415.

— (2022b). "Selection and sorting of heterogeneous firms through competitive pressures". *CEPR Discussion Paper No. DP17092*.

Newey, W. K. (1997). "Convergence rates and asymptotic normality for series estimators". *Journal of Econometrics* 79.1, pp. 147–168.

Newey, W. K. and McFadden, D. (1994). "Large sample estimation and hypothesis testing". *Handbook of econometrics* 4, pp. 2111–2245.

Novy, D. (2013). "International Trade without CES: Estimating Translog Gravity". *Journal of International Economics* 89.2, pp. 271–282.

Ossa, R. (2015). "Why trade matters after all". *Journal of International Economics* 97.2, pp. 266–277.

Pfaffermayr, M. (2020). "Constrained Poisson Pseudo Maximum Likelihood Estimation of Structural Gravity Models". *International Economics* 161, pp. 188–198.

Piveteau, P. and Smagghue, G. (2024). *Gravity beyond CES*. Working Paper 976. Banque de France.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). "Semiparametric estimation of index coefficients". *Econometrica* 57.6, pp. 1403–1430.

Silva, J. S. and Tenreyro, S. (2006). "The log of gravity". *The Review of Economics and statistics* 88.4, pp. 641–658.

Timmer, M., Erumban, A. A., Gouma, R., Los, B., Temurshoev, U., Vries, G. J. de, Arto, I., Genty, V. A. A., Neuwahl, F., Francois, J., and others (2012). *The world input-output database (WIOD): contents, sources and methods*. Tech. rep. Institue for International and Development Economics.

Vaart, A. W. van der and Wellner, J. A. (1996). "Weak Convergence and Empirical Processes: With Applications to Statistics". In: *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, pp. 284–308.

Wang, J. and Ghosh, S. K. (2012). "Shape restricted nonparametric regression with Bernstein polynomials". *Computational Statistics & Data Analysis* 56.9, pp. 2729–2741.

Weidner, M. and Zylkin, T. (2021). "Bias and Consistency in Three-Way Gravity Models". *Journal of International Economics* 132, p. 103513.

Yotov, Y. V. (2022). "On the Role of Domestic Trade Flows for Estimating the Gravity Model of Trade". *Contemporary Economic Policy* 40.3, pp. 526–540.

— (2024). "The Evolution of Structural Gravity: The Workhorse Model of Trade". *Contemporary Economic Policy* 42.4, pp. 578–603.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A., and Larch, M. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. Geneva: World Trade Organization, United Nations Conference on Trade, and Development.

A Proof for Properties of General Equilibrium Mappings

A.1 The Feasible Parameter Space

Lemma 1. *Under Assumption 3, the feasible parameter space Θ is nonempty, is a convex cone, and satisfies* $\dim(\Theta) = d$.

First we show that Θ_1 is nonempty. Without loss of generality, say the first element $W_{ij}^{(1)}$ of W_{ij} is strictly positive for all i, j. Then, letting $e_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^d$ be the 1st basis vector, $W_{ij}' e_1 = W_{ij}^{(1)} > 0$ for all i, j. Hence $e_1 \in \Theta_1$. Similarly, we have $(W_{ij} - W_{i\ell} - W_{\ell j})' e_1 = W_{ij}^{(1)} - W_{i\ell}^{(1)} - W_{\ell j}^{(1)} < 0$ so $e_1 \in \Theta_2$. This combined yields $e_1 \in \Theta \neq \emptyset$.

Full dimensionality $(\dim(\Theta) = d)$ is implied from the fact that there exists an open ball with radii $r > 0$ around $e_1 \in \Theta$ such that the ball $B(e_1, r) \subset \Theta \subset \mathbb{R}^d$. This can be seen by (1) all of the constraints are linear, and (2) inner product is continuous.

The set Θ is clearly a convex cone. For any $\alpha > 0$, $\alpha\theta \in \Theta$ if $\theta \in \Theta$ as multiplication for any $\alpha > 0$ does not break the inequality constraints imposed upon the set Θ. Convexity follows from the fact that Θ is an intersection of two linear half spaces Θ_1 and Θ_2.

A.2 Proof of Proposition 1

The existence of the equilibrium follows from Theorem 5.5 in Jehle and Reny (2011). Uniqueness follows from gross substitutability of the excess demand function for any $\mathcal{I}$ and β, as verified below.

We begin by proving the second assertion in Proposition 1 concerning the uniqueness of the equilibrium, as the first assertion is a simple application of Theorem 5.5 in Jehle and Reny (2011).

The uniqueness of the equilibrium can be shown by establishing that the vector of excess demand function satisfies gross substitutability. Let us prepare some notation.

$$\mathbb{Z}(\mathbf{p}, \beta, \mathcal{I}) \equiv (\mathbb{Z}_1(\mathbf{p}, \beta, \mathcal{I}), \mathbb{Z}_2(\mathbf{p}, \beta, \mathcal{I}), \ldots, \mathbb{Z}_N(\mathbf{p}, \beta, \mathcal{I}))' \in \mathbb{R}^N,$$

where $\mathbb{Z}_i(\cdot)$ is the excess demand function defined in 9.

Let us define the matrix J by

$$J(\mathbf{p}, \beta, \mathcal{I}) \equiv \begin{bmatrix} \nabla_{\mathbf{p}'} \mathbb{Z}_1(\mathbf{p}, \beta, \mathcal{I}) \\ \nabla_{\mathbf{p}'} \mathbb{Z}_2(\mathbf{p}, \beta, \mathcal{I}) \\ \vdots \\ \nabla_{\mathbf{p}'} \mathbb{Z}_N(\mathbf{p}, \beta, \mathcal{I}) \end{bmatrix}.$$

We have to show that the off diagonal elements of J is non-negative for all equilibrium price vectors $\mathbf{p} = \mathbf{p}(\beta, \mathcal{I})$.

Take any $\mathcal{I} \in \mathbb{S}(\mathcal{I})$ and $\beta \in \mathcal{B}$. Then, Theorem 5.5 in Jehle and Reny (2011) ensures that $\mathbf{p}^* \equiv \mathbf{p}(\beta, \mathcal{I})$ exists. Without loss of generality, consider the $(2, 1)$ entry of the matrix $J(\mathbf{p}^*, \beta, \mathcal{I})$.

We have

$$\begin{aligned}\frac{\partial \mathbb{Z}_2}{\partial p_1}(\mathbf{p},\beta,\mathcal{I}) &= \sum_{\ell=1}^{N}\frac{\partial}{\partial p_1}\left[S_2\left(\frac{p_2\tau_{2\ell}}{A_\ell(\mathbf{p}^*,\beta,\mathcal{I})}\right)\right]\frac{p_\ell L_\ell}{p_2},\\ &= \frac{\partial}{\partial p_1}\left[S_2\left(\frac{p_2\tau_{2\ell}}{A_1(\mathbf{p}^*,\beta,\mathcal{I})}\right)\frac{p_1 L_1}{p_2}\right]\\ &\quad+\sum_{\ell>1}^{N}\frac{\partial}{\partial p_1}\left[S_2\left(\frac{p_2\tau_{2\ell}}{A_\ell(\mathbf{p}^*,\beta,\mathcal{I})}\right)\right]\frac{p_\ell L_\ell}{p_2}.\end{aligned} \tag{26}$$

Inspecting the first term, we have

$$\begin{aligned}\frac{\partial}{\partial p_1}\left[S_2\left(\frac{p_2\tau_{21}}{A_1(\mathbf{p}^*,\beta,\mathcal{I})}\right)\frac{p_1 L_1}{p_2}\right] &= \dot{S}_2\left(\frac{p_2^*\tau_{21}}{A_1(\mathbf{p}^*,\beta,\mathcal{I})}\right)\left[\frac{\partial}{\partial p_1}\left(\frac{1}{A_1(\mathbf{p}^*,\beta,\mathcal{I})}\right)\right]\frac{p_1^* L_1}{p_2^*}\\ &\quad+S_2\left(\frac{p_2^*\tau_{21}}{A_1(\mathbf{p}^*,\beta,\mathcal{I})}\right)\frac{L_1}{p_2^*},\end{aligned} \tag{27}$$

where we used the Newton's notation $(d/dz)S_2(z) \equiv \dot{S}_2(z)$ to denote the derivative.

Since the single price aggregator exists for all strictly positive prices $\mathbf{p}$ and $(\beta,\mathcal{I}) \in \mathcal{B}\times \mathbb{S}(\mathcal{I})$, by the Implicit Function Theorem applied to

$$G_1(\mathbf{p}, A_1 \mid \beta,\mathcal{I}) \coloneqq \sum_{\ell=1}^{N} S_\ell\left(\frac{p_\ell\tau_{\ell i}}{A_1}\right) - 1$$

tells us that

$$\nabla_{\mathbf{p}} A_1(\mathbf{p}\mid\beta,\mathcal{I}) = \frac{A_1(\mathbf{p}\mid\beta,\mathcal{I})}{\sum_{\ell=1}^{N}\dot{S}_\ell\left(\frac{p_\ell\tau_{\ell 1}}{A_1(\mathbf{p}|\beta,\mathcal{I})}\right)p_\ell\tau_{\ell 1}}\begin{pmatrix}\dot{S}_1(\frac{p_1\tau_{11}}{A_1})\tau_{11}\\ \dot{S}_2(\frac{p_2\tau_{21}}{A_1})\tau_{21}\\ \vdots\\ \dot{S}_N(\frac{p_N\tau_{N1}}{A_1})\tau_{N1}\end{pmatrix} > 0,$$

where $>$ is interpreted as a componentwise inequality. We have

$$\nabla_A G_1(\mathbf{p}, A\mid\beta,\mathcal{I}) = \frac{-1}{A^2}\sum_{\ell=1}^{N}\dot{S}_\ell\left(\frac{p_\ell\tau_{\ell 1}}{A}\right)p_\ell\tau_{\ell 1} > 0,$$

and each element of the gradient $\nabla_{\mathbf{p}} G_1$ is given by:

$$\frac{\partial}{\partial p_i}G_1(\mathbf{p}, A\mid\beta,\mathcal{I}) = \dot{S}_i\left(\frac{p_i\tau_{i1}}{A}\right)\frac{\tau_{i1}}{A}, \text{ for all } i.$$

Hence,

$$\begin{aligned}\frac{\partial}{\partial p_1}\frac{1}{A_1(\mathbf{p},\beta,\mathcal{I})} &= -\frac{1}{A_1^2(\mathbf{p},\beta,\mathcal{I})}\frac{\partial}{\partial p_1}A(\mathbf{p},\beta,\mathcal{I}) \\ &= -\frac{\dot{S}_1(\frac{p_1\tau_{11}}{A_1})\tau_{11}}{A_1(\mathbf{p},\beta,\mathcal{I})\sum_{\ell=1}^{N}\dot{S}_\ell\left(\frac{p_\ell\tau_{\ell 1}}{A_1(\mathbf{p}|\beta,\mathcal{I})}\right)p_\ell\tau_{\ell 1}} < 0. \end{aligned} \tag{28}$$

Thus, the first term in (27) is strictly positive because $z \mapsto S_i(z)$ is strictly decreasing.

Inspecting the second term in (27), since p_1 only appears in the single price aggregator $A_\ell(\cdot)$, we have

$$\sum_{\ell>1}^{N}\frac{\partial}{\partial p_1}\left[S_2\left(\frac{p_2\tau_{2\ell}}{A_\ell(\mathbf{p}^*,\beta,\mathcal{I})}\right)\right]\frac{p_\ell L_\ell}{p_2} = \sum_{\ell>1}^{N}\dot{S}_2\left(\frac{p_2\tau_{2\ell}}{A_\ell(\mathbf{p}^*,\beta,\mathcal{I})}\right)\left[\frac{\partial}{\partial p_1}\frac{1}{A_\ell(\mathbf{p}^*,\beta,\mathcal{I})}\right]p_\ell L_\ell\tau_{2\ell}$$

which is strictly positive by (28) being negative and S_ℓ being strictly decreasing.

Hence, the off-diagonal elements of J is strictly positive for all $(\beta,\mathcal{I})$, which proves the assertion for all equilibrium prices $\mathbf{p}^* = \mathbf{p}(\beta,\mathcal{I})$.

A.3 Continuity of Equilibrium Prices and Aggregators

Lemma 2. *For the fixed $\mathcal{W}$, let $\mathcal{I} \equiv (\mathcal{W},\ell)$ with $\ell \in \mathbb{S}(\mathbf{L})$. The equilibrium-price mapping $\mathcal{I} \mapsto \mathbf{p}(\beta,\mathcal{I})$ is continuous for every $\beta \in \mathcal{B}$. The mapping*

$$\mathcal{I} \mapsto A_j(\mathbf{p}(\beta,\mathcal{I}),\beta \mid \mathcal{W})$$

is also continuous for every $\beta \in \mathcal{B}$ and $j = 1,\ldots,N$. Finally, there exist constants $0 < \underline{A}_i < \overline{A}_i < \infty$ such that

$$A_i(\mathbf{p}(\beta_0,\mathcal{I}),\beta \mid \mathcal{W}) \in [\underline{A}_i,\overline{A}_i]$$

uniformly over $(\beta,\mathcal{I}) \in \mathcal{B}^C \times \mathbb{S}(\mathcal{I})$.

First we prove assertion (1).

Take any sequence $\mathcal{I}_{(r)} \in \mathbb{S}(\mathcal{I})$ such that $\mathcal{I}_{(r)} \to \mathcal{I}^* \in \mathbb{S}(\mathcal{I})$. Fix any $\beta \in \mathcal{B}$ and let $\mathbf{p}_{(r)} = \mathbf{p}(\beta,\mathcal{I}_{(r)})$. Proposition 1 implies that every $\mathbf{p}_{(r)}$ exists and belongs to the interior of the normalized price simplex $\Delta_{\mathbf{P}}$. Sequential compactness of $\Delta_{\mathbf{P}}$ implies that every subsequence of $\{\mathbf{p}_{(r)}\}$ has a further subsequence, denoted without relabeling by $\{\mathbf{p}_{(r_k)}\}$, such that $\mathbf{p}_{(r_k)} \to \bar{\mathbf{p}} \in \Delta_{\mathbf{P}}$.

We first show that $\bar{\mathbf{p}}$ cannot lie on the boundary of $\Delta_{\mathbf{P}}$. We adapt the boundary argument in the proof of Theorem 5.4 of Jehle and Reny (2011). Suppose, toward a contradiction, that $\bar{p}_j = 0$ for at least one j. The aggregate endowments remain uniformly bounded away from zero because $\mathbb{S}(\mathbf{L})$ is a compact subset of $\mathbb{R}^N_{++}$. Moreover, $\mathcal{I}_{(r_k)} \to \mathcal{I}^*$ while the design points are fixed. The corresponding iceberg trade costs therefore remain bounded above and away from zero. The preferences indexed by the fixed β are continuous and strongly increasing. Consequently, the same revealed-preference argument used in Theorem 5.4 implies that, along

a further subsequence, excess demand for some good j' with $\bar{p}_{j'} = 0$ is unbounded above. This is impossible because every $\mathbf{p}_{(r_k)}$ is an equilibrium and hence

$$\mathbb{Z}_{j'}(\mathbf{p}_{(r_k)}, \beta, \mathcal{I}_{(r_k)}) = 0$$

for every k. Therefore $\bar{\mathbf{p}} \gg \mathbf{0}$.

The excess demand function is continuous in $(\mathbf{p}, \mathcal{I})$ on $\mathbb{R}^N_{++} \times \mathbb{S}(\mathcal{I})$. We may therefore pass to the limit in the market-clearing equations to obtain

$$\mathbb{Z}(\bar{\mathbf{p}}, \beta, \mathcal{I}^*) = \lim_{k\to\infty} \mathbb{Z}(\mathbf{p}_{(r_k)}, \beta, \mathcal{I}_{(r_k)}) = \mathbf{0}.$$

Thus $\bar{\mathbf{p}}$ is a normalized equilibrium price vector at $\mathcal{I}^*$. Uniqueness from Proposition 1 gives $\bar{\mathbf{p}} = \mathbf{p}(\beta, \mathcal{I}^*)$. Since every subsequence has a further subsequence converging to this same limit, the original sequence satisfies $\mathbf{p}(\beta, \mathcal{I}_{(r)}) \to \mathbf{p}(\beta, \mathcal{I}^*)$. This proves assertion (1).

For the second assertion, take any sequence $\mathcal{I}_{(r)} \in \mathbb{S}(\mathcal{I})$ such that $\mathcal{I}_{(r)} \to \mathcal{I}^* \in \mathbb{S}(\mathcal{I})$, and any $\beta \in \mathcal{B}$. By the first assertion, we have a sequence of equilibrium prices $\mathbf{p}_{(r)} \equiv \mathbf{p}(\beta, \mathcal{I}_{(r)})$ converging to $\mathbf{p}^* = \mathbf{p}(\beta, \mathcal{I}^*)$. Now let us define

$$G_j(A, \mathbf{p}(\beta, \mathcal{I}), \mathcal{I} \mid \mathbf{S}) \equiv \sum_{\ell=1}^{N} S_\ell \left(\frac{p_\ell(\beta, \mathcal{I}) \tau_{\ell j}(\theta)}{A} \right) - 1,$$

for all $j = 1, \ldots, N$. We show that the sequence of the solution $A_{(r)}$ obtained as $G_j(A_{(r)}, \mathbf{p}_{(r)}, \mathcal{I}_{(r)} \mid \mathbf{S}) = 0$ converges to A^* defined by $G_j(\mathbf{A}^*, \mathbf{p}^*, \mathcal{I}^* \mid \mathbf{S}) = 0$ to establish the continuity of the implicit function $\mathcal{I} \mapsto A_j(\mathbf{p}(\beta, \mathcal{I}), \mathcal{I}, \beta)$.

The endpoint conditions in Assumption 2 and the Intermediate Value Theorem imply that the root exists, while strict monotonicity implies that it is unique. For the fixed β, assertion (1) and compactness of $\mathbb{S}(\mathcal{I})$ imply that the equilibrium prices range over a compact subset of the interior of $\Delta_{\mathbf{P}}$. The same construction used in assertion (3) below therefore supplies positive lower and upper bounds for $A_{(r)}$. Every subsequence of $\{A_{(r)}\}$ consequently has a convergent further subsequence. If $A_{(r_k)} \to \bar{A}$, continuity of G_j gives

$$G_j(\bar{A}, \mathbf{p}^*, \mathcal{I}^* \mid \mathbf{S}) = 0.$$

Uniqueness of the root implies $\bar{A} = A_j(\mathbf{p}^*, \beta \mid \mathcal{W}^*)$. Hence the entire sequence converges to this value, which proves assertion (2).

Finally, we prove the third assertion. By Proposition 1, the equilibrium price vector $\mathbf{p}_0 \equiv \mathbf{p}(\beta_0, \mathcal{I})$ is an element of the unit simplex, and is strictly positive. Continuity of $\mathcal{I} \mapsto \mathbf{p}(\beta_0, \mathcal{I})$ and compactness of $\mathbb{S}(\mathcal{I})$ imply that there exist constants $0 < \underline{p} \leq \bar{p} < \infty$ such that $p_{0\ell} \in [\underline{p}, \bar{p}]$ uniformly over ℓ and $\mathcal{I}$. Compactness of Θ^C and the fixed finite collection of design points similarly give constants $0 < \underline{\tau} \leq \bar{\tau} < \infty$ such that $\tau_{\ell i}(\theta) \in [\underline{\tau}, \bar{\tau}]$ uniformly over $(\ell, i, \theta, \mathcal{W})$.

For each ℓ, let $\underline{z}_\ell$ and $\bar{z}_\ell$ be the constants in Assumption 6, and define $\underline{z} = \min_\ell \underline{z}_\ell$ and

$\bar{z} = \max_\ell \bar{z}_\ell$. Strict monotonicity implies, uniformly over $S_\ell \in \mathcal{S}_\ell$,

$$S_\ell(\underline{z}) > \frac{1}{N} \quad \text{and} \quad S_\ell(\bar{z}) < \frac{1}{N}.$$

Choose $\bar{A}_i$ sufficiently large that

$$\frac{p_{0\ell}\tau_{\ell i}(\theta)}{\bar{A}_i} \leq \underline{z}$$

for every ℓ, uniformly over $(\beta, \mathcal{I})$. Then $G_i(\bar{A}_i, \mathbf{p}_0, \beta, \mathcal{I}) > 0$. Likewise, choose $\underline{A}_i > 0$ sufficiently small that

$$\frac{p_{0\ell}\tau_{\ell i}(\theta)}{\underline{A}_i} \geq \bar{z}$$

for every ℓ, uniformly over $(\beta, \mathcal{I})$. Then $G_i(\underline{A}_i, \mathbf{p}_0, \beta, \mathcal{I}) < 0$. Since G_i is continuous and strictly increasing in A, its unique root satisfies

$$0 < \underline{A}_i < A_i(\mathbf{p}_0, \beta \mid \mathcal{W}) < \bar{A}_i < \infty$$

uniformly over $(\beta, \mathcal{I}) \in \mathcal{B}^C \times \mathbb{S}(\mathcal{I})$. The argument applies to every $i = 1, \ldots, N$, which proves assertion (3).

A.4 Proof of Proposition 2

By assertion (1) of Lemma 2, the map $\mathcal{I} \mapsto \mathbf{p}_0(\mathcal{I})$ is continuous. Its domain $\mathbb{S}(\mathcal{I})$ is compact and its image lies in the interior of the normalized price simplex. Hence there exist constants $0 < \underline{p} \leq \bar{p} < \infty$ such that

$$\underline{p} \leq p_{0i}(\mathcal{I}) \leq \bar{p}$$

uniformly over i and $\mathcal{I} \in \mathbb{S}(\mathcal{I})$. Compactness of Θ^C and the fixed design $\mathcal{W}$, together with $\tau_{ij}(\theta) = \exp(W_{ij}'\theta)$, similarly implies that there exist $0 < \underline{\tau} \leq \bar{\tau} < \infty$ satisfying

$$\underline{\tau} \leq \tau_{ij}(\theta) \leq \bar{\tau}$$

uniformly over $(i, j, \theta, \mathcal{W})$. Finally, assertion (3) of Lemma 2 gives, for each j, constants $0 < \underline{A}_j \leq \bar{A}_j < \infty$ such that

$$\underline{A}_j \leq A_j(\mathbf{p}_0(\mathcal{I}), \beta \mid \mathcal{W}) \leq \bar{A}_j$$

uniformly over $(\beta, \mathcal{I}) \in \mathcal{B}^C \times \mathbb{S}(\mathcal{I})$. Therefore,

$$0 < \frac{\underline{p}\,\underline{\tau}}{\bar{A}_j} \leq \frac{p_{0i}(\mathcal{I})\tau_{ij}(\theta)}{A_j(\mathbf{p}_0(\mathcal{I}), \beta \mid \mathcal{W})} \leq \frac{\bar{p}\,\bar{\tau}}{\underline{A}_j} < \infty.$$

The result follows by taking

$$K_j = \left[\frac{\underline{p}\,\underline{\tau}}{\bar{A}_j}, \frac{\bar{p}\,\bar{\tau}}{\underline{A}_j}\right] \subset \mathbb{R}_{++}.$$

B Identification with Known Share Functions

Suppose that the share functions belong to a known finite-dimensional family $\{S_i(\cdot;\gamma) : \gamma \in \Gamma \subset \mathbb{R}^{d_\gamma}\}$ and that the true demand parameter γ_0 is known to the researcher. Consequently, the vector of true share functions

$$\mathbf{S}_0(\cdot;\gamma_0) = (S_1(\cdot;\gamma_0), \ldots, S_N(\cdot;\gamma_0))$$

is known. The only unknown structural parameter in this section is the trade-cost parameter θ_0. We write the population objective in Equation (13) as $Q(\theta \mid \mathbf{S}_0)$ to emphasize that the demand parameter is held fixed at its known value γ_0.

Proposition 2 and the bounded second moments of $\varepsilon_{(r)}$ imply that $Q(\theta \mid \mathbf{S}_0)$ is bounded. Lemma 2 and Assumption 5 then imply

$$\sup_{\theta \in \Theta^C} Q(\theta \mid \mathbf{S}_0) < \infty.$$

We impose the following rank condition on the design points.

Assumption 19. *There exists a destination j^* such that*

$$\operatorname{span}\left\{W_{ij^*} - W_{\ell j^*} \,\middle|\, i, \ell \neq j^*,\ i \neq \ell\right\} = \mathbb{R}^d.$$

Assumption 19 requires sufficient variation in the bilateral trade-cost variables across exporters for at least one destination. In particular, it rules out a constant component that is common to every exporter within a destination. The assumption is sufficient to establish identification, as shown in the next proposition.

Proposition 11. *Suppose that γ_0 is known and Assumption 19 holds. Then the population objective $Q(\theta \mid \mathbf{S}_0)$ is uniquely minimized at $\theta = \theta_0$.*

Proposition 11 applies directly when γ_0 is known from outside the model. If γ_0 is calibrated and treated as known, the identification result is conditional on that calibration. If γ_0 is estimated from an external sample, the same argument gives identification conditional on its value. Its estimation uncertainty must then be incorporated when conducting inference on θ_0.

As a concrete example, suppose that preferences are CES, so that $S_i(z;\gamma) = z^{1-\gamma}$, where the scalar parameter $\gamma > 1$ is the elasticity of substitution. The researcher may calibrate γ to a value $\bar{\gamma}$ taken from an external empirical study and estimate θ_0 while holding $\gamma = \bar{\gamma}$ fixed. The known share functions in Proposition 11 are then

$$S_{0i}(z) = z^{1-\bar{\gamma}}, \qquad i = 1, \ldots, N.$$

The proposition identifies θ_0 conditional on the chosen calibration. It does not establish that $\bar{\gamma}$ equals the true elasticity. Uncertainty or misspecification in the calibration is not reflected in inference that treats $\bar{\gamma}$ as fixed.

B.1 Proof of Lemma 3

Lemma 3. *Let $\theta, \tilde{\theta} \in \Theta^C$ satisfy $\theta \neq \tilde{\theta}$. For every $\mathcal{I} = (\mathcal{W}, \ell) \in \mathbb{S}(\mathcal{I})$, there exists an exporter-destination pair (i, j) such that*

$$S_{0i}\left(\frac{p_{0i}\exp(W_{ij}'\theta)}{A_j(\mathbf{p}_0, \theta, \mathbf{S}_0 \mid \mathcal{W})}\right) \neq S_{0i}\left(\frac{p_{0i}\exp(W_{ij}'\tilde{\theta})}{A_j(\mathbf{p}_0, \tilde{\theta}, \mathbf{S}_0 \mid \mathcal{W})}\right),$$

where $\mathbf{p}_0 = \mathbf{p}(\beta_0, \mathcal{I})$.

Fix $\mathcal{I} \in \mathbb{S}(\mathcal{I})$ and suppose, toward a contradiction, that the two predicted shares are equal for every (i, j). Strict monotonicity of S_{0i} implies

$$\frac{p_{0i}\exp(W_{ij}'\theta)}{A_j(\mathbf{p}_0, \theta, \mathbf{S}_0 \mid \mathcal{W})} = \frac{p_{0i}\exp(W_{ij}'\tilde{\theta})}{A_j(\mathbf{p}_0, \tilde{\theta}, \mathbf{S}_0 \mid \mathcal{W})}$$

for every (i, j). Because $p_{0i} > 0$, this gives

$$\exp\left(W_{ij}'(\theta - \tilde{\theta})\right) = \frac{A_j(\mathbf{p}_0, \theta, \mathbf{S}_0 \mid \mathcal{W})}{A_j(\mathbf{p}_0, \tilde{\theta}, \mathbf{S}_0 \mid \mathcal{W})}.$$

The ratio on the right is common across exporters within destination j. Therefore, for the destination j^* in Assumption 19,

$$\left(W_{ij^*} - W_{\ell j^*}\right)'(\theta - \tilde{\theta}) = 0$$

for every $i, \ell \neq j^*$ with $i \neq \ell$. The rank condition then implies $\theta - \tilde{\theta} = 0$, contradicting $\theta \neq \tilde{\theta}$.

B.2 Proof of Proposition 11

The zero-mean and independence conditions in Assumption 4 give

$$Q(\theta \mid \mathbf{S}_0) - Q(\theta_0 \mid \mathbf{S}_0) \quad = \quad \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left\{g_{ij,(r)}(\theta, \mathbf{S}_0, \mathbf{P}_{0,(r)}) - g_{ij,(r)}(\theta_0, \mathbf{S}_0, \mathbf{P}_{0,(r)})\right\}^2\right].$$

The right-hand side is nonnegative. If $\theta \neq \theta_0$, Lemma 3 implies that its integrand is strictly positive for every $\mathcal{I} \in \mathbb{S}(\mathcal{I})$. Hence the expectation is strictly positive, proving unique minimization at θ_0.

C Proof of CES Estimator

C.1 Proof of Lemma 4 (Separation along the Curve)

Lemma 4. *Suppose Assumption 8 holds. For any $\sigma \neq \sigma_0$ such that $\theta_\sigma \in \Theta^C$, where $\theta_\sigma = (1 - \sigma_0)\theta_0/(1 - \sigma)$, the vectors $\{g^{CES}_{ij,(r)}(\theta_\sigma, \sigma)\}_{i,j}$ and $\{g^{CES}_{ij,(r)}(\theta_0, \sigma_0)\}_{i,j}$ differ almost surely.*

Let $a = 1 - \sigma$ and $a_0 = 1 - \sigma_0$. Along the curve in the lemma, $a\theta_\sigma = a_0\theta_0$. Suppose, toward a contradiction, that the two vectors of predicted shares are identical. Using the domestic share

as the denominator and the normalization $W_{jj} = 0$, equation (14) gives, for every i, j,

$$\log \frac{g^{CES}_{ij,(r)}(\theta_\sigma, \sigma)}{g^{CES}_{jj,(r)}(\theta_\sigma, \sigma)} = a \log \frac{P_{0i,(r)}}{P_{0j,(r)}} + W'_{ij} a\theta_\sigma$$

and the analogous expression at the truth. Equality of the predicted shares and $a\theta_\sigma = a_0\theta_0$ therefore imply

$$(a - a_0) \log \frac{P_{0i,(r)}}{P_{0j,(r)}} = 0 \qquad \text{for every } i, j.$$

Because $\sigma \neq \sigma_0$, we have $a \neq a_0$. Assumption 8 supplies, almost surely, a pair with unequal prices, contradicting the last display. The restriction $\theta_\sigma \in \Theta^C$ ensures that the alternative considered by the lemma is feasible; no closure of Θ^C under positive rescaling is required.

C.2 Proof of Lemma 5 (Identification)

Lemma 5. *Suppose Assumption 9 holds. For any $(\theta, \sigma) \in \mathcal{B}^{CES}$ satisfying $(1-\sigma)\theta \neq (1-\sigma_0)\theta_0$, the vectors $\{g^{CES}_{ij,(r)}(\theta, \sigma)\}_{i,j}$ and $\{g^{CES}_{ij,(r)}(\theta_0, \sigma_0)\}_{i,j}$ cannot be identical.*

Put $a = 1 - \sigma$, $a_0 = 1 - \sigma_0$, $\phi = a\theta$, and $\phi_0 = a_0\theta_0$. Suppose, toward a contradiction, that the two vectors of predicted shares are identical. Log share odds relative to the domestic supplier give

$$(a - a_0) \log \frac{P_{0i,(r)}}{P_{0j,(r)}} + W'_{ij}(\phi - \phi_0) = 0 \qquad \text{for every } i, j.$$

For three distinct indices i, j, ℓ, subtracting the equations for (i, j) and (j, ℓ) from the equation for (i, ℓ) eliminates relative prices because

$$\log(P_{0i,(r)}/P_{0\ell,(r)}) = \log(P_{0i,(r)}/P_{0j,(r)}) + \log(P_{0j,(r)}/P_{0\ell,(r)}).$$

It follows that

$$(W_{i\ell} - W_{ij} - W_{j\ell})'(\phi - \phi_0) = 0$$

for every such triplet. Assumption 9 therefore implies $\phi = \phi_0$, contradicting the hypothesis $(1-\sigma)\theta \neq (1-\sigma_0)\theta_0$.

C.3 Proof of Proposition 3

Every alternative in $\mathcal{B}^{CES}$ either lies on the curve $(1-\sigma)\theta = (1-\sigma_0)\theta_0$ or lies outside it. Lemma 4 separates every feasible alternative on the curve, and Lemma 5 separates every alternative outside it. Under Assumption 4, the usual population least-squares decomposition then implies that every alternative has a strictly larger criterion value than (θ_0, σ_0).

D Proofs of HSA Estimator

We will introduce a shorthand which we will use throughout Appendix C.

Let us define a shorthand $z_{ij}(\mathbf{p},\beta) = p_i\tau_{ij}(\theta)/A_j(\mathbf{p},\beta \mid \mathcal{W})$, and further put $z_{ij,(r)}(\beta) = z_{ij}(\mathbf{P}_{0,(r)},\beta)$. The former is a functional $(\mathbf{p},\beta) \mapsto z_{ij}(\mathbf{p},\beta)$, while the latter $z_{ij,(r)}(\beta)$ is a random object as it is the functional $z_{ij}(\cdot,\beta)$ evaluated at a random object $\mathbf{P}_{0,(r)}$.

D.1 Identification

We first impose an appropriate normalization on the Sieve coefficient. Observe that for $\beta = (\theta, f_{1,\kappa},\ldots,f_{N,\kappa}) \in \mathcal{B}_\kappa$, and some constant $c > 0$, defining $\tilde{\alpha_0^{(i)}} = \alpha_0^{(i)}$, and $\tilde{\alpha_q^{(i)}} = c^q\alpha_q^{(i)}$ for all $(i,q) \in \{1,\ldots,N\}\times\{1,\ldots,\kappa\}$, we have

$$\begin{aligned}
\tilde{f_{i,\kappa}}(z) &= \sum_{q=0}^{\kappa} \tilde{\alpha_q^{(i)}}\,(z)^{-q}, \\
&= \alpha_0^{(i)} + \sum_{q=1}^{\kappa} c^q\alpha_q^{(i)}\,(z)^{-q}, \\
&= f_{i,\kappa}\left(\frac{z}{c}\right).
\end{aligned}$$

Letting $\tilde{\beta} = (\theta, \tilde{f_{1,\kappa}},\ldots,\tilde{f_{N,\kappa}})$, if the single aggregator $A_j(\mathbf{p},\beta)$ solves the HSA normalization equation, then uniqueness of the aggregator implies that we have $A_j(\mathbf{p},\tilde{\beta}) = A_j(\mathbf{p},\beta)/c$.

Indeed, we have

$$z_{ij}(\mathbf{p},\tilde{\beta}) = \frac{p_i\tau_{ij}(\theta)}{A_j(\mathbf{p},\beta)/c} = cz_{ij}(\mathbf{p},\beta),$$

and therefore

$$\tilde{f_{i,\kappa}}\left(z_{ij}(\mathbf{p},\tilde{\beta})\right) = f_{i,\kappa}\left(\frac{cz_{ij}(\mathbf{p},\beta)}{c}\right) = f_{i,\kappa}\left(z_{ij}(\mathbf{p},\beta)\right),$$

which implies that two parameters $\beta,\tilde{\beta}$ generates the same shares.

We therefore work with the symmetric normalized parameter spaces $\mathcal{B}^*$ and $\mathcal{B}_\kappa^*$ defined in Subsection 3.2. In particular, both the candidate and true parameters satisfy $A_N(\bar{\mathbf{p}},\beta) = A_N(\bar{\mathbf{p}},\beta_0) = 1$. We shall work the coefficient space Ξ_κ^* defined as

$$\Xi_\kappa^* \equiv \{\xi \in \Xi_\kappa \mid \beta(\xi) \in \mathcal{B}_\kappa^*\}.$$

We establish Condition 3.1(ii) in Chen (2007) to show identification. The population objective function is

$$\mathcal{B}^* \ni \beta \mapsto Q(\beta) \equiv \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left|\lambda_{ij,(r)} - f_i\left(z_{ij,(r)}(\beta)\right)\right|^2\right].$$

First note that we have, for all $\beta \in \mathcal{B}^*$,

$$\begin{aligned}
Q(\beta) &= Q(\beta_0) + \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left|f_i\left(z_{ij,(r)}(\beta)\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right|^2\right], \\
&= Q(\beta_0) + M^2(\beta,\beta_0),
\end{aligned}$$

where in line 1, we used the definition of $z_{ij,(r)}(\beta) \equiv P_{0i,(r)}\tau_{ij}(\theta)/A_j(\mathbf{P}_{0,(r)},\beta \mid \mathcal{W})$, and used the

fact that (1) $\mathbb{E}\left(\varepsilon_{ij,(r)} \mid \mathbf{P}_{0,(r)}\right) = 0$, and (2) $0 < f_i(z_{ij,(r)}(\beta)) < 1$ almost surely uniformly over $\beta \in \mathcal{B}$. Line 2 follows from the definition of our metric $M(\cdot,\cdot)$. Over the constrained space $\beta \in \mathcal{B}^*_\kappa$, we have $M^2(\beta,\beta_0) = 0$ implying that $\beta = \beta_0$. We show that for every $\beta \in \mathcal{B}^*_\kappa$ satisfying $\beta \neq \beta_0$, we have $M^2(\beta,\beta_0) > 0$.

To establish separation, let us first define $\Delta(\delta) \equiv \{\beta \in \mathcal{B}^* \mid d(\beta,\beta_0) \geq \delta\}$. We show that

$$\inf_{\beta\in\Delta(\delta)} M^2(\beta,\beta_0) \geq \eta(\delta) > 0.$$

The separation follows if (1) M^2 is continuous, (2) $\mathcal{B}^*_\kappa$ is compact, and (3) $M^2(\beta,\beta_0) = 0$ only at the truth β_0. Compactness and continuity is satisfied.

D.1.1 Step 1 (Convert Probabilistic Statement to a Deterministic Statement)

Suppose that we have, for some $\beta \in \mathcal{B}^*$ satisfying $\beta \neq \beta_0$, we have $M^2(\beta,\beta_0) = 0$. Let us denote $\beta = (\theta, f_1, \ldots, f_N)$, and $\beta_0 = (\theta_0, S_{0i}, \ldots, S_{0N})$. Since the integrand is a sum of squared components, it must be the case that we have $f_i\left(z_{ij,(r)}(\beta)\right) = S_{0i}\left(z_{ij,(r)}(\beta_0)\right)$ for all i, j, μ-almost surely.

Now define a deterministic map

$$g_{ij}(\mathbf{p}) \equiv f_i\left(z_{ij}(\mathbf{p},\beta)\right) - S_{0i}(z_{ij}(\mathbf{p},\beta_0)).$$

By (1) continuity of the share functions $\{f_i, S_{0i}\}_{i=1}^N$ with respect to it's single argument z, (2) the continuity of the single aggregator $\{A_j(\mathbf{p},\beta)\}_{j=1}^N$ with respect to $\mathbf{p}$, (3) continuity of $\mathbf{p} \mapsto z_{ij}(\mathbf{p},\beta)$ at each $\beta \in \mathcal{B}^*$ implies continuity of $\mathbf{p} \mapsto g_{ij}(\mathbf{p})$. By Assumption 10(1), we have $g_{ij}(\mathbf{p}) = 0$ for every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. This see this, suppose toward a contradiction that we have $g_{ij}(\tilde{\mathbf{p}}) \neq 0$ for some $\tilde{\mathbf{p}} \in \mathbb{S}(\mathbf{P}_0)$. Then, there exists an open set (relative to $\mathbb{S}(\mathbf{P}_0)$) neighborhood of $\tilde{\mathbf{p}}$, say U, and some $\delta > 0$ such that we have $|g_{ij}(\mathbf{p})| \geq \delta$ for every $\mathbf{p} \in U$. The support is convex and has a nonempty relative interior, every relatively open subset of $\mathbb{S}(\mathbf{P}_0)$ must have positive $(N-1)$ Lebesgue measure. Since μ admits a density $f_{\mathbf{P}_0}$ that is strictly positive on $\mathbb{S}(\mathbf{P}_0)$, we have $\int_U f_{\mathbf{P}}(p)\mathrm{d}p > 0$. It thus follows that $\Pr(|g_{ij}(\mathbf{P}_{0,(r)})| \geq \delta) \geq \mu(U) > 0$. This is a contradiction to $g_{ij}(\mathbf{P}_{0,(r)}) = 0$, μ-almost surely. Thus, $g_{ij}(\mathbf{p}) = 0$ for every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$.

D.1.2 Step 2 (Exploit Strict Monotonicity)

Let us fix the exporter i without loss of generality. Take any $\mathbf{p}, \tilde{\mathbf{p}} \in \mathbb{S}(\mathbf{P}_0)$. Then, by strict monotonicity of f_i, S_{0i}, it follows that we have

$$\begin{aligned} z_{ij}(\mathbf{p},\beta) < z_{i\ell}(\tilde{\mathbf{p}},\beta) &\iff f_i\left(z_{ij}(\mathbf{p},\beta)\right) > f_i\left(z_{i\ell}(\tilde{\mathbf{p}},\beta)\right), \\ &\iff S_{0i}\left(z_{ij}(\mathbf{p},\beta_0)\right) > S_{0i}\left(z_{i\ell}(\tilde{\mathbf{p}},\beta_0)\right), \\ &\iff z_{ij}(\mathbf{p},\beta_0) < z_{i\ell}(\tilde{\mathbf{p}},\beta_0), \end{aligned}$$

where we used $g_{ij}(\mathbf{p}) = 0$ for all indices i, j and $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$ in line 2 and strict monotonicity in line 3.

Strict monotonicity of S_{0i} admits an inverse. Thus, let us put the composite map as $h_i = S_{0i}^{-1} \circ f_i$. From $g_{ij}(\mathbf{p}) = 0$, we have $z_{ij}(\mathbf{p}, \beta_0) = h_i\left(z_{ij}(\mathbf{p}, \beta)\right)$.

D.1.3 Step 3 (Symmetric Case)

Under the symmetry restriction in Assumption 10, fix a destination j and two exporters $i \neq \ell$. Then, we have

$$\begin{aligned} z_{ij}(\mathbf{p}, \beta) < z_{\ell j}(\mathbf{p}, \beta) &\iff f\left(z_{ij}(\mathbf{p}, \beta)\right) > f\left(z_{\ell j}(\mathbf{p}, \beta)\right), \\ &\iff S_0\left(z_{ij}(\mathbf{p}, \beta_0)\right) > S_0\left(z_{\ell j}(\mathbf{p}, \beta_0)\right), \\ &\iff z_{ij}(\mathbf{p}, \beta_0) < z_{\ell j}(\mathbf{p}, \beta_0), \end{aligned}$$

where the first line uses the fact that f is strictly decreasing, the second line uses $g_{ij}(\mathbf{p}) = g_{\ell j}(\mathbf{p}) = 0$ as this holds for every exporters i, ℓ, every destination j, and every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$, and the third line uses the fact that S_0 is strictly decreasing.

Within destination j, the single aggregator is common across exporters, and by the functional-form assumption on the trade costs $\tau_{ij}(\theta) = \exp(W_{ij}'\theta)$, we have

$$\begin{aligned} z_{ij}(\mathbf{p}, \beta) < z_{\ell j}(\mathbf{p}, \beta) &\iff \frac{p_i \tau_{ij}(\theta)}{A_j(\mathbf{p}, \beta \mid \mathcal{W})} < \frac{p_\ell \tau_{\ell j}(\theta)}{A_j(\mathbf{p}, \beta \mid \mathcal{W})}, \\ &\iff p_i \tau_{ij}(\theta) < p_\ell \tau_{\ell j}(\theta) \\ &\iff \ln(p_i / p_\ell) < (W_{\ell j} - W_{ij})'\theta. \end{aligned}$$

Similarly, we obtain $z_{ij}(\mathbf{p}, \beta_0) < z_{\ell j}(\mathbf{p}, \beta_0) \iff \ln(p_i/p_\ell) < (W_{\ell j} - W_{ij})'\theta_0$. Thus from the equivalence

$$z_{ij}(\mathbf{p}, \beta) < z_{\ell j}(\mathbf{p}, \beta) \iff z_{ij}(\mathbf{p}, \beta_0) < z_{\ell j}(\mathbf{p}, \beta_0),$$

we have

$$\ln(p_i/p_\ell) < (W_{\ell j} - W_{ij})'\theta \iff \ln(p_i/p_\ell) < (W_{\ell j} - W_{ij})'\theta_0.$$

Consequently, we have

$$I\left\{\ln(p_i/p_\ell) < (W_{\ell j} - W_{ij})'\theta\right\} = I\left\{\ln(p_i/p_\ell) < (W_{\ell j} - W_{ij})'\theta_0\right\},$$

where $I\{A\}$ is an indicator function which equals one when the event A is true.

Let $R_{i\ell} = \{\ln(p_i/p_\ell) : \mathbf{p} \in \mathbb{S}(\mathbf{P}_0)\}$ be the collection of the log relative prices between i and ℓ. The support of the price vector is convex which implies connectedness of $\mathbb{S}(\mathbf{P}_0)$. And continuity of $\mathbf{p} \mapsto \ln(p_i/p_\ell)$ implies that the image $R_{i\ell}$ is also connected, which gives us an interval. Assumption 10(2) places every candidate threshold in the interior of this interval. Let $a = (W_{\ell j} - W_{ij})'\theta$ and $a_0 = (W_{\ell j} - W_{ij})'\theta_0$. Suppose that $a \neq a_0$. Without loss of generality, let $a_0 <$

a. Since $R_{i\ell}$ is an interval that contains both thresholds a and a_0, it contains every $x \in (a_0, a)$. Therefore, there exists a price vector $\tilde{\mathbf{p}}$ in the support $\mathbb{S}(\mathbf{P}_0)$ satisfying $a_0 < \ln(\tilde{p}_i/\tilde{p}_\ell) < a$. At this price vector, we have $I\{\ln(\tilde{p}_i/\tilde{p}_\ell) < a\} = 1$, and $I\{\ln(\tilde{p}_i/\tilde{p}_\ell) < a_0\} = 0$. This contradicts the equality of the indicators. Therefore, it must be the case that $a = a_0$. This implies that we have:

$$\left(W_{\ell j} - W_{ij}\right)' \theta = \left(W_{\ell j} - W_{ij}\right)' \theta_0.$$

The span condition $\text{span}\{W_{\ell j} - W_{ij} : i \neq \ell, j = 1, \ldots, N\} = \mathbb{R}^d$ implies that we have $\theta = \theta_0$.

D.1.4 Step 4

From Step 1, we have $g_{ij}(\mathbf{p}) = 0$ for every i, j and every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. Therefore, we have $f(z_{ij}(\mathbf{p}, \beta)) = S_0(z_{ij}(\mathbf{p}, \beta_0))$. Using the strict monotonicity of S_0, we can apply the inverse S_0^{-1} to both sides of the equality. This yields:

$$z_{ij}(\mathbf{p}, \beta_0) = h\left(z_{ij}(\mathbf{p}, \beta)\right),$$

where we defined the composite map $h = S_0^{-1} \circ f$.

From Step 3, we have $\theta = \theta_0$. Thus, we have

$$\frac{p_i \tau_{ij}(\theta_0)}{A_j(\mathbf{p}, \beta_0)} = h\left(\frac{p_i \tau_{ij}(\theta_0)}{A_j(\mathbf{p}, \beta)}\right).$$

Now let us define $c_j(\mathbf{p}) \equiv A_j(\mathbf{p}, \beta)/A_j(\mathbf{p}, \beta_0)$. Note that we have $c_j(\mathbf{p}) > 0$ uniformly over $\mathcal{B}^*$ and $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. Using $c_j(\mathbf{p})$. we can express $z_{ij}(\mathbf{p}, \beta_0) = c_j(\mathbf{p}) z_{ij}(\mathbf{p}, \beta)$ which gives us

$$c_j(\mathbf{p}) z_{ij}(\mathbf{p}, \beta) = h\left(z_{ij}(\mathbf{p}, \beta)\right),$$

which equivalently gives us

$$r\left(z_{ij}(\mathbf{p}, \beta)\right) = c_j(\mathbf{p}), \qquad r(z) \equiv \frac{h(z)}{z},$$

for each exporter i. Thus r is constant on every set

$$\mathcal{Z}_j(\mathbf{p}, \beta) \equiv \{z_{ij}(\mathbf{p}, \beta) : i = 1, \ldots, N\}.$$

To propagate this equality across destinations and price vectors, let

$$\mathcal{Z}(\beta) = \bigcup_{j=1}^{N} \bigcup_{\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)} \mathcal{Z}_j(\mathbf{p}, \beta).$$

At this stage, $c_j(\mathbf{p})$ may depend on both the destination and the price vector. The objective of

the propagation argument is to show that there is one constant c such that

$$r\left(z_{ij}(\mathbf{p},\beta)\right) = c$$

for every exporter i, every destination j, and every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$.

Take any two intervals I_a and I_b and the finite sequence supplied by Assumption 11. For any two consecutive intervals in this sequence and any (x, y) in their product, the assumption provides a price vector for which $x = z_{i_q j_q}(\mathbf{p},\beta)$ and $y = z_{\ell_q j_q}(\mathbf{p},\beta)$. The within-destination equality therefore gives $r(x) = r(y)$. To see why this makes r constant, fix y° in the second interval and take any x_1, x_2 in the first interval. The joint support condition may provide different price vectors for (x_1, y°) and (x_2, y°), but the within-destination equality gives

$$r(x_1) = r(y^\circ) = r(x_2).$$

Thus, r is constant on the first interval. The same argument shows that it is constant on the second interval, and $r(x) = r(y)$ shows that the two constants agree. Hence, r is constant on the union of the two consecutive intervals. Repeating this argument along the finite sequence propagates the same constant from I_a to I_b. Because the two intervals were arbitrary and the intervals cover K, there is a single $c > 0$ such that

$$r(z) = c \qquad \text{for every } z \in K.$$

Every effective relative price $z_{ij}(\mathbf{p},\beta)$ belongs to the common support K. Therefore,

$$c_j(\mathbf{p}) = r\left(z_{ij}(\mathbf{p},\beta)\right) = c$$

for every i, j, and $\mathbf{p}$. Consequently, $h(z) = cz$ on K and

$$z_{ij}(\mathbf{p},\beta_0) = c\, z_{ij}(\mathbf{p},\beta)$$

for every i, j and every $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. By the definition of the two effective relative prices and the already established equality $\theta = \theta_0$, this implies

$$A_j(\mathbf{p},\beta) = cA_j(\mathbf{p},\beta_0)$$

for every j and $\mathbf{p}$. Evaluating the last equality at $(j, \mathbf{p}) = (N, \bar{\mathbf{p}})$ and using the normalization of both parameters gives $1 = c \cdot 1$, hence $c = 1$. It follows that $h(z) = z$ on K. Therefore,

$$f(z) = S_0(z) \qquad \text{for every } z \in K.$$

Under the convention that share functions are distinguished by their restrictions to K, we have therefore shown that $M(\beta,\beta_0) = 0$ implies $\theta = \theta_0$ and $f = S_0$ on the common support. Hence

$\beta = \beta_0$ in the normalized parameter space, completing the identification argument for the symmetric HSA model.

D.2 Proof of Proposition 4 (Error Bound of the Sieve Estimator)

The proof is obtained by a simple adaptation of Lemma 10.2 in Györfi et al. (2002) to the present criterion, which jointly uses all bilateral shares in each observation. The argument requires only the existence of the sample minimizer. Uniqueness is not required. The derivation follows.

For $\beta \in \mathcal{B}_\kappa^*$, let

$$\ell_\beta(D_{(r)}) \equiv \frac{1}{N^2} \sum_{i,j} \left| \lambda_{ij,(r)} - f_i(z_{ij,(r)}(\beta)) \right|^2,$$

so that $Q_R(\beta) = P_R \ell_\beta$ and $Q(\beta) = P\ell_\beta$.

First, using $\lambda_{ij,(r)} = S_0(z_{ij,(r)}(\beta_0)) + \varepsilon_{ij,(r)}$, expand the population criterion as

$$\begin{aligned} Q(\beta) = \mathbb{E}\left[\frac{1}{N^2} \sum_{i,j} \left| S_0(z_{ij,(r)}(\beta_0)) - f_i(z_{ij,(r)}(\beta)) \right|^2 \right] \\ + \mathbb{E}\left[\frac{1}{N^2} \sum_{i,j} \varepsilon_{ij,(r)}^2 \right] \\ + 2\mathbb{E}\left[\frac{1}{N^2} \sum_{i,j} \varepsilon_{ij,(r)} \left\{ S_0(z_{ij,(r)}(\beta_0)) - f_i(z_{ij,(r)}(\beta)) \right\} \right]. \end{aligned}$$

The final term is zero under Assumption 4, while the second term equals $Q(\beta_0)$. Hence the population least-squares decomposition is

$$Q(\beta) - Q(\beta_0) = M^2(\beta, \beta_0).$$

Second, take any $\mathbf{f} \in \mathcal{S}_{\kappa,0}^*$ and put $\beta_\mathbf{f} = (\theta_0, \mathbf{f}) \in \mathcal{B}_\kappa^*$. Add and subtract the empirical criteria evaluated at $\hat{\beta}_\kappa$ and $\beta_\mathbf{f}$:

$$\begin{aligned} M^2(\hat{\beta}_\kappa, \beta_0) &= Q(\hat{\beta}_\kappa) - Q(\beta_0) \\ &= \underbrace{Q(\hat{\beta}_\kappa) - Q_R(\hat{\beta}_\kappa)}_{A_{1\kappa}} \\ &\quad + \underbrace{Q_R(\hat{\beta}_\kappa) - Q_R(\beta_\mathbf{f})}_{A_{2\kappa}} \\ &\quad + \underbrace{Q_R(\beta_\mathbf{f}) - Q(\beta_\mathbf{f})}_{A_{3\kappa}} \\ &\quad + \underbrace{Q(\beta_\mathbf{f}) - Q(\beta_0)}_{A_{4\kappa}}. \end{aligned}$$

Third, because $\hat{\beta}_\kappa$ minimizes Q_R over $\mathcal{B}_\kappa^*$ and $\beta_{\mathbf{f}} \in \mathcal{B}_\kappa^*$,

$$A_{2\kappa} = Q_R(\hat{\beta}_\kappa) - Q_R(\beta_{\mathbf{f}}) \leq 0.$$

Moreover,

$$A_{1\kappa} \leq T_{1\kappa}, \qquad A_{3\kappa} \leq T_{1\kappa},$$

and the population least-squares decomposition gives

$$A_{4\kappa} = M^2(\beta_{\mathbf{f}}, \beta_0).$$

It follows that

$$M^2(\hat{\beta}_\kappa, \beta_0) \leq 2T_{1\kappa} + M^2(\beta_{\mathbf{f}}, \beta_0).$$

Finally, taking the infimum over $\mathbf{f} \in \mathcal{S}_{\kappa,0}^*$ yields

$$M^2(\hat{\beta}_\kappa, \beta_0) \leq 2T_{1\kappa} + T_{2\kappa},$$

which proves the proposition.

D.3 Proof of Lipschitz Continuity of The Single Aggregator over the Sieve Space

Let us fix $\kappa \geq 1$, and take any $\xi, \xi' \in \Xi_\kappa$ where Ξ_κ is the coefficient space defined in Appendix D.4. Let us write $S_{\ell,\xi}$ and $S_{\ell,\xi'}$ for the share function for good $\ell \in \{1, \ldots, N\}$ corresponding to the sieve coefficients $\{\alpha_q^{(\ell)}\}_{q=0}^{\kappa}$ in ξ and ξ', respectively. We show that for any $j \in 1, \ldots, N$, for some metric $d(\cdot, \cdot)$, we have

$$\left|A_j(\mathbf{p}, \xi \mid \mathcal{W}) - A_j(\mathbf{p}, \xi' \mid \mathcal{W})\right| \leq L_A d(\xi, \xi'),$$

for any $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$ where L_A is a Lipschitz constant.

Take any $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. Since $\mathcal{S}_\kappa \subset \mathcal{S}$, it follows that for any $\xi \in \Xi_\kappa$, there exists unique constant $A_\xi \equiv A_j(\mathbf{p}, \xi \mid \mathcal{W})$ and $A_{\xi'} \equiv A_j(\mathbf{p}, \xi' \mid \mathcal{W})$, such that HSA normalization (2) is satisfied. That is, we have, $G_j(A_\xi, \mathbf{p} \mid \xi, \mathcal{W}) = G_j(A_{\xi'}, \mathbf{p} \mid \xi', \mathcal{W}) = 0$ where

$$G_j(A, \mathbf{p} \mid \xi, \mathcal{W}) := \sum_{\ell=1}^{N} S_{\ell,\xi}\left(\frac{p_\ell \tau_{\ell j}(\theta)}{A}\right) - 1.$$

We obtain the following decomposition:

$$\begin{aligned} 0 \;=\; & G_j(A_\xi, \mathbf{p} \mid \xi, \mathcal{W}) - G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W}) \\ & + G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W}) - G_j(A_{\xi'}, \mathbf{p} \mid \xi', \mathcal{W}), \end{aligned}$$

which gives us:

$$\left|G_j(A_\xi, \mathbf{p} \mid \xi, \mathcal{W}) - G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W})\right| = \left|G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W}) - G_j(A_{\xi'}, \mathbf{p} \mid \xi', \mathcal{W})\right|. \tag{29}$$

To derive the Lipschitz constant, we bound the right hand side of (29) from below, and the left hand side of (29) from above.

First inspect the right hand side of (29). Since $A_\xi, A_{\xi'} \in [\underline{A_j}, \bar{A}_j]$ where $\underline{A_j} > 0$, a mean value expansion gives:

$$S_{\ell,\xi'}\left(\frac{p_\ell \tau_{\ell j}(\theta')}{A_\xi}\right) = S_{\ell,\xi'}\left(\frac{p_\ell \tau_{\ell j}(\theta')}{A_{\xi'}}\right) + \dot{S_{\ell,\xi'}}(\zeta_{\ell j}) p_\ell \tau_{\ell j}(\theta') \frac{A_{\xi'} - A_\xi}{A_\xi A_{\xi'}}$$

for some intermediate value $\zeta_{\ell j}$. Since $p_\ell \tau_{\ell j}(\theta)/A_j(\mathbf{p}, \xi \mid \mathcal{W}) \in K_j$ uniformly over the parameter space, the intermediate value is also an element of K_j. Hence, we obtain:

$$\begin{aligned}
\left|G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W}) - G_j(A_{\xi'}, \mathbf{p} \mid \xi', \mathcal{W})\right| &= \left|\sum_{\ell=1}^{N}\left[S_{\ell,\xi'}\left(\frac{p_\ell \tau_{\ell j}(\theta')}{A_\xi}\right) - S_{\ell,\xi'}\left(\frac{p_\ell \tau_{\ell j}(\theta')}{A_{\xi'}}\right)\right]\right|, \\
&= \left|\sum_{\ell=1}^{N} \dot{S_{\ell,\xi'}}(\zeta_{\ell j}) p_\ell \tau_{\ell j}(\theta') \frac{A_{\xi'} - A_\xi}{A_\xi A_{\xi'}}\right|, \\
&\geq \left|\sum_{\ell=1}^{N} \dot{S_{\ell,\xi'}}(\zeta_{\ell j}) p_\ell \tau_{\ell j}(\theta')\right| \left|A_{\xi'} - A_\xi\right| \left(\bar{A}_j\right)^{-2}, \\
&\geq \rho_R^L \left|A_{\xi'} - A_\xi\right| \left(\bar{A}_j\right)^{-2},
\end{aligned} \tag{30}$$

where the third inequality follows from $A_\xi, A_{\xi'} \leq \bar{A}_j < \infty$, and the fourth line uses the direct derivative restriction, $\sum_\ell p_\ell = 1$, and $\tau_{\ell j}(\theta') \geq 1$.

We will next bound the left hand side of (29) from above. Let $z_{ij}(\theta, A_\xi) = p_i \tau_{ij}(\theta)/A_\xi$.

$$\begin{aligned}
S_{\ell,\xi}\left(z_{\ell j}(\theta, A_\xi)\right) - S_{\ell,\xi'}\left(z_{\ell j}(\theta', A_\xi)\right) &= \sum_{q=0}^{\kappa} (\alpha_{q,\xi}^{(\ell)} - \alpha_{q,\xi'}^{(\ell)}) z_{\ell j}(\theta, A_\xi)^{-q} \\
&\quad - \sum_{q=1}^{\kappa} \alpha_{q,\xi}^{(\ell)} (-q) \zeta_{\ell j}^{-q-1} \frac{p_\ell}{A_\xi} \left(\tau_{\ell j}(\theta') - \tau_{\ell j}(\theta)\right), \\
&= \sum_{q=0}^{\kappa} (\alpha_{q,\xi}^{(\ell)} - \alpha_{q,\xi'}^{(\ell)}) z_{\ell j}(\theta, A_\xi)^{-q} \\
&\quad + \sum_{q=1}^{\kappa} \alpha_{q,\xi}^{(\ell)} q \zeta_{\ell j}^{-q-1} \frac{p_\ell}{A_\xi} \left(\tau_{\ell j}(\theta') - \tau_{\ell j}(\theta)\right), \\
&= B_{1\ell} + B_{2\ell}, \text{ say},
\end{aligned}$$

where we used $\zeta_{\ell j}$ to denote the intermediate value between $z_{ij}(\theta, A_\xi)$ and $z_{ij}(\theta', A_\xi)$.

Let $C_{z,\kappa} = \max_{q \in \{1,\ldots,\kappa\}} \sup_{z \in K_j} z^{-q}$. Then, we have $|B_{1\ell}| \leq C_{z,\kappa} \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(\ell)} - \alpha_{q,\xi'}^{(\ell)}|$.

By compactness of Θ^C and finiteness of the fixed design coupled with the parametrization $\theta \mapsto \tau_{\ell j}(\theta) = \exp(W_{\ell j}'\theta)$, we have $|\tau_{\ell j}(\theta) - \tau_{\ell j}(\theta')| \leq C_\tau ||\theta - \theta'||_1$, for some constant $C_\tau = \sup_{\theta, W}\{\exp(W'\theta) \cdot ||W||_1\} < \infty$, where $||\cdot||_1$ is the L1 norm. Therefore, the second term $B_{2\ell}$ is

bounded by:

$$\begin{aligned}|B_{2\ell}| &\leq \rho_R^U \frac{p_\ell}{A_\xi} \left(\sum_{q=1}^{\kappa} q \sup_{z \in K_j} z^{-q-1} \right) C_\tau ||\theta - \theta'||_1, \\ &\leq (\rho_R^U \kappa / \underline{A_j}) (\max_{q \in \{1,\ldots,\kappa\}} q \sup_{z \in K_j} z^{-q-1}) C_\tau ||\theta - \theta'||_1,\end{aligned}$$

as $p_\ell < 1$ due to normalization.

Thus, we have

$$\sum_{\ell=1}^{N} |B_{1\ell}| \leq C_{z,\kappa} \sum_{\ell=1}^{N} \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(\ell)} - \alpha_{q,\xi'}^{(\ell)}|,$$

and

$$\sum_{\ell=1}^{N} |B_{2\ell}| \leq (\rho_R^U \kappa / \underline{A_j}) (\max_{q \in \{1,\ldots,\kappa\}} q \sup_{z \in K_j} z^{-q-1}) N C_\tau ||\theta - \theta'||_1.$$

$$\begin{aligned}&\left| G_j(A_\xi, \mathbf{p} \mid \xi, \mathcal{W}) - G_j(A_\xi, \mathbf{p} \mid \xi', \mathcal{W}) \right| \\ &\leq \max\{C_{z,\kappa}, \frac{\rho_R^U \kappa}{\underline{A_j}} N C_\tau (\max_{q \in \{1,\ldots,\kappa\}} q \sup_{z \in K_j} z^{-q-1}) \} ||\xi - \xi'||_1. \end{aligned} \tag{31}$$

Hence, combining (31) and (30), we obtain

$$\left| A_{\xi'} - A_\xi \right| \leq L_{A,j}(\kappa, \rho_R^U, \rho_R^L) \cdot ||\xi - \xi'||_1,$$

where

$$L_{A,j}(\kappa, \rho_R^U, \rho_R^L) = \frac{\bar{A}_j^2}{\rho_R^L} \max \left\{ C_{z,\kappa}, \frac{\rho_R^U \kappa}{\underline{A}_j} N C_\tau \left(\max_{q=1,\ldots,\kappa} q \sup_{z \in K} z^{-q-1} \right) \right\}.$$

for each $\kappa \geq 1$. Since K_j is compact, and $\rho_R^U < \infty$, and $\rho_R^L > 0$, the Lipschitz constant satisfies $0 < L_A(\kappa, \rho_R^U, \rho_R^L) < \infty$ at each κ. We have therefore established Lipschitz continuity of $\xi \mapsto A_j(\mathbf{p}, \xi \mid \mathcal{W})$ for each $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$.

D.4 Empirical-Process Bound under Assumption 13

Define

$$\ell_\beta(\Lambda, \mathbf{p}) \equiv \frac{1}{N^2} \sum_{i,j} \left[f_i \left(\frac{p_i \tau_{ij}(\theta)}{A_j(\mathbf{p}, \beta \mid \mathcal{W})} \right) - \lambda_{ij} \right]^2,$$

and let

$$\mathcal{G}_\kappa^* \equiv \{\ell_\beta : \beta \in \mathcal{B}_\kappa^*\}, \qquad \overline{\mathcal{G}}_\kappa \equiv \{\ell_\beta : \beta \in \overline{\mathcal{B}}_\kappa\}.$$

For any measurable function g, define

$$\nu_R(g) = \frac{1}{R} \sum_{r=1}^{R} \left\{ g(\Lambda_{(r)}, \mathbf{P}_{0,(r)}) - \mathbb{E}[g(\Lambda, \mathbf{P}_0)] \right\}.$$

Then $\mathcal{G}_\kappa^* \subseteq \overline{\mathcal{G}}_\kappa$ and

$$T_{1\kappa} = \sup_{g \in \mathcal{G}_\kappa^*} |\nu_R(g)| \leq \sup_{g \in \overline{\mathcal{G}}_\kappa} |\nu_R(g)|.$$

The entropy calculation below is for the larger class $\overline{\mathcal{G}}_\kappa$; thus it provides a valid upper bound for $T_{1\kappa}$. In the remainder of this argument, $\mathcal{G}_\kappa$ is understood as this larger class.

Each $f_i \in \mathcal{S}_{i,\kappa}$ is indexed by the coefficient vector $\alpha^{(i)}$. Collect the finite-dimensional parameters in

$$\Xi_\kappa = \left\{ \xi = (\theta', \alpha^{(1)'}, \ldots, \alpha^{(N)'})' \;\middle|\; (\theta, g_{\alpha^{(1)}}, \ldots, g_{\alpha^{(N)}}) \in \overline{\mathcal{B}}_\kappa \right\}.$$

This set is compact and has dimension $d + N(\kappa + 1)$. For each $\xi \in \Xi_\kappa$, let g_ξ denote the corresponding loss in $\overline{\mathcal{G}}_\kappa$.

To show that $T_{1\kappa} = o_p(1)$, we first show that the class $\mathcal{G}_\kappa$ is Lipschitz in the index parameter $\xi \in \Xi_\kappa$, with a constant envelope $G < \infty$. This, by an application of Theorem 2.7.11 in van der Vaart and Wellner (1996), gives us an upper bound on the bracketing number of the class $\mathcal{G}_\kappa$ given by a covering number of the class Ξ_κ. Then, by applying Theorem 3.7 in van de Geer (2006), we conclude that $\mathcal{G}_\kappa$ is Glivenko-Cantelli, for each $\kappa \geq 1$.

Take any $\xi, \tilde{\xi} \in \Xi_\kappa$. Let $\beta_\xi = (\theta, f_1, \ldots, f_N)$ and $\beta_{\tilde{\xi}} = (\tilde{\theta}, h_1, \ldots, h_N)$ denote the corresponding elements of $\overline{\mathcal{B}}_\kappa$. Let us introduce a shorthand $z_{ij}(\xi) = p_i \tau_{ij}(\theta) / A_j(\mathbf{p}, \beta_\xi \mid \mathcal{W})$ to lighten the notation. Then, for any Λ and $\mathbf{p}$, we have

$$\begin{aligned} \left| g_\xi(\Lambda, \mathbf{P}) - g_{\tilde{\xi}}(\Lambda, \mathbf{P}) \right| &\leq \frac{1}{N^2} \sum_{i,j} \left| f_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\tilde{\xi})\right) \right| \\ &\qquad \times \left\{ \left| f_i\left(z_{ij}(\xi)\right) \right| + \left| h_i\left(z_{ij}(\tilde{\xi})\right) \right| + 2|\lambda_{ij}| \right\}, \qquad (32) \\ &\leq (4 + 2C_\varepsilon) \frac{1}{N^2} \sum_{i,j} \left| f_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\tilde{\xi})\right) \right|, \end{aligned}$$

where the first line uses $|a^2 - b^2| = |a - b||a + b|$. The second line uses $|\lambda_{ij}| \leq |\lambda_{ij}^*| + |\varepsilon_{ij}| \leq 1 + C_\varepsilon$ almost surely due to $\varepsilon_{ij,(r)}$ having compact support, and the adding-up condition in (2), which implies $0 < f_i(z_{ij}(\xi)) \leq 1$ for every candidate in $\overline{\mathcal{B}}_\kappa$.

Let us examine each term in the summation in the last inequality in (32):

$$\begin{aligned} \left| f_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\tilde{\xi})\right) \right| &\leq \left| f_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\xi)\right) \right| + \left| h_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\tilde{\xi})\right) \right|, \\ &= t_1 + t_2, \text{ say}. \end{aligned}$$

We inspect the two components t_1 and t_2 in order. Let us denote by $\{\alpha_{q,\xi}^{(i)}\}_{q=0}^{\kappa}$ the sieve coeffi-

cients corresponding to f_i. For t_1, we have

$$\begin{aligned} |t_1| &= \left| \sum_{q=0}^{\kappa} (\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}) z_{ij}(\xi)^{-q} \right|, \\ &\leq \left(\max_{q \in \{0,\ldots,\kappa\}} z_{ij}(\xi)^{-q} \right) \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}|, \\ &\leq \left(\max_{q \in \{0,\ldots,\kappa\}} \sup_{z \in K_j} z^{-q} \right) \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}|. \end{aligned}$$

For t_2, by a mean value expansion $h_i(z_{ij}(\tilde{\xi})) = h_i(z_{ij}(\xi)) + \dot{h}_i(\zeta_{ij})(z_{ij}(\tilde{\xi}) - z_{ij}(\xi))$ which is valid as $z_{ij}(\xi) \in K_j$ uniformly over $\xi \in \Xi_\kappa$ (Proposition 2),

$$|t_2| \leq \left| \dot{h}_i(\zeta_{ij}) \right| \left| z_{ij}(\tilde{\xi}) - z_{ij}(\xi) \right|.$$

The derivative evaluated at an intermediate value ζ_{ij} can be uniformly bounded as follows:

$$\begin{aligned} \left| \dot{h}_i(\zeta_{ij}) \right| &\leq \left| - \sum_{q=1}^{\kappa} \alpha_{q,\tilde{\xi}}^{(i)} q \zeta_{ij}^{-q-1} \right| \\ &\leq \kappa \rho_R^U \max_{q \in \{1,\ldots,\kappa\}} \sup_{z \in K_j} q z^{-q-1}. \end{aligned}$$

Thus, what remains to be shown is the Lipschitz continuity of the components $|z_{ij}(\tilde{\xi}) - z_{ij}(\xi)|$. Let us abbreviate $A_{j,\xi} \equiv A_j(\mathbf{p}, \xi \mid \mathcal{W})$. Then we have,

$$\begin{aligned} \left| z_{ij}(\tilde{\xi}) - z_{ij}(\xi) \right| &\leq \left| \frac{p_i \tau_{ij}(\theta)}{A_{j,\xi}} - \frac{p_i \tau_{ij}(\tilde{\theta})}{A_{j,\tilde{\xi}}} \right|, \\ &\leq \underline{\frac{p_i}{A_j^2}} \left| \tau_{ij}(\theta) A_{j,\tilde{\xi}} - \tau_{ij}(\tilde{\theta}) A_{j,\xi} \right|, \\ &\leq \underline{\frac{1}{A_j^2}} \left| \tau_{ij}(\theta) - \tau_{ij}(\tilde{\theta}) \right| \bar{A}_j + \underline{\frac{1}{A_j^2}} \left| A_{j,\xi} - A_{j,\tilde{\xi}} \right| \bar{\tau}, \end{aligned}$$

where we used $A_j(\mathbf{p}, \xi \mid \mathcal{W}) \in [\underline{A_j}, \bar{A}_j]$ with $\underline{A_j} > 0$ uniformly over $(\mathbf{p}, \xi) \in \mathbb{S}(\mathbf{P}_0) \times \Xi_\kappa$ in line 2, and line 3 uses $p_i < 1$ and $\tau_{ij}(\theta) \leq \bar{\tau} < \infty$ due to the support $\mathbb{S}(\mathbf{P}_0)$ of $\mathbf{p}$ being a subset of the unit simplex $\Delta_{\mathbf{P}} \setminus \partial\Delta_{\mathbf{P}}$ (net of the boundary), compactness of Θ^C, and finiteness of the fixed design, respectively.

Hence, we have:

$$
\begin{aligned}
&\sum_{i,j} \left| f_i\left(z_{ij}(\xi)\right) - h_i\left(z_{ij}(\tilde{\xi})\right)\right| \\
&\leq \sum_{i,j} \left(\max_{q\in\{0,\ldots,\kappa\}} \sup_{z\in K_j} z^{-q} \right) \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}| \\
&+ \sum_{i,j} \kappa \rho_R^U \max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \left| z_{ij}(\tilde{\xi}) - z_{ij}(\xi)\right|, \\
&\leq N \sum_{j=1}^{N} \left(\max_{q\in\{0,\ldots,\kappa\}} \sup_{z\in K_j} z^{-q} \right) \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}| \\
&+ N\kappa\rho_R^U \sum_{j=1}^{N} \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \frac{\bar{A}_j}{\underline{A_j^2}} \left| \tau_{ij}(\theta) - \tau_{ij}(\tilde{\theta})\right| \\
&+ N\kappa\rho_R^U \sum_{j=1}^{N} \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \frac{\bar{\tau}}{\underline{A_j^2}} \left| A_{j,\xi} - A_{j,\tilde{\xi}}\right|, \\
&= N^2 \max_{j\in\{1,\ldots,N\}} \left(\max_{q\in\{0,\ldots,\kappa\}} \sup_{z\in K_j} z^{-q} \right) \sum_{j=1}^{N} \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}| \\
&+ N^2\kappa\rho_R^U\bar{\tau} \left(\max_{i,j} \|W_{ij}\|_1 \right) \max_{j\in\{1,\ldots,N\}} \left\{ \frac{\bar{A}_j}{\underline{A_j^2}} \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right\} \\
&\qquad \times ||\theta - \tilde{\theta}||_1 \\
&+ N^2\kappa\rho_R^U\bar{\tau} \max_{j\in\{1,\ldots,N\}} \left(\frac{L_{A,j}(\kappa,\rho_R^U,\rho_R^L)}{\underline{A_j^2}} \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right) \\
&\qquad \times ||\xi - \tilde{\xi}||_1, \\
&\leq N^2 \max_{j\in\{1,\ldots,N\}} \left(\max_{q\in\{0,\ldots,\kappa\}} \sup_{z\in K_j} z^{-q} \right) \sum_{j=1}^{N} \sum_{q=0}^{\kappa} |\alpha_{q,\xi}^{(i)} - \alpha_{q,\tilde{\xi}}^{(i)}| \\
&+ N^2\kappa\rho_R^U\bar{\tau} \left(\max_{i,j} \|W_{ij}\|_1 \right) \frac{\bar{A}^*}{(\underline{A^*})^2} \max_{j\in\{1,\ldots,N\}} \left\{ \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right\} \\
&\qquad \times ||\theta - \tilde{\theta}||_1 \\
&+ N^2\kappa\rho_R^U\bar{\tau} \frac{1}{(\underline{A^*})^2} \max_{j\in\{1,\ldots,N\}} \left(L_{A,j}(\kappa,\rho_R^U,\rho_R^L) \left(\max_{q\in\{1,\ldots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right) \\
&\qquad \times ||\xi - \tilde{\xi}||_1.
\end{aligned}
$$

Since $A_j \in [\underline{A_j}, \bar{A}_j]$ for all $j \in \{1, \ldots, N\}$, there exists $0 < \underline{A^*} = \min_j \underline{A_j}$, and $\bar{A}^* = \max_j \bar{A}_j$, we have

$$\left| g_{\xi}(\Lambda, \mathbf{P}) - g_{\xi'}(\Lambda, \mathbf{P}) \right| \leq G(\kappa, \rho_R^U, \rho_R^L) \times \|\xi - \tilde{\xi}\|_1,$$

where

$$G(\kappa, \rho_R^U, \rho_R^L) = (4 + 2C_\varepsilon) \max\left\{ I_\kappa, II_\kappa, III_\kappa \right\},$$

with

$$
\begin{aligned}
I_\kappa &\equiv \max_{j\in\{1,\dots,N\}} \left(\max_{q\in\{0,\dots,\kappa\}} \sup_{z\in K_j} z^{-q} \right), \\
II_\kappa &\equiv \kappa \rho_R^U \bar{\tau} \left(\max_{i,j} \|W_{ij}\|_1 \right) \frac{\bar{A}^*}{(\underline{A^*})^2} \max_{j\in\{1,\dots,N\}} \left\{ \left(\max_{q\in\{1,\dots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right\}, \\
III_\kappa &\equiv \kappa \rho_R^U \bar{\tau} (\underline{A^*})^{-2} \max_{j\in\{1,\dots,N\}} \left(L_{A,j}(\kappa, \rho_R^U, \rho_R^L) \left(\max_{q\in\{1,\dots,\kappa\}} \sup_{z\in K_j} q z^{-q-1} \right) \right).
\end{aligned}
$$

Let us introduce a shorthand $G_\kappa = G(\kappa, \rho_R^U, \rho_R^L)$ to conserve space.

Let us write $a_j = \min_{z\in K_j} z$ and $b_j = \max_{z\in K_j} z$ for the minimum and the maximum of the elements in our compact sets K_j. Note that $K_j \subset \mathbb{R}_{++}$ implies that $a_j > 0$. Let us further denote the minimum/maximum over a finite number of countries by $a = \min_{j\in\{1,\dots,N\}} a_j$ and $b = \min_{j\in\{1,\dots,N\}} b_j$. We have $I_\kappa = O(a^{-\kappa} \vee 1)$ and $II_\kappa = O(\kappa^2 \rho_R^U) O((a^{-\kappa-1} \vee 1))$. Using the corrected Lipschitz constant in Appendix C.3 gives

$$
III_\kappa = O\left(\kappa^4 (\rho_R^U)^3 (\rho_R^L)^{-1}\right) O\left(a^{-2(\kappa-1)} \vee 1\right).
$$

The additional polynomial factor contributes only $O(\log \kappa)$ to $\log G_\kappa$ and therefore does not alter the entropy rate.

By Theorem 2.7.11 in van der Vaart and Wellner (1996), we have, for any $\delta > 0$ and any norm $\|\cdot\|$,

$$
\mathcal{N}_{[]}\left(\delta, \mathcal{G}_\kappa, ||\cdot||\right) \leq \mathcal{N}\left(\frac{\delta}{2\|G_\kappa\|}, \Xi_\kappa, d_e\right),
$$

where $\mathcal{N}_{[]}\left(\delta, \mathcal{G}_\kappa, \|\cdot\|\right)$ is the δ-bracketing number of a class $\mathcal{G}_\kappa$ (see Definition 2.1.6 in van der Vaart and Wellner (1996)) with respect to the norm $\|\cdot\|$, and $\mathcal{N}(\delta', \Xi_\kappa, d_e)$ is the δ'-covering number (see Definition 2.1.5 in van der Vaart and Wellner (1996)) of the parameter space Ξ_κ with respect to the Euclidean metric d_e.

Let D_Θ denote the Euclidean diameter of Θ^C. The coefficient restrictions imply that Ξ_κ has Euclidean diameter at most $D_\Theta + 2\sqrt{N}\rho_R^U$. Therefore,

$$
\mathcal{N}\left(\frac{\delta}{2\|G_\kappa\|}, \Xi_\kappa, d_e\right) \lesssim \left(1 + \frac{C(1+\rho_R^U)\|G_\kappa\|}{\delta}\right)^{N(\kappa+1)+d},
$$

where C does not depend on R or κ.

Thus the $\delta/2G_\kappa$ entropy of the class Ξ_κ is of the order:

$$
(N(\kappa+1)+d) \ln\left(1 + \frac{C(1+\rho_R^U)G_\kappa}{\delta}\right) = O\left(N\kappa \ln\left(1 + \frac{G_\kappa \rho_R^U}{\delta}\right)\right), \tag{33}
$$

as $G_\kappa > 0$, and $\rho_R^U \in \mathbb{R}_{++}$ with $\rho_R^U \to \infty$, the log term is strictly positive for any $\delta > 0$.

To invoke the LLN uniformly over the class $\mathcal{G}_\kappa$ for each $\kappa \geq 1$, we need to verify that we have $N\kappa \ln(1 + G(\kappa, \rho_R^U, \rho_R^L)\rho_R^U/\delta) = o(R)$ for all $\delta > 0$. The quantities I_κ, II_κ, and III_κ need not

themselves be $o(R)$. What enters the entropy bound is their logarithm. Since $a > 0$ is fixed, the displayed bounds above imply

$$\log\left(1 + G(\kappa, \rho_R^U, \rho_R^L)\rho_R^U/\delta\right) = O\left(\kappa + \log(\kappa\rho_R^U/\rho_R^L)\right)$$

for every fixed $\delta > 0$. Therefore,

$$N\kappa \log\left(1 + G(\kappa, \rho_R^U, \rho_R^L)\rho_R^U/\delta\right) = O\left(\kappa^2 + \kappa\log(\kappa\rho_R^U/\rho_R^L)\right).$$

By Assumption 13, $R^{-1}\kappa^2\log(\kappa\rho_R^U/\rho_R^L) \to 0$. Because the logarithm is eventually bounded away from zero, this also implies $\kappa^2/R \to 0$ and $\kappa\log(\kappa\rho_R^U/\rho_R^L)/R \to 0$. Hence the required entropy is $o(R)$.

We therefore have the following:

$$\begin{aligned}
\ln\mathcal{N}\left(\delta, \mathcal{G}_\kappa, L_1(\mathbb{P}_R)\right) &\leq \ln\mathcal{N}_{[]}\left(\delta, \mathcal{G}_\kappa, L_1(\mathbb{P}_R)\right), \\
&\leq \ln\mathcal{N}\left(\frac{\delta}{2\|G_\kappa\|}, \Xi_\kappa, d_e\right), \\
&= o(R), \ (\forall\delta > 0)
\end{aligned}$$

where the first inequality follows from Lemma 2.1 in van de Geer (2006), the second inequality follows from Theorem 2.7.11 in van der Vaart and Wellner (1996), and the third line follows from the preceding entropy calculation. Therefore, an application of Theorem 3.7 in van de Geer (2006) establishes that $\{\mathcal{G}_\kappa\}_\kappa$ is Glivenko-Cantelli.

The rate of convergence of $T_{1\kappa}$ follows from applying the bracketing maximal inequality recorded as Theorem 2.14.2 in van der Vaart and Wellner (1996). Formally, their Theorem 2.14.2 gives us:

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}_\kappa}|\nu_R(g)|\right] \lesssim \frac{1}{\sqrt{R}} J_{[]}\left(1, \mathcal{G}_\kappa, L_2(\mathbb{P})\right) \cdot \|G\|_{2,P}, \tag{34}$$

with

$$J_{[]}\left(\delta, \mathcal{G}_\kappa, ||\cdot||_{2,\mathbb{P}}\right) \equiv \int_0^\delta \sqrt{1 + \ln\mathcal{N}_{[]}\left(\eta\|G\|_{2,P}, \mathcal{G}_\kappa, L_2(\mathbb{P})\right)}\mathrm{d}\eta,$$

defined as a bracketing integral of a class of function $\mathcal{G}_\kappa$.

Application of Theorem 2.14.2 requires that we have $\|g\|_{2,P} < \|G\|_{2,P}$ for all $g \in \mathcal{G}_\kappa$. The envelope $||G||_{2,\mathbb{P}}$ is obtained by exploiting the property of HSA aggregation as follows. We

have, almost surely, :

$$\begin{aligned}
g(\Lambda_{(r)}, \mathbf{P}_{0,(r)}) &= \frac{1}{N^2}\sum_{i,j}\left(\lambda_{ij,(r)} - f_i\left(\frac{P_{0i,(r)}\tau_{ij}(\theta)}{A_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W})}\right)\right)^2, \\
&\leq \frac{2}{N^2}\sum_{i,j}\lambda_{ij,(r)}^2 + \frac{2}{N^2}\sum_{i,j}\left\{f_i\left(\frac{P_{0i,(r)}\tau_{ij}(\theta)}{A_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W})}\right)\right\}^2, \\
&= \frac{2}{N^2}\sum_{i,j}\left(\lambda^*_{ij,(r)} + \varepsilon_{ij,(r)}\right)^2 \\
&\quad + \frac{2}{N^2}\sum_{i,j}\left\{f_i\left(\frac{P_{0i,(r)}\tau_{ij}(\theta)}{A_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W})}\right)\right\}^2, \\
&\leq \frac{2}{N^2}\sum_{i,j}\left(\lambda^*_{ij,(r)}\right)^2 + \frac{2}{N^2}\sum_{i,j}\varepsilon^2_{ij,(r)} \\
&\quad + \frac{2}{N^2}\sum_{i,j}\left\{f_i\left(\frac{P_{0i,(r)}\tau_{ij}(\theta)}{A_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W})}\right)\right\}^2, \\
&\leq \frac{2}{N^2}\sum_{i,j}\varepsilon^2_{ij,(r)} + 4,
\end{aligned}$$

where we used the inequality $(x+y)^2 \leq 2x^2 + 2y^2$ in lines two and four, and we used that the HSA normalization holds uniformly over the parameter space to bound the shares $0 < |\lambda^{*2}_{ij,(r)}| < 1$ and $0 < f_i(P_{0i}\tau_{ij}(\theta)/A_j(\mathbf{P}_0, \beta \mid \mathcal{W})) < 1$ to obtain the final bound. Thus, provided that we have $\mathbb{E}\varepsilon^4_{ij,(r)} < \infty$ we have a finite constant (across κ) envelope $\|G\|_{2,\mathbb{P}} < \infty$ which holds for any $\kappa \geq 1$. Clearly, $\|g\| < \|G\|_{2,P}$ holds for any $g \in \mathcal{G}_\kappa$. This validates the application of maximal inequality given in (34).

Theorem 2.14.2 of van der Vaart and Wellner (1996) gives us:

$$\begin{aligned}
\mathbb{E}\left[\sup_{g\in\mathcal{G}_\kappa}|\nu_R(g)|\right] &\lesssim \frac{1}{\sqrt{R}} J_{[]}\left(1, \mathcal{G}_\kappa, L_2(\mathbb{P})\right) \cdot \|G\|_{2,P}, \\
&= \frac{\|G\|_{2,P}}{\sqrt{R}}\int_0^1 \sqrt{1 + \ln \mathcal{N}_{[]}\left(\eta\|G\|_{2,P}, \mathcal{G}_\kappa, L_2(\mathbb{P})\right)}\mathrm{d}\eta
\end{aligned}$$

We will further bound the bracketing integral $J_{[]}\left(1, \mathcal{G}_\kappa, L_2(\mathbb{P})\right)$. By applying Theorem 2.7.11 of van der Vaart and Wellner (1996) for the δ-bracketing number with respect to the $L_2(\mathbb{P})$ norm, we have:

$$\ln \mathcal{N}_{[]}\left(\delta, \mathcal{G}_\kappa, L_2(\mathbb{P})\right) \leq \ln \mathcal{N}\left(\frac{\delta}{2||G_\kappa||}, \Xi_\kappa, d_e\right),$$

for any $\delta > 0$. Hence we obtain:

$$\begin{aligned}
J_{[]}\left(1, \mathcal{G}_\kappa, L_2(\mathbb{P})\right) &= \int_0^1 \sqrt{1 + \ln \mathcal{N}_{[]}(\eta\|G\|_{2,P}, \mathcal{G}_\kappa, L_2(\mathbb{P})}\mathrm{d}\eta, \\
&\leq \int_0^1 \sqrt{1 + \ln \mathcal{N}(\eta\|G\|_{2,P}/2\|G_\kappa\|, \Xi_\kappa, d_e}\mathrm{d}\eta.
\end{aligned}$$

Since $\|G\|_{2,P}$ is a constant, the bound is of the same order as the one obtained earlier in (33).

Therefore, we obtain:

$$\begin{aligned}\mathbb{E}\left[\sup_{g\in\mathcal{G}_\kappa}|\nu_R(g)|\right] &\lesssim \frac{1}{\sqrt{R}}\sqrt{N\kappa\ln(1+G(\kappa,\rho_R^U,\rho_R^L)\rho_R^U)},\\ &= O\left(\kappa\sqrt{\frac{\ln(\kappa\rho_R^U/\rho_R^L)}{R}}\right),\end{aligned}$$

which by Markov's Inequality, tells us that

$$\sup_{g\in\mathcal{G}_\kappa}|\nu_R(g)| = O_p\left(\kappa\sqrt{\frac{\ln(\kappa\rho_R^U/\rho_R^L)}{R}}\right),$$

for all $\kappa \geq 1$.

D.5 Local Lipschitz Continuity of the Single Aggregator

Let us write $\beta_\kappa = (\theta_0, f_1, \ldots, f_N) \in \mathcal{B}_\kappa$, and $\beta_0 = (\theta_0, S_{01}, \ldots, S_{0N}) \in \mathcal{B}$. Conditional on the fixed $\mathcal{W}$, we show that uniformly over $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$, we have

$$|A_j(\mathbf{p}, \beta_\kappa \mid \mathcal{W}) - A_j(\mathbf{p}, \beta_0 \mid \mathcal{W})| \leq C_j \times \sum_{\ell=1}^N \sup_{z\in K_j}|f_\ell(z) - S_{0\ell}(z)|,$$

for some constant C_j which does not depend on κ, for any $j \in \{1, \ldots, N\}$.

Let us fix j and take any $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. Let us abbreviate the single aggregator by $A_{j,\mathbf{f}}(\mathbf{p}) = A_j(\mathbf{p}, \beta_\kappa \mid \mathcal{W})$ and $A_{j,\mathbf{S_0}}(\mathbf{p}) = A_j(\mathbf{p}, \beta_0 \mid \mathcal{W})$. Since the single aggregator exists uniquely uniformly over $\mathbf{S} \in \mathcal{S}$, it follows that we have

$$G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_\kappa \mid \mathcal{W}\right) - G_j\left(A_{j,\mathbf{S}_0}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right) = 0$$

which by adding and subtracting $G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right)$ implies:

$$\begin{aligned}&\left|G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right) - G_j\left(A_{j,\mathbf{S}_0}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right)\right|\\ &= \left|G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_\kappa \mid \mathcal{W}\right) - G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right)\right|.\end{aligned} \tag{35}$$

Inspecting the right hand side of (35), by linearity of $\mathbf{S} \mapsto G_j(a, \mathbf{p}, \mathbf{S} \mid \mathcal{W})$, we obtain

$$\begin{aligned}&\left|G_j\left(A_{j,\mathbf{S}_0}(\mathbf{p}), \mathbf{p}, \beta_\kappa \mid \mathcal{W}\right) - G_j\left(A_{j,\mathbf{S}_0}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right)\right|\\ &= \left|\sum_{\ell=1}^N f_\ell\left(\frac{p_\ell\tau_{\ell j}(\theta_0)}{A_{j,\mathbf{f}}(\mathbf{p})}\right) - S_{0\ell}\left(\frac{p_\ell\tau_{\ell j}(\theta_0)}{A_{j,\mathbf{f}}(\mathbf{p})}\right)\right|,\\ &\leq \sum_{\ell=1}^N \sup_{z\in K_j}|f_\ell(z) - S_{0\ell}(z)|,\end{aligned}$$

where we used $A_j(\mathbf{p}, \beta \mid \mathcal{W}) \in [\underline{A_j}, \bar{A}_j]$ uniformly over $(\mathbf{p}, \beta)$ and the ratio $p_i\tau_{ij}(\theta)/A_j(\mathbf{p}, \beta \mid$

$\mathcal{W}) \in K_j$ as a result of Proposition 2. Note that the supremum bound does not depend on $\mathbf{p}, \mathcal{W}$.

Next inspect the left hand side of (35), by a mean value expansion with respect to the first argument, we have, for some intermediate value $a \in [\underline{A}_j, \bar{A}_j]$:

$$\left|G_j\left(A_{j,\mathbf{f}}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right) - G_j\left(A_{j,\mathbf{s}_0}(\mathbf{p}), \mathbf{p}, \beta_0 \mid \mathcal{W}\right)\right| = \left|\partial G_j\left(a, \mathbf{p}, \beta_0 \mid \mathcal{W}\right)\right| \cdot \left|A_{j,\mathbf{s}_0}(\mathbf{p}) - A_{j,\mathbf{f}}(\mathbf{p})\right|,$$

where $A \mapsto \partial_A G_j(A, \mathbf{p}, \beta_0 \mid \mathcal{W})$ is the derivative of $G_j(\cdot)$ with respect to the first argument. Let

$$\underline{s}_j \equiv \inf_{z \in K_j} \{-S_0'(z)\} > 0,$$

which follows from the uniform-slope condition stated in the main text. For every intermediate $a \in [\underline{A}_j, \overline{A}_j]$,

$$\begin{aligned} \partial_A G_j(a, \mathbf{p}, \beta_0 \mid \mathcal{W}) &= -\sum_{\ell=1}^{N} S_0'\left(\frac{p_\ell \tau_{\ell j}(\theta_0)}{a}\right) \frac{p_\ell \tau_{\ell j}(\theta_0)}{a^2} \\ &\geq \frac{\underline{s}_j}{\overline{A}_j^2} \sum_{\ell=1}^{N} p_\ell \tau_{\ell j}(\theta_0) \\ &\geq \frac{\underline{s}_j}{\overline{A}_j^2}, \end{aligned}$$

where the final inequality uses $\sum_\ell p_\ell = 1$ and $\tau_{\ell j}(\theta_0) \geq 1$. Combining this lower bound with (35) gives

$$\left|A_j(\mathbf{p}, \beta_\kappa \mid \mathcal{W}) - A_j(\mathbf{p}, \beta_0 \mid \mathcal{W})\right| \leq \frac{\overline{A}_j^2}{\underline{s}_j} \sum_{\ell=1}^{N} \sup_{z \in K_j} |f_\ell(z) - S_0(z)|.$$

Thus the valid constant is $C_j = \overline{A}_j^2 / \underline{s}_j$, which is finite and does not depend on κ, $\mathbf{p}$, or $\mathcal{W}$.

D.6 Rate of Convergence of the Deterministic Approximation Error

We show that the rate of convergence of the deterministic component $T_{2\kappa}$ given in Proposition 4.

Let us write $\mathbf{f} = (f_1, \ldots, f_N) \in \mathcal{S}_{\kappa,0}^*$ and abbreviate $A_j(\mathbf{p}, \mathbf{f}) = A_j(\mathbf{p}, \mathbf{f}, \theta_0 \mid \mathcal{W})$. We also put $\tau_{ij0} = \tau_{ij}(\theta_0)$ for all i, j. First note that we have, uniformly over $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$ conditional on the

fixed $\mathcal{W}$,

$$\begin{aligned}
\left|f_i\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{f})}\right) - S_{0i}\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{S}_0)}\right)\right| &\leq \left|f_i\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{f})}\right) - S_{0i}\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{f})}\right)\right| \\
&\quad + \left|S_{0i}\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{f})}\right) - S_{0i}\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{S}_0)}\right)\right|, \\
&\leq \sup_{z\in K_j}|f_i(z) - S_{0i}(z)| \\
&\quad + \left\{\sup_{z\in K_j}|\dot{S}_{0i}(z)|\right\} p_i\tau_{ij0}\left|\frac{1}{A_j(\mathbf{p},\mathbf{f})} - \frac{1}{A_j(\mathbf{p},\mathbf{S}_0)}\right|, \\
&\leq \sup_{z\in K_j}|f_i(z) - S_{0i}(z)| \\
&\quad + \left\{\sup_{z\in K_j}|\dot{S}_{0i}(z)|\right\} \frac{\bar{\tau}}{\underline{A_j^2}}\left|A_j(\mathbf{p},\mathbf{f}) - A_j(\mathbf{p},\mathbf{S}_0)\right|
\end{aligned}$$

for all i, j. Note that the second inequality follows from a mean value expansion, and the third inequality is obtained by using $\bar{\tau} = \max_{i,j}\sup_{\theta\in\Theta^C}\tau_{ij}(\theta)$ and $A_j(\mathbf{p},\beta \mid \mathcal{W}) \in [\underline{A_j}, \bar{A}_j]$ uniformly over $(\mathbf{p},\beta,\mathcal{W})$.

By Lipschitz continuity of the single aggregator locally around β_0 established in Appendix D.5, we have:

$$\left|A_j(\mathbf{p},\mathbf{f}) - A_j(\mathbf{p},\mathbf{S}_0)\right| \leq c_j \cdot d_{\mathcal{S},j}(\mathbf{f},\mathbf{S}_0),$$

where

$$c_j \equiv \bar{A}_j^2/\underline{s}_j > 0,$$

and

$$d_{\mathcal{S},j}(\mathbf{f},\mathbf{g}) \equiv \sum_{\ell=1}^{N}\sup_{z\in K_j}|f_\ell(z) - g_\ell(z)|,$$

is a pseudo metric defined on $\mathcal{B}$.

Hence, we have

$$\left|f_i\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{f})}\right) - S_{0i}\left(\frac{p_i\tau_{ij0}}{A_j(\mathbf{p},\mathbf{S}_0)}\right)\right| \lesssim \sup_{z\in K_j}|f_i(z) - S_{0i}(z)| + \sum_{\ell=1}^{N}\sup_{z\in K_j}|f_\ell(z) - S_{0\ell}(z)|$$

for all i, j, uniformly over $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$.

Since the final bound does not depend on $(\mathbf{p},\mathcal{W})$, we obtain:

$$\inf_{\mathbf{f}\in\mathcal{S}^*_{\kappa,0}} \mathbb{E}\left[\sum_{i,j}\left|f_i\left(\frac{P_{0i}\tau_{ij0}}{A_j(\mathbf{P}_0,\mathbf{f})}\right) - S_{0i}\left(\frac{P_{0i}\tau_{ij0}}{A_j(\mathbf{P}_0,\mathbf{S}_0)}\right)\right|^2\right] \lesssim \inf_{\mathbf{f}\in\mathcal{S}^*_{\kappa,0}}\left(\sum_{j=1}^{N} d_{\mathcal{S},j}(\mathbf{f},\mathbf{S}_0)\right)^2,$$

Let $[a,b] \subset \mathbb{R}_{++}$ be a compact interval whose interior contains K, and put $X = [1/b, 1/a]$. Define

$$H_0(x) = S_0(1/x).$$

The function H_0 is positive and strictly increasing on X. Since S_0 is continuously differentiable

with a strictly negative derivative on a compact neighborhood of K,

$$\inf_{x\in X} H_0(x) > 0, \qquad \inf_{x\in X} H_0'(x) > 0.$$

When S_0 admits $\gamma \geq 2$ continuous derivatives, simultaneous Jackson approximation gives a polynomial P_κ satisfying

$$\sup_{x\in X} |P_\kappa(x) - H_0(x)| = O(\kappa^{-\gamma})$$

and

$$\sup_{x\in X} |P_\kappa'(x) - H_0'(x)| = O(\kappa^{-(\gamma-1)}).$$

Consequently, P_κ is positive and strictly increasing for all sufficiently large κ. Therefore,

$$f_\kappa^u(z) = P_\kappa(1/z)$$

is positive and strictly decreasing on $[a, b]$, and

$$\sup_{z\in K} |f_\kappa^u(z) - S_0(z)| = O(\kappa^{-\gamma}).$$

To impose the aggregator normalization, let

$$A_\kappa^u = A_N(\bar{\mathbf{p}}, \theta_0, (f_\kappa^u, \dots, f_\kappa^u)).$$

The corrected local Lipschitz bound in Appendix D.5 gives

$$|A_\kappa^u - 1| \lesssim \sup_{z\in K} |f_\kappa^u(z) - S_0(z)| = O(\kappa^{-\gamma}).$$

Define the normalized approximant by

$$f_\kappa(z) = f_\kappa^u(z/A_\kappa^u).$$

The scale property gives the required aggregator normalization for $(f_\kappa, \dots, f_\kappa)$. Writing $P_\kappa(x) = \sum_{q=0}^{\kappa} a_{q,\kappa} x^q$ gives

$$f_\kappa(z) = \sum_{q=0}^{\kappa} a_{q,\kappa} (A_\kappa^u)^q z^{-q}.$$

Thus, the normalized approximant remains in the inverse-polynomial linear span. Define its coefficient norm by

$$c_\kappa^J = \sum_{q=0}^{\kappa} \left| a_{q,\kappa} (A_\kappa^u)^q \right|.$$

For every $z \in K$,

$$\begin{aligned} |f_\kappa(z) - S_0(z)| \leq & \quad |f^u_\kappa(z/A^u_\kappa) - S_0(z/A^u_\kappa)| \\ + & \quad |S_0(z/A^u_\kappa) - S_0(z)| \\ = O(\kappa^{-\gamma}), & \end{aligned}$$

where the final equality uses $A^u_\kappa = 1 + O(\kappa^{-\gamma})$ and the bounded derivative of S_0 on a compact neighborhood of K. The simultaneous derivative approximation also implies

$$\inf_{z \in K} f_\kappa(z) > \rho^L_R, \qquad \inf_{z \in K}\{-f'_\kappa(z)\} > \rho^L_R$$

for all sufficiently large R. The condition $\rho^U_R \geq c^J_{\kappa_R}$ eventually means that ρ^U_R dominates the coefficient norm. Together with the two preceding inequalities, this is the approximation-compatibility requirement in Assumption 12. It places the normalized approximant in the sieve. Substituting this approximant into the preceding display gives $T_{2\kappa} = O(\kappa^{-2\gamma})$.

D.7 Proof of Proposition 5

Under Assumption 13, the empirical-process argument in Appendix D.4 establishes that

$$T_{1\kappa} = O_p\left(\kappa\sqrt{\frac{\log(\kappa\rho^U_R/\rho^L_R)}{R}}\right).$$

The approximation-error argument in Appendix D.6 uses simultaneous Jackson approximation and the local continuity of the single aggregator to give

$$T_{2\kappa} = O(\kappa^{-2\gamma}).$$

The result now follows directly from Proposition 4.

E Proofs for Specification Testing

E.1 Linear Representation of the Sieve Estimator around the Pseudo Truth

The proof is broken down into two steps. We first establish the linear representation for a fixed sieve dimension. We then verify that the remainder remains negligible along the sequence $\kappa = \kappa_R$. The fixed-dimensional argument follows from the uniform law of large numbers and a mean-value expansion of the first order condition. For the increasing-dimensional argument, Lemma 3.10 of Geer (2006) is applied after verifying that the relevant class is Lipschitz. Section 2.7 of Vaart and Wellner (1996) then gives the entropy bound. The tuning restrictions ensure that the resulting remainder is negligible. The details follow.

We provide a linear representation of our Sieve estimator β_κ. Let us first define the pseudo true value β_κ^* as

$$\beta_\kappa^* \in \arg\min_{\beta \in \mathcal{B}_\kappa^*} Q(\beta)$$

which exists because the restrictions imposed on the sieve space $\mathcal{B}_\kappa^*$ ensure compactness of the sieve coefficients and continuity of the relevant mappings.

We use the reduced coordinate space Ξ_κ defined in Section 4. Its elements are

$$\xi_\kappa = (\theta', \alpha_1, \ldots, \alpha_\kappa)'$$

, while $\alpha_0(\xi_\kappa)$ is determined by the normalization. The common function is $f_\kappa(\cdot; \xi_\kappa)$ and the coordinate dimension is $d + \kappa$. Let us write $\xi_\kappa, \xi_\kappa^* \in \Xi_\kappa$ which corresponds to the minimizer of the sample objective function $\hat{\beta}_\kappa$ and the pseudo true value β_κ^*.

First note that we have, for all $\beta_\kappa \in \mathcal{B}_\kappa^*$,

$$\begin{aligned} Q(\beta_\kappa) &= Q(\beta_0) + \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left|f_i\left(z_{ij,(r)}(\beta_\kappa)\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right|^2\right], \\ &= Q(\beta_0) + M^2(\beta_\kappa, \beta_0), \end{aligned}$$

where in line 1, we defined

$$z_{ij,(r)}(\beta) \equiv P_{0i,(r)}\tau_{ij}(\theta) / A_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W})$$

, and used the fact that (1)

$$\mathbb{E}\left(\varepsilon_{ij,(r)} \mid \mathbf{P}_{0,(r)}\right) = 0$$

, and (2) $0 < f_i(z_{ij,(r)}(\beta)) < 1$ almost surely uniformly over $\beta \in \mathcal{B}$. Line 2 follows from the definition of our metric $M(\cdot,\cdot)$.

Since β_κ^* minimizes the population loss $Q(\beta)$ over the normalized symmetric sieve space $\mathcal{B}_\kappa^*$, it follows that

$$\begin{aligned} Q(\beta_\kappa^*) &= \min_{\beta \in \mathcal{B}_\kappa^*} \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left|f_i\left(z_{ij,(r)}(\beta)\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right|^2\right], \\ &\leq \min_{\mathbf{f} \in \mathcal{S}_{\kappa,0}^*} \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j}\left|f_i\left(z_{ij,(r)}(\theta_0, \mathbf{f})\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right|^2\right], \end{aligned}$$

where we fixed the finite-dimensional component to the true value θ_0.

We have, almost surely, the following bound for any $\mathbf{f} \in \mathcal{S}_\kappa$:

$$\begin{aligned}
\left| f_i\left(z_{ij,(r)}(\theta_0, \mathbf{f})\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right| &\leq \left| f_i\left(z_{ij,(r)}(\theta_0, \mathbf{f})\right) - S_{0i}\left(z_{ij,(r)}(\theta_0, \mathbf{f})\right)\right| \\
&\quad + \left| S_{0i}\left(z_{ij,(r)}(\theta_0, \mathbf{f})\right) - S_{0i}\left(z_{ij,(r)}(\beta_0)\right)\right|, \\
&\leq \sup_{z \in K_j} |f_i(z) - S_{0i}(z)| \\
&\quad + \left\{ \sup_{z \in K_j} \left|\dot{S_{0i}}(z)\right| \right\} \left| z_{ij,(r)}(\theta_0, \mathbf{f}) - z_{ij,(r)}(\beta_0)\right|, \\
&\leq \sup_{z \in K_j} |f_i(z) - S_{0i}(z)| \\
&\quad + \left\{ \sup_{z \in K_j} \left|\dot{S_{0i}}(z)\right| \right\} C_j \sum_{\ell=1}^{N} \sup_{z \in K_j} |f_i(z) - S_{0i}(z)|,
\end{aligned}$$

where in the third line we used the Lipschitz continuity of the single aggregator obtained in Appendix D.5 with a bounded Lipschitz constant $C_j > 0$.

Hence, by Jackson's Theorem for polynomials, we obtain the deterministic approximation rate given by:

$$|Q(\beta_\kappa^*) - Q(\beta_0)| = M^2(\beta_\kappa^*, \beta_0) \lesssim O(\kappa^{-2\gamma}),$$

which depends on the number of continuous derivatives (γ) the true share functions $S_{01}, \ldots, S_{0N}$ possesses.

Since we know that $M^2(\hat{\beta_\kappa}, \beta_0) = o_p(1)$ by Proposition 5, the triangle inequality and $(a+b)^2 \leq 2a^2 + 2b^2$ give

$$M^2(\hat{\beta_\kappa}, \beta_\kappa^*) \leq 2M^2(\hat{\beta_\kappa}, \beta_0) + 2M^2(\beta_0, \beta_\kappa^*) = o_p(1)$$

. Thus, the Sieve estimator is a consistent estimator for the pseudo true value with respect to $M(\cdot, \cdot)$. Since our restriction on the Sieve space ensures that (Ξ_κ, d_e) is a compact metric space, any Cauchy sequence is convergent in Ξ_κ under $d_e(\cdot, \cdot)$. Thus, completeness of (Ξ_κ, d_e) implies that for any sequence $\{\xi_{(n)}\}_{n \in \mathbb{N}_+} \subset \Xi_\kappa$ such that

$$M^2\left(\beta(\xi_{(n)}), \beta(\xi_\kappa^*)\right) \to 0$$

, we have $d_e(\xi_{(n)}, \xi_\kappa^*) \to 0$. This validates the Taylor expansions used in the latter part of the proof.

Now let us turn our attention to the sample minimizer $\hat{\beta_\kappa}$. Let us use $\hat{\xi_\kappa}, \xi_\kappa^*$ instead of $\hat{\beta_\kappa}$ and β_κ^*. The sample minimizer $\hat{\xi_\kappa}$ sets the FOC to zero, implying:

$$\begin{aligned}
0 &= \frac{1}{R} \sum_{r=1}^{R} \nabla_\xi Q_{(r)}(\hat{\xi_\kappa}), \\
&= \frac{1}{R} \sum_{r=1}^{R} \nabla_\xi Q_{(r)}(\xi_\kappa^*) + \left[\frac{1}{R} \sum_{r=1}^{R} \nabla_{\xi'} \nabla_\xi Q_{(r)}(\tilde{\xi_\kappa}) \right] (\hat{\xi_\kappa} - \xi_\kappa^*),
\end{aligned}$$

where the second line used the mean value expansion around $\xi = \xi_\kappa^*$ assuming that $\xi_\kappa^* \in \text{int}(\Xi_\kappa)$.

Assuming invertibility of the Hessian, we obtain:

$$\hat{\xi}_\kappa - \xi_\kappa^* = \left[\frac{1}{R}\sum_{r=1}^{R} \nabla_{\xi'}\nabla_\xi Q_{(r)}(\tilde{\xi}_\kappa)\right]^{-1} \left[\frac{-1}{R}\sum_{r=1}^{R} \nabla_\xi Q_{(r)}(\xi_\kappa^*)\right].$$

Define the observation-level residual at the pseudo truth by

$$u_{ij,(r)}^* \equiv \lambda_{ij,(r)} - f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*).$$

The score satisfies

$$-\frac{1}{2}\nabla_\xi Q_{(r)}(\xi_\kappa^*) = \psi_{\kappa,(r)} = \frac{1}{N^2}\sum_{i,j} u_{ij,(r)}^* \nabla_\xi f_\kappa(z_{ij,(r)}(\xi_\kappa^*); \xi_\kappa^*).$$

The residual contains both the measurement error and the discrepancy between the true function and its sieve pseudo truth. Since ξ_κ^* is an interior minimizer of the population objective,

$$\mathbb{E}[\psi_{\kappa,(r)}] = 0.$$

This first-order condition is why the sieve approximation component is part of the centered observation-level score. It is not a deterministic bias that is repeated inside the sum over observations.

We will next show that we have ULLN of the Hessian. We first validate the application of ULLN at each $\kappa \geq 1$ by checking the sufficient conditions to invoke Lemma 3.10 in van de Geer (2006), and then show that the entropy condition is met for moving sequence $\kappa = \kappa_R \to \infty$. This is verified in Appendix E.1.2.

We have

$$\sup_{\xi_\kappa \in \Xi_\kappa} \left\| \frac{1}{R}\sum_{r=1}^{R} \nabla_{\xi'}\nabla_\xi Q_{(r)}(\xi_\kappa) - \mathbb{E}\left[\nabla_{\xi'}\nabla_\xi Q_{(r)}(\xi_\kappa)\right] \right\| = o_p(1).$$

Together with consistency and the eigenvalue condition, this gives

$$\sqrt{R}\left(\hat{\xi}_\kappa - \xi_\kappa^*\right) = H_\kappa^{-1} \frac{1}{\sqrt{R}}\sum_{r=1}^{R} \psi_{\kappa,(r)} + o_p(1).$$

Furthermore, inspecting the Hessian component, we have

$$H_\kappa = \mathbb{E}\left[H_{2,\kappa}(\xi_\kappa^*, \mathbf{P}_{0,(r)}) + H_{3,\kappa}(\xi_\kappa^*, \mathbf{P}_{0,(r)})\right],$$

where

$$H_{2,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) \equiv -\frac{1}{N^2}\sum_{i,j}\left(S_0\left(z_{ij,(r)}(\beta_0)\right) - f_\kappa\left(z_{ij,(r)}(\xi_\kappa);\xi_\kappa\right)\right)\nabla_{\xi'}\nabla_\xi f_\kappa\left(z_{ij,(r)}(\xi_\kappa);\xi_\kappa\right),$$

$$H_{3,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) \equiv \frac{1}{N^2}\sum_{i,j}\nabla_{\xi'} f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\nabla_\xi f_i\left(z_{ij,(r)}(\xi_\kappa)\right).$$

The measurement-error component of the residual-weighted second derivative has conditional expectation zero. The approximation component $H_{2,\kappa}$ generally remains and is therefore included in H_κ.

E.1.1 Regularity Conditions of the Gradient Component

We show that the regularity condition such as continuity, boundedness of $\nabla_\xi f_i^*\left(z_{ij,(r)}(\xi_\kappa^*)\right)$. Let us take any $\xi_\kappa \in \text{int}(\Xi_\kappa)$, and put $f_i(z) = f_\kappa(z;\xi_\kappa)$ for every i. This is the common sieve approximation of the symmetric share function under ξ_κ.

We have

$$\nabla_\xi f_i\left(z_{ij,(r)}(\xi_\kappa)\right) = \left(\nabla_{\theta'} f_i\left(z_{ij,(r)}(\xi_\kappa)\right), \nabla_{(\xi\setminus\theta)'} f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\right)'$$

where we used the notation $\xi \setminus \theta$ to denote the Sieve coefficients of ξ, that is, the components that excludes the finite-dimensional parameters θ.

Inspecting the components $\nabla_\theta f_i(z)$ and

$$\nabla_{(\xi\setminus\theta)} f_i\left(z_{ij,(r)}(\xi_\kappa)\right)$$

, we have

$$\nabla_\theta f_i\left(z_{ij,(r)}(\xi_\kappa)\right) = \dot{f}_i(z_{ij,(r)}(\xi_\kappa))P_{0i,(r)}\nabla_\theta\frac{\tau_{ij}(\theta_\kappa)}{A_j\left(\mathbf{P}_{0,(r)}, \xi_\kappa \mid \mathcal{W}\right)} + \partial_\theta f_i\left(z_{ij,(r)}(\xi_\kappa)\right), \quad (36)$$

$$\nabla_{(\xi\setminus\theta)} f_i\left(z_{ij,(r)}(\xi_\kappa)\right) = \dot{f}_i(z_{ij,(r)})P_{0i,(r)}\tau_{ij}(\theta_\kappa)\nabla_{(\xi\setminus\theta)}\frac{1}{A_j\left(\mathbf{P}_{0,(r)}, \xi_\kappa \mid \mathcal{W}\right)} + \partial_{(\xi\setminus\theta)} f_i\left(z_{ij,(r)}\right) \quad (37)$$

where we used the chain rule to break the gradient into components involving differentiation with respect to $\partial f/\partial z = \dot{f}_i(z)\nabla z$ and the differentiation of the function f_i itself with respect to the Sieve parameters

$$(\partial_{\theta'}, \partial_{\xi'})' f_i$$

. The normalization makes $\alpha_0(\xi_\kappa)$ depend on both θ and $(\alpha_1, \ldots, \alpha_\kappa)$. Consequently,

$$\partial_\theta f_i(z;\xi_\kappa) = \frac{1}{N}\sum_{q=1}^{\kappa}\alpha_q q\sum_{m=1}^{N}[\bar{p}_m\tau_{mN}(\theta)]^{-q}W_{mN},$$

$$\partial_{\alpha_q} f_i(z;\xi_\kappa) = z^{-q} - \frac{1}{N}\sum_{m=1}^{N}[\bar{p}_m\tau_{mN}(\theta)]^{-q}$$

for $q = 1, \ldots, \kappa$. These expressions are common across i. Since z belongs to K_j, the normalization points are bounded away from zero, and the coefficient norm is bounded by ρ_R^U, these partial derivatives admit finite bounds uniformly over Ξ_κ for each κ.

We know that $z_{ij,(r)} \in K_j$ almost surely, and $z \mapsto \dot{f}_i(z)$ is at least twice continuously differentiable in a superset of K_j, and the parametric component $\theta \mapsto \tau_{ij}(\theta)$ is continuous. To establish that the derivative $\nabla_\xi f_i(z_{ij,(r)}(\xi))$ is continuous, we argue first that the components $\dot{f}_i(z_{ij,(r)})P_{0i,(r)}$ and $\dot{f}_i(z_{ij,(r)})P_{0i,(r)}\tau_{ij}(\theta_\kappa)$ are continuous almost surely by establishing that $\xi \mapsto z_{ij,(r)}(\xi)$ is continuous. Then, we argue that the gradient $\nabla_\xi(1/A_j(\mathbf{p}, \xi \mid \mathcal{W})$ is also continuous μ almost everywhere $\mathbf{p}$. Then the conclusion follows from preservation of continuity.

Let us fix $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. Let us take any sequence $\{\xi_{(n)}\}_{n\in\mathbb{N}_+} \subset \Xi_\kappa$ such that $\xi_{(n)} \to \xi_{(*)} \in \Xi_\kappa$. Let us define

$$z_{ij}(\mathbf{p}, \xi) = p_i\tau_{ij}(\theta)/A_j(\mathbf{p}, \xi \mid \mathcal{W})$$

. We show that we have

$$z_{ij}(\mathbf{p}, \xi_{(n)}) \to z_{ij}(\mathbf{p}, \xi_{(*)})$$

. Since the numerator of $z_{ij}(\mathbf{p}, \xi)$ is $p_i\tau_{ij}(\theta)$ which is clearly continuous, the preservation of continuity of ratios implies that we need only establish continuity of the implicit function

$$A_j(\mathbf{p}, \xi_{(n)}) \to A_j(\mathbf{p}, \xi_{(*)})$$

. Since Ξ_κ is sequentially compact by construction, it follows that there exists a convergent subsequence $\xi_{(m_n)} \to \xi_{(m_*)} \in \Xi_\kappa$. Since the original sequence converges to $\xi_{(*)}$ by hypothesis, it follows that $\xi_{(m_*)} = \xi_{(*)}$. We know that the single aggregator exists uniquely for all $(\mathbf{p}, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$ conditional on the fixed $\mathcal{W}$. Thus, we have:

$$\begin{aligned} G_j(a_{m_n}, \mathbf{p}, \xi_{(m_n)} \mid \mathcal{W}) &\overset{(\text{def})}{=} \sum_{\ell=1}^{N} f_{\ell,(m_n)}\left(\frac{p_\ell\tau_{\ell j}(\theta_{(m_n)})}{a_{m_n}}\right) - 1 \\ &\to \sum_{\ell=1}^{N} f_{\ell,(m_*)}\left(\frac{p_\ell\tau_{\ell j}(\theta_{(m_*)})}{a_{m_*}}\right) - 1, \\ &\overset{(\text{def})}{=} G_j(a_{(m_*)}, \mathbf{p}, \xi_{(m_*)} \mid \mathcal{W}), \end{aligned}$$

where the $\to$ in the second line follows from continuity of the composite map $\xi \mapsto (f_\ell \circ z_{ij})(\mathbf{p}, \xi)$. Uniqueness of the solution to $G_j(a_{(m_*)}, \mathbf{p}, \xi_{(m_*)} \mid \mathcal{W}) = 0$ implies that we have $a_{(m_*)} = a_{(*)}$ implying

$$A_j(\mathbf{p}, \xi_{(m_*)} \mid \mathcal{W}) = A_j(\mathbf{p}, \xi_{(*)} \mid \mathcal{W})$$

. Therefore, we have

$$A_j(\mathbf{p}, \xi_{(n)} \mid \mathcal{W}) \to A_j(\mathbf{p}, \xi_{(*)} \mid \mathcal{W})$$

as desired. Clearly, this holds for any $\mathbf{p}, \mathcal{W}$.

What remains to be shown is the continuity of the gradient $\nabla_\xi(1/A_j(\mathbf{p}, \xi \mid \mathcal{W}))$. Since the

single aggregator exists for all $(\mathbf{p}, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$ conditional on the fixed $\mathcal{W}$, and $\partial_A G_j(A, \mathbf{p}, \xi \mid \mathcal{W}) \neq 0$ for all $(\mathbf{p}, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$, the Implicit Function Theorem tells us that we have:

$$\nabla_\xi A_j(\mathbf{p}, \xi \mid \mathcal{W}) = \left[\partial_A G_j\left(A_j(\mathbf{p}, \xi \mid \mathcal{W}), \mathbf{p}, \xi\right)\right]^{-1} \left[-\partial_\xi G_j\left(A_j(\mathbf{p}, \xi \mid \mathcal{W}), \mathbf{p}, \xi\right)\right] \tag{38}$$

for any $(\mathbf{p}, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$.

Since we have

$$\nabla_\xi \frac{1}{A_j(\mathbf{p}, \xi \mid \mathcal{W})} = -\frac{1}{A_j^2(\mathbf{p}, \xi \mid \mathcal{W})} \nabla_\xi A_j(\mathbf{p}, \xi \mid \mathcal{W}), \tag{39}$$

continuity of $\xi \mapsto \nabla_\xi(1/A_j(\mathbf{p}, \xi \mid \mathcal{W}))$ follows from continuity of $\nabla_\xi A_j(\mathbf{p}, \xi \mid \mathcal{W})$. Clearly, by our choice of polynomials, both $\partial_A G_j(\cdot)$ and $\partial_\xi G_j(\cdot)$ are continuous. Thus, we conclude that $\xi \mapsto \nabla_\xi f_i\left(z_{ij}(\mathbf{p}, \xi)\right)$ is continuous for μ almost all $\mathbf{p}$.

Next we show that there exists an envelope

$$F_{ij,\kappa}(\mathbf{P}_{0,(r)}) = \sup_{\xi \in \Xi_\kappa} \|\nabla_\xi f_i\left(z_{ij,(r)}(\xi)\right)\|_1$$

such that we have $\mathbb{E}\left[F_{ij,\kappa}(\mathbf{P}_{0,(r)})\right] < \infty$. Inspecting both (36) and (37), compactness of $\mathbb{S}(\mathbf{P}_0)$ and Θ, finiteness of the fixed design, and continuity of $\theta \mapsto \tau_{ij}(\theta)$ imply that the component $P_{0,(r)}\tau_{ij}(\theta) \leq \bar{\tau}$ almost surely. Thus the only relevant components for envelope calculation is $\dot{f}_i\left(z_{ij,(r)}(\xi)\right)$, and (39).

The first derivative is of order:

$$\begin{aligned}
\dot{f}_i\left(z_{ij,(r)}(\xi)\right) &\leq \sup_{z \in K_j} |\dot{f}_i(z)|, \\
&\lesssim \rho_R^U \kappa^2 \max\{a_j^{-\kappa}, a_j^{-2}\},
\end{aligned}$$

where $a_j = \min_{z \in K_j} z$ and the final bound depends whether $a_j \leq 1$ or $a_j > 1$.

We inspect the two components in (38). We have:

$$\begin{aligned}
\partial_A G_j\left(A_j(\mathbf{p}, \xi \mid \mathcal{W}), \mathbf{p}, \xi\right) &= \sum_{\ell=1}^N \left\{-\dot{f}_\ell(z_{\ell j}(\mathbf{p}, \xi))\right\} \frac{p_\ell \tau_{\ell j}(\theta)}{A_j^2(\mathbf{p}, \xi \mid \mathcal{W})}, \\
&\geq \frac{p_\varepsilon}{\left(\bar{A}_j\right)^2} \sum_{\ell=1}^N \inf_{z \in K_j} \left\{-\dot{f}_\ell(z)\right\}, \\
&\gtrsim \rho_R^L \sum_{\ell=1}^N \min\{b_j^{-2}, b_j^{-\kappa-1}\}, \\
&> \rho_R^L b_j^*,
\end{aligned}$$

where in line 2 we used compactness of $\mathbb{S}(\mathbf{P}_0)$ which ensures the existence of $p_\varepsilon > 0$,

$$A_j(\mathbf{p}, \xi \mid \mathcal{W}) \in [\underline{A_j}, \bar{A}_j]$$

for some $0 < \underline{A_j} < \bar{A}_j < \infty$, and in line 3 we used

$$\begin{aligned}
\inf_{z \in K_j} -\dot{f}_\ell(z) &= \inf_{z \in K_j} \sum_{q=1}^{\kappa} q \alpha_q^{(\ell)} z^{-q-1}, \\
&\geq \rho_R^L \inf_{z \in K_j} \sum_{q=1}^{\kappa} q z^{-q-1}, \\
&\geq \rho_R^L \min\{b_j^{-2}, b_j^{-\kappa-1}\},
\end{aligned}$$

and we defined $b_j^* = \min_\ell \min\{b_j^{-2}, b_j^{-\kappa-1}\} > 0$.

Also, we have

$$\begin{aligned}
\partial_\xi G_j \left(A_j(\mathbf{p}, \xi \mid \mathcal{W}), \mathbf{p}, \xi \right) &= \sum_{\ell=1}^{N} \partial_\xi f_\ell \left(z_{\ell j}(\mathbf{p}, \xi) \right), \\
&= \sum_{\ell=1}^{N} \dot{f}_\ell \left(z_{\ell j}(\mathbf{p}, \xi) \right) \frac{p_\ell}{A_j(\mathbf{p}, \xi \mid \mathcal{W})} \nabla_\theta \tau_{\ell j}(\theta), \\
&\leq \frac{\bar{\tau} \bar{W}}{\underline{A_j}} \sum_{\ell=1}^{N} \sup_{z \in K_j} |\dot{f}_\ell(z)|, \\
&\lesssim \rho_R^U N \kappa^2 \max_{j=1,\ldots,N} \max\{a_j^{-\kappa}, a_j^{-2}\},
\end{aligned}$$

where in the third line we used the fact that the fixed design is finite, which ensures that there exists some $\bar{W} < \infty$ such that we have $\|W_{ij}\|_1 \leq \bar{W}$ for all i, j.

Let us define

$$a_j^* \equiv \max_{j=1,\ldots,N} \max\{a_j^{-\kappa}, a_j^{-2}\}$$

. Then, we have

$$\nabla_\xi \frac{1}{A_j(\mathbf{p}, \xi \mid \mathcal{W})} = O\left(\kappa^2 \frac{a^* \rho_R^U}{b^* \rho_R^L} \right).$$

Since our bounds are uniform across entries of the vector $\nabla_\xi f_i$ and $\mathbf{p}$, it follows that

$$F_{ij,\kappa}(\mathbf{P}_{0,(r)}) = O(\kappa) \times O\left(\rho_R^U \kappa^2 a_j^* \right) \times O\left(\frac{\rho_R^U}{\rho_R^L} \kappa^2 \frac{a_j^*}{b_j^*} \right) + O(\kappa) \times O(\rho_R^U a_j^*) < \infty,$$

where the envelope does not depend on the random variable $\mathbf{P}_{0,(r)}$, nor ξ.

By taking a maximum $\bar{F}_\kappa \equiv \max_{i,j} F_{ij,\kappa}(\mathbf{P}_{0,(r)})$, we have

$$\bar{F}_\kappa = O\left(\kappa^5 \frac{\left(\rho_R^U a_j^* \right)^2}{\rho_R^L b_j^*} \right),$$

which is finite for each $\kappa \geq 1$.

E.1.2 Regularity Conditions of the Hessian Component

We verify the sufficient conditions to invoke ULLN at each $\kappa \geq 1$ recorded as Lemma 3.10 in van de Geer (2006). Let us first inspect the structure of the Hessian. We have

$$\frac{1}{2R}\sum_{r=1}^{R}\nabla_{\xi'}\nabla_{\xi}Q_{(r)}(\xi_\kappa) = \frac{1}{R}\sum_{r=1}^{R}\left[H_{1,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) + H_{2,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) + H_{3,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right)\right],$$

where

$$\begin{aligned}
H_{1,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) &\equiv \frac{1}{N^2}\sum_{i,j}\varepsilon_{ij,(r)}\left[-\nabla_{\xi'}\nabla_{\xi}f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\right], \\
H_{2,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) &\equiv \frac{1}{N^2}\sum_{i,j}\left(S_{0i}\left(z_{ij,(r)}(\beta_0)\right) - f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\right)\left[-\nabla_{\xi'}\nabla_{\xi}f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\right], \\
H_{3,\kappa}\left(\xi_\kappa, \mathbf{P}_{0,(r)}\right) &\equiv \frac{1}{N^2}\sum_{i,j}\nabla_{\xi'}f_i\left(z_{ij,(r)}(\xi_\kappa)\right)\nabla_{\xi}f_i\left(z_{ij,(r)}(\xi_\kappa)\right).
\end{aligned}$$

Since we have $0 < f_i^*(z_{ij,(r)}(\xi_\kappa)) < 1$ almost surely due to the HSA normalization, and bounded second moment $\sigma_\varepsilon^2 < \infty$ of $\varepsilon_{ij,(r)}$, the envelope condition can be verified upon showing that we have

$$\sup_{\xi\in\Xi_\kappa}\max_{i,j}\sup_{z\in K_j}\|\nabla_{\xi'}\nabla_{\xi}f_i(z)\|_2 < \infty$$

for each $\kappa \geq 1$ as this guarantees an L_1 integrable envelope for terms $H_{1,\kappa}(\cdot)$ and $H_{2,\kappa}(\cdot)$ by the Cauchy Schwartz inequality.

For the continuity condition, we only inspect the continuity of

$$\xi \mapsto \nabla_{\xi'}\nabla_{\xi}f_i\left(z_{ij}(\mathbf{p}, \xi)\right)$$

for μ almost all $\mathbf{p}$ as components $H_{1,\kappa}$ and $H_{2,\kappa}$ has a similar structure. We have:

$$
\begin{aligned}
&\nabla_{\theta'}\nabla_{\theta} f_i\left(z_{ij,(r)}(\xi_\kappa)\right) \\
&= \ddot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)}^2 \nabla_{\theta'}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right)\nabla_{\theta}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right) \\
&+ \dot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)} \nabla_{\theta'}\nabla_{\theta}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j\left(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W}\right)}\right), \\
&\nabla_{(\xi\setminus\theta)'}\nabla_{\theta} f_i\left(z_{ij,(r)}(\xi_\kappa)\right) \\
&= \ddot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)}^2 \tau_{ij}(\theta_\kappa)\nabla_{(\xi\setminus\theta)'}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right)\nabla_{\theta}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right) \\
&+ \dot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)} \nabla_{(\xi\setminus\theta)'}\nabla_{\theta}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j\left(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W}\right)}\right), \\
&\nabla_{(\xi\setminus\theta)'}\nabla_{(\xi\setminus\theta)} f_i\left(z_{ij,(r)}(\xi_\kappa)\right) \\
&= \ddot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)}^2 \tau_{ij}^2(\theta_\kappa)\nabla_{(\xi\setminus\theta)'}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right)\nabla_{(\xi\setminus\theta)}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right) \\
&+ \dot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)} \tau_{ij}(\theta_\kappa)\nabla_{(\xi\setminus\theta)'}\nabla_{(\xi\setminus\theta)}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right), \\
&\nabla_{\theta'}\nabla_{(\xi\setminus\theta)} f_i\left(z_{ij,(r)}(\xi_\kappa)\right) \\
&= \ddot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)}^2 \tau_{ij}(\theta_\kappa)\nabla_{\theta'}\left(\frac{\tau_{ij}(\theta_\kappa)}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right)\nabla_{(\xi\setminus\theta)}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right) \\
&+ \dot{f}_i\left(z_{ij,(r)}(\xi_\kappa)\right) P_{0i,(r)} \tau_{ij}(\theta_\kappa)\nabla_{\theta'}\nabla_{(\xi\setminus\theta)}\left(\frac{1}{A_j(\mathbf{P}_{0,(r)},\xi_\kappa \mid \mathcal{W})}\right).
\end{aligned}
$$

For every $\kappa \geq 1$, the inverse-polynomial sieve functions are twice continuously differentiable on a neighborhood of K_j for every $j = 1, \ldots, N$ because $K_j \subset \mathbb{R}_{++}$. The finite-dimensional component $\theta \mapsto \tau_{ij}(\theta)$ is an analytic function. We therefore need to verify the continuity of the Hessian of the single aggregator functional $\xi \mapsto A_j(\mathbf{p}, \xi \mid \mathcal{W})$. Note that both $(a, \xi) \mapsto \partial_A G_j(a, \mathbf{p}, \xi \mid \mathcal{W})$, and

$$(a, \xi) \mapsto \partial_{\xi} G_j(a, \mathbf{p}, \xi \mid \mathcal{W})$$

is at least twice continuously differentiable with respect to (a, ξ) in a neighborhood of $(A_j(\mathbf{p}, \xi \mid \mathcal{W}), \xi)$. The Implicit Function Theorem therefore applies for every $\kappa \geq 1$. Hence, for almost all $\mathbf{p}$, every entry of the component $H_{3,\kappa}(\xi, \mathbf{p})$ is continuous with respect to $\xi \in \operatorname{int}(\Xi_\kappa)$.

Next we verify the envelope condition. We need only inspect

$$\nabla_{\xi'}\nabla_{\xi}(1/A_j(\mathbf{p}, \xi \mid \mathcal{W}))$$

and the second derivative $\ddot{f}_i\left(z_{ij}(\mathbf{p}, \xi)\right)$, because $\tau_{ij}(\theta)$ is analytic given the exponential function

specification. From

The component $H_{3,\kappa}(\xi_\kappa, \mathbf{P}_{0,(r)})$ satisfied the sufficient conditions to invoke ULLN at each $\kappa \geq 1$. We already have continuity from Appendix E.1.1 and a bound for

$$\max_{i,j} \|\nabla_\xi f_i \left(z_{ij,(r)}(\xi) \right)\|_1 \leq \bar{F}_\kappa$$

for a constant $\bar{F}_\kappa < \infty$ almost surely. Since every element of the components in the summand of $H_{3,\kappa}(\xi_\kappa, \mathbf{P}_{0,(r)})$ is bounded by $(\bar{F}_\kappa)^2$, Lemma 3.10 of van de Geer (2006) applies and we have

$$R^{-1} \sum_{r=1}^{R} H_{3,\kappa}(\xi_\kappa, \mathbf{P}_{0,(r)}) \xrightarrow{p} \mathbb{E}\left[H_{3,\kappa}(\xi_\kappa, \mathbf{P}_{0,(r)}) \right]$$

uniformly over $\xi_\kappa \in \Xi_\kappa$.

To extend this to a moving $\kappa \to \infty$, first define

$$\mathcal{L}_\kappa^H \equiv \sup_{\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)} \sup_{\xi \in \Xi_\kappa} \|D_\xi \nabla_{\xi'} \nabla_\xi Q_{(r)}(\xi)\|_1,$$

where the operator $D_\xi = (\nabla_\xi, \nabla_\xi, \ldots, \nabla_\xi)$ is the differentiation operator constructed by stacking $(d + \kappa)$ gradient operators ∇_ξ. By twice continuous differentiability of $\xi \mapsto \nabla_\xi Q_{(r)}(\xi)$ which implies continuous differentiability of the Hessian $\xi \mapsto \nabla_{\xi'} \nabla_\xi Q_{(r)}(\xi)$, and compactness of $\mathbb{S}(\mathbf{P}_0)$, Θ and $[\underline{A_j}, \bar{A}_j]$ for all j, it follows that for any $\xi, \tilde{\xi} \in \Xi_\kappa$, we have

$$\|\nabla_{\xi'} \nabla_\xi Q_{(r)}(\xi) - \nabla_{\xi'} \nabla_\xi Q_{(r)}(\tilde{\xi})\|_1 \leq \mathcal{L}_\kappa^H \|\xi - \tilde{\xi}\|_1.$$

By a similar argument to Appendix D.2.1, we shall work with a rough bound on the Lipschitz constant $\mathcal{L}_\kappa^H$. In addition to

$$\dot{f}_\ell, \ddot{f}_\ell, \nabla_\xi (1/A_\ell(\cdot)), \nabla_\xi \nabla_\xi (1/A_\ell(\cdot))$$

, The relevant components are additional derivatives $\dddot{f}_\ell$ and

$$D_\xi \left(\nabla_{\xi'} \nabla_\xi A_\ell(\cdot) \right)$$

. Let $a_j \equiv \min_{z \in K_j} z$ and $b_j \equiv \max_{z \in K_j} z$ for every $j = 1, \ldots, N$ which exists because of compactness of $K_j \subset \mathbb{R}_{++}$. Let us further define

$$a_\kappa^* = \max_{j=1,\ldots,N} \max\{a_j^{-\kappa-1}, a_j^{-2}\}$$

and

$$b_\kappa^* = \min_{j=1,\ldots,N} \min\{b_j^{-2}, b_j^{-\kappa-1}\}$$

.We have

$$\dddot{f}_\ell(z) \leq \kappa^3 \rho_R^U a_\kappa^*,$$

thus we only inspect the third derivative of $1/A_j(\mathbf{p},\xi)$. From Appendix D.2.1, we have

$$\nabla_{\xi'}\nabla_\xi(1/A_j(\mathbf{p},\xi)) = \frac{1}{A_j^2(\mathbf{p},\xi)}\left(\nabla_{\xi'}A_j(\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right) - \frac{1}{A_j(\mathbf{p},\xi)}\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi),$$

which by taking an additional derivative yields:

$$\begin{aligned}
&D_\xi\nabla_{\xi'}\nabla_\xi(1/A_j(\mathbf{p},\xi)) \\
&= D_\xi\left\{\left(\nabla_{\xi'}A_j(\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right)\right\}\frac{1}{A_j(\mathbf{p},\xi)^2} \\
&- \left(\nabla_{\xi'}A_j(\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right)D_\xi\left(A_j(\mathbf{p},\xi)^2\right)\frac{1}{A_j(\mathbf{p},\xi)^4} \\
&- D_\xi\left\{\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi)\right\}\frac{1}{A_j(\mathbf{p},\xi)} + \left\{\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi)\right\}D_\xi A_j(\mathbf{p},\xi)\frac{1}{A_j(\mathbf{p},\xi)},
\end{aligned}$$

thus the only relevant component for the entropy calculation is

$$D_\xi\left(\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi)\right)$$

. Again, from Appendix D.2.1, we have

$$\begin{aligned}
&D_\xi\left(\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi)\right) \\
&= -\frac{1}{\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)}\left\{D_\xi\left(\nabla_{\xi'}\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\right\}\mathbf{1} \\
&+ \frac{1}{\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)^2}\left(\nabla_{\xi'}\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)D_\xi\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right) \\
&- \frac{1}{\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)}D_\xi\left\{\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right)\right\}\mathbf{1} \\
&+ \frac{1}{\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)^2}\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right)D_\xi\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right),
\end{aligned}$$

where $\mathbf{1}$ is a matrix of ones of dimension $(d+\kappa, d+\kappa)$. We have:

$$\left|\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right| \gtrsim \rho_R^L b_\kappa^*,$$

uniformly over indices j and

$$(\mathbf{p},\xi) \in \mathbb{S}(\mathbf{P}_{0,(r)}) \times \Xi_\kappa$$

. Hence, we only inspect the order of

$$D_\xi\left(\nabla_{\xi'}\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right), D_\xi\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)$$

and

$$D_{\xi}\left\{\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\left(\nabla_{\xi}A_j(\mathbf{p},\xi)\right)\right\}$$

.

We have

$$D_{\xi}\left(\nabla_{\xi'}\partial_{\xi}G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right) \lesssim O\left(\kappa^8\frac{\left(\rho_R^U a_\kappa^*\right)^4}{\left(\rho_R^L b_\kappa^*\right)^3}\right),$$

$$D_{\xi}\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right) \lesssim O\left(\kappa^4\frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}\right),$$

$$\begin{aligned}
&D_{\xi}\left\{\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\left(\nabla_{\xi}A_j(\mathbf{p},\xi)\right)\right\}\\
&\lesssim O\left(\nabla_{\xi}\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\right)O(\nabla_{\xi'}\nabla_{\xi}A_j(\mathbf{p},\xi))\\
&\lesssim O\left(\kappa^{15}\frac{\left(\rho_R^U a_\kappa^*\right)^7}{\left(\rho_R^L b_\kappa^*\right)^6}\right).
\end{aligned}$$

Therefore, the worst case bound is given by

$$\mathcal{L}_\kappa^H \lesssim O\left(\kappa^{c_1}(\rho_R^U a_\kappa^*)^{c_2}(\rho_R^L b_\kappa^*)^{-c_3}\right),$$

for some constant $c_1, c_2, c_3 > 0$ and bounded.

E.1.3 Propagation of the Elementwise Bound to the Hessian Matrix with Varying Dimension

The preceding entrywise entropy calculation is not by itself sufficient for convergence in matrix norm when κ increases. Let $p_\kappa = d + \kappa$. We use ∂_{ξ_a} to denote differentiation with respect to the ath component of ξ. Define the uniform entrywise envelope by

$$B_\kappa^H \equiv \sup_{\mathbf{p}\in\mathbb{S}(\mathbf{P}_0)}\sup_{\xi\in\Xi_\kappa}\max_{1\le a,b\le p_\kappa}\left|\frac{1}{2}\partial_{\xi_a}\partial_{\xi_b}Q_{(r)}(\xi)\right|.$$

The compact support of the errors and the bounds derived above imply $B_\kappa^H < \infty$ for each κ. We can also express this envelope using the Lipschitz constant already derived. Choose any fixed $\xi_\kappa^\circ \in \Xi_\kappa$ and define

$$B_{\kappa,0}^H \equiv \sup_{\substack{\mathbf{p}\in\mathbb{S}(\mathbf{P}_0)\\ \mathbf{e}\in\mathbb{S}(\mathcal{E})}}\max_{1\le a,b\le p_\kappa}\left|\frac{1}{2}\partial_{\xi_a}\partial_{\xi_b}Q(\xi_\kappa^\circ;\mathbf{p},\mathbf{e})\right|.$$

For notational convenience, let

$$H_{ab,\kappa,(r)}(\xi) \equiv \frac{1}{2}\partial_{\xi_a}\partial_{\xi_b}Q_{(r)}(\xi).$$

The Lipschitz property first gives, for every $\xi \in \Xi_\kappa$,

$$\left|H_{ab,\kappa,(r)}(\xi)\right| \leq \left|H_{ab,\kappa,(r)}(\xi_\kappa^\circ)\right| + \mathcal{L}_\kappa^H \|\xi - \xi_\kappa^\circ\|_1.$$

Taking the supremum over the compact support of an observation, the maximum over a, b, and the supremum over $\xi \in \Xi_\kappa$ gives

$$B_\kappa^H \leq B_{\kappa,0}^H + \mathcal{L}_\kappa^H \operatorname{diam}(\Xi_\kappa).$$

Let

$$D_\Theta \equiv \sup_{\theta,\widetilde{\theta} \in \Theta^C} \|\theta - \widetilde{\theta}\|_1 < \infty.$$

The coefficient restriction and compactness of Θ^C imply

$$\operatorname{diam}(\Xi_\kappa) \leq D_\Theta + 2\rho_R^U.$$

Consequently, we may take

$$B_\kappa^H \leq B_{\kappa,0}^H + \mathcal{L}_\kappa^H (D_\Theta + 2\rho_R^U).$$

The derivative bounds established above ensure that $B_{\kappa,0}^H$ is finite for each κ.

Let $\mathcal{H}_{ab,\kappa}$ be the class formed by the (a, b) entry of one half of the Hessian. Since the Hessian is symmetric, it is sufficient to consider $1 \leq a \leq b \leq p_\kappa$. The Lipschitz bound above and the Euclidean covering number of Ξ_κ give

$$\log N\left(\delta, \mathcal{H}_{ab,\kappa}, L_2(P)\right) \lesssim p_\kappa \log\left(1 + \frac{\rho_R^U \mathcal{L}_\kappa^H}{\delta}\right).$$

The preceding display gives the entropy bound for each element of the Hessian. The Hessian has dimension $(d + \kappa) \times (d + \kappa)$. By symmetry, it contains $p_\kappa(p_\kappa + 1)/2$ distinct elements. Taking the maximum over these elements adds at most

$$\log(p_\kappa^2) = 2 \log p_\kappa = 2 \log(d + \kappa)$$

to the entropy bound. Applying Bernstein's inequality recorded as Lemma 2.2.10 in Vaart and Wellner (1996) to a finite cover of each Hessian-element class, followed by a union bound over the distinct Hessian elements and the Lipschitz extension from the cover to Ξ_κ, gives

$$\max_{1 \leq a,b \leq p_\kappa} \sup_{\xi \in \Xi_\kappa} \left|(P_R - P)\frac{1}{2}\partial_{\xi_a}\partial_{\xi_b} Q_{(r)}(\xi)\right| = O_p\left(B_\kappa^H \left[\sqrt{\frac{v_{R,\kappa}}{R}} + \frac{v_{R,\kappa}}{R}\right]\right),$$

where

$$v_{R,\kappa} \equiv p_\kappa \log\left(1 + \frac{\rho_R^U \mathcal{L}_\kappa^H}{B_\kappa^H}\right) + 2 \log p_\kappa.$$

For any $p_\kappa \times p_\kappa$ matrix $C = (C_{ab})$, define

$$\|C\|_{\text{op}} \equiv \sup_{x \in \mathbb{R}^{p_\kappa},\, \|x\|_2 = 1} \|Cx\|_2.$$

This is the operator norm induced by the Euclidean norm. Since the Hessian is symmetric, this norm equals the largest absolute eigenvalue of C. Let $\max_{1 \le a,b \le p_\kappa} |C_{ab}|$ denote the largest absolute value among the elements of C. These quantities satisfy

$$\|C\|_{\text{op}} \le p_\kappa \max_{1 \le a,b \le p_\kappa} |C_{ab}|.$$

We use the operator norm because it directly controls the difference between the sample and population Hessian eigenvalues. It follows that the sufficient tuning condition

$$p_\kappa B_\kappa^H \left[\sqrt{\frac{v_{R,\kappa}}{R}} + \frac{v_{R,\kappa}}{R} \right] \longrightarrow 0 \tag{D.1}$$

implies

$$\sup_{\xi \in \Xi_{\kappa_R}} \left\| \frac{1}{2R} \sum_{r=1}^{R} \nabla_{\xi'} \nabla_{\xi} Q_{(r)}(\xi) - \frac{1}{2} \mathbb{E} \left[\nabla_{\xi'} \nabla_{\xi} Q_{(r)}(\xi) \right] \right\|_{\text{op}} = o_p(1).$$

The quantities B_κ^H and $\mathcal{L}_\kappa^H$ contain the dependence on ρ_R^U, ρ_R^L, a_κ^*, and b_κ^* displayed above. Condition (D.1) therefore records the additional restriction needed to pass from entrywise convergence to matrix-norm convergence. The original two entropy restrictions alone do not establish this conclusion.

We verify that Condition (D.1) holds for the tuning sequence stated before Proposition 6. Since K is a fixed compact subset of $\mathbb{R}_{++}$, there is a finite constant $C_K > 0$ such that

$$a_\kappa^* \le C_K \exp(C_K \kappa), \qquad (b_\kappa^*)^{-1} \le C_K \exp(C_K \kappa).$$

Under

$$\rho_R^U = \exp(C_U \kappa), \qquad \rho_R^L = \exp(-C_L \kappa),$$

the bound obtained above for the Hessian Lipschitz constant implies

$$\mathcal{L}_\kappa^H \le C \kappa^C \exp(C\kappa)$$

for a finite constant $C > 0$ that does not depend on R. The same derivative bounds imply

$$B_{\kappa,0}^H \le C \kappa^C \exp(C\kappa).$$

Using $B_\kappa^H \le B_{\kappa,0}^H + \mathcal{L}_\kappa^H (D_\Theta + 2\rho_R^U)$, we obtain

$$B_\kappa^H \le C \kappa^C \exp(C\kappa),$$

after increasing C if necessary. We may take the envelope to be at least one. It then follows that

$$v_{R,\kappa} \leq p_\kappa \log\left(1 + \rho_R^U \mathcal{L}_\kappa^H\right) + 2\log p_\kappa = O(\kappa^2).$$

Now take $\kappa_R = \lfloor \log\log R \rfloor$. The first component of Condition (D.1) satisfies

$$p_\kappa B_\kappa^H \sqrt{\frac{v_{R,\kappa}}{R}} \leq \frac{C\kappa^C \exp(C\kappa)}{\sqrt{R}} = \frac{C(\log\log R)^C (\log R)^C}{\sqrt{R}} \longrightarrow 0.$$

Similarly,

$$p_\kappa B_\kappa^H \frac{v_{R,\kappa}}{R} \leq \frac{C\kappa^C \exp(C\kappa)}{R} = \frac{C(\log\log R)^C (\log R)^C}{R} \longrightarrow 0.$$

Thus, the proposed tuning sequence satisfies Condition (D.1).

E.2 CES Null Hypothesis

Let us put the parameters under CES as $\beta_{CES} = (\theta_{CES}, \sigma_{CES})$ where we set the share function for all $\ell \in \{1, \ldots, N\}$ to $S_\ell(z) = z^{1-\sigma_{CES}}$. Let us put the parameters under HSA as $\beta_0 = (\theta_0, \mathbf{S}_0) \in \mathcal{B}$. Clearly, under our assumption of compact parameter space for σ_{CES}, we have $\beta_{CES} \in \mathcal{B}$.

From Proposition 1 (Assertion (ii)) in Matsuyama and Ushchev (2017), uniformly over

$$(\mathbf{p}_0, \beta) \in \mathbb{S}(\mathbf{P}_0) \times \mathcal{B}$$

, the ideal price index $P_j^I(\mathbf{p}_0, \beta \mid \mathcal{W})$ is related to the single aggregator $A_j(\mathbf{p}_0, \beta \mid \mathcal{W})$ by a deterministic relationship:

$$\ln P_j^I(\mathbf{p}_0, \beta \mid \mathcal{W}) = \ln A_j(\mathbf{p}_0, \beta \mid \mathcal{W}) + \sum_{\ell=1}^N \int_c^{p_\ell \tau_{lj}(\theta)/A_j(\mathbf{p}_0, \beta \mid \mathcal{W})} \frac{1}{z} S_\ell(z)\, \mathrm{d}z. \tag{40}$$

The constant $c > 0$ is fixed and known. It determines the lower limit of integration and only shifts the two sides by the corresponding normalization constant.

Also by Assertion (iii) of the same proposition in Matsuyama and Ushchev (2017), when $N \geq 3$, we have

$$\sum_{\ell=1}^N \int_c^{p_\ell \tau_{lj}(\theta)/A_j(\mathbf{p}_0, \beta \mid \mathcal{W})} \frac{1}{z} S_\ell(z)\, \mathrm{d}z = \Gamma(c, \sigma_{CES}), \tag{41}$$

where $\Gamma(c, \sigma_{CES}) = (1 - Nc^{1-\sigma_{CES}})/(1 - \sigma_{CES})$ is a constant that does not depend on the price vector $\mathbf{p}_0$ if and only if $S_\ell(z) = z^{1-\sigma_{CES}}$ for all ℓ.

Equations (40) and (41) provides a basis for testing the null hypothesis of CES against an HSA alternative.

Let us fix $c > 0$. Let us define

$$\Phi_j(\mathbf{p}, \beta \mid \mathcal{W}) \equiv \sum_{\ell=1}^N \int_c^{p_\ell \tau_{lj}(\theta)/A_j(\mathbf{p}_0, \beta \mid \mathcal{W})} \frac{1}{z} S_\ell(z)\, \mathrm{d}z.$$

Let us define :

$$T_R(\beta) \equiv \frac{1}{RN} \sum_{r=1}^{R} \sum_{j=1}^{N} \Phi_j(\mathbf{P}_{0,(r)}, \beta \mid \mathcal{W}),$$

and study the asymptotic expansion of

$$T_R(\hat{\beta_\kappa}) - \mathbb{E}\left[T_R(\beta_0)\right]$$

, where $\hat{\beta_\kappa}$ is our Sieve estimator.

Let β_κ^* be the pseudo true value. Then, we have:

$$T_R(\hat{\beta_\kappa}) - \mathbb{E}T_R(\beta_0) \quad = \quad I_R + II_R + III_R,$$

where we defined

$$I_R \equiv T_R(\hat{\beta_\kappa}) - T_R(\beta_\kappa^*),$$

$II_R \equiv T_R(\beta_\kappa^*) - T_R(\beta_0)$, and $III_R \equiv T_R(\beta_0) - \mathbb{E}T_R(\beta_0)$.

The term I_R is an estimation error of $\hat{\beta_\kappa}$ relative to the pseudo true parameter β_κ^*, this can be analyzed using the usual delta method. The second term II_R requires the use of functional delta method. The last component is simply an average of mean zero IID random variables

$$\frac{1}{N} \sum_{j=1}^{N} \left(\Phi_j(\mathbf{P}_{0,(r)}, \beta_0 \mid \mathcal{W}) - \mathbb{E}\Phi_j(\cdot) \right)$$

. We analyze the three terms in Appendix E.2.1, E.2.3, and E.2.4, respectively.

E.2.1 Term I_R

Since both β_κ^* and $\hat{\beta_\kappa}$ is finite dimensional, with components $\xi_\kappa^*, \hat{\xi_\kappa} \in \Xi_\kappa$, we will put $T_R(\xi_\kappa^*)$ and $T_R(\hat{\xi_\kappa})$.

Let us fix $\mathbf{p} \in \mathbb{S}(\mathbf{P}_0)$. By a mean value expansion of $T_R(\hat{\xi_\kappa})$around $\xi = \xi_\kappa^*$, we have:

$$\Phi_j(\mathbf{p}, \hat{\xi_\kappa}) = \Phi_j(\mathbf{p}, \xi_\kappa^*) + \nabla_{\xi'} \Phi_j(\mathbf{p}, \tilde{\xi_\kappa}) \left(\hat{\xi_\kappa} - \xi_\kappa^*\right).$$

Let us define the class of functions

$$\mathcal{H}_\kappa \equiv \left\{ h_\xi(\cdot) := \frac{1}{N} \sum_{j=1}^{N} \nabla_\xi \Phi_j(\cdot, \xi) : \xi \in \Xi_\kappa \right\},$$

for each $\kappa \geq 1$. We first show that for each $\kappa \geq 1$ fixed, a componentwise ULLN applies to the class $\mathcal{H}_\kappa$ by verifying the sufficient conditions of Lemma 3.10 in van de Geer (2006).

Let us fix j and inspect the gradient of $\Phi_j(\mathbf{p}, \xi)$:

$$\begin{aligned}
\nabla_\xi \Phi_j(\mathbf{p}, \xi) &= \nabla_\xi \sum_{\ell=1}^{N} \int_c^{z_{\ell j}(\mathbf{p},\xi)} \frac{1}{x} f_\ell(x) \mathrm{d}x, \\
&= \sum_{\ell=1}^{N} \nabla_\xi \int_c^{z_{\ell j}(\mathbf{p},\xi)} \frac{1}{x} f_\ell(x) \mathrm{d}x, \\
&= \sum_{\ell=1}^{N} \frac{f_\ell\left(z_{\ell j}(\mathbf{p}, \xi)\right)}{z_{\ell j}(\mathbf{p}, \xi)} \nabla_\xi z_{\ell j}(\mathbf{p}, \xi) \\
&\quad + \sum_{\ell=1}^{N} \int_c^{z_{\ell j}(\mathbf{p},\xi)} \frac{1}{x} \nabla_\xi f_\ell(x) \mathrm{d}x, \\
&= \phi_{1,\kappa,j}(\mathbf{p}, \xi) + \phi_{2,\kappa,j}(\mathbf{p}, \xi), \text{ say,}
\end{aligned}$$

by the Leibniz rule. Continuity of $\xi \mapsto \phi_{1,\kappa,j}(\mathbf{p}, \xi)$ for μ almost all $\mathbf{p}$ follows from continuity of the derivative of

$$\nabla_\xi (\tau_{ij}(\theta) / A_j(\mathbf{p}, \xi \mid \mathcal{W}))$$

which is established in Appendix E.1.1. Also, the non-zero components of $\nabla_\xi f_\ell(x)$ is given by the vector

$$\left[\sum_{q=0, q \neq e}^{\kappa} \alpha_q^{(\ell)} x^{-q} + x^{-j} \right]_{0 \leq j \leq \kappa+1}$$

which is continuous almost everywhere x. This gives us continuity of $\xi \mapsto \phi_{2,\kappa,j}(\mathbf{p}, \xi)$.

The envelope condition required to invoke ULLN is also satisfied as both

$$\sup_{\xi \in \Xi_\kappa} \sup_{z \in K_j} f_\ell(z)/z \leq C_\kappa < \infty$$

holds for some bound C_κ only depending on κ. Thus, Lemma 3.10 in van der Geer (2006) applies at each $\kappa \geq 1$.

We will next establish the ULLN along $\kappa \to \infty$. Let us define

$$\mathcal{L}_\kappa \equiv \sup_{\mathbf{p} \in \mathbf{S}(\mathbf{P}_0)} \sup_{\xi \in \Xi_\kappa} \| \nabla_{\xi'} \nabla_\xi \Phi_j(\mathbf{p}, \xi) \|_1,$$

which is guaranteed to exist because of compactness of $\mathbf{S}(\mathbf{P}_0)$, Ξ_κ, and continuous differentiability of $\xi \mapsto h_\xi(\mathbf{p})$. By the mean value theorem, we have

$$\left| \nabla_\xi \Phi_j(\mathbf{p}, \xi) - \nabla_\xi \Phi_j(\mathbf{p}, \tilde{\xi}) \right| \leq \mathcal{L}_\kappa \| \xi - \tilde{\xi} \|_1,$$

componentwise. By an application of Theorem 2.7.14 in van der Vaart and Wellner (1996), we obtain the following upper bound on the covering number.

$$\ln \mathcal{N}_{[]}(\delta, \mathcal{H}_\kappa, \| \|_1) \lesssim O(\kappa) \ln \left(1 + \frac{\rho_R^U \mathcal{L}_\kappa}{\delta} \right), \text{ for all } \delta > 0.$$

Thus, to justify the ULLN along $\kappa \to \infty$, we verify that we have

$$\frac{\kappa}{R} \ln \left(1 + \rho_R^U \mathcal{L}_\kappa\right) \to 0.$$

We work with a rough bound on $\mathcal{L}_\kappa$. We inspect every second derivative term $\nabla_{\xi'} \nabla_\xi \Phi_j(\mathbf{p}, \xi)$. Let $a_j \equiv \min_{z \in K_j} z$ and $b_j \equiv \max_{z \in K_j} z$ for every $j = 1, \ldots, N$ which exists because of compactness of $K_j \subset \mathbb{R}_{++}$. Let us further define

$$a_\kappa^* = \max_{j=1,\ldots,N} \max\{a_j^{-\kappa-1}, a_j^{-2}\}$$

and

$$b_\kappa^* = \min_{j=1,\ldots,N} \min\{b_j^{-2}, b_j^{-\kappa-1}\}$$

.

The basic worst-case rates under our sieve parametrization are as follows.

$$\begin{aligned}
\sup_{z \in K_j} |f_\ell(z)| &\lesssim \rho_R^U \max\{1, a_j^{-\kappa}\}, \\
\sup_{z \in K_j} |\dot{f}_\ell(z)| &\lesssim \kappa \rho_R^U a_\kappa^*, \\
\sup_{z \in K_j} |\ddot{f}_\ell(z)| &\lesssim \kappa^2 \rho_R^U a_\kappa^*,
\end{aligned}$$

and from Appendix D.1, the derivatives of

$$z_{\ell j}(\mathbf{p}, \xi) \equiv p_\ell \tau_{\ell j}(\theta) / A_j(\mathbf{p}, \xi)$$

is bounded by:

$$\max_\ell \sup_{\xi \in \Xi_\kappa} \left| \nabla_\xi \frac{\tau_{\ell j}(\theta)}{A_j(\mathbf{p}, \xi)} \right| \leq O\left(\frac{\rho_R^U}{\rho_R^L} \kappa^2 \frac{a_\kappa^*}{b_\kappa^*}\right),$$

and inspecting the second derivatives of $\nabla_{\xi'} \nabla_\xi z_{\ell j}(\mathbf{p}, \xi)$, the relevant component for entropy calculation is

$$\nabla_{\xi'} \nabla_\xi \frac{1}{A_j(\mathbf{p}, \xi)},$$

because the finite-dimensional component $\tau_{\ell j}(\theta)$ is a continuous function on the compact set Θ for the fixed design. We have,

$$\begin{aligned}
\nabla_{\xi'} \nabla_\xi \frac{1}{A_j(\mathbf{p}, \xi)} &= \frac{1}{A_j^2(\mathbf{p}, \xi)} \left(\nabla_{\xi'} A_j(\mathbf{p}, \xi)\right) \left(\nabla_\xi A_j(\mathbf{p}, \xi)\right) \\
&\quad - \frac{1}{A_j(\mathbf{p}, \xi)} \nabla_{\xi'} \nabla_\xi A_j(\mathbf{p}, \xi) \\
&\lesssim O\left(\left(\kappa^2 \frac{\rho_R^U}{\rho_R^L} \frac{a_\kappa^*}{b_\kappa^*}\right)^2\right) + \nabla_{\xi'} \nabla_\xi A_j(\mathbf{p}, \xi),
\end{aligned}$$

where the bound ($\lesssim$) in the third line is componentwise, and follows from $A_j(\mathbf{p},\xi) \in [\underline{A_j}, \bar{A}_j]$ with $\underline{A_j} > 0$ for any $(\mathbf{p},\xi) \in \mathbb{S}(\mathbf{P}_0) \times \Xi_\kappa$. We have, by the chain rule,

$$\begin{aligned}
\nabla_{\xi'}\nabla_\xi A_j(\mathbf{p},\xi) &= -\frac{1}{\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)}\nabla_{\xi'}\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi) \\
&\quad -\frac{1}{\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)}\left(\nabla_{\xi'}\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right)\left(\nabla_\xi A_j(\mathbf{p},\xi)\right), \\
&\leq O\left(\frac{1}{\rho_R^L b_\kappa^*}\right)\nabla_{\xi'}\left(\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right) \\
&\quad +O\left(\frac{\rho_R^U}{(\rho_R^L)^2}\kappa^2\frac{a_\kappa^*}{(b_\kappa^*)^2}\right)\nabla_{\xi'}\left(\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)\right),
\end{aligned}$$

where we used the bound

$$-\partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi) \gtrsim \rho_R^L b_\kappa^*$$

for all j and any $(\mathbf{p},\xi) \in \mathbb{S}(\mathbf{P}_0) \times \Xi_\kappa$.

As ξ enters the normalization equations $G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi)$ through the common sieve function f_κ and $z_{\ell j}(\mathbf{p},\xi)$, by the chain rule, we have

$$\partial_\xi G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi) = \sum_{\ell=1}^N \dot{f}_\kappa(z_{\ell j}(\mathbf{p},\xi);\xi)\frac{\nabla_\xi p_\ell \tau_{\ell j}(\theta)}{A_j(\mathbf{p},\xi)} + V_j(\mathbf{p},\xi),$$

where we define the $d+\kappa$ dimensional vector

$$V_j(\mathbf{p},\xi) \equiv \sum_{\ell=1}^N \partial_\xi f_\kappa(z_{\ell j}(\mathbf{p},\xi);\xi).$$

Here ∂_ξ holds the effective relative price fixed. Its components are the normalization-adjusted derivatives displayed in Appendix E.1.1. with $\nu_\kappa(x) \equiv (x^{-0}, x^{-1}, \ldots, x^{-\kappa})'$ defined as the basis transformation. By compactness of Θ^C and finiteness of the fixed design, the functional form $\tau_{\ell j}(\theta)$,

$$\|\nabla_{\xi'}\nabla_\xi \tau_{\ell j}(\theta)\|_1 \leq C$$

which does not depend on ℓ, j. Let us therefore ignore the $\nabla_\xi \tau_{\ell j}(\theta)$ and bound

$$\nabla_{\xi'}\sum_{\ell=1}^N \dot{f}_\ell(z_{\ell j}(\mathbf{p},\xi))\frac{1}{A_j(\mathbf{p},\xi)} + \nabla_{\xi'}V_j(\mathbf{p},\xi).$$

We have:

$$\begin{aligned}
\nabla_{\xi'} \sum_{\ell=1}^{N} \dot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \frac{1}{A_j(\mathbf{p},\xi)} &= \frac{1}{A_j(\mathbf{p},\xi)} \sum_{\ell=1}^{N} \left(\ddot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \nabla_{\xi'} z_{\ell j}(\mathbf{p},\xi)\right) \\
&\quad - \frac{1}{A_j^2(\mathbf{p},\xi)} \sum_{\ell=1}^{N} \dot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \nabla_{\xi'} A_j(\mathbf{p},\xi), \\
&\lesssim O\left(\kappa^2 \rho_R^U a_\kappa^*\right) O\left(\kappa^2 \frac{\rho_R^U}{\rho_R^L} \frac{a_\kappa^*}{b_\kappa^*}\right) \\
&\quad + O\left(a_\kappa^*\right) + O(\kappa \rho_R^U a_\kappa^*) O\left(\frac{\rho_R^U}{\rho_R^L} \kappa^2 \frac{a_\kappa^*}{b_\kappa^*}\right), \\
&\lesssim O\left(\kappa^4 \frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}\right).
\end{aligned}$$

Similarly, we can bound

$$\begin{aligned}
\nabla_{\xi'} \partial_A G_j(A_j(\mathbf{p},\xi),\mathbf{p},\xi) &= -\nabla_{\xi'} \sum_{\ell=1}^{N} \dot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \frac{p_\ell \tau_{\ell j}(\theta)}{A_j^2(\mathbf{p},\xi)}, \\
&\lesssim -\frac{1}{A_j^2(\mathbf{p},\xi)} \sum_{\ell=1}^{N} \left(\ddot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \nabla_{\xi'} z_{\ell j}(\mathbf{p},\xi) + V_j(\mathbf{p},\xi)\right) \\
&\quad + \frac{1}{A_j^4(\mathbf{p},\xi)} \sum \dot{f}_\ell(z_{\ell j}(\mathbf{p},\xi)) \nabla_{\xi'} A_j(\mathbf{p},\xi), \\
&\lesssim O\left(\kappa^4 \frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}\right).
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\nabla_{\xi'} \nabla_\xi \frac{1}{A_j(\mathbf{p},\xi)} &\lesssim O\left(\left(\kappa^2 \frac{\rho_R^U}{\rho_R^L} \frac{a_\kappa^*}{b_\kappa^*}\right)^2\right) + O\left(\frac{1}{\rho_R^L b_\kappa^*}\right) O\left(\kappa^4 \frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}\right) \\
&\quad + O\left(\frac{\rho_R^U}{(\rho_R^L)^2} \kappa^2 \frac{a_\kappa^*}{(b_\kappa^*)^2}\right) O\left(\kappa^4 \frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}\right) \\
&\lesssim O\left(\kappa^6 \left(\frac{\rho_R^U a_\kappa^*}{\rho_R^L b_\kappa^*}\right)^3\right).
\end{aligned}$$

The terms in $\nabla_{\xi'} \nabla_\xi \Phi_j(\mathbf{p},\xi)$ involves finite sums of products of

$$\dot{f}_\ell, \ddot{f}_\ell, \nabla_\xi z_{\ell j}(\mathbf{p},\xi)$$

, and $\nabla_{\xi'} \nabla_\xi z_{\ell j}(\mathbf{p},\xi)$ with the dominant term coming from the product of $\ddot{f}_\ell$ and $\nabla_{\xi'} \nabla_\xi z_{\ell j}(\mathbf{p},\xi)$. Hence, we have

$$\mathcal{L}_\kappa \lesssim O\left(\kappa^2 \rho_R^U a_\kappa^*\right) \times O\left(\kappa^6 \left(\frac{\rho_R^U a_\kappa^*}{\rho_R^L b_\kappa^*}\right)^3\right).$$

By using Theorem 2.7.11 in van der Vaart and Wellner (1996) and using the same argument as

in the proof of D.4, the entropy is bounded as follows:

$$\begin{aligned}\ln \mathcal{N}_{[]}(\delta, \mathcal{H}_\kappa, \|\cdot\|_1) &\lesssim O(\kappa) \ln\left(1 + \rho_R^U \mathcal{L}_\kappa\right), \\ &= o(R),\end{aligned}$$

where the final line is true under conditions $\kappa^2/R \to 0$ and $(\kappa/R)\ln(\kappa\rho_R^U/\rho_R^L) \to 0$. This justifies the application of ULLN along $\kappa = \kappa_R \to \infty$.

The regularity condition on the tuning parameters $(\kappa, \rho_R^U, \rho_R^L)$ ensures that we have

$$\ln \mathcal{N}_{[]}(\delta, \mathcal{H}_\kappa, d_e) = o(R)$$

which validates the approximation:

$$\lim_{R\to\infty} \sup_{\xi_\kappa \in \Xi_\kappa} \left| \frac{1}{RN} \sum_{r=1}^{R} \sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{P}_{0,(r)}, \xi_\kappa) - \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{P}_{0,(r)}, \xi_\kappa)\right] \right| = o_p(1).$$

Define

$$G_\kappa \equiv \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{P}_{0,(r)}, \xi_\kappa^*)\right]$$

and

$$\zeta_{\kappa,(r)} \equiv G_\kappa' H_\kappa^{-1} \psi_{\kappa,(r)}.$$

Put $p_\kappa = d + \kappa$ and define

$$A_\kappa^2 \equiv 1 + \mathbb{E}\left[\left\| H_\kappa^{-1} \psi_{\kappa,(r)} \right\|_2^2\right].$$

Also define

$$B_\kappa^G \equiv \max_{1\le a\le p_\kappa} \sup_{\mathbf{p}\in \mathbb{S}(\mathbf{P}_0)} \left| \left[\frac{1}{N}\sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{p}, \xi_\kappa^*)\right]_a \right|$$

and

$$B_\kappa^\Phi \equiv \sup_{\mathbf{p}\in \mathbb{S}(\mathbf{P}_0)} \sup_{\xi\in\Xi_\kappa} \left\| \frac{1}{N}\sum_{j=1}^{N} \nabla_{\xi'}\nabla_\xi \Phi_j(\mathbf{p}, \xi) \right\|_{\text{op}}.$$

We now make the remainder calculation explicit. Write $\Delta_\kappa = \hat{\xi}_\kappa - \xi_\kappa^*$ and

$$h_\kappa(\mathbf{p}, \xi) = \frac{1}{N}\sum_{j=1}^{N} \nabla_\xi \Phi_j(\mathbf{p}, \xi).$$

A second-order expansion gives

$$I_{R,\kappa} = G_\kappa' \Delta_\kappa + (\mathbb{P}_R - P) h_\kappa(\cdot, \xi_\kappa^*)' \Delta_\kappa + \frac{1}{2}\Delta_\kappa' \left\{ \mathbb{P}_R \nabla_{\xi'} h_\kappa(\cdot, \widetilde{\xi}_\kappa) \right\} \Delta_\kappa,$$

where $\widetilde{\xi}_\kappa$ lies between $\hat{\xi}_\kappa$ and ξ_κ^*. Proposition 6 and the definition of A_κ imply

$$\|\Delta_\kappa\|_1 = O_p\left(\frac{\sqrt{p_\kappa}A_\kappa}{\sqrt{R}}\right).$$

Applying the scalar Bernstein inequality component by component and taking a union bound over the p_κ entries gives

$$\|(\mathbb{P}_R - P)h_\kappa(\cdot, \xi_\kappa^*)\|_\infty = O_p\left(B_\kappa^G\left\{\sqrt{\frac{\log p_\kappa}{R}} + \frac{\log p_\kappa}{R}\right\}\right).$$

Consequently, the empirical-gradient remainder multiplied by $\sqrt{R}$ is

$$O_p\left(B_\kappa^G\sqrt{p_\kappa}A_\kappa\left\{\sqrt{\frac{\log p_\kappa}{R}} + \frac{\log p_\kappa}{R}\right\}\right) = o_p(1)$$

under the first rate calculation below. The quadratic remainder satisfies

$$\sqrt{R}\left|\frac{1}{2}\Delta_\kappa'\left\{\mathbb{P}_R\nabla_{\xi'}h_\kappa(\cdot, \widetilde{\xi}_\kappa)\right\}\Delta_\kappa\right| \leq \frac{\sqrt{R}}{2}B_\kappa^\Phi\|\Delta_\kappa\|_2^2 = O_p\left(\frac{B_\kappa^\Phi A_\kappa^2}{\sqrt{R}}\right) = o_p(1).$$

The second equality is controlled by the second rate calculation below. Therefore, Proposition 6 yields

$$\sqrt{R}I_{R,\kappa} = \frac{1}{\sqrt{R}}\sum_{r=1}^{R}\zeta_{\kappa,(r)} + o_p(1).$$

We finally verify that the two rate restrictions are compatible with the tuning sequence used above. The derivative calculations preceding the expansion give the bounds

$$B_\kappa^G \leq C\kappa^3\frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}$$

and

$$B_\kappa^\Phi \leq C\kappa^8\rho_R^U a_\kappa^*\left(\frac{\rho_R^U a_\kappa^*}{\rho_R^L b_\kappa^*}\right)^3.$$

The first bound follows from the two terms in $\nabla_{\xi}\Phi_j$. The direct derivative of the sieve function is of smaller order than the product of $\dot{f}_\kappa$ and the derivative of the effective relative price. The second bound is the bound for $\mathcal{L}_\kappa$ obtained above, and the operator norm is bounded by the matrix one-norm.

The residual in $\psi_{\kappa,(r)}$ is bounded because both the observed and predicted shares are bounded. The same derivative bound for a predicted share and Assumption 14 therefore give

$$A_\kappa \leq 1 + C\sqrt{p_\kappa}\kappa^3\frac{(\rho_R^U a_\kappa^*)^2}{\rho_R^L b_\kappa^*}.$$

For the proposed sequence

$$\kappa_R = \lfloor \log\log R \rfloor, \qquad \rho_R^U = \exp(C_U \kappa_R), \qquad \rho_R^L = \exp(-C_L \kappa_R),$$

compactness of K implies that each of B_κ^G, B_κ^Φ, and A_κ is bounded by $C\kappa^C \exp(C\kappa)$ after increasing C when necessary. It follows that the empirical-gradient remainder is bounded by

$$\frac{C\kappa^C \exp(C\kappa)\sqrt{\log p_\kappa}}{\sqrt{R}} + \frac{C\kappa^C \exp(C\kappa)\log p_\kappa}{R},$$

which converges to zero because it is a polynomial in $\log R$ and $\log\log R$ divided by $\sqrt{R}$. Similarly, the second expression is bounded by

$$\frac{C\kappa^C \exp(C\kappa)}{\sqrt{R}} \longrightarrow 0.$$

Thus, both remainders are $o_p(1)$ after multiplication by $\sqrt{R}$.

E.2.2 Sufficient Conditions for Scalar Functional Continuity

We give sufficient conditions for the application of Assumption 15 to the pseudo-true sequence. Let $\beta_\kappa^u = \beta(\xi_\kappa^u)$ denote the normalized shape-preserving Jackson approximant constructed in Appendix D.6. Define the vector of predicted shares

$$m_\kappa(\mathbf{p}, \xi) = \left[f_\kappa(z_{ij}(\mathbf{p}, \xi); \xi) \right]_{i,j}$$

and the scalar functional

$$\overline{\Phi}_\kappa(\mathbf{p}, \xi) = \frac{1}{N} \sum_{j=1}^{N} \Phi_j(\mathbf{p}, \xi).$$

Suppose there is a neighborhood $\mathcal{U}_\kappa$ containing ξ_κ^* and ξ_κ^u such that the following bounds hold uniformly over $\xi \in \mathcal{U}_\kappa$

$$\lambda_{\min}\left(\mathbb{E}\left[\frac{1}{N^2} \nabla_\xi m_\kappa(\mathbf{P}_{0,(r)}, \xi)' \nabla_\xi m_\kappa(\mathbf{P}_{0,(r)}, \xi) \right] \right) \geq \underline{\lambda} > 0$$

and

$$\mathbb{E}\left[\left\| \nabla_\xi \overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \xi) \right\|^2 \right] \leq \overline{C} < \infty.$$

Suppose also that the second derivatives of the two maps are uniformly bounded on $\mathcal{U}_\kappa$ and that this neighborhood is small enough for their Taylor remainders to be dominated by their linear terms. These conditions imply the local norm comparison

$$\left\| \overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \xi) - \overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \widetilde{\xi}) \right\|_{L^2} \leq C M(\beta(\xi), \beta(\widetilde{\xi}))$$

for $\xi, \widetilde{\xi} \in \mathcal{U}_\kappa$, where C does not depend on κ.

To see this, a mean value expansion of the predicted-share map and the first eigenvalue bound give

$$M(\beta(\xi), \beta(\widetilde{\xi})) \geq \frac{\sqrt{\lambda}}{2}\|\xi - \widetilde{\xi}\|.$$

The derivative bound for $\overline{\Phi}_\kappa$ and another mean value expansion give

$$\left\|\overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \xi) - \overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \widetilde{\xi})\right\|_{L^2} \leq 2\sqrt{C}\|\xi - \widetilde{\xi}\|.$$

Combining the two displays proves the local norm comparison. The optimality of β_κ^* and the approximation property of β_κ^u imply

$$M(\beta_\kappa^*, \beta_\kappa^u) \leq M(\beta_\kappa^*, \beta_0) + M(\beta_\kappa^u, \beta_0) = O(\kappa^{-\gamma}).$$

The local norm comparison therefore controls the functional difference between the pseudo truth and the normalized Jackson approximant. The uniform approximation of the share function and the local continuity of the single aggregator directly give

$$\left\|\overline{\Phi}_\kappa(\mathbf{P}_{0,(r)}, \xi_\kappa^u) - \overline{\Phi}(\mathbf{P}_{0,(r)}, \beta_0)\right\|_{L^2} = O(\kappa^{-\gamma}).$$

The triangle inequality then gives the scalar functional bound required for the pseudo-true sequence in Assumption 15. The eigenvalue condition has a direct interpretation. Any local parameter direction that changes the Matsuyama–Ushchev functional must also generate detectable variation in the equilibrium shares across observations.

E.2.3 Term II_R

Define

$$D_{\kappa,(r)} \equiv \frac{1}{N}\sum_{j=1}^{N}\left\{\Phi_j(\mathbf{P}_{0,(r)}, \beta_\kappa^* \mid \mathcal{W}) - \Phi_j(\mathbf{P}_{0,(r)}, \beta_0 \mid \mathcal{W})\right\}.$$

Then $II_{R,\kappa} = R^{-1}\sum_{r=1}^{R} D_{\kappa,(r)}$. Assumption 15 and the approximation rate for the pseudo truth give

$$\left\{\mathbb{E}[D_{\kappa,(r)}^2]\right\}^{1/2} \leq C_\Phi M(\beta_\kappa^*, \beta_0) = O(\kappa^{-\gamma}).$$

By the Cauchy–Schwarz inequality,

$$|\mathbb{E}[D_{\kappa,(r)}]| \leq \left\{\mathbb{E}[D_{\kappa,(r)}^2]\right\}^{1/2} = O(\kappa^{-\gamma}).$$

The IID assumption also gives

$$\mathrm{Var}\left(\frac{1}{R}\sum_{r=1}^{R} D_{\kappa,(r)}\right) \leq \frac{1}{R}\mathbb{E}[D_{\kappa,(r)}^2] = O\left(R^{-1}\kappa^{-2\gamma}\right).$$

Consequently,

$$II_{R,\kappa} = O(\kappa^{-\gamma}) + O_p(R^{-1/2}\kappa^{-\gamma}) = O_p(\kappa^{-\gamma}).$$

E.2.4 Term III_R

Under the CES null, Proposition 1 of Matsuyama and Ushchev (2017) implies

$$\Phi_j(\mathbf{P}_{0,(r)}, \beta_0) = \Gamma(c, \sigma_0)$$

for every destination j and every observation r. Therefore,

$$T_R(\beta_0) = \Gamma(c, \sigma_0)$$

for any realization of the price vectors in the sample. Hence, $III_R = 0$ almost surely.

E.3 Asymptotic Expansion of the CES Component

We first show that $\Gamma(c, \hat{\sigma}) = \Gamma(c, \sigma_0) + O_p(R^{-1/2})$ provided that $\sqrt{R}(\hat{\beta}_{CES} - \beta_{CES,0}) = O_p(1)$. From the standard extremum-estimation results in Newey and McFadden (1994), the NLS estimator for the CES model is asymptotically normal under the usual regularity conditions and admits the observation-level influence representation

$$\sqrt{R}(\hat{\sigma} - \sigma_0) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \varphi_{\sigma,(r)} + o_p(1).$$

A mean value expansion of $\Gamma(c, \hat{\sigma})$ around the truth σ_0 gives

$$\Gamma(c, \hat{\sigma}) = \Gamma(c, \sigma_0) + \dot{\Gamma}(c, \bar{\sigma})(\hat{\sigma} - \sigma_0),$$

where

$$\dot{\Gamma}(c, \sigma) = \frac{1 - Nc^{1-\sigma} + N(1-\sigma)c^{1-\sigma} \ln c}{(1-\sigma)^2},$$

and $\bar{\sigma}$ is an intermediate value between σ_0 and $\hat{\sigma}$. Note that the parameter space for the elasticity of substitution σ is $[\sigma_L, \sigma_U]$ with $\sigma_L > 1$, and $\sigma_0 \in \text{int}([\sigma_L, \sigma_U])$ which validates the Taylor expansion.

By the continuous mapping theorem, we have $\dot{\Gamma}(c, \bar{\sigma}) \xrightarrow{p} \dot{\Gamma}(c, \sigma_0) < \infty$. The delta method therefore gives

$$\sqrt{R}\left(\Gamma(c, \hat{\sigma}) - \Gamma(c, \sigma_0)\right) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \dot{\Gamma}(c, \sigma_0)\varphi_{\sigma,(r)} + o_p(1).$$

E.4 Combined Expansion of the Test Statistic

The expansion for Term $I_{R,\kappa}$ and the expansion for the CES component imply

$$\sqrt{R}\,(I_{R,\kappa} + IV_R) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \left\{ \zeta_{\kappa,(r)} - \dot{\Gamma}(c, \sigma_0)\varphi_{\sigma,(r)} \right\} + o_p(1).$$

Term III_R is zero almost surely under the CES null. It follows that

$$\sqrt{R}\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \chi_{\kappa,(r)} + \sqrt{R} II_{R,\kappa} + o_p(1),$$

which proves Proposition 7.

The variables $\chi_{\kappa,(1)}, \ldots, \chi_{\kappa,(R)}$ are IID within each row indexed by R. Their distribution may change with R through $\kappa = \kappa_R$. Assumption 16 is the Lyapunov condition for this scalar triangular array. Therefore,

$$\frac{1}{\sqrt{RV_{\mathcal{T},\kappa}}} \sum_{r=1}^{R} \chi_{\kappa,(r)} \xrightarrow{d} N(0,1).$$

If $\sqrt{R} II_{R,\kappa} = o_p(V_{\mathcal{T},\kappa}^{1/2})$, the corollary follows from Slutsky's theorem.

E.5 Feasible Variance Estimation

We verify the relative consistency of the variance estimator. The Lyapunov condition in Assumption 16 implies the triangular-array law of large numbers

$$\frac{1}{RV_{\mathcal{T},\kappa}} \sum_{r=1}^{R} \chi_{\kappa,(r)}^2 \xrightarrow{p} 1.$$

Indeed, after dividing $\chi_{\kappa,(r)}^2$ by $V_{\mathcal{T},\kappa}$, the required moment condition for this law of large numbers is precisely

$$\frac{\mathbb{E}|\chi_{\kappa,(r)}|^{2+\delta}}{R^{\delta/2} V_{\mathcal{T},\kappa}^{1+\delta/2}} \longrightarrow 0.$$

Assumption 17 and the reverse triangle inequality in the empirical L^2 norm give

$$\left| \left\{ \frac{1}{R} \sum_{r=1}^{R} \widehat{\chi}_{\kappa,(r)}^2 \right\}^{1/2} - \left\{ \frac{1}{R} \sum_{r=1}^{R} \chi_{\kappa,(r)}^2 \right\}^{1/2} \right| = o_p(V_{\mathcal{T},\kappa}^{1/2}).$$

It follows that

$$\frac{1}{RV_{\mathcal{T},\kappa}} \sum_{r=1}^{R} \widehat{\chi}_{\kappa,(r)}^2 \xrightarrow{p} 1.$$

Moreover,

$$\overline{\widehat{\chi}}_\kappa = \frac{1}{R} \sum_{r=1}^{R} \chi_{\kappa,(r)} + o_p(V_{\mathcal{T},\kappa}^{1/2}) = o_p(V_{\mathcal{T},\kappa}^{1/2}).$$

Therefore,

$$\frac{\widehat{V}_{\mathcal{T},\kappa}}{V_{\mathcal{T},\kappa}} = \frac{R^{-1}\sum_{r=1}^{R}\widehat{\chi}_{\kappa,(r)}^2 - \overline{\widehat{\chi}}_{\kappa}^2}{V_{\mathcal{T},\kappa}} \xrightarrow{p} 1.$$

This proves the feasible part of the corollary.

E.6 Proof of the Infeasible Test Central Limit Theorem

The result follows from Proposition 7 and the scalar Lyapunov central limit theorem.

E.7 Proof under a Fixed Alternative

Consistency of the HSA estimator, consistency of the CES estimator for its pseudo-true parameter, and continuity of the two functionals give

$$\mathcal{T}_R(\hat{\beta}_\kappa, \hat{\sigma}) \xrightarrow{p} \Delta_0.$$

Since $\Delta_0 \neq 0$, the absolute value of the numerator of the root-R statistic diverges. The stated result follows when its estimated standard deviation is bounded in probability and bounded away from zero in probability.

E.8 Proof of the Feasible Test Central Limit Theorem

Relative consistency of the variance estimator is established in Appendix E.5. The result follows from the infeasible limiting distribution and Slutsky's theorem.

F Proofs for Counterfactual Analysis

Throughout this Section, we will use $\hat{\beta}_\kappa = (\hat{\theta}_\kappa, \hat{f}_\kappa, \ldots, \hat{f}_\kappa) \in \mathcal{B}_\kappa^*$ for the Sieve estimator, and $\beta_\kappa^* = (\theta_\kappa^*, f_\kappa^*, \ldots, f_\kappa^*) \in \mathcal{B}_\kappa^*$ for the pseudo true parameters. We will also use $\hat{\xi}_\kappa, \xi_\kappa^* \in \Xi_\kappa$ to denote the sieve parameters corresponding the $\hat{\beta}_\kappa$ and β_κ^*, respectively.

F.1 Weak Consistency of a Marginal Change in Trade Costs

Let $\mathcal{D}_{-1}$ denote the data from observations $r = 2, \ldots, R$. The estimator $\hat{\beta}_{\kappa,-1}$ is measurable with respect to $\mathcal{D}_{-1}$. Since the observations are IID, $\mathbf{P}_{0,(1)}$ is independent of $\hat{\beta}_{\kappa,-1}$. Proposition 5, with $R-1$ in place of R, gives

$$M^2(\hat{\beta}_{\kappa,-1}, \beta_0) = o_p(1).$$

Put $C_i \equiv \sum_{j \neq i} c_{ji}^2$. By the definition of the marginal counterfactual and the Cauchy–Schwarz

inequality,

$$\left|\hat{\Delta}_{i,(1)} - \Delta^*_{i,(1)}\right|^2 \leq C_i \sum_{j \neq i} \left|\hat{f}_{\kappa,-1}\left(z_{ji,(1)}(\hat{\beta}_{\kappa,-1})\right) - S_0\left(z_{ji,(1)}(\beta_0)\right)\right|^2.$$

Conditional on $\mathcal{D}_{-1}$, the held-out price vector has distribution μ. The definition of M and the nonnegativity of its componentwise squared errors therefore imply

$$\begin{aligned}
&\Pr\left(\left|\hat{\Delta}_{i,(1)} - \Delta^*_{i,(1)}\right| \geq \eta \;\middle|\; \mathcal{D}_{-1}\right) \\
&\quad \leq \frac{C_i}{\eta^2}\mathbb{E}\left[\sum_{j \neq i}\left|\hat{f}_{\kappa,-1}\left(z_{ji,(1)}(\hat{\beta}_{\kappa,-1})\right) - S_0\left(z_{ji,(1)}(\beta_0)\right)\right|^2 \;\middle|\; \mathcal{D}_{-1}\right] \\
&\quad \leq \frac{C_i N^2}{\eta^2} M^2(\hat{\beta}_{\kappa,-1}, \beta_0) = o_p(1).
\end{aligned}$$

The conditional probability is bounded by one. Its convergence in probability to zero therefore also implies convergence of its expectation to zero. This proves

$$\left|\hat{\Delta}_{i,(1)} - \Delta^*_{i,(1)}\right| \xrightarrow{p} 0.$$

F.2 Weak Consistency of Counterfactual Analysis with General Equilibrium Adjustment

Define the estimated counterfactual equilibrium price vector by

$$\hat{\mathbf{p}}_{(1)}(c) \equiv \mathbf{p}_0\left(\hat{\beta}_{\kappa,-1}, \mathbf{L}_{(1)} \mid \mathbf{T}^{CF}(\hat{\theta}_{\kappa,-1} \mid c)\right).$$

We similarly define the estimated initial equilibrium price vector by

$$\hat{\mathbf{p}}_{(1)}(0) \equiv \mathbf{p}_0\left(\hat{\beta}_{\kappa,-1}, \mathbf{L}_{(1)} \mid \mathbf{T}^{CF}(\hat{\theta}_{\kappa,-1} \mid 0)\right).$$

Proposition 1 implies that $\hat{\mathbf{p}}_{(1)}(c)$ exists uniquely almost surely. By Definition 1, it satisfies

$$\mathbb{Z}_j^{CF}\left(\hat{\mathbf{p}}_{(1)}(c), \hat{\beta}_{\kappa,-1} \mid c, \mathbf{L}_{(1)}\right) = 0$$

for every $j = 1, \ldots, N$, almost surely. Proposition 5, with $R - 1$ in place of R, gives

$$M(\hat{\beta}_{\kappa,-1}, \beta_0) \xrightarrow{p} 0.$$

By the subsequence characterization of convergence in probability, for any subsequence $\{R_s\}_{s \geq 1}$, there exists a further subsequence $\{R_{s_t}\}_{t \geq 1}$ such that

$$M\left(\hat{\beta}_{\kappa_{R_{s_t}},-1}, \beta_0\right) \longrightarrow 0$$

almost surely as $t \to \infty$. Proposition 1 places the corresponding normalized equilibrium price

vectors in the compact unit simplex $\Delta_{\mathbf{P}}$. By sequential compactness, we may therefore take a further convergent subsequence, which we do not relabel, such that

$$\hat{\mathbf{p}}_{(1),R_{s_t}}(c) \longrightarrow \widetilde{\mathbf{p}}_{(1)}(c) \in \Delta_{\mathbf{P}}$$

almost surely. Assumption 18 then gives

$$\mathbb{Z}_j^{CF}\left(\widetilde{\mathbf{p}}_{(1)}(c), \beta_0 \mid c, \mathbf{L}_{(1)}\right) = 0$$

for every $j = 1, \ldots, N$. Proposition 1 establishes existence and uniqueness of the normalized equilibrium for every $\mathbf{L}_{(1)} \in \mathbb{S}(\mathbf{L})$. Hence, it must be the case that we have

$$\widetilde{\mathbf{p}}_{(1)}(c) = \mathbf{p}_{0,(1)}(c).$$

The initial subsequence $\{R_s\}_{s\geq 1}$ was arbitrary. We have therefore shown that every subsequence of $\{\hat{\mathbf{p}}_{(1),R}(c)\}_{R\geq 1}$ admits a further subsequence that converges almost surely to the same limit $\mathbf{p}_{0,(1)}(c)$. By the subsequence characterization of convergence in probability, this implies

$$\hat{\mathbf{p}}_{(1)}(c) \xrightarrow{p} \mathbf{p}_{0,(1)}(c).$$

Assumption 18 also holds at $c = 0$. Repeating the same argument shows that every subsequence of $\{\hat{\mathbf{p}}_{(1),R}(0)\}_{R\geq 1}$ admits a further subsequence that converges almost surely to the unique initial equilibrium $\mathbf{p}_{0,(1)}(0)$. Applying the subsequence characterization again gives

$$\hat{\mathbf{p}}_{(1)}(0) \xrightarrow{p} \mathbf{p}_{0,(1)}(0).$$

Proposition 1 implies that the unique limiting equilibrium price vector is in the interior of the normalized price simplex. Thus, every component of the limiting price vector is strictly positive almost surely. All logarithms appearing in the welfare functional are therefore well defined with probability one. The continuity results for A_j and Φ_j established in the analysis of the Matsuyama–Ushchev test, together with the two price convergence results above, imply

$$\hat{\Delta}_{j,(1)}(c) - \Delta^*_{j,(1)}(c) \xrightarrow{p} 0.$$

Finally, $\Delta^*_{j,(1)}(c)$ has a distribution that does not depend on R. Slutsky's theorem gives

$$\hat{\Delta}_{j,(1)}(c) \xrightarrow{d} \Delta^*_{j,(1)}(c).$$

G Simulation Study : CES Model

Given the importance of the CES model, we provide a simulation study for the CES case. Our main contributions to the estimation of CES gravity models are the identification of the elastic-

ity of substitution σ_0 and the use of statistical testing for counterfactual analysis. The simulation study therefore focuses on the finite-sample performance of $\hat{\sigma}$ and the statistical procedure for the gains-from-trade measure.

Our simulation design is as follows.

$$\lambda_{ij,(r)} = \lambda^*_{ij,(r)} + \varepsilon_{ij,(r)}, \tag{42}$$

$$\lambda^*_{ij,(r)} = \left(\frac{p_i \tau_{ij}(\theta_0)}{A_j(\mathbf{p}, \beta_0, \mathcal{I})} \right)^{1-\sigma_0}, \tag{43}$$

$$A_j(\mathbf{p}, \beta_0, \mathcal{I}) = \left(\sum_{\ell=1}^{N} \left(p_\ell \tau_{\ell j}(\theta_0) \right)^{1-\sigma_0} \right)^{\frac{1}{1-\sigma_0}}, \tag{44}$$

where $\mathcal{I} = (\{W_{ij}\}_{i,j}, \mathbf{L})$ with $\mathbf{L}$ being a vector whose elements equal 500, and $\{W_{ij}\}$ is generated by drawing N points independently from the uniform distribution on a two-dimensional square with side length 0.2 and computing the Euclidean distance between the two points. We set $\tau_{ij}(\theta_0) = \exp(W_{ij}\theta_0)$ for $i \neq j$ and 1 for $i = j$. The price vector is computed according to Definition 1, where the price of country N is normalized to 1. We set $N = 28$, which is the number of countries in the European Union, and draw $\varepsilon_{ij,(r)} \sim N(0, \sigma_\varepsilon^2)$ independently with $\sigma_\varepsilon \in \{0.005, 0.01, 0.02\}$. Note that we are keeping the distribution of the error fixed across r. The noise level σ_ε is chosen based on the signal-to-noise ratio defined by the ratio of the mean of λ^*_{ij} $(i \neq j)$ and σ_ε. The signal-to-noise ratios for the three levels of σ_ε are 6.32, 3.16, and 1.58, respectively. The observed data for each $r = 1, \ldots, R$ are constructed according to Definition 2, and we observe $\{D_{(r)}\}_{r=1}^R = \{\{\Lambda_{(r)}, \mathbf{p}, \mathcal{I}\}\}_{r=1}^R$, where $R \in \{5, 10, 20\}$. For each configuration (R, σ_ε), we conduct 1,000 replications.

We set the true parameter $\beta_0 = (\theta_0, \sigma_0)$ to $(5.0, 4.5)$ and search for the minimizer of the sample least-squares criterion over the compact rectangle $(\sigma, \theta) \in [1.5, 8.0] \times [10^{-10}, 10.0]$. The point estimates are obtained by a two-step procedure. First, we conduct a grid search over a relatively coarse grid consisting of $(\sigma, \theta) \in \{3, 4, 5, 6\} \times \{1, 2.5, 4, 5.5\}$. Note that the grid does not include the true parameter values. In the second stage, we use the minimizer of the grid search as a starting value to solve the FOC of the NLS criterion function.

Two key conditions for identification are satisfied in our setup. First, all elements of the equilibrium price vector differ from 1 except for the last normalized element. The mean of the price vector is 0.993, and the standard deviation is 0.027. It turns out that this dispersion in the equilibrium prices is sufficient numerically to produce accurate estimates of the parameters. Second, the triple difference $W_{ij} - W_{i\ell} - W_{\ell j}$ for all distinct indices (i, j, ℓ) is nonzero by construction. The true shares have mean 0.036, with the mean of the (diagonal) domestic shares being roughly 4.6 times larger than international (off-diagonal) shares. We report precise summary statistics in Table 8.

We also assess the performance of the gains-from-trade counterfactual. We focus on the

welfare effect for country N. The counterfactual is

$$\Delta(\theta_0 \mid \sigma_0, \mathcal{I}) = 1 - (\lambda^*_{NN}(\mathbf{p}_0, \theta_0, \sigma_0, \mathcal{I}))^{\frac{1}{\sigma_0 - 1}} \approx 0.4678, \tag{45}$$

which is estimated by $\hat{\Delta} \equiv \Delta(\hat{\theta} \mid \hat{\sigma}, \mathcal{I})$ by replacing the true parameters with their estimators in equations (43) and (44). The asymptotic distribution for $\hat{\Delta}$ is obtained by the delta method.

In Table 9, we report the absolute bias and mean squared error (MSE) of the point estimates $\hat{\beta}$ and the counterfactual $\hat{\Delta}$. The results are summarized as follows. First, all point estimates regardless of the noise level σ_ε improve substantially as R increases. When R increases from 5 to 10, the MSE typically declines by 50% to 80%. When R increases from 10 to 20, the MSE falls by a further 40% to 60%. This suggests that the NLS estimator becomes more stable as the sample size increases. Second, doubling the noise σ_ε roughly quadruples the MSE. For example, the MSE of $\hat{\theta}$ at $R = 10$ is 0.007 at $\sigma_\varepsilon = 0.005$, but this increases by a factor of 3.9 when $\sigma_\varepsilon = 0.01$, to 0.0282, and further increases by a factor of 5.5 when $\sigma_\varepsilon = 0.02$, to 0.1548. Third, in our simulation setup, the performance of $\hat{\sigma}$ is much better than the performance of $\hat{\theta}$. The MSE for the latter is consistently 2 to 5 times larger than the former. At the same time, the gains-from-trade counterfactual $\hat{\Delta}$ is remarkably stable across configurations (R, σ_ε). Finally, we comment on the bias of our estimates. For $\hat{\sigma}$, the absolute bias is generally small (< 0.02) relative to the MSE. Thus, the bias contributes very little to the MSE reported. For $\hat{\theta}$, we see a relatively large bias at $R = 5$, but it declines for larger R. For example, at $R = 5$, when $\sigma_\varepsilon = 0.05$, the absolute bias is 0.001 but this jumps up by a factor of 68.25 to 0.087 for $\sigma_\varepsilon = 0.02$. At $R = 20$, the bias only increases by a factor of 5.40. This suggests that the estimates for θ_0 struggle in environments with low R and a low signal-to-noise ratio. The gains-from-trade statistic $\hat{\Delta}$ shows negligible bias (< 0.004) across all specifications of (R, σ_ε).

In Table 10, we report the empirical coverage of the two-sided test of the null hypothesis $H_0 : \rho = \rho_0$, where $\rho \in (\theta, \sigma, \Delta)$ is reported. The coverage of the t-statistics is based on the studentization implied by the large-sample theory of the NLS estimator. Coverage based on the asymptotic approximation indicates systematic overrejection of the null hypothesis. Compared with the 5% nominal size, the empirical coverage is approximately 89% and improves slightly with larger R. For instance, the empirical coverage hovers around 91% when $R = 20$ compared to $R = 5$ where we have the coverage falling below 90% for some noise levels σ_ε. The coverage of the gains from trade statistic is the worst out of the three estimates. This is likely due to the fact that the asymptotic distribution is derived using the Delta Method, which involves a first-order Taylor approximation. For example, the worst coverage out of all estimates, and all configurations (R, σ_ε) is the coverage of $\hat{\Delta}$ at $R = 5$ and $\sigma_\varepsilon = 0.01$ at 0.885. This is a significant overrejection of the null hypothesis by 6.6 percentage points. The t-statistics for all three estimates across simulations are roughly distributed around zero, with a standard deviation of roughly 1.14, which is slightly larger than $N(0, 1)$.

We suspect that the systematic overrejection is due to the Taylor expansion used to derive the asymptotic distribution, which induces finite-sample bias in the estimator. To investigate

this problem, we also conduct the residual bootstrap of the test statistic. Since the t-statistic is asymptotically pivotal, the bootstrap should, in large samples, give us a better approximation. We report the bootstrapped empirical coverage for each combination (R, σ_ε). For each simulation replication, we conduct 499 bootstrap replications and reject the null hypothesis using the $(499+1)(1-0.05) = 475$th highest (in absolute value) bootstrapped t-statistic as the critical value.

The bootstrap provides a much better coverage rate. Figure 4 shows the bootstrap distribution of the t-statistic for the point estimates and the gains-from-trade counterfactual, based on one realization with $R = 10$ and $\sigma_\varepsilon = 0.01$. For all three plots, we see that the critical value implied by the bootstrap is larger in absolute value than the critical value implied by the asymptotic approximation. Table 10 reports the empirical coverage using the residual bootstrap for each (R, σ_ε). We see a much better empirical coverage across all configurations. Across all (R, σ_ε), the bootstrap improves empirical coverage. The average coverage over R and σ_ε for $\hat{\sigma}, \hat{\theta}, \hat{\Delta}$ is 0.948, 0.948, and 0.947, respectively. This is a substantial improvement over the asymptotic approximation, which gives 0.906, 0.911, and 0.902 for the three estimates, respectively. The most notable improvement occurs at the small sample size $R = 5$. The asymptotic coverage ranges from 0.885 to 0.914, while the bootstrap coverage ranges from 0.929 to 0.953. Since practitioners often do not have access to large R, we suggest the use of bootstrap in applied work. It is also worth noting that the bootstrap provides a much better performance across all noise levels σ_ε. The gains-from-trade estimate $\hat{\Delta}$ benefits the most from bootstrap. At $R = 5, \sigma_\varepsilon = 0.01$, the performance improved from 0.885 to 0.929, which is a reasonable performance. At $R = 10$, the bootstrap coverage ranges from 0.948 to 0.954. This is again a substantial improvement over the asymptotic approximation (0.907 to 0.913).

Table 8: Summary Statistics of Equilibrium Prices and Shares

Variable	Mean	Std. Dev.	Q1	Median	Q3	# of obs.
Price Vector						
Price $\mathbf{p}$	0.993309	0.026561	0.974520	0.997459	1.014119	28
Trade Shares						
All Elements Λ^*	0.035714	0.035313	0.010357	0.024256	0.048264	784
Diagonal (Domestic)	0.146684	0.042327	0.110118	0.135144	0.166111	28
Off-Diagonal (Intl.)	0.031604	0.027488	0.009938	0.023126	0.045446	756

Notes: The price vector $\mathbf{p}$ has $N = 28$ elements, and the true trade shares Λ^* has $N^2 = 784$ elements. The diagonal elements (domestic shares) and off-diagonal elements (international bilateral shares) are reported separately.

Table 9: Comprehensive Simulation Results

		$\hat{\sigma}$		$\hat{\theta}$		Gains From Trade	
R	σ_{noise}	| Bias |	MSE	| Bias |	MSE	| Bias |	MSE
5	0.0050	0.00074	0.00750	0.00127	0.01597	0.00009	0.00007
5	0.0100	0.00068	0.03109	0.01228	0.06651	0.00051	0.00029
5	0.0200	0.01654	0.12662	0.08668	0.39349	0.00408	0.00144
10	0.0050	0.00049	0.00342	0.00168	0.00718	0.00012	0.00003
10	0.0100	0.00013	0.01314	0.00422	0.02822	0.00030	0.00012
10	0.0200	0.01055	0.06117	0.04290	0.15478	0.00218	0.00062
20	0.0050	0.00239	0.00178	0.00433	0.00374	0.00026	0.00002
20	0.0100	0.00097	0.00717	0.00195	0.01531	0.00003	0.00007
20	0.0200	0.00870	0.02766	0.02339	0.06141	0.00138	0.00026

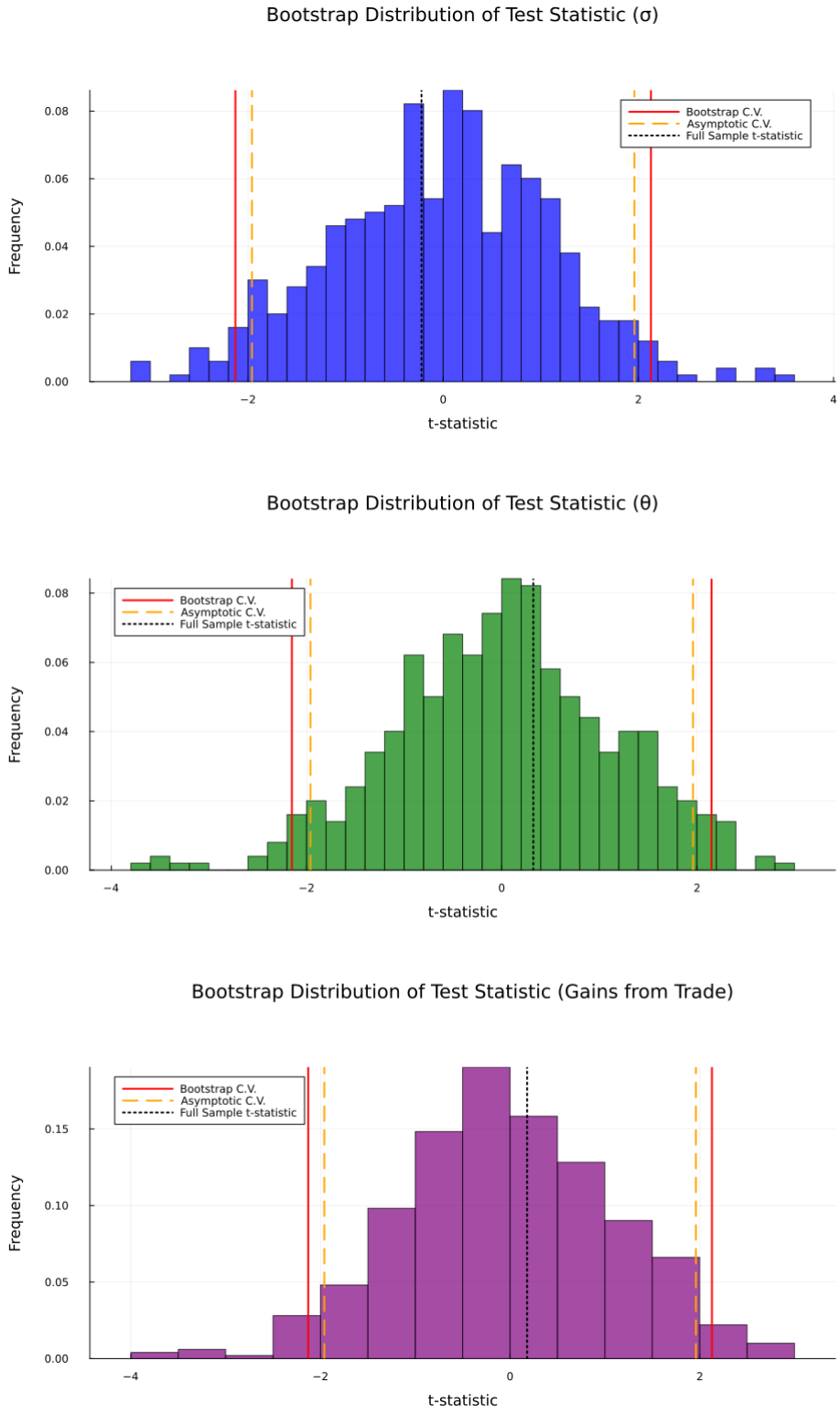
Notes: We report the absolute bias, the Mean Squared Error (MSE), both centered at the truth, for the three estimators. Each configuration uses 1000 replications.

Table 10: Bootstrap vs Asymptotic Coverage Rates

		$\hat{\sigma}$		$\hat{\theta}$		Gains from Trade	
R	σ_{noise}	Bootstrap	Asymptotic	Bootstrap	Asymptotic	Bootstrap	Asymptotic
5	0.0050	0.935	0.887	0.933	0.897	0.935	0.890
5	0.0100	0.938	0.892	0.929	0.899	0.929	0.885
5	0.0200	0.948	0.913	0.953	0.914	0.947	0.913
10	0.0050	0.954	0.911	0.961	0.918	0.948	0.908
10	0.0100	0.956	0.917	0.949	0.916	0.954	0.913
10	0.0200	0.951	0.906	0.958	0.916	0.953	0.907
20	0.0050	0.951	0.903	0.951	0.914	0.954	0.898
20	0.0100	0.952	0.910	0.947	0.906	0.952	0.904
20	0.0200	0.950	0.913	0.952	0.922	0.950	0.910

Notes: Empirical Coverage rates for a two-sided t-test with a 5% significance level based on 1,000 simulation replications. The asymptotic critical values are 1.96, and bootstrap critical values were computed using 499 bootstrap replications per simulation.

Figure 4: Residual Bootstrap of T-statistics



Notes: The histogram reports the critical values based on the asymptotic and bootstrap approximations. The bootstrap uses 499 replications. The figure is based on one realization with $R = 10$ and $\sigma_\varepsilon = 0.01$.

H Implementation Details for the HSA Simulation

H.1 Normalized Sieve Estimation

The simulation uses $\bar{\mathbf{p}} = \mathbf{1}_N$ and destination N as the reference destination. Under symmetry, write

$$f_\kappa(z;\xi_\kappa) = \alpha_0(\xi_\kappa) + \sum_{q=1}^{\kappa} \alpha_q z^{-q}, \qquad \xi_\kappa = (\theta, \alpha_1, \ldots, \alpha_\kappa)'.$$

At every evaluation of the criterion function, we set

$$\alpha_0(\xi_\kappa) = \frac{1}{N}\left\{1 - \sum_{q=1}^{\kappa} \alpha_q \sum_{i=1}^{N} [\bar{p}_i \tau_{iN}(\theta)]^{-q}\right\}. \tag{46}$$

This imposes $A_N(\bar{\mathbf{p}}, \beta) = 1$. The optimization is therefore conducted over ξ_κ. The value of α_0 is recovered from the normalization and is not an additional estimated coordinate.

We estimate the normalized sieve and the CES benchmark on the same simulated observations. We use multiple starting values and retain the solution with the smallest criterion value. For every accepted solution, we verify the normalization numerically and check the shape restrictions over the realized compact support of the relative prices.

H.2 Feasible Matsuyama–Ushchev Test

Define

$$b(\theta, \sigma) = \left\{\sum_{i=1}^{N} [\bar{p}_i \tau_{iN}(\theta)]^{1-\sigma}\right\}^{-1}.$$

Under $A_N(\bar{\mathbf{p}}, \beta) = 1$, the normalized CES share function is $S(z) = b(\theta, \sigma) z^{1-\sigma}$. Its price-index functional is

$$\Gamma_{\bar{p}}(c, \theta, \sigma) = \frac{1 - N b(\theta, \sigma) c^{1-\sigma}}{1 - \sigma}. \tag{47}$$

The unstudentized discrepancy is

$$\widehat{\mathcal{T}}_{R,\kappa} = T_R(\hat{\beta}_\kappa) - \Gamma_{\bar{p}}(c, \hat{\theta}_{CES}, \hat{\sigma}).$$

The feasible variance follows directly from the parametric delta method. We differentiate T_R with respect to the reduced sieve coordinates ξ_κ. We differentiate $\Gamma_{\bar{p}}$ with respect to both CES parameters. We then combine these gradients with the observation-level influence functions of the two NLS estimators. Because both estimators use the same observations, their covariance is retained. The feasible test rejects at the 5 percent level when

$$\left| \frac{\sqrt{R}\widehat{\mathcal{T}}_{R,\kappa}}{\widehat{V}_{\mathcal{T},\kappa}^{1/2}} \right| > 1.96.$$

I Empirical Application

This appendix gives the detailed CES exercise and the corrected-data implementation for the comparison in Subsection 7.1.

I.1 CES Model

The CES model provides a parametric benchmark for the HSA estimates compared in Table 4. We estimate both models using the corrected annual trade matrices. Each complete share matrix is one observation. Appendix I describes the data checks and numerical implementation.

The trade-cost parameter is restricted to the cone Θ. The cone has nine extreme rays in this application. Since θ is three dimensional, we search over nonnegative combinations of triples of rays, using multiple starting values. We first optimize over σ at each initial value of θ and then jointly refine the best candidate within each triple. The search restricts σ to $[1.01, 4]$ and the ray scale to $[0, 60]$.

Table 11 reports the corrected full-sample estimate. The estimated elasticity of substitution is 1.0393, implying a trade elasticity of approximately 0.0393. This point estimate is below the trade elasticity of five often used in quantitative trade applications, which corresponds to $\sigma = 6$ in this model. See Section 2.3 of Costinot and Rodríguez-Clare (2014). This comparison concerns point estimates. We do not report a hypothesis test of the conventional value using the earlier, invalidated standard errors or objective profiles.

Table 11: CES Estimates Using the Corrected Annual Data

$\hat{\sigma}$	Distance	Language	Border	Criterion
1.0393	17.1161	-34.7247	21.8525	0.000251659

Notes. Estimates use the 20 annual observations from 1995 through 2014. The selected optimizer reports convergence and neither the elasticity nor ray scale reaches its search bound. Multiple starting values do not certify a global minimum. Entries are point estimates; standard errors and gains-from-trade inference are not reported.

Distance raises the fitted trade costs and common language lowers them. The border coefficient is positive, despite the cone restrictions. These restrictions impose the iceberg-cost axioms without separately restricting the signs of individual coefficients. The comparison with HSA in the main text therefore concerns both the demand specification and the associated trade-cost estimates.

I.2 Annual Data and Numerical Search

We rebuild each annual trade matrix from its corresponding raw WIOT file. We verify that the 20 annual matrices are distinct, that each contains all 41^2 bilateral pairs, and that expenditure totals and shares agree. These checks correct an earlier cleaning error that repeated the 2014 matrix across multiple years. Estimates and inference based on that earlier panel are not used.

The CES criterion averages the squared bilateral share residuals over all countries and annual observations. At every candidate (θ, σ), the CES aggregator is recomputed. We normalize the nine extreme rays of the trade-cost cone and enumerate their 84 triples. For each triple, the starting weights are $(1/3, 1/3, 1/3)$, the three permutations of $(1/2, 1/4, 1/4)$, the three equally weighted pairs, and the three simplex vertices. We multiply these combinations by scales in $\{0.25, 0.50, \ldots, 10\}$. For each initial θ, we minimize over $\sigma \in [1.01, 4]$. We retain the best candidate in each triple and jointly refine the elasticity, scale, and simplex coordinates. The scale upper bound is 60. Full-sample and training estimates are obtained separately.

I.3 HSA Sieve and Aggregators

Write

$$f_\kappa(z) = \alpha_0 + \sum_{q=1}^{\kappa} \alpha_q z^{-q}, \qquad F_\kappa(y) = f_\kappa(1/y), \qquad y = 1/z.$$

The main specification imposes nonnegative coefficients and uses $\kappa = 1$. Signed coefficients for $\kappa = 2, 3$ provide robustness comparisons on a fixed domain. The reference price vector is $\bar{\mathbf{p}} = (1, \ldots, 1)$ and the reference destination is N. At the reference trade-cost design, the normalization $A_N(\bar{\mathbf{p}}, \beta) = 1$ determines

$$\alpha_0 = \frac{1}{N}\left[1 - \sum_{q=1}^{\kappa} \alpha_q \sum_{i=1}^{N} \tau_{iN}(\theta)^{-q}\right].$$

When the trade-cost characteristics vary across years, the reference design uses the last year of the estimation sample.

For the main specification, $\alpha_0 \geq 0$, $\alpha_1 > 0$, and $N\alpha_0 < 1$. For any positive delivered-price vector q, the aggregator has the explicit solution

$$A_j = \frac{1 - N\alpha_0}{\alpha_1 \sum_i q_i^{-1}}.$$

This gives positive shares that add to one and decrease in relative prices globally. Higher positive-coefficient orders yield essentially the same estimate: the additional coefficients reach their numerical lower bounds of 10^{-10}.

For the signed robustness fits, each estimation sample first supplies a positive-coefficient $\kappa = 1$ estimate. Its fitted inverse prices and the inverse prices at the normalization point determine a minimum and maximum. We divide the minimum by two and multiply the maximum by two to form a fixed interval $[a, b]$. For training estimates this construction uses 1995–2013 only. The interval is not expanded during optimization or evaluation in 2014.

To improve numerical scaling, let $v = y/b$ and $\widetilde{\alpha}_q = \alpha_q b^q$. We impose

$$F_\kappa(y) > 0, \qquad F'_\kappa(y) > 0, \qquad y \in [a, b].$$

Numerically, the share and derivative with respect to v are bounded below by 10^{-10}. For degree

at most three, the derivative is a polynomial of degree at most two. Its minimum on the interval is computed from the endpoints and any interior vertex. Once the derivative is positive, the minimum share occurs at a. These restrictions are checked over the entire interval.

For delivered prices $q_i = p_i\tau_{ij}(\theta)$, all inverse prices A_j/q_i lie in $[a, b]$ precisely when

$$a \max_i q_i \leq A_j \leq b \min_i q_i.$$

On this bracket, the adding-up function

$$G_j(A) = \sum_{i=1}^{N} F_\kappa(A/q_i) - 1$$

is strictly increasing. We require the bracket to be nonempty and its endpoint values to straddle zero. Bisection then gives the unique admissible aggregator. Roots outside this bracket are not used. A candidate must admit an aggregator for every estimation column. During validation, an inadmissible column is marked as uncovered.

These conditions certify the numerical demand system only on $z \in [1/b, 1/a]$. They do not establish its global admissibility on $\mathbb{R}_{++}$. Applying the global economic model or extrapolating to counterfactual prices outside this interval would require an admissible extension or additional restrictions.

I.4 Signed Multistart Refinement

The positive-coefficient fits provide initial values for signed robustness estimation. For each order, we also include the lower-order estimate with a zero coefficient on the additional term. Each feasible anchor is represented in all triples of extreme rays that contain it. We perturb the highest coefficient in both directions and retain the feasible starting points.

Within each such cone chart, we jointly optimize the ray scale, two simplex coordinates, and the scaled coefficients. The bounds are

$$0 \leq s \leq 60, \qquad 0 \leq u_1, u_2 \leq 1, \qquad |\widetilde{\alpha}_q| \leq 100.$$

The search uses Nelder–Mead with hard feasibility rejection. An invalid or worse optimizer result is replaced by its feasible starting value. The smallest retained criterion determines the estimate. We use 3,000 iterations and 180 seconds per start. The reported full-sample estimates each use 70 starts. This procedure searches locally from multiple initial values and does not certify a global minimum.

I.5 Validation and Inference Limitations

The full sample contains 820 destination-year columns. For validation, parameters and the admissible interval are estimated using 1995–2013, containing 779 columns. All 41 columns in 2014 are covered by both signed fits. Foreign-share error comparisons exclude domestic shares

and use the same covered observations for HSA and CES. The 2014 sample informed comparisons among the signed specifications and is treated as validation. The paper uses positive $\kappa = 1$ as the main specification because it provides global shape restrictions and the higher positive orders reduce to essentially the same fit.

The positive main estimate lies on a boundary of the trade-cost cone. The higher positive orders also have active coefficient restrictions. Conventional interior influence-function formulas therefore do not automatically justify standard errors or test p-values for these constrained estimates. Section 7.2 reports the resulting standardized statistic as a descriptive calculation.

The diagnostic uses the fitted normalization $A_N(\bar{\mathbf{p}}) = 1$. Under this normalization, the CES share function is $S_{\mathrm{CES}}(z) = b(\theta, \sigma) z^{1-\sigma}$, where

$$b(\theta, \sigma) = \left[\sum_i \tau_{iN}(\theta)^{1-\sigma} \right]^{-1}.$$

The corresponding CES component of the discrepancy is

$$\Gamma(\theta, \sigma; c) = \frac{1 - N b(\theta, \sigma) c^{1-\sigma}}{1 - \sigma}.$$

Its derivative includes the dependence of b on both θ and σ. HSA aggregator derivatives are obtained by implicit differentiation. The full Hessian includes residual curvature; the Gauss–Newton calculation provides a sensitivity comparison. The centered variance combines HSA and CES contributions within each of the 20 annual observations. The numerical statistic is stable across finite-difference step sizes of 10^{-5}, 3×10^{-6}, and 10^{-6}, but this stability does not resolve the constraint-boundary issue or establish the normal approximation.

For counterfactuals, producer prices are normalized by fixing the last country's price to one. We solve all remaining market-clearing equations and check the omitted equation. Policy continuation and direct solutions from baseline prices agree within 10^{-7} in log prices. Demand checks verify adding up, price homogeneity, and agreement between price-index derivatives and expenditure shares. Independent existing HSA excess-demand and price-index-integral routines reproduce the main model's equilibrium and welfare calculations. The largest absolute log market-clearing residual across the reported HSA and CES equilibria is below 10^{-9}. The calculations use the annual input manifest and saved full-sample estimates; no parameters are re-estimated for the policy experiments. Signed HSA equilibrium calculations retain their fixed domain and are not extrapolated outside it.