Industrial policies for multi-stage production: The battle for battery-powered vehicles*

Keith Head[1], Thierry Mayer[2], Marc Melitz[3], and Chenying Yang[4]

[1]University of British Columbia
[2]Sciences Po
[3]Harvard University
[4]Singapore Management University

October 16, 2024

Draft version, under revision

Abstract

Many countries have set ambitious targets for transitioning away from fossil fuels. The plans generally involve switching from combustion engines to electric vehicles (EVs). As batteries constitute around 40% of the cost of EVs, firms need to establish low-cost battery supply chains in order to make EVs attractive to consumers. At the same time, governments increasingly use tax and subsidy schemes to induce firms to locate more stages of the supply chain within their jurisdictions. We specify a multi-stage supply chain for EVs from battery cell production to vehicle distribution. Each car producer selects where to open facilities at each stage considering production costs, transport costs, tariffs and subsidies. This is a difficult combinatorial choice problem, but we leverage a mixed integer linear program formulation which can be solved in under a minute for a single firm. We estimate the parameters of our model—which include the variable production costs and fixed plant/model activation costs—using observed sourcing decisions for all production stages over the period 2021 to 2023. The next step is a set of counterfactuals that compute how policy interventions affect the final pattern of production and trade in this sector.

*We thank internal grants from our respective institutions for helping to finance the data purchased in this project. There are no conflicts of interest to disclose.

1 Introduction

The rise of global supply chains has dramatically changed the landscape of the international organization of production. Firms slice up their production process, retaining only a subset of the stages within their domestic economies. However, this international organization of production does not come with international cooperation in government policies, especially in growing industries, like electric vehicles (EV). As more countries accelerate their transition to green economies and announce their ban on internal combustion engine vehicle sales by 2035, countries compete via subsidies to draw in local EV supply chains.

The 2022 US Inflation Reduction Act (IRA) awards consumer and production subsidies that do not extend to European-made cars. European governments are considering measures to counter those policies and to stem the rising tide of Chinese imports. Canada, meanwhile has promised to match US IRA production subsidies for plants established in Canada. Deputy prime minister Freeland defended the Canadian government's decision to spend roughly $30 billion on incentives to induce Volkswagen and Stellantis to build factories in Ontario, saying "Our government is absolutely determined that Canada gets its fair share of those green jobs."

The value chain for EVs is economically interesting because of countervailing forces at work. First, there are China's cost advantages in all stages of the EV value chain from refining of minerals to final assembly. The IRA contains a provision that prevents subsidies from being applied to vehicles containing Chinese minerals and other components. Second, there are large transport costs because batteries are heavy, bulky once arranged in packs, and the EVs themselves are challenging to transport because of their weight and fire risk. Finally, there are the subsidies and protectionist rules designed to pull EV and battery production into the consuming countries. With large fixed costs for each new facility, these forces interact in complex ways to determine the equilibrium locations of production and the distribution of sales.

This paper studies how firms endogenously form supply chains over space and how these chains shape the spatial distribution of the EV industry. Firms decide where to build plants and from whom to source inputs at every stage along a supply chain. Characterizing such allocation of production stages to countries is a challenging computational problem for even moderate number of alternative locations. Even with just one stage of production, the facility location problem (as it is referred to in the operations research literature) is already known to be an "NP-hard" problem. Loosely speaking, this means that there are no algorithms guaranteed to solve it in polynomial time, or put more simply

the problem-solving time blows up as the number of locations increases. The difficulty is compounded under multi-stage production because the optimal location of a given stage is not only a function of the marginal cost and fixed cost at that production stage, but is also shaped by the proximity of that location to the desired locations of production of the upstream and downstream stages. One contribution of this paper is to adapt techniques from operations research (OR) to solve for the geography of global supply chains with fixed costs. We extend the multi-stage production cost minimization problem considered in OR to allow for endogenous demand and market entry.

We show in simulations that the way firms respond to government policies depends in complex ways on the geographic structure and parameter choices. Various types of industrial policy can be effective, ineffective, or even counterproductive. Therefore, one cannot make policy recommendations without bringing in detailed data on the costs of distance, borders, and trade agreements. Variable and fixed costs of production need to be estimated. We propose a methodology to do so that uses different decisions the firms make to extract the implied parameters.

Once the model is quantified through data and estimation, we perform counterfactual exercises to determine how various industry policies currently in use are predicted to affect the extent to which various countries participate in domestic, regional, or global supply chains. We also quantify the welfare consequences of these policies. Specifically, we examine the effects of the US Inflation Reduction Act (IRA) which subsidize the US-assembled EVs with batteries sourced from its trade partners. The conditions to qualify for the EV tax credit have raised concerns from major battery makers, including those in China, Japan, South Korea and the EU. In response, the European Union is also considering loosening its rules to allow governments to provide more subsidies for EV manufacturers, leading to a subsidy war that may potentially be wasteful. In addition, the restrictions on cheap input sources also increase the cost of US car manufacturers, leading to a potential welfare loss.

US government officials argue that the transformation of green technology and the global urgency of decarbonization imply that the more subsidies—on both sides of the Atlantic—the better.[1] Analyzing the welfare effects of US IRA alone is not sufficient unless the policy is compared to an alternative with international cooperation. To answer this question, we plan on deriving the global optimal subsidy and predicting the optimal spatial distribution of the EV supply chains.

[1]The *Financial Times* (Feb. 24, 2023) quotes Biden's senior clean energy adviser John Podesta saying "The challenge of dealing with the climate crisis requires ... a transformation of the global economy on a size and scale that's never occurred in human history so there's plenty of room for everybody to participate in that."

The paper is organized as follows. Section 2 positions this paper within the literature on multi-stage production and the role of fixed costs in location choices and sourcing along the value chain. Section 3 describes three key facts about the EV industry that inform core assumptions in the model. Section 4 presents the model of multi-stage production market entry, and equilibrium. In section 5, we verify the computational feasibility and illustrate how policies work within a stripped down model with a single firm. We specify the variable profits associated with each potential path in section 6. The next step is to use discrete choice estimation to quantify the path-level profits. Section 7 develops the estimation framework and presents the estimated and calibrated parameters needed to solve the model using real world data. Section 9 carries out counterfactual policy exercises. Section 10 concludes.

2 Literature

This paper is positioned methodologically at the confluence of two streams in the literature. The first is the modelling of global value chains (GVCs) which we refer to in this paper as multi-stage production. The second literature considers the role of interdependencies in sourcing/location choices at an given stage of production.

Global value chains with constant returns The pioneering papers in the modern GVC literature are Yi (2003) and Yi (2010). With recent important contributions by Antràs and De Gortari (2020), Tyazhelnikov (2022), and Johnson and Moxnes (2023), this literature tackles a first type of interdependence: the presence of trade costs makes production stages interdependent such that simple stage-by-stage minimization of costs does not solve the global problem. These papers all consider multi-stage production with transport costs and comparative advantage and develop recursive methods to find the optimal assignment solution, where cost minimization at one stage accounts for the fact that all preceding stages have already made optimal choices. Those methods work with a constant-returns assumption. Opening a new factory implies some form of fixed cost, and therefore of increasing returns. Hence, to allow for optimal location of factories along the value chain, one needs to tackle a new spatial interdependence *within* each stage.

Plant location/sourcing with fixed costs Tintelnot (2017) pioneers work on multinational location choice with fixed costs. Antràs et al. (2017) consider a sourcing decision but the fixed cost of adding a new supplier country as a source is not so different from the fixed cost of adding a supplier plant in a new country. Given our emphasis on fixed

costs, we focus on that literature. However, in the short run there is a second form of interdependence that can be very important. Castro-Vincenzi (2024) shows how to solve single-stage location models with capacity constraints. This is important in their empirical application: floods that shut down factories. Our paper is oriented towards long run impacts and does not incorporate capacity constraints. We will present evidence to justify this approach in our context.

Plant location with fixed costs is a hard combinatorial problem as the number of potential locations (N) grows. Even with just a single production stage, brute force requires evaluating 2^N alternative solutions. Jia (2008) pioneered using supermodularity of the profit function to reduce computational difficulty. Antràs et al. (2017) pointed out that the global sourcing problem can be supermodular or submodular depending on the relative magnitudes of the key demand and supply parameters, σ and θ. With σ denoting the elasticity of substitution between varieties in the demand function, and θ denoting the Frechet shape parameter on the supply side, supermodularity obtains if and only if $\sigma > 1+\theta$. This restriction holds in Antràs et al. (2017): $\sigma = 3.9$, $\theta = 1.8$. However, Arkolakis et al. (2023) find $\sigma = 4$, $\theta = 4.5$ implying submodularity. They extend the Jia (2008) method to exploit submodularity with a "generalized squeezing" algorithm. Arkolakis et al. (2023) propose a solution that covers cases that are either sub or super-modular *globally*.[2]

Our structure requires a different approach because one mechanism for submodularity and two mechanisms pushing for supermodularity. The submodularity mechanism is that plants at the same stage k substitute for each other. The supermodularity mechanism is that active upstream or downstream production facilities at stages $k \leq K$ complement each other. Furthermore, activated distribution ($K + 1$) facilities increase the profitability of adding production plants. We do not think parametric restrictions to ensure global sub- or super- exist for this setup. One attractive feature of our approach is that it is not necessary to rely on them.

Combining multi-stage production with fixed costs is a challenging problem and there have so far been only a few attempts. de Gortari (2020) and Antràs et al. (2024b) stand out in this respect. The first paper proposes an algorithm for solving GVC models with fixed costs. The proposed solution is elegant and fast, but limits the analysis to a single final good sold in a single destination, reducing considerably the extent of interdependencies. The presence of fixed costs makes it impossible to separate the computation of equilibrium for each product-market combination: The problem must be considered globally to

[2]Oberfield et al. (2024) reduce considerably the computational burden by developing a limit solution where firms choose a density of plants over continuous space rather than making discrete plant location decisions. This limit approach is however more relevant for retail/services where firms have many "plants" serving small catchment areas.

account for the interdependencies. Antràs et al. (2024b) extend Antràs et al. (2017) to allow for many markets to be served by the firm with its potentially many plants. Firms choose three global strategies associated with three *firm-level fixed costs:* one for assembly plants, one for sourcing plants and one for marketing. This enables rich interdependencies *between* each of those margins. However, the authors impose a strong parametric restriction that neutralizes the interdependencies *within* each margin. [3]

Our methodological contribution is to adapt techniques from operations research to solve for the geography of global supply chains with fixed costs, allowing for a range of interdependencies substantially wider than the existing literature. To be very precise, we extend the Multi-level Uncapacitated Facility Location Problem (MUFLP), (Ortiz-Astorquiza et al., 2018, provide a comprehensive survey.) to allow for endogenous demand, multi-product firms and market entry.

3 Facts

This section describes our data and presents some useful facts about the battery EV industry.

3.1 Data on batteries and electric vehicles

We use three main sources of data in this project.

1. The first one is a dataset on the value chain of batteries for electric vehicles produced by the consultancy firm IHS-Markit (now part of S&P). For each vehicle produced in a given assembly plant, the dataset provides the source of the cells, modules, pack with associated respective plants and suppliers (LG, CATL, Panasonic etc.). Additional details are provided regarding the shape of the cell (prismatic, pouch or cylinder), the detailed chemistry (various forms of two main types: NMC and LFP), and the battery capacity in KWh. This is production-based data (years 2015 to 2022), with total volumes of cars assembled associated battery elements, but no information on where the cars are sold.

[3] A recent paper by the same set of co-authors (Antràs et al., 2024a) relax that restriction, allowing for richer interdependencies, but do not propose a method to solve for the firm's optimal strategy. The three papers de Gortari (2020), Antràs et al. (2024b) and (Antràs et al., 2024a) consider a setup in which firms only produce one variety per assembly plant, which sidesteps the thorny question of optimal sharing of production facilities by multi-product firms.

2. The second source is the updated version of the sales data (also from IHS-Markit) used in Head and Mayer (2019). It provides bilateral volumes at the car model level for each of 75 markets, and providing information on the plant of assembly. It also provides information on the manufacturing firm at that assembly plant (which can differ from the brand name, i.e. GM vs Chevrolet). importantly, the car model information does not include fuel type: when a flow of Hyundai Konas appears in a dataset, it sums all fuel types, Gas, Diesel, Hybrid, Plug-in hybrid and BEVs in that case. This is not an issue for models that are only produced as BEVs (the Nissan Leaf for instance), but there is no way to identify those in this source.

3. The third source, still produced by IHS-Markit is called New Registration. This is a destination-based dataset available for 24 markets, where sales (registrations) are reported, with more detail on the car model but no information on the assembly plant. We use the fuel type indicator to identify i) models that are only BEVs, which means that the allocation of production in data source 1) can be safely done with data source 2), ii) models that have mixed fuel types, for which we compute a share of BEVs in total sales over the markets in order to allocate production of the BEV version.

The combination of those three sources help us construct a novel type of micro-level GVC dataset where we can track where the final product is sold (and how much), where it is assembled, and where do the elements of its core value chain (the battery pack) are themselves produced. All plants along the value chain have been geo-coded with latitude and longitude. The data is available for the quasi universe of the nascent BEV industry (none of the major destinations or countries of production are missing), over the period 2015–2022. Frictions that will be used in estimation stages use the usual sources in this literature (CEPII gravity data for distance and RTAs), WITS for tariffs.

3.2 Main markets of BEVs

Table 1 shows the top 15 markets for BEVs in 2023. China is by far the largest market in all dimensions except the share of EVs. Relative to the US it has more than double the number of active firms and almost four times as many models. Many major ICE vehicle-making countries (US, Japan, Canada, Italy) have EV shares under 10%. In the model it will be important to capture both market size in terms of car buyers and pro-EV policies (Three countries with high EV shares—Norway, Sweden, and China—offered early aggressive inducements for buyers to switch to EVs).

Table 1: The 15 top markets of BEVs in 2023

Rank	Country	# Models	# Firms	Sales (000')	EV share (%)
1	China	206	38	5353.0	21.0
2	United States	57	16	1264.3	8.1
3	Germany	124	25	625.8	20.2
4	United Kingdom	104	20	389.7	17.3
5	France	108	20	383.6	17.9
6	South Korea	48	10	181.8	10.7
7	Canada	53	15	160.3	9.4
8	Netherlands	120	24	147.3	33.3
9	Sweden	111	23	143.7	43.1
10	Norway	113	26	132.0	83.7
11	Belgium	111	21	109.4	20.0
12	Japan	49	13	101.0	2.1
13	Australia	51	16	99.8	8.4
14	Italy	105	18	90.5	5.1
15	Thailand	30	13	80.8	10.7
16	Rest Of World	38	11	743.6	7.2

Note: Rest of world row reports averages across 59 countries for numbers of models and EV share and the sum for Sales.

3.3 Location of plants and sourcing along the value chain

Figures 1, 2, and 3 show the impressive growth in the numbers of plants for cells, packs, and vehicles from 2015 to 2022—in all the main regions. The size of plant symbols corresponds to total output in gigawatt-hours for cells and packs and total sales for vehicles.

East Asia remains dominant in cell production but the number of plants in the US and Europe grow by factors of three and two respectively and the capacity in GWH rises by two orders of magnitude. While the growth of cells is impressive, it is also worth noting that many of countries in Figure 3 with multiple vehicle plants (Italy, Spain, Portugal, and Turkey) lack local cell production.

Figure 2 shows packs are intermediate: there are more pack plants than cell plants in every region in both years but fewer pack plants than vehicle plants. We will see that this "fanning out" is a natural feature of the multi-stage production model and does not rely on differences in either transport costs or fixed costs across stages.

The maps show the geographic dispersion of plants but they do not tell us which plants trade with which. Table 2 provides some insight into the distances that components travel. By showing medians and means we see the huge skewness in the distribution of distances that the output of a stage travels before becoming an input. For comparison

Figure 1: Cell plants in the major regions

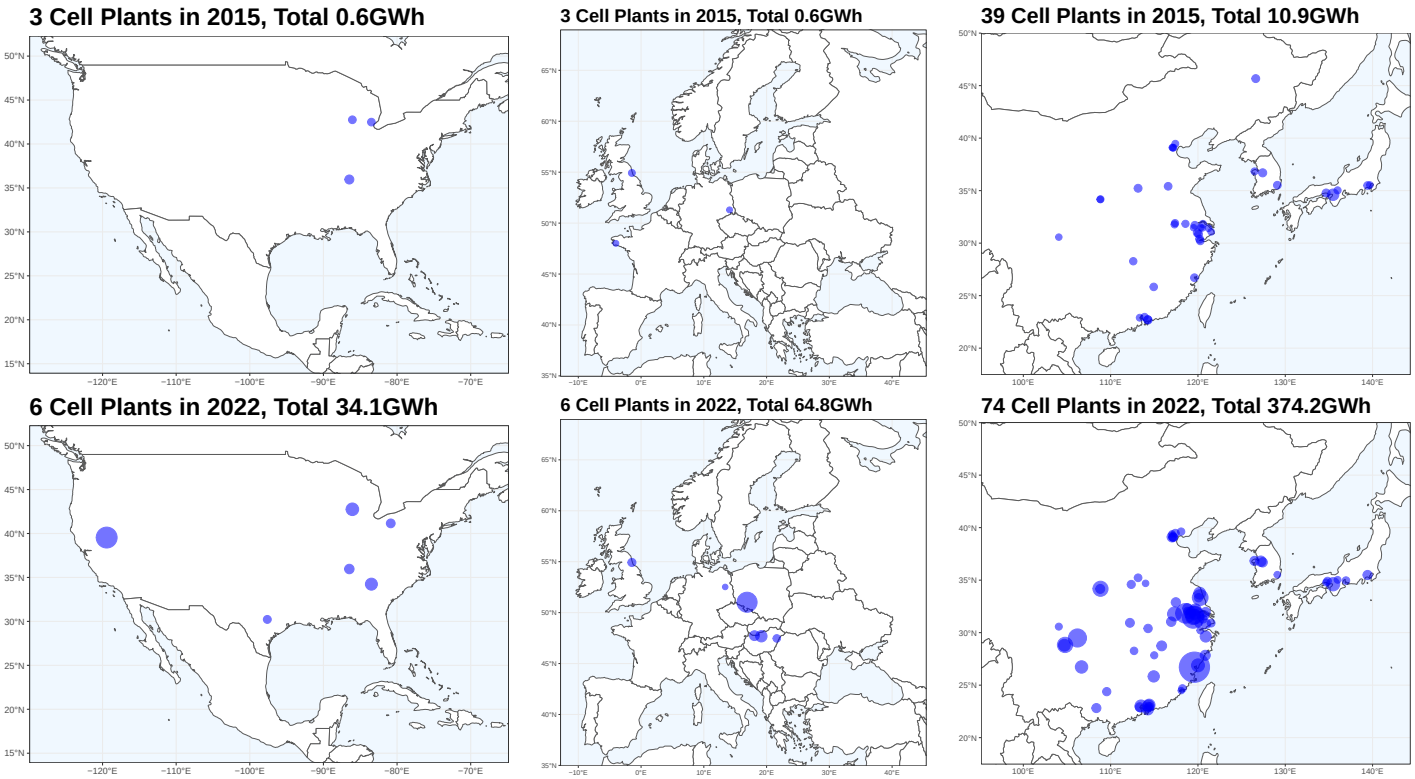


Figure 2: Pack plants in the major regions

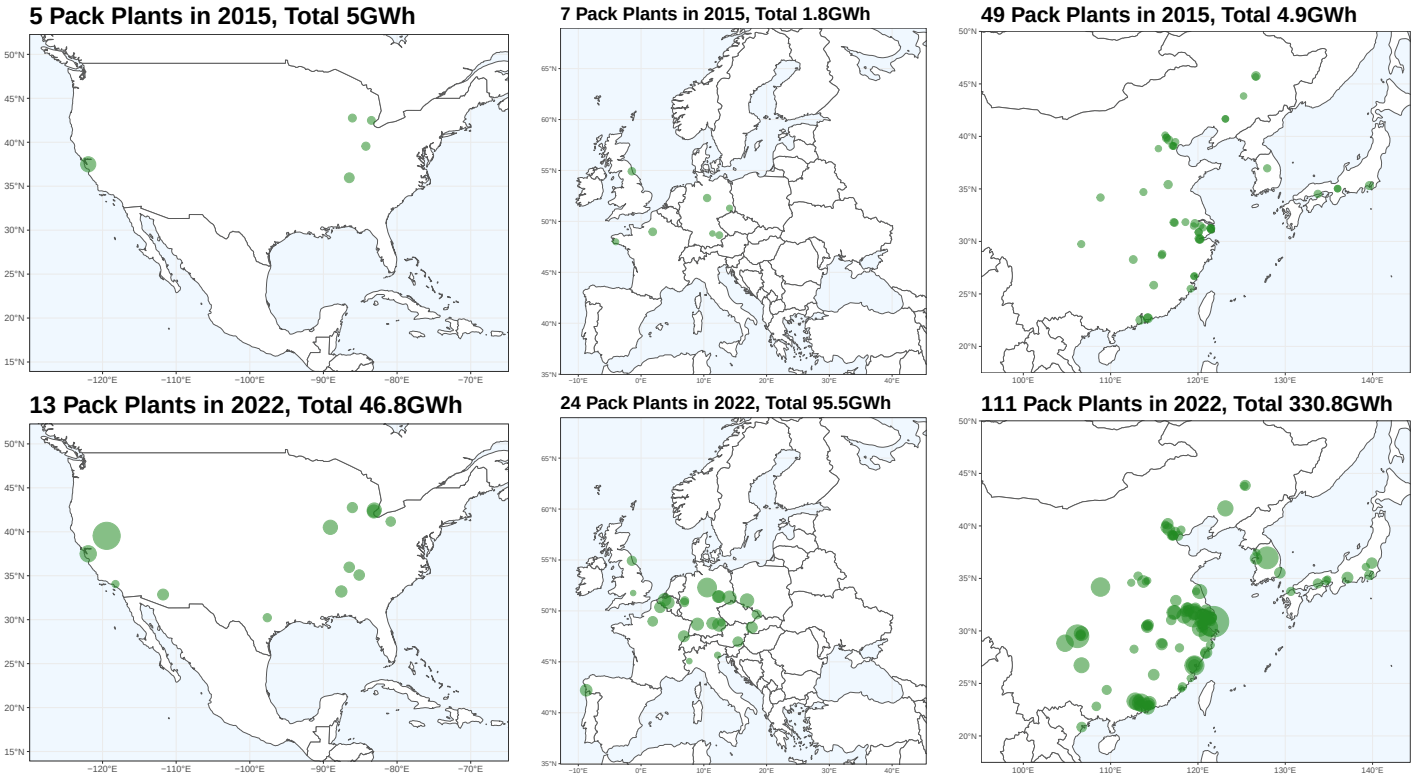


Figure 3: EV plants in the major regions

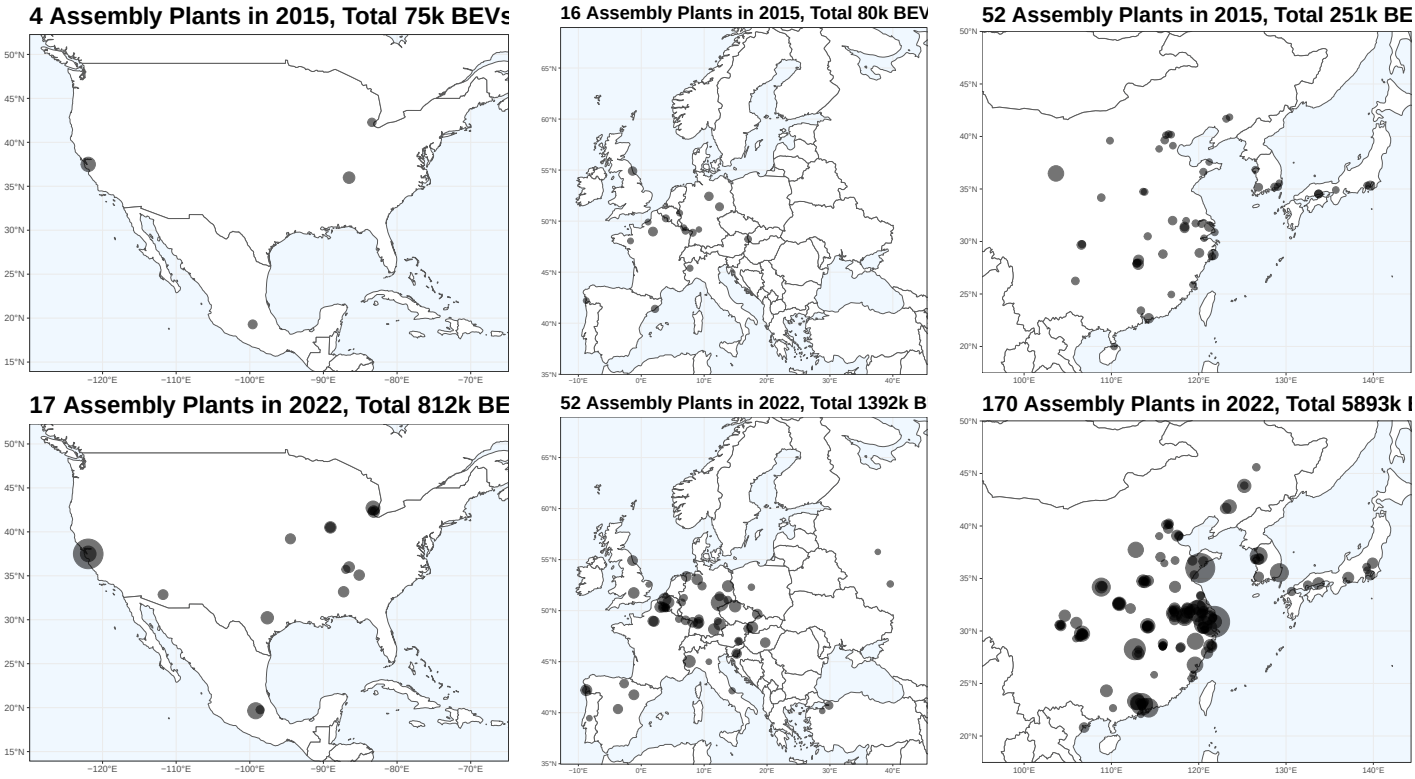


Table 2: How much do components travel?

Link	Year	Distance in km median	mean
Battery Electric Vehicle			
Pack to Assembly	2015	299	819
	2022	215	641
Module to Pack	2015	1	994
	2022	1	830
Cell to Module	2015	13	1782
	2022	1	477
Internal Combustion Engine Vehicle			
Engine to Assembly	2018	133	1034
Transmission to Assembly	2018	681	2184

purposes we also show distance for two of the most important parts of the ICE vehicle: the engine and the transmission. The finding that over 50% of modules travel 1km reflects our decision to code intra-plant distances as 1 to make them easy to recognize. Of the modules that travel long distances a very large share of them are put together in the Cell factory. Also many cells skip the module stage and go straight into packs. For this reason, we will omit modules from our representation of the value chain.

Table 3: Multi-sourcing and multi-supplying are rare

	% multi-sourcing by			% multi-supplying by		
Stage	firms	models	specs	firms	models	specs
Cells	10	0.63	0.17	52.33	21.59	1.71
Packs	8.57	2.86	0.00	52.38	11.49	1.40
Assembly	17.46	2.44	NA	43.06	2.40	NA

Note: The numbers in the first 3 columns are percentages of firms, models or specs that source from more than one country at each stage, conditional on the set of countries used in subsequent stages. A spec for the battery stages consists of a combination pack size (in kWh), cell material (NMC vs LFP) and cell shape (cylinder, pouch or prismatic). The numbers in the last three columns are percentages for regional multi-supplying, i.e. when the firm, model or spec has several plants supplying it for the cells, pack and vehicle. There are three regions in this computation: Asia, North America and Europe.

Table 3 shows that when we disaggregate as much as the data allow, a component in the value chain is almost always single-sourced. This is not the case at more aggregate levels of models and firms. The reason is that there are different types of battery components and a purchaser may obtain one type of cell from one plant and a different type from another plant. Even at the final assembly stage we see that market countries tend to single source, with a narrowly defined vehicle only coming from two different plants 2.44% of the cases. In this respect, EVs do not differ notably from traditional vehicles. Head and Mayer (2019) find multi-sourcing (measured in terms of countries of origin rather than plants) for 2.3% of all model-market years from 2000 to 2018.

This fact is very important for thinking about how to formalize the firm's profit maximization problem. It is common in the trade literature that is applied to more aggregate data to think of each source country as supplying a distinct variety (Armington assumption) with firms using a CES aggregator to combine them. This is not appropriate at the detailed micro level. Instead, our setup realistically assumes that firms choose the lowest cost source for a narrowly defined component or final vehicle.

The last three columns of Table 3 examine a related but different pattern of the data

regarding the frequency of multi-supplying inside a region. For instance, the third column reports that around half of the firms have more than one plant supplying cells, packs and vehicles inside a region (defined as Asia, North America and Europe). This should not be surprising, since even a recently born firm like Tesla has two assembly plants in the US, and uses two cell plants in each of the regions.[4] The figures drop drastically at the car model level. For instance, Renault has 3 plants assembling electric cars in Europe, but each of them supplies the whole of European markets with a different car model.[5] Overall only 2.4 % of EVs are assembled in more than one plant inside a region (the Tesla Model Y in both Fremont and Austin for North America being an example). At the most detailed level (spec accounting notably for battery material), the pattern is the same for cells and packs: Debrecen is the only plant supplying prismatic LMFP cells, and Berlin the only plant supplying cylinder NCM cells intended for two different versions of the battery pack that can equip the German-made Tesla Model Y. Regional multi-supplying is extremely rare at this level of detail.

There is another explanation for multi-sourcing/supplying which is also important to consider in setting up the firm's profit maximization problem: capacity constraints. We might see multi-sourcing/supplying not because of the love of variety common in trade models but simply because the preferred source is unable to meet the demand. However, the numbers in Table 3 suggest that capacity constraints are not a major concern. In line with the data patterns exhibited in the data, we employ an uncapacitated model in this paper.

3.4 Investment costs

The development of the EV industry has generated staggering levels of capital investment in new plants along the value chain. There are no official statistics on these investments, but we have collected everything we could find from news reports. Unfortunately, most articles tend to refer to battery plants without differentiating between cell and pack plants. Also, some reports lump investments in the assembly factories together with the battery investments.

Table 4 reports, based on hundreds of news articles, that on average EV assembly plants involve a $0.79 billion investment and battery plants require $1.85 billion. There is substantial heterogeneity as captured by standard deviations that are similar to the means.

[4] Austin and Sparks in North America, Nanjing and Shanghai in Asia, Berlin and Debrecen in Europe.
[5] The Zoe is supplied by Flins, The Twingo by Novo-Mesto, and the Megane by Douai.

Table 4: Investment costs for battery and EV plants

Stage	Cost ($US bn) Mean	Std. Dev.	Articles Count
Cells	2.53	2.60	83
Packs	0.67	0.71	20
Vehicle Assembly	0.79	0.94	198
Battery and Vehicle	4.09	3.54	13

Source: Authors' calculations based on collected news articles.

Taking the "Battery" investment costs as including both the fixed costs of the cell and pack facilities allows us to produce a ballpark estimate of total investment. There are 234 EV assembly plants and 145 pack plants in the IHS EV supply chain data set. Multiplying by the reported means, total EV investment from 2015 to 2022 amounts to $452 billion.

These massive investment costs strongly suggest that there are indivisible fixed costs that we need to incorporate into the model because they are core features of industry. Firms face a fundamental dilemma regarding activating more plants or incurring higher distance and border costs. The fixed costs relevant for the static model described in the next section differ from these investment costs in two main respects. First, the investments lead to long-lived plants, so we need to multiply by the capital cost (interest plus depreciation) to obtain annual fixed costs. Second, governments have often given large subsidies to these firms. From the firm's point of view the relevant fixed costs are the ones net of subsidies.

4 Multi-product, Multi-stage, Multinational Production

A firm is characterized by a set of models (varieties), $\{m : m \in M_f\}$.[6] A global value chain for a given model has levels of production $\{1, ..., K\}$, where level 1 is the furthest production stage from consumers and level K the closest stage. The set of potential locations where firms can open facilities, $\mathbf{L}$ is partitioned into K levels: $\mathbf{L} \equiv \{L_1, ..., L_K\}$. A particular production location at level k is denoted as ℓ_k. A complete path of production (or value chain) is denoted $\boldsymbol{\ell} \equiv \{\ell_1, \ell_2, \ldots, \ell_K\}$. Final consumption happens at level $K+1$, with the location of final consumers (which we also refer to as markets) denoted $n \in N$. Therefore, a path for a specific market n for model m is given by $\boldsymbol{\ell}_{mn} \equiv \{\ell_1, ..., \ell_K\}_{mn}$.

[6] We do not consider adding or dropping products in this paper.

4.1 The firm's problem

Our model has two types of decisions that a firm must take. The first decision we refer to as an *activation* problem. As in the OR literature firm chooses where to open facilities at each level from $k = 1$ to K. In our framework, the set of clients for the facilities is also another kind of activation decision in which firms decide which models to offer in which markets (stage $K + 1$). The sets of nodes at each stage which paths can follow is given by $\boldsymbol{\mathcal{L}}_f = \{\mathcal{L}_1(f), ..., \mathcal{L}_K(f)\}$. Because of fixed costs, firms do not open facilities in every element of L_k nor serve all markets in N. Constrained by the active facilities and markets, the firm takes its second type of decision: choosing *paths* for each car model from cells to final consumers. We refer to this as the *assignment* problem. The chosen path $\boldsymbol{\ell}_{mn}$ for model m sold in n is constrained by $\boldsymbol{\mathcal{L}}_f$.

Figure 4: Schematic of a supply chain with $K = 3$

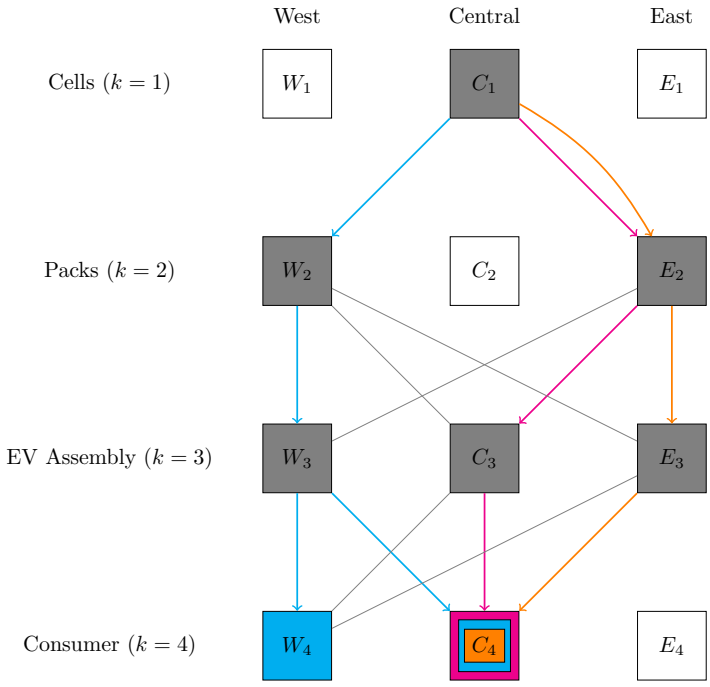


Figure 4 depicts a simplified version of our model to clarify the decisions we are modeling. As in our application, we have $K = 3$, that is three stages of production. In addition, there is the decision to enter a consumer market. The schematic portrays a single firm that makes three models (identified by the colors Cyan, Magenta, and Orange). There are three potential locations: East, Central, and West. In figure 4, Cyan is offered in West and Central; Magenta and Orange are only offered in Central. No model was profitable

to offer in East. The firm simultaneously decides the locations of facilities along the value chain. In this example, only one location (Central) has an active cell plant; there are two Pack plants (Central is not activated at $k = 2$). In our model once a firm activates a production facility, it can be used by more than one model, as we see for E_2. The lines shown in gray are edges that are available but not selected. The decisions of which plants and markets to activate depend on quantities, which are in turn determined by the marginal costs incurred at each node and along each edge.

The firm's problem is to maximize variable profits net of activation costs. Our method relies on the following assumption that ensures that the firm's variable profits are determined by the path it selects.

Assumption 1. *The choice of path $\boldsymbol{\ell}_{mn}$, combined with a market-level aggregator of the paths of its competitors, A_n, are sufficient to determine the variable profits of model m in market n.*

Here the path enters profits via the variable cost function, $c(\boldsymbol{\ell}_{mn})$, so variable profits can be expressed as $\pi(c(\boldsymbol{\ell}_{mn}), A_n)$.[7] Aggregators, often specified as price indexes, exist for linear, logit, and CES demand curves under monopolistic competition. Note that it is also possible to extend this model to oligopoly. [8] Furthermore, the path could affect variable profits through channels such as demand (if consumers care directly about sourcing).

For a chosen path $\boldsymbol{\ell}_{mn}$, the firm also incurs fixed costs $\phi_{\ell k}$ in location ℓ for each stage of production k (presumably, a substantial portion of these are sunk once a facility is open). Note that these fixed costs are not indexed by model: Once a facility is open in ℓ for production stage k, we assume that it can accommodate any of the firm's models. We denote this facility location choice with a binary variable $y_{\ell_k} \in \{0, 1\}$ (the firm identifier f is suppressed to simplify notation).

Let ϕ_{mn} be the fixed market entry cost of selling model m in market n. This includes the cost of marketing, sales training, and space for display. The firm will then sell all models for which the variable profit $\pi(c(\boldsymbol{\ell}_{mn}), A_n)$ exceeds this cost. We denote those market entry choices with binary variables $z_{mn} \in \{0, 1\}$. Firm's total profits sums variable profits over all of its models m and served markets n, and subtracts the fixed costs of activating facilities and entering markets.

In the following section, we describe how the firm maximizes global profits by finding an optimal set of paths $\boldsymbol{\ell}_{mn}$, together with determining all the binary $y_{\ell k}$ facility location choices and z_{mn} market entry choices.

[7] For simplicity, we express profits as a function of costs only. Later on we will add a model-market mn demand shifter, but its impact on variable profits will be isomorphic to a cost shifter.

[8] In this case, markups become a function of the market share which depends on the paths of all other models sold in market n.

4.2 Mixed integer linear programming formulation

The problem of the firm described in the previous subsection can be expressed as a linear optimization problem with integer constraints, for which the Operations Research (OR) literature has developed very efficient algorithms. More precisely, the OR framework closest to ours is the Multi-level Uncapacitated Facility Location Problem (MUFLP),[9] to which we add two features: multi-product firms and endogenous choice of market entry. In a typical MUFLP, there are two types of decisions involved. One is the facility location decision y_{ℓ_k} at each level of production. The other is the assignment of chosen paths across those locations and stages of production for a given product to a given set of clients. In the optimized decision, $x_{mn\boldsymbol{\ell}}$ equals one for the path $\boldsymbol{\ell} = \{\ell_1, \ell_2, \ldots, \ell_K\}$ chosen for model m to market n and is zero for all alternative paths. If there are L_k potential locations with production stages $k = 1...K$, then there are a total of $L_1 \times ... \times L_K$ indicator variables $x_{mn\boldsymbol{\ell}}$ for a given product m and location n. Multiplying by the number of products and number of locations yields the total number of x variables.[10] On top of considering multiple products for each firm, we further adapt this optimization problem to incorporate the market entry decisions z_{mn}. Because our problem has Multi-stage, Multi-product and Multi-national features we refer to it as MMM-UFLP. Setting $K = 3$ as in our application, profit maximization for a given firm f can be written as the following mixed integer linear program:

$$\max \sum_{m \in M_f} \sum_{n \in N} \sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} \pi(c(\boldsymbol{\ell}_{mn}), A_n) x_{mn\ell_1\ell_2\ell_3} - \sum_{k=1}^{3} \sum_{\ell_k \in L_k} \phi_{\ell_k} y_{\ell_k} - \sum_{m \in M_f} \sum_{n} \phi_{mn} z_{mn} \quad (1)$$

[9] We draw on the Ortiz-Astorquiza et al. (2018) survey of MUFLP formulations.

[10] The operations research literature refers to this MUFLP formulation as a path-based optimization problem. There is also an alternative arc-based formulation, which defines the x variables for "arcs" from one stage k to stage $k+1$. We use the path-based formulation as this delivers a linear optimization problem in the x and y variables, even when the variable profit is non-linear in delivered marginal cost $c(\boldsymbol{\ell}_{mn})$ and aggregator A_n.

subject to

$$\sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq z_{mn}, \quad n \in N, m \in M_f \tag{2}$$

$$\sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_3}, \quad n \in N, m \in M_f, \ell_3 \in L_3 \tag{3}$$

$$\sum_{\ell_1 \in L_1} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_2}, \quad n \in N, m \in M_f, \ell_2 \in L_2 \tag{4}$$

$$\sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_1}, \quad n \in N, m \in M_f, \ell_1 \in L_1 \tag{5}$$

$$x_{mn\ell_1\ell_2\ell_3} \geq 0, \quad n \in N, m \in M_f, \ell_k \in L_k, k = 1, 2, 3 \tag{6}$$

$$y_{\ell_k} \in \{0, 1\}, \quad \ell_k \in L_k, k = 1, 2, 3 \tag{7}$$

$$z_{mn} \in \{0, 1\}, \quad n \in N, m \in M_f. \tag{8}$$

The inequalities defined in (2) are *activity constraints* that govern market entry: the left-hand-side sum will be equal to 1 if the model m is sold in market n (only a single $x_{mn\ell_1\ell_2\ell_3}$ path is set to 1 at the optimum), or will be equal to 0 otherwise (all the $x_{mn\ell_1\ell_2\ell_3}$ paths are set to 0). In the standard OR formulation of MUFLP with cost minimization and no market entry, these constraints are defined with an equal sign and named as flow-conservation constraint, meaning that the outflow (unit demand) of each model to a market must equal to its inflow via supply chains. However, in our setting of profit maximization, the equality constraints can be relaxed as inequalities because if the firm has paid the fixed market entry cost, i.e., $z_{mn} = 1$, it is only profit maximizing to actually produce and sell this model in the market via some supply chains to gain variable profit, even if the firm has the option to not do so.[11]

The inequalities (3)-(5) are *activity constraints* that govern whether a given facility is open: the left hand side sum will be equal to 1 if a path through a facility in ℓ_k is used (only a single $x_{mn\ell_1\ell_2\ell_3}$ path is set to 1 at the optimum), or will be equal to 0 otherwise (all the $x_{mn\ell_1\ell_2\ell_3}$ paths are set to 0). These constraints guarantee that a facility must be active to be used in supply chains.

Equations (7) and (8) restrict facility activation and market entry to be binary. Some statements of the uncapacitated FLP explicitly constrain the $x_{mn\ell_1\ell_2\ell_3}$ to be binary. However, this is not needed here. The x need to be non-negative to ensure that paths with negative variable profits would not be chosen. Given the other constraints, it is suffi-

[11]Computationally, Gurobi handles inequality constraints more efficiently because it invokes a "sifting" algorithm which can be effective on problems with many more variables than equations. Sifting solves a sequence of LP subproblems where the results from one subproblem are used to select columns from the original model for inclusion in the next subproblem.

cient in equation (6) to restrict the assignment variables to be positive. While fractional assignments are not explicitly excluded, the x are binary at the optimum.[12]

5 Single-model Policy Simulation

We simulate a production chain of a single firm with one car model. There are three stages of production (cell, pack, vehicle assembly for concreteness) in a world of 40 locations[13] (production and consumption sites). The coordinates of each site are randomly drawn from the unit interval. Demand for the representative model is Logit (with an outside good: ICE vehicles). The number of vehicle buyers in each consumption location rises with the square root of latitude, thus the further north the more buyers.

To make costs observable in the maps we draw, the delivered marginal cost of the final good is given by its Euclidean distance from stage 1 to the final consumer: $c(\boldsymbol{\ell}_{mn}) = \tilde{d}_{\ell_1,\ell_2} + \tilde{d}_{\ell_2,\ell_3} + \tilde{d}_{\ell_3,n}$, where $\tilde{d}_{ab} = [(\text{lat}_a - \text{lat}_b)^2 + (\text{lat}_a - \text{lat}_b)^2]^{1/2}$. Fixed costs of establishing plants at any stage are set to be the same for all locations and stages. Fixed costs of market access are drawn randomly from a uniform distribution.

The low and high fixed costs to revenue ratios implied by the parameter settings in the simulation are 6.2% and 16%. Converting the up-front investments from Table 4 into fixed costs per year at rates or 5% or 10%, we compute fixed cost to revenue ratios for the EV industry as 6% to 11%. Our revenue calculation assumes an average price of $50,000 per EV for the 8.1 million EVs assembled worldwide in 2022. These "back of the envelope" calculations suggest that the fixed cost to revenue ratios in the simulation are in the right ballpark.

Each of the mapped simulation below takes about 20–40 seconds[14] to solve the firm's location problem, depending on the computer and the levels of fixed costs and subsidies. This establishes that the mixed integer programming solver can handle realistically sized problems in a reasonable amount of time.[15]

We begin the investigation of the MUFLP by simulating the model without any subsidies for EVs. The left panel of Figure 5 shows the case of low fixed costs. We see the three key features of the model: (i) there are more plants in the North, reflecting the higher de-

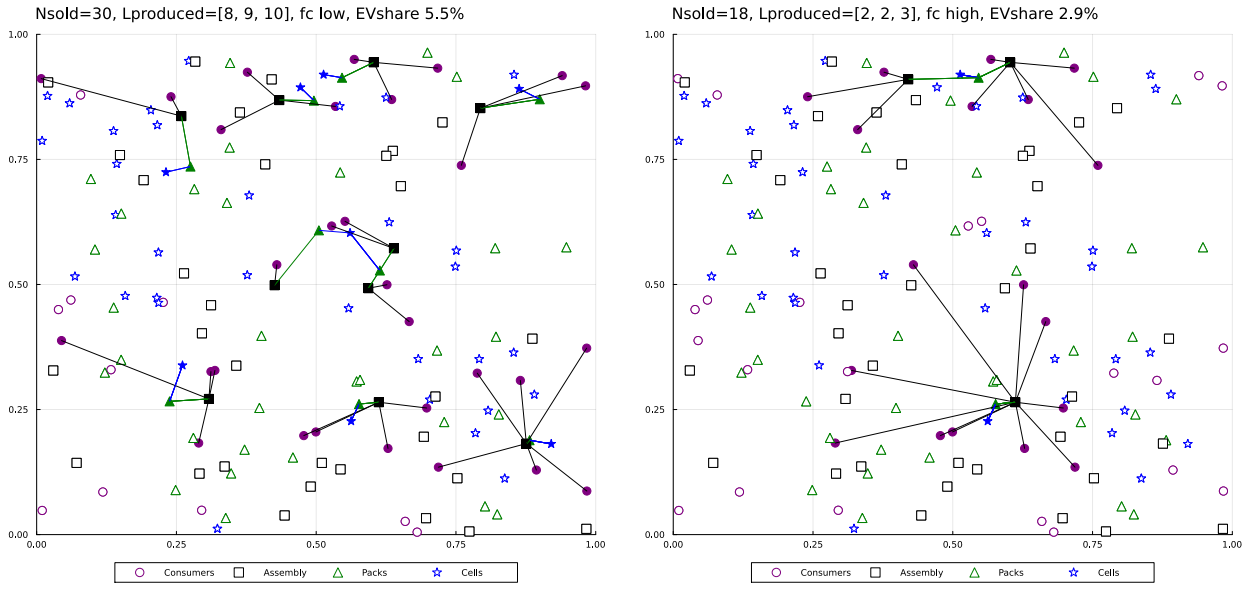
[12] Intuitively, it makes no sense under constant marginal costs to source from a high (delivered) cost plant when a lower cost plant is available. Formally the assignment problem is *totally unimodular*.

[13] As of 2021, 38 countries are involved in one or multiple stages of the production of EVs.

[14] The program can be solved an order of magnitude more quickly in the arc-based formulation (ABF) described in Appendix **??**. Unfortunately, while this method minimizes cost given quantity and market entry decisions, it is not guaranteed to maximize profits.

[15] The code, which will be made available online, is written in Julia using the JuMP package to interface with the solver.

Figure 5: Simulation: no interventions



(a) Low fixed costs

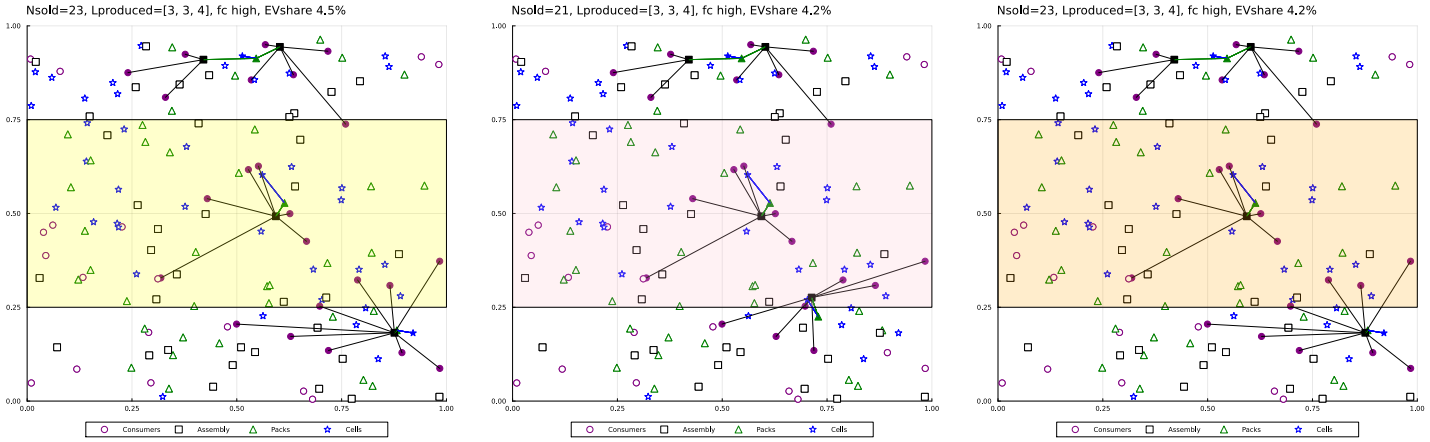
(b) High fixed costs

mand, (ii) for any given latitude, plants tend to be central to be closer to more consumers, (iii) peripheral and southern consumers are less likely to be served (one quarter of the consumers do not have sufficient demand to justify the fixed costs of serving them). In the right panel, we multiply fixed costs by five. This reduces the number of cell plants by a factor of four and the number of assembly plants by a factor of three. The greater distances from assembly to consumer lower the number markets served from 30 to 18. The combination of higher marginal costs and reduced number of markets lowers the share of EVs in the market from 5.5% to 2.9%.

Figure 6 brings consumer subsidies into the picture, continuing the high fixed cost case from Figure 5. The three frames correspond to three different types of rules for which vehicles qualify for the subsidy. In panel (a) any consumer in the designated region—shown in yellow—qualifies for the $5,000 dollar subsidy. This resembles the way credits are allocated in Canada and many European countries. In panel (b), the subsidy is only available for vehicles assembled in the region. This is similar to the situations in China and the US where there are consumer subsidies coupled with large tariffs on imported EVs. Panel (c) takes its inspiration from the IRA provisions requiring the whole value chain of batteries to come from countries with whom the US has a free trade agreement.

The unconditional subsidy in Middle (panel a) leads to a new assembly plant being

Figure 6: Simulation: $5,000 subsidies in Middle



(a) Unconditional subsidy (b) Final-assembly (c) Full-path

opened in the South, and the existing plant in the Middle is replaced by a more centrally located plant. Five new consumer sites are activated, and the EV share rises from 2.9% to 4.5%. Making the subsidy eligible only to regionally produced vehicles leads assembly site serving Southern consumers to be relocated to the lower part of Middle. Two markets in the South are closed due to higher marginal costs. The EV share falls to 4.2%. Panel (c) makes qualification rules even stricter. The pack plant used in panel (b) would disqualify vehicle sales from the plant in the southern part of Middle. It is replaced by a new plant in the South, which leads to reactivating two Southern consumers. There are conflicting effects on affordabilty that leave the overall market share of EVs unchanged. This case illustrates the Laffer curve effect identified by Head et al. (2024): stricter rules can be counterproductive in terms of domestic production.

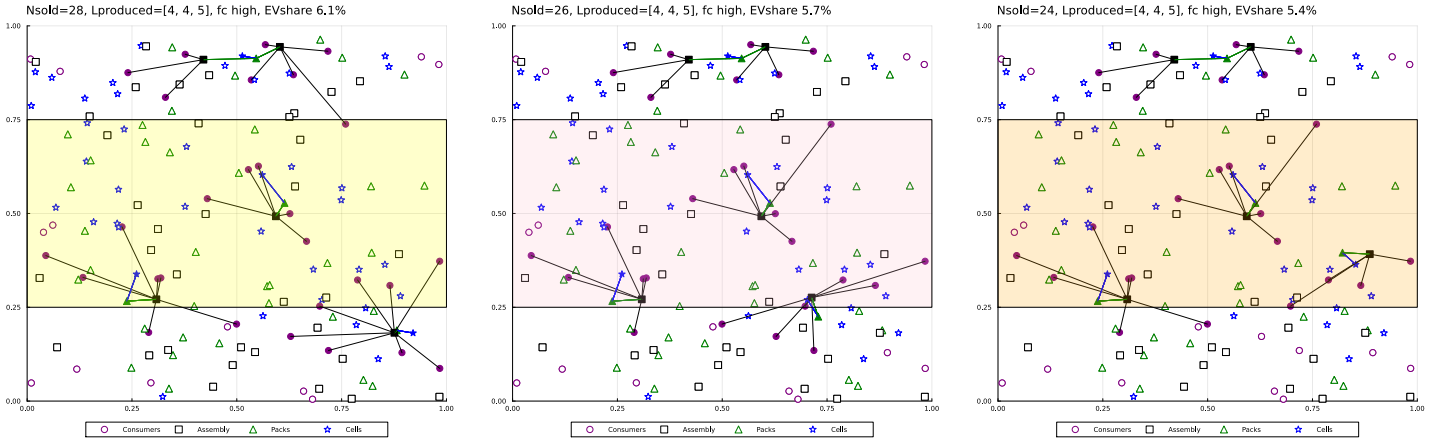
The takeaway from this series of simulations is that in a world where fixed costs are important, subsidies can reshape the geography of production but do so in ways that are not easy to predict, either quantitatively or even qualitatively. To obtain a better idea of the effects of the gamut of recent policies, we are going to need parameter estimates grounded in the data.

6 Mapping paths to profits

The MMM-UFLP setup from section 4 is very general but we need to make more specific modelling choices before we can estimate the parameters and conduct realistic policy simulations.

One important simplification we make for estimation is to reduce it to just the two

Figure 7: Simulation: $7,500 subsidies in Middle



(a) Unconditional subsidy (b) Final-assembly (c) Full-path (IRA)

most important stages of production: cell manufacture and vehicle assembly. While we have seen in flat-world that MILP is fully capable of solving three-stage production problems, estimation of simulated moments via a global solver (TikTak) requires us to solve the problem over and over again. Another simplification we make here is to take model entry (z) as given by the data. This is just a short-cut to try to get a first set of estimates and counterfactuals.

6.1 Demand

While much of the literature on the car industry uses nested logit or mixed logit demands, we use a constant elasticity of substitution (CES) form. The main difference between the two is that while logit demand has price responsiveness α parameter that multiplies price, CES has an elasticity η that multiplies log price. In the discrete choice derivation of logit demand, α corresponds to the marginal utility of income and $1/\alpha$ is the limiting value of the *additive* markup as market shares become small. There are two problems with this here: first, poorer countries should have higher marginal utilities of income and second markups of say $10,000 seem reasonable for a Tesla priced at $50,000 but not for Chinese model priced at $15,000. These issues could be addressed with a household-specific α as is common in mixed logit models, but we would lose the tractability needed for our facility-location problem. One convenient feature of CES is that the average own-price elasticity is widely reported in the literature. Table 7 provides 18 different estimates, taken mainly from the Industrial Organization literature. We calibrate $-\eta$ to the median of these estimates.

Each country spends R_n^{EV} on battery-powered EVs, with R_n^{EV} determined by *inter alia* charging stations, local regulations regarding emissions, gas prices, etc. Market shares of model m depend on the ratio of the model's quality-adjusted price p_{mn}/ξ_{mn} to the price index for EVs, P_n^{EV}.[16] The price of model m in country n is adjusted to incorporate any tariffs or consumer subsidies (such as tax credits). The resulting demand is given by

$$q_{mn} = R_n^{\text{EV}} \left(P_n^{\text{EV}}\right)^{\eta-1} \xi_{mn}^{\eta-1} p_{mn}^{-\eta}, \tag{9}$$

where P_n^{EV} is a CES index of firm-level quality-adjusted price index of models offered in n (models for which an entry dummy z_{mn} turns 1):

$$P_n^{\text{EV}} = \left(\sum_f \left(P_{fn}^{\text{EV}}\right)^{1-\eta}\right)^{1/(1-\eta)}, \text{ with } P_{fn}^{\text{EV}} \equiv \left(\sum_{m \in M_f} z_{mn} (p_{mn}/\xi_{mn})^{1-\eta}\right)^{1/(1-\eta)}. \tag{10}$$

Assume that ICE varieties substitute for EV varieties using the same CES η as other EV varieties. Then we have

$$\frac{R_n^{\text{EV}}}{R_n} = \frac{(P_n^{\text{EV}})^{1-\eta}}{(P_n^{\text{EV}})^{1-\eta} + (P_n^{\text{ICE}})^{1-\eta}}, \tag{11}$$

where R_n is total spending on vehicles of all fuel types and is treated as exogenous and P_n^{ICE} is the price index for ICE vehicles, defined analogously to P_n^{EV}.

Under monopolistic competition, each firm is sufficiently small that it takes the aggregator, P_n to be unaffected by its price (which itself depends on the choice of path). With CES demand (9), firms apply a constant markup $\mu_{mn} = \mu = \eta/(\eta-1), \ \forall \ n, m$ to their delivered marginal cost $c(\boldsymbol{\ell}_{mn})$ for all their models: $p_{mn} = \mu c(\boldsymbol{\ell}_{mn})$. The resulting variable profit for each model is given by

$$\pi_{mn} = \frac{c(\boldsymbol{\ell}_{mn}) q_{mn}}{\eta - 1}, \tag{12}$$

where q_{mn} is defined in (9) as a function of $c(\boldsymbol{\ell}_{mn})$ and P_n^{EV}. This completes the specification of $\pi(c(\boldsymbol{\ell}_{mn}), P_n^{\text{EV}})$ and establishes that variable profits depend only the path and the aggregator ($A_n = P_n^{\text{EV}}$), thus complying with Assumption 1. The next step is to flesh out the cost function $c(\boldsymbol{\ell}_{mn})$.

[16] We show how to calibrate the ξ_{mn} in section 8.3.

6.2 Path costs

Going back to the general specification of the firm's problem, our first step is to choose functional forms for the costs that will determine assignment and location decisions. Any production function that gives final costs as function of the chosen path can work for the firm's optimization problem. Leontief and Cobb-Douglas yield cost functions that are readily suitable for estimation by discrete choice sourcing regressions because of their separability. These cases can be seen as first-order approximations for more general cost (or log cost in the case of Cobb-Douglas) functions. We follow Yi (2003), Yi (2010), Antràs and De Gortari (2020), and Johnson and Moxnes (2023) in using Cobb-Douglas across stages here, but the appendix shows the Leontief version. We prefer Cobb-Douglas because it estimates cost differences as percentages, and it works well with *ad valorem* duties (Leontief's additive structure is more compatible with specific duties).[17]

The variable costs for a particular chain depend on three key aspects: (i) a location-stage production cost shifter, $w^k_{\ell_k}$, comprising wages and other local material costs, (ii) trade costs $\tau_{\ell_K n}$ between consumers and assembly, and $\tau_{\ell_k \ell_{k+1}}$ between two adjacent levels of production, (iii) model-specific cost shocks attributable to bilateral locations $\varepsilon^k_{m\ell_k\ell_{k+1}}$.[18]

The following two equations provide the marginal costs cumulating up to stage k:

$$c^1_{m\ell_1\ell_2} = w^1_{\ell_1}\tau^1_{\ell_1\ell_2}\varepsilon^1_{m\ell_1\ell_2}, \tag{13}$$

$$c(\boldsymbol{\ell}_{mn}) = \left(w^2_{\ell_2}\right)^{\alpha_{22}} \left(c^1_{m\ell_1\ell_2}\right)^{\alpha_{12}} \tau^2_{\ell_2 n}\varepsilon^2_{m\ell_2 n}, \tag{14}$$

where $\alpha_{kk} \in (0,1)$ denotes the cost share of local factor at stage k and $\alpha_{(k-1)k} \in (0,1)$ denotes the cost share of input from last stage $(k-1)$ to this stage k. The above costs consider those incurred at production facilities and along transport edges. Costs at the final consumption stage—sales taxes, buyer subsidies, variable costs of distribution—can be incorporated in $\tau^3_{\ell_3 n}$.

Equation (12) translates an estimate of path costs—derived from (14)—into profits, which is the key requirement of the MUFLP. The drawback of this formulation of the

[17] Leontief would seem the natural choice when combining stages of production. Thus, each car needs two axles and four wheels. Cobb-Douglas allows for labor at an upstream stage to substitute for labor at a downstream stage. One way to justify this is in terms of subassemblies. In our case cells are usually combined into modules that are then combined into packs. The module and packs stage could occur at the cell factory, at some intermediate location, or the vehicle factory. The former case would increase labor input upstream relative to the last case which requires more labor at the assembly factory. Cobb-Douglas also implies constant cost shares across stages. This is appealing for the case of BEVs because as the size, power and range of a vehicle rises, its battery capacity requirements increase proportionately.

[18] The subscript m appears in the shock term and explains why sourcing decisions vary across models. We will provide parametric assumptions for the ε later when we estimate these cost equations using discrete choice.

profit equation is that it also depends on a price index P_n. If we calibrated P_n with actual prices, it would be in different units from the path costs (which are only identified up to a scalar since they are based on differences in log costs). We substitute path costs into equation (10) to express the price index in terms of an index of the path costs of all competing models:

$$P_n^{\text{EV}} = \mu C_n^{\text{EV}}, \quad \text{where} \quad C_n^{\text{EV}} \equiv \left(\sum_{m' \in \text{EV}_n} [c(\boldsymbol{\ell}_{m'n})/\xi_{m'n}]^{1-\eta} \right)^{1/(1-\eta)}, \tag{15}$$

where EV_n is the set of EV models sold in market n (that is: all m such that $z_{mn} = 1$). Now profits can be expressed as a function of the cost index:

$$\pi(c(\boldsymbol{\ell}_{mn}), C_n^{\text{EV}}) = \left(\frac{c(\boldsymbol{\ell}_{mn})/\xi_{mn}}{C_n^{\text{EV}}} \right)^{1-\eta} \frac{R_n^{\text{EV}}}{\eta}. \tag{16}$$

The factor R_n^{EV}/η represents EV industry variable profits in market n. The fraction in front of it shows how relative quality-adjusted cost determines each EV model's market share. The unidentified scalar cancels out in this fraction.

In the next section, we delineate a method to estimate the parameters for variable costs in equations (13)–(14). Applying a system of sourcing regressions, we can get the estimates of $w_{\ell_k}^k$ and $\tau_{\ell_k \ell_{k+1}}^k$ for all locations at each level of production (up to a normalization), and estimates of the parameters driving the distribution of the random cost terms ε^k. These estimates for $k = 1, ..., K$ will be used, together with simulated ε draws, to construct equation (16), and then solve the optimal GVC for each firm-model-consumer, given guesses of fixed costs.

7 Estimation method

To estimate the parameters of the model we divide the problem into two main pieces. First, we use the assignment (or sourcing) decisions in the data (which are conditional on open markets and facilities) to estimate the variable costs of production. The second stage estimates a parsimonious set of fixed-cost parameters to match moments in the MMM-UFLP solution to corresponding moments in the data. Our focus here is the cost-side of the firm's problem, so we calibrate the demand side shifters to fit the sales data, drawing on the literature for the own-price elasticity.

7.1 Conditional choice estimation of variable costs

Taking logs and substituting each upstream cost into the next stage's cost equation, the cost of the complete production path is

$$\ln c(\boldsymbol{\ell}_{mn}) = \alpha_{22} \ln w^2_{\ell_2} + \ln \tau^2_{\ell_2 \ell_3} + \alpha_{12}(\ln w^1_{\ell_1} + \ln \tau^1_{\ell_1 \ell_2}) + u(\boldsymbol{\ell}_{mn}), \tag{17}$$

where $u(\boldsymbol{\ell}_{mn}) = \alpha_{12} \ln \varepsilon^1_{\ell_1 \ell_2} + \ln \varepsilon^2_{\ell_2 \ell_3}$ is the structural interpretation of the error term. The probability of choosing path $\boldsymbol{\ell}_{mn}$ is given by the product of the conditional probabilities at each stage:

$$\mathbb{P}(\boldsymbol{\ell}_{mn}) = \mathbb{P}_{\ell_1|\ell_2} \times \mathbb{P}_{\ell_2|n}.$$

Breaking down the probability into these logarithmically separable probabilities greatly facilitates estimation.

The literature has taken three approaches to deriving the choice probabilities at each stage. The most straightforward approach is to ignore the common components in the path-level error and model $u(\boldsymbol{\ell}_{mn})$ as the log of a Fréchet shock, as in Antràs and De Gortari (2020) and Allen and Arkolakis (2022). The best known method is based on McFadden (1978), modelling $u(\boldsymbol{\ell}_{mn})$ as single shock with a generalized extreme value distribution that allows for correlated errors. Finally, one can model the components of $u(\boldsymbol{\ell}_{mn})$ directly. Cardell (1997) and (Ben-Akiva and Lerman, 1985, p. 287) take this approach, and it was also adopted by Berry (1994). Regardless of which set of error term assumptions we use, the stage-level choice probabilities can be expressed using the same functional forms that we show below.

7.1.1 Estimating equations

Conditional on the set of level-1 facilities established by a firm, $\mathcal{L}_1(f)$, it chooses for each model the location with the lowest cost to serve ℓ_2 (due to constant returns to scale in the variable costs of production). We can therefore use equation (13) to derive the conditional probability of ℓ_1 being the cost-minimizing location of supplying to ℓ_2 as

$$\mathbb{P}^1_{\ell_1|\ell_2} = \frac{\left(w^1_{\ell_1} \tau^1_{\ell_1 \ell_2}\right)^{-\theta_1}}{\Phi^1_{\ell_2}}, \text{ with } \Phi^1_{\ell_2} = \sum_{\ell_1 \in \mathcal{L}_1(f)} \left(w^1_{\ell_1} \tau^1_{\ell_1 \ell_2}\right)^{-\theta_1}. \tag{18}$$

Our estimating equation is

$$\mathbb{P}^1_{\ell_1|\ell_2} = \exp\left[\text{FE}^1_{\ell_1} + \text{FE}^1_{\ell_2} + \beta^1_D \ln D_{\ell_1 \ell_2} + \beta^1_t \ln\left(1 + t^1_{\ell_1 \ell_2}\right)\right], \tag{19}$$

where $D_{\ell_1\ell_2}$ is a measure of bilateral "distance" (any observable bilateral friction), $t^1_{\ell_1\ell_2}$ is the bilateral tariff for stage 1, with β^1_D and β^1_t being their respective regression coefficients. The term $\text{FE}^1_{\ell_1}$ is an ℓ_1 fixed effect, and

$$\text{FE}^1_{\ell_2} = -\ln\left\{\sum_{\ell_1\in\mathcal{L}_1(f)} \exp\left[\text{FE}^1_{\ell_1} + \beta^1_D \ln D_{\ell_1\ell_2} + \beta^1_t \ln\left(1+t^1_{\ell_1\ell_2}\right)\right]\right\}$$

is the chooser fixed effect. The correspondence between estimated coefficients and structural variables is

$$\begin{aligned} w^1_{\ell_1} &= \exp\left[-\text{FE}^1_{\ell_1}/\theta_1\right] \\ \tau^1_{\ell_1\ell_2} &= \exp\left\{-\left[\beta^1_D \ln D_{\ell_1\ell_2} + \beta^1_t \ln\left(1+t^1_{\ell_1\ell_2}\right)\right]/\theta_1\right\} \\ \Phi^1_{\ell_2} &= \exp\left[-\text{FE}^1_{\ell_2}\right] \end{aligned}$$

Note that since tariffs are direct cost shifters, they can be used to provide a direct estimate of θ_1 (as in Head and Mayer, 2019). Here, we do not restrict β^1_t to be equal to $-\theta_1$, a flexibility allowing for unmeasured bilateral trade costs that are correlated with the tariff $t^1_{\ell_1\ell_2}$.

An important component of the nesting approach to estimating production costs along the value chain is the integration of the cost of upstream parts in downstream decisions. The Nested Logit approach implied from the assumptions of Ben-Akiva and Lerman (1985) replaces the actual cost of cells (which includes the error term) with the expected cost of the optimal source of the cells. With extreme value distributions, this expected cell cost for pack plant ℓ_2 is

$$\mathcal{E}c^1_{.,\ell_2} = \left(\Phi^1_{\ell_2}\right)^{-1/\theta_1}. \tag{20}$$

We refer to this term as the *inclusive cost* which depends on the sourcing potential of ℓ_2 and differs across firms due to various choice sets $\mathcal{L}_1(f)$.

Moving down the supply chain and using (14), the probability that ℓ_2 is selected conditional on ℓ_3 being chosen is

$$\mathbb{P}^2_{\ell_2|\ell_3} = \frac{\left[\left(w^2_{\ell_2}\right)^{\alpha_{22}}\left(\mathcal{E}c^1_{.,\ell_2}\right)^{\alpha_{12}}\tau^2_{\ell_2\ell_3}\right]^{-\theta_2}}{\Phi^2_{\ell_3}}, \text{ with } \Phi^2_{\ell_3} = \sum_{\ell_2\in\mathcal{L}_2(f)}\left[\left(w^2_{\ell_2}\right)^{\alpha_{22}}\left(\mathcal{E}c^1_{.,\ell_2}\right)^{\alpha_{12}}\tau^2_{\ell_2\ell_3}\right]^{-\theta_2}. \tag{21}$$

The estimating equation of assembly plants when choosing among potential pack facilities is

$$\mathbb{P}^2_{\ell_2|\ell_3} = \exp\left[\text{FE}^2_{\ell_2} + \text{FE}^2_{\ell_3} + \beta^2_D \ln D_{\ell_2\ell_3} + \beta^2_t \ln\left(1+t^2_{\ell_2\ell_3}\right) + \beta^2_\Phi \text{FE}^1_{\ell_2}\right], \tag{22}$$

where $\text{FE}^1_{\ell_2}$ was estimated in equation (19), and accounts for determinants of the upstream (cells) costs.[19]

The second decision to consider is the one made by a market n when sourcing the EV from an assembly plant. The probability that ℓ_2 is selected as the location of assembly is[20]

$$\begin{aligned}
\left(w^2_{\ell_2}\right)^{\alpha_{22}} &= \exp\left(-\text{FE}^2_{\ell_2}/\theta_2\right) \\
\tau^2_{\ell_2 n} &= \exp\left\{-\left[\beta^2_D \ln D_{\ell_2 n} + \beta^2_t \ln\left(1 + t^2_{\ell_2 n}\right)\right]/\theta_2\right\} \\
\alpha_{12} &= -\theta_1 \beta^2_\Phi / \theta_2 \\
\Phi^2_n &= \exp\left[-\text{FE}^2_n\right]
\end{aligned}$$

Again, note that we do not restrict β^2_t to be equal to θ_2 due to unmeasured bilateral trade costs that are potentially correlated with the tariff $t^2_{\ell_2 n}$.

The Antràs and De Gortari (2020) approach of a single independent path-shock would imply $\theta_1 = \theta_2$ in the two probability equations.[21] As we estimate the three equations separately, we do not impose this coefficient restriction. It would be difficult to test in any case due to the unknown α parameters that go into the the β_Φ estimates. We use the equivalence (in terms of likelihood functions) between Poisson (with fixed effects) and multinomial logit estimators (Guimaraes et al., 2003), which enables us to estimate equations (19) and (22) as high-dimensional fixed effects PPML regressions.

Define a vector $\boldsymbol{\kappa} = \{\kappa_1, \kappa_2\}$. Parameter κ_2 scales up and down the relative importance of marginal costs relative to fixed costs, whereas κ_1 provides the weights to use on the cell stage relative to vehicle assembly.

$$\begin{aligned}
\ln c(\boldsymbol{\ell}_{mn}) = -\kappa_2 \Big\{ &\text{FE}^2_{\ell_2} + \beta^2_D \ln D_{\ell_2 \ell_3} + \beta^2_t \ln\left(1 + t^2_{\ell_2 \ell_3}\right) \\
&+ \kappa_1 \left[\text{FE}^1_{\ell_1} + \beta^1_D \ln D_{\ell_1 \ell_2} + \beta^1_t \ln\left(1 + t^1_{\ell_1 \ell_2}\right)\right] \Big\},
\end{aligned} \tag{23}$$

where $\kappa_2 = -1/\beta^2_t = 1/\theta_2$, and $\kappa_1 = \beta^3_\Phi \beta^2_\Phi = \alpha_{23}\alpha_{12}\left(\theta_3/\theta_1\right)$. We expect for reasons outlined Ben-Akiva and Lerman (1985), that $\theta_2 > \theta_1$. Consequently, κ_2 will be bounded between 0 and α_{23}. Thus, κ_1 will be bounded between 0 and $\alpha_{23}\alpha_{12}$. This upper bound is just the

[19] For stages $k = \{2, 3\}$, we can write more generally that $\text{FE}^{k-1}_{\ell_k}(m) = -\ln \Phi^{k-1}_{\ell_k}(m) = -\ln \sum_{\ell \in L_{k-1}(m)} \exp[\text{FE}^{k-1}_\ell + \boldsymbol{X}'_{\boldsymbol{\ell \ell_k}} \boldsymbol{\beta^{k-1}_\tau} + \beta^{k-1}_\Phi \text{FE}^{k-2}_\ell(m)]$.

[20] Based on the stage-2 Logit specification above, we will recover $\beta^2_\Phi \text{FE}^1_{\ell_2} = -\alpha_{12}\theta_2 \ln \mathcal{E}c^1_{\cdot,\ell_2}$. Substituting $\ln \mathcal{E}c^1_{\cdot,\ell_2} = -\ln \Phi^1_{\ell_2}/\theta_1$ and $\text{FE}^1_{\ell_2} = -\ln \Phi^1_{\ell_2}$ derived from the stage-1 specification, this yields $\alpha_{12} = -\theta_1 \beta^2_\Phi / \theta_2$.

[21] As noted in the appendix of Antràs and De Gortari (2020) the IID path shock also implies that the product of each stage's conditional probability collapses to a single conditional logit formula, referred to as the joint logit in Ben-Akiva and Lerman (1985).

cost share of the battery cell in the total cost of the vehicle.

We can estimate the $\boldsymbol{\kappa}$ in equation (23) by repeatedly solving MMM-UFLP for different values of this vector until a loss function of the data is minimized. Alternatively, or as starting values, we can use the vehicle tariff to estimate $\theta_1 = \theta_2 = -\beta_t^2$ and take reported values of costs, $\alpha_{23} = 0.4$ and $\alpha_{12} = 0.75$, yielding $\boldsymbol{\kappa} = \{0.4 \times 0.75, 0.4, -1/\beta_t^2\}$.

8 Estimation sample and results

8.1 Filters applied before estimation

The simulated method of moments is a computationally intensive and greedy in terms of data requirements. To keep the time manageable (less than 24 hours), we estimate the model on a subset of the data.

1. Restrict to 15 large multinationals. This removes mainly Chinese firms that serve the Chinese market only.
2. Restrict to 24 countries that have good data on EV sales and prices. This is needed for calibrating demand parameters.
3. Restrict to 139 models offered by the 15 large firms, with price data available in 2022. This eliminates some light trucks for which we lack price data.
4. Restrict estimation to 2021–2023, the most recent years we have complete data.

Our filter of keeping only the top 15 firms retains a large share of the world's sales. In our data there are 50 firms that assemble EVs in 2022 for whom we have sales by market. Of those, 12 sell exclusively in China. Table 5 shows that the top 15 firms—ranked by number of sales destinations—account for 99% of all EVs sold outside China (but inside the 24 countries used in the simulation, which account for 97% of world sales). Because the Chinese market is so large, these firms account for a smaller share (90%) of the 24-country world. With the exception of three firms, the top firms sell in 17 or more markets in the 24-country world.

8.2 Sourcing estimation results

Table 6 combines estimation results for the two stages considered in our estimation of cost parameters for the BEV industry: battery cells and assembly. These estimates show

Table 5: 15 top firms in 2022

No.	Manufacturer	# Markets	# Models	Sales Cum. Share (%)	Sales-exCHN Cum. Shr (%)
1	Tesla	23	4	20.5	28.0
2	Volkswagen	24	23	30.0	42.4
3	Hyundai	23	15	35.6	54.1
4	Stellantis	18	19	40.3	63.7
5	BMW	24	7	44.0	69.7
6	Renault	19	6	46.7	75.3
7	Mercedes-Benz	24	9	49.3	80.0
8	Geely	24	15	55.6	84.4
9	Ford	22	4	57.5	88.3
10	Nissan-Mitsubishi	23	8	60.1	91.9
11	SAIC	17	21	72.8	95.0
12	General Motors	7	7	74.1	96.6
13	Toyota	23	8	74.9	97.8
14	Rivian	3	3	75.3	98.6
15	BYD	10	14	89.6	98.9

Note: Market shares defined over 24 countries used in simulations.

variable costs of production and delivery of cells to assembly sites, and of assembled vehicles to dealerships in the 24 markets.

Figure 8 uses the estimates of country fixed effects to reveal relative costs for each stage. This is obtained as $w_{\ell_k} = -\text{FE}_{\ell_k}/\theta_k$, and normalized with respect to the USA in each k. The competitiveness of China is visible for cells. Mexico, Korea, and Belgium are identified as very good places for EV assembly as opposed with China being revealed as the least cost competitive in assembly. This is due to the fact that the electric Volvos and Audis assembled in Belgium sell in many places in the world (not only EU), when most Chinese-assembled EVs are sold locally.

Consider two examples of countries that have high estimated variable costs in regressions that include all source countries. The Hyundai electric Kona produced in India is sold only in India, when the Korean-made one is sold in 59 destinations, and the figure for the Czech-made one in 39. The case of Malaysia is also quite telling since it assembles electric Volvos (the same as in Belgium) since 2022 that are sold only in Malaysia, and a few in Thailand.[22] In 2022, the Volkswagen electric SUV ID.4 is produced in 4 plants: 2 in Germany, 1 in the USA, 2 in China. Each of the German plants serves at least 29 markets, the US plant 2 markets, and the 2 Chinese plants serve only China. Those FE estimates

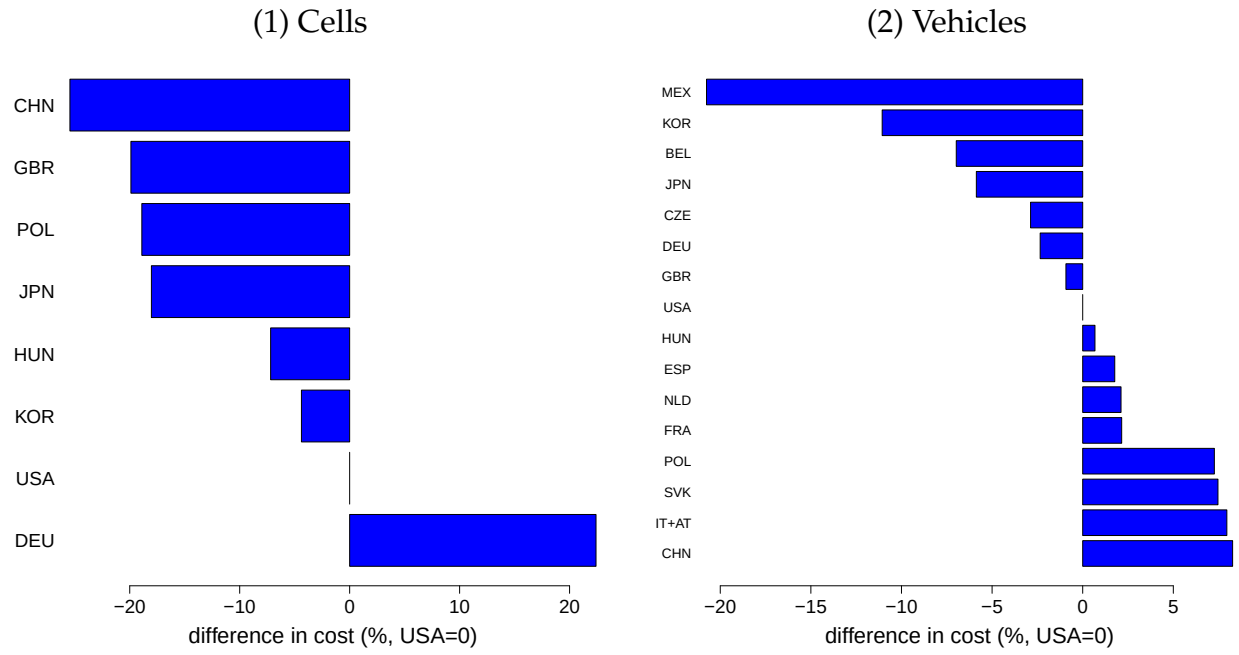
[22] `https://paultan.org/2023/02/03/` "Volvo Car Malaysia to begin exporting..."

Table 6: Combined sourcing regression results

Stage:	(1) Cells	(2) Vehicle
border	-0.746^b	-0.426^a
	(0.347)	(0.139)
log distance	-0.287^b	0.099
	(0.121)	(0.062)
intraplant	0.129	
	(1.02)	
log(1+tariff)	-5.45^b	-10.3^a
	(2.15)	(1.65)
cross-continent	-1.35^b	-0.634^a
	(0.571)	(0.192)
log GDP per capita	0.594^a	0.116
	(0.179)	(0.079)
EC(cells)		-0.217^a
		(0.083)
Observations	6,688	34,776
Squared Correlation	0.165	0.080

Note: EC(cells) are the expected costs of cells for assemblers, $\mathcal{E}c^1_{\cdot,\ell_2}$, defined in equation (20).

Figure 8: Relative costs in the value chain of BEVs



reflect those patterns: while the situation is evolving rapidly for China, in our 2015–2022 sample most of the Chinese plants still serve only China. Other consumers in the world prefer to source their BEVs from alternative plants, which translates into high estimated costs.

8.3 Calibration of demand parameters

We calibrate the demand parameters using an estimate of η from the literature. The calibration of ξ_{mn} demand shifters uses IHS model-level car prices in each market, as well as the sales of EV and ICE cars.

The following inversion[23] gives us ξ_{mn} as a ratio of appeal relative to the geometric mean of EV models (as always in this paper we exclude plug-in hybrids) in market n. Let $\tilde{p}_n$ and $\tilde{q}_n$ denote the geometric means of EV prices and quantities in market n. Then inversion of (9) yields:

$$\breve{\xi}_{mn} \equiv \frac{\xi_{mn}}{\tilde{\xi}_n} = \left(\frac{p_{mn}}{\tilde{p}_n}\right)^{\frac{\eta}{\eta-1}} \left(\frac{q_{mn}}{\tilde{q}_n}\right)^{\frac{1}{\eta-1}}. \tag{24}$$

With an estimate of η, we can then compute the relative appeal, $\breve{\xi}_{mn}$ of each EV model in each market. Table 7 provides a number of average elasticities from a set of papers that estimate demand for cars. The left column gives figures for papers focusing on EVs, while the right column does not distinguish between EVs and traditional gas-powered vehicles, and therefore mostly covers ICEs. In table 7, we see that several recent papers estimate price elasticities. The median over all 18 studies is 4.

We can now back out C_n^{EV} from the data using the $c(\boldsymbol{\ell}_{mn})$ obtained from the sourcing regressions. We plug in the $\breve{\xi}_{mn} \equiv \xi_{mn}/\tilde{\xi}_n$ into equation (15) and divide by $\tilde{\xi}_n$. Along with our estimates of the $c(\boldsymbol{\ell}_{mn})$ costs and η, we can now compute variable profits for every potential path: $\pi(c(\boldsymbol{\ell}_{mn}), C_n^{\text{EV}})$.

8.4 SMM estimation of fixed facility costs

In this section, we estimate fixed costs parameters $\boldsymbol{\rho}$. This is a vector of six parameters which allows for fixed costs to be determined at the continent-stage level. We parameter-

[23] Our approach here is closest to Hottman et al. (2016). It is an intellectual descendant of the Berry (1994) inversion but here we used quantities relative to an inside good. Khandelwal (2010) also backs out appeal in a logit demand, but he uses a regression approach with instruments for prices. Khandelwal et al. (2013) imposes a price responsiveness parameter as we do, but then estimates appeal across markets as a residual from a fixed-effects regression.

Table 7: Price responsiveness of car demand from recent literature

EV-only	**elas.**	**Mainly ICEV**	**elas.**
Barwick et al. (2024)	4.2	Beresteanu and Li (2011)	8.4
Kwon (2023)	4.4	Castro-Vincenzi (2024)	4.3
Li et al. (2022)	3.7	Colon and Gortmaker (2020)	3.9
Li (2023)	3.7	Coşar et al. (2018)	14.9
Li et al. (2017)	1.3	Goldberg (1995)	3.3
Linn (2022)	5.3	Goldberg and Verboven (2001)	5.2
Muehlegger and Rapson (2022)	2.1	Grieco et al. (2024)	5.4
Springel (2021)	1.8	Head and Mayer (2019)	3.9
Xing et al. (2021)	2.8	Li (2018)	9.5
Median	3.7	**Median**	5.2
Overall Median	4.0		

ize fixed costs of plant activation as draws from log normal distributions, $\phi_{\ell_k} \sim \text{LN}(\boldsymbol{\rho}, \widehat{\sigma_k})$ We currently use the empirical distribution of investment costs from the collected news articles as our estimate of the standard deviation of log fixed costs at each stages of production.

This section also incorporates estimation of the relative marginal cost parameters $\boldsymbol{\kappa} = \{\kappa_1, \kappa_2, \kappa_3\}$ as in equation (23). We use tariff-based θ_k estimates from table 6 and cost shares from a consultancy to provide the starting values of these parameters.

We draw u shocks and fixed costs $S = 10$ times. For each set of draws, we can compute the objective function of MMM-UFLP in its estimable version, $\Pi_f(\boldsymbol{\kappa}, \boldsymbol{\rho})$, given estimated and calibrated parameters from previous steps indexed by $\hat{}$ and normalized variables indexed by $\breve{}$. Inside $\Pi_f(\boldsymbol{\kappa}, \boldsymbol{\rho})$, the variable profit component is

$$\pi_{mn\boldsymbol{\ell}}(\boldsymbol{\kappa}) = \left(\frac{\breve{c}_{mn\boldsymbol{\ell}}/\xi_{mn}}{\breve{C}_n^{\text{EV}}} \right)^{1-\eta} \frac{R_n^{\text{EV}}}{\eta}, \tag{25}$$

where

$$\begin{aligned} \breve{c}_{mn\boldsymbol{\ell}} = \exp\Bigg[\kappa_2 \Big\{ & \left[\widehat{\text{FE}_{\ell_2}^2} + \widehat{\beta_D^2} \ln D_{\ell_2\ell_3} + \widehat{\beta_t^2} \ln\left(1 + t_{\ell_2\ell_3}^2\right) \right] \\ & + \kappa_1 \left[\widehat{\text{FE}_{\ell_1}^1} + \widehat{\beta_D^1} \ln D_{\ell_1\ell_2} + \widehat{\beta_t^1} \ln\left(1 + t_{\ell_1\ell_2}^1\right) \right] \Big\} \Bigg] \end{aligned} \tag{26}$$

$$\breve{C}_n^{\text{EV}} \equiv \left(\sum_f [\breve{C}_{fn}^{\text{EV}}]^{1-\eta} \right)^{1/(1-\eta)} \tag{27}$$

$$\breve{C}_{fn}^{\text{EV}} \equiv \left(\sum_{m' \in M_f} z_{m'n} [\breve{c}_{m'n\ell} / \xi_{m'n}]^{1-\eta} \right)^{1/(1-\eta)}, \tag{28}$$

and $z_{m'n}$ uses *observed entry* of models sold in destination markets, R_n^{EV} also from observed data, and $\eta = 4$. The tariff at last stage is m-specific because some cars are PC and some are LCV. However, all PC models who has the same ξ_{mn} and the same origin country will predict to have identical supply chains.

For one set of parameter guess $\{\boldsymbol{\kappa}, \boldsymbol{\rho}\}$, we solve MMM-UFLP S times and obtain a simulated estimate of the model outcome. Stacking them into a vector of $g(\boldsymbol{\kappa}, \boldsymbol{\rho})$, these moments are distances between the respective model outcome and the observed outcome in the data. The calibration's objective is to

$$\min_{\boldsymbol{\kappa}, \boldsymbol{\rho}} g(\boldsymbol{\kappa}, \boldsymbol{\rho})' g(\boldsymbol{\kappa}, \boldsymbol{\rho}). \tag{29}$$

To identify $\boldsymbol{\kappa}$ which are determined by the ratio of trade elasticities across stages, we match the trade share in quantities across regions at different stages. At the cell stage, it is the share of BEVs assembled in each region that used cells produced from respective regions. At the pack stage, it is the share of BEVs assembled in each region that used packs produced from respective regions. At the car stage, it is the share of BEVs sold in each region that are assembled from respective regions.

To identify the overall magnitude of fixed costs $\boldsymbol{\rho}$, we match the share of firms out of the top 15 that have established facilities of any stage in all regions, and the share of cars out of the 138 models that are sold to all regions. If we also need to estimate the dispersion parameter of the fixed market entry cost in this stage, then we can use the share of models with quality below the median that are sold to all regions to pin it down. The intuition is that if there are zero dispersion in fixed market entry costs across models, firms would decide market entry decision entirely based on their models' variable profits. For models that have higher quality (defined by the geometric mean of $\breve{\xi}_{mn}$ across all markets n), variable profits will be higher holding price fixed, and hence, entering in more markets. An increase in dispersion of fixed costs draws, even if i.i.d., would break the ordering of market entry, leading to disproportionately high or low share of models in one bin to the total share.

Figure 9: Fit of the estimated model: production lines by continent

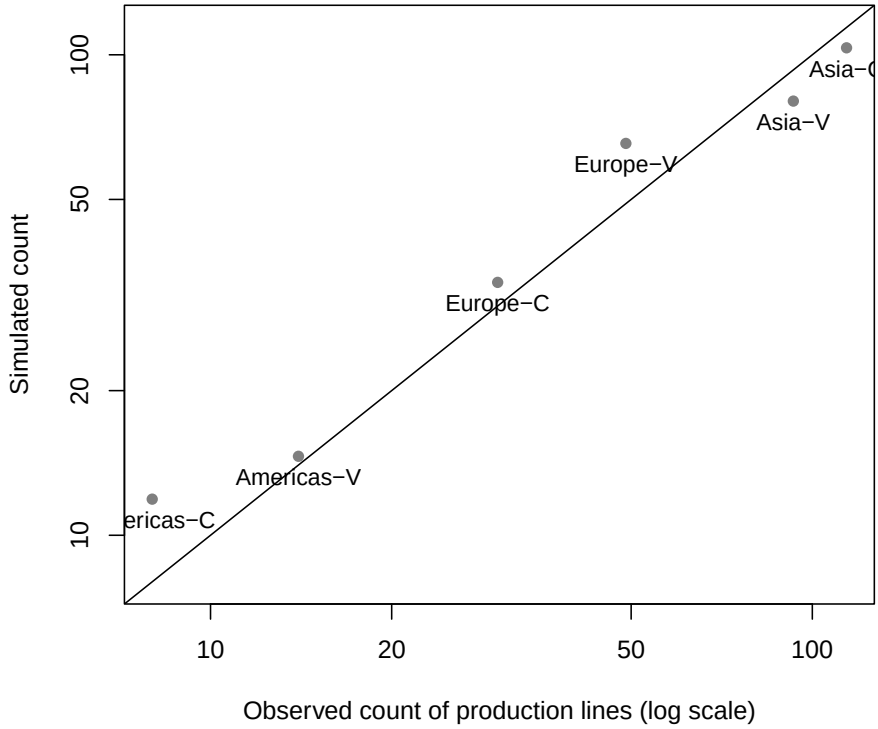
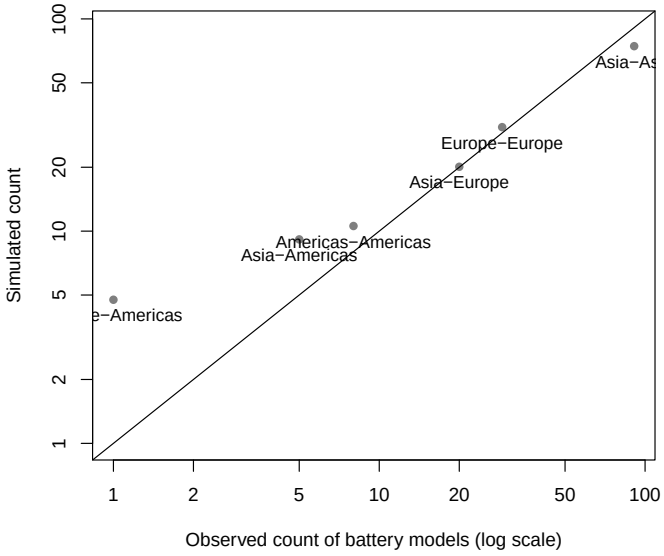
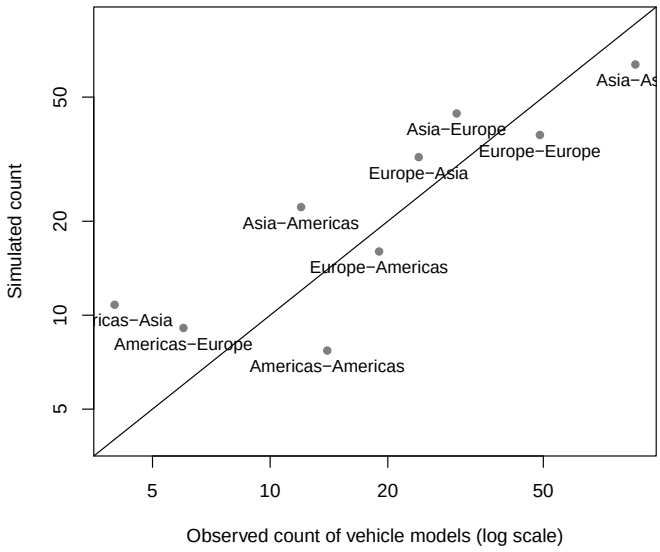


Figure 10: Fit of the estimated model: origin-destination model sourcing

(1) battery cells



(2) vehicle assembly



Notes: Targeted moments in SMM.

9 Counterfactuals

There are a large number of policies currently in use to incentivize EV purchases and to shift the location of production. While we plan to expand the set of policies evaluated in future versions of this paper, we currently consider the same three policies we investigated in the more stylized "flat world." Recall that those policies are

1. **Subsidy to North American EV buyers** The policy we consider is a 20% subsidy to buyers (the government rebates 20% of the cost of the car). These credits apply to all consumers from a given continent, and are available for all vehicles regardless of the locations of assembly and battery manufacture.

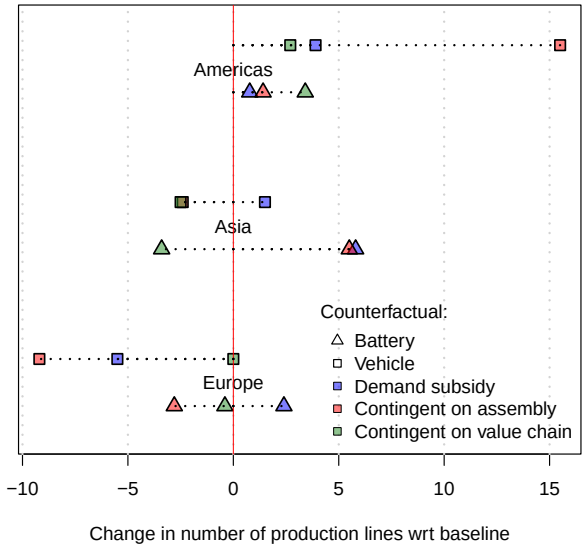
2. **Subsidy conditional on domestic assembly** Only cars assembled in the same continent as the buyer are eligible for the subsidy.

3. **Subsidy conditional on domestic cell-vehicle supply chain** To qualify both cell production and vehicle assembly must be in the same continent as the buyer.

9.1 Policy Simulation Results

Figure 11: Counterfactual results with 20% subsidy by continent

(a) Continental production lines



(b) Expenditures on EVs

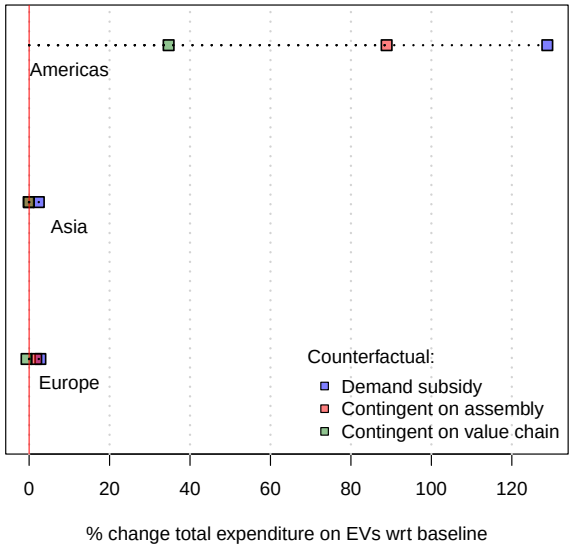
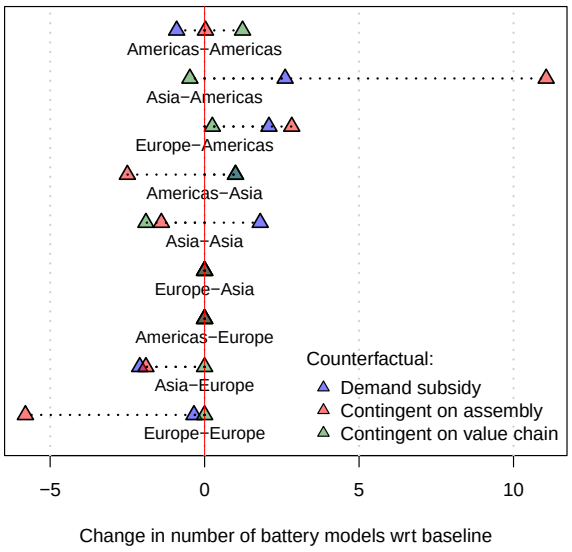
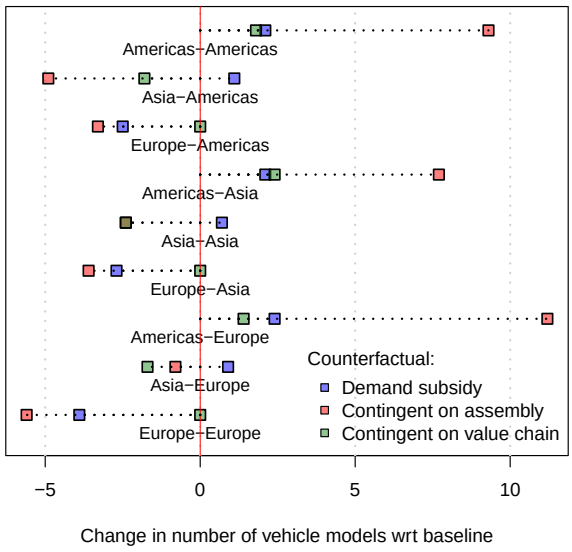


Figure 12: Counterfactual results with 20% subsidy: by origin-destination

(1) Battery Cells



(2) Vehicle assembly



10 Conclusion

To be added when the estimation and counterfactuals are completed.

References

Allen, T. and C. Arkolakis (2022). Welfare effects of transportation infrastructure improvements. *Review of Economic Studies 89*(6), 2911–2957.

Antràs, P. and A. De Gortari (2020). On the geography of global value chains. *Econometrica 88*(4), 1553–1598.

Antràs, P., T. C. Fort, and F. Tintelnot (2017). The margins of global sourcing: Theory and evidence from US firms. *American Economic Review 107*(9), 2514–64.

Antràs, P., E. Fadeev, T. C. Fort, and F. Tintelnot (2024a, January). Export-platform FDI: Cannibalization or complementarity? Working Paper 32081, National Bureau of Economic Research.

Antràs, P., E. Fadeev, T. C. Fort, and F. Tintelnot (2024b, 04). Exporting, global sourcing, and multinational activity: Theory and evidence from the United States. *The Review of Economics and Statistics*, 1–48.

Arkolakis, C., F. Eckert, and R. Shi (2023, November). Combinatorial discrete choice: A quantitative model of multinational location decisions. Working Paper 31877, National Bureau of Economic Research.

Ben-Akiva, M. E. and S. R. Lerman (1985). *Discrete choice analysis: theory and application to travel demand*, Volume 9. MIT press.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, 185–213.

Castro-Vincenzi, J. (2024). Climate hazards and resilience in the global car industry. mimeo.

de Gortari, A. (2020). Global value chains and increasing returns. Online note.

Guimaraes, P., O. Figueirdo, and D. Woodward (2003). A tractable approach to the firm location decision problem. *The Review of Economics and Statistics 85*(1), 201–204.

Head, K. and T. Mayer (2019). Brands in motion: How frictions shape multinational production. *American Economic Review 109*(9), 3073–3124.

Head, K., T. Mayer, and M. Melitz (2024). The Laffer curve for rules of origin. *Journal of International Economics 150*, 103911.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

Jia, P. (2008). What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Johnson, R. C. and A. Moxnes (2023). GVCs and trade elasticities with multistage production. *Journal of International Economics 145*, 103796.

Khandelwal, A. (2010). The long and short (of) quality ladders. *The Review of Economic Studies 77*(4), 1450–1476.

Khandelwal, A. K., P. K. Schott, and S.-J. Wei (2013). Trade liberalization and embedded institutional reform: evidence from chinese exporters. *American Economic Review 103*(6), 2169–95.

McFadden, D. (1978). Modeling the choice of residential location. In A. Karlqvist (Ed.), *Spatial Interaction Theory and Planning Models*, pp. 75–96. Elsevier.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2024). Plants in space. *Journal of Political Economy 132*(3), 867–909.

Ortiz-Astorquiza, C., I. Contreras, and G. Laporte (2018). Multi-level facility location problems. *European Journal of Operational Research 267*(3), 791–805.

Tintelnot, F. (2017). Global production with export platforms. *The Quarterly Journal of Economics 132*(1), 157–209.

Tyazhelnikov, V. (2022). Production clustering and offshoring. *American Economic Journal: Microeconomics 14*(3), 700–732.

Yi, K.-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of political Economy 111*(1), 52–102.

Yi, K.-M. (2010). Can multistage production explain the home bias in trade? *The American Economic Review 100*(1), 364–393.

A Data Appendix

A.1 IHS HVBD

A.2 IHS Sales

A.3 IHS New Registrations (NRQ,NRP)

We can obtain ξ_{mn} only for models with prices (the NRP data). The raw data from IHS often express model names slightly differently from the HVBD and Sales data. After some effort to match the model names across IHS modules, we have the merged data has 99% overlap in sales for the set of 15 big firms. Another problem is that some models in the NRQ have Model recorded as the brand name with "Unspecified" as the model name. There are also some Brand names (Make) that are unspecified. Collectively, these cases account for very small shares of the total sales in NRQ, under 2% in all markets in 2022, and 0.5% in all markets except China and Germany. These observations are dropped.

B Additional empirical results

Table B1: Cell Sourcing decision

Dependent Variable:	Cell sourcing decision					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
border	-0.973^a	-0.662^b	-0.642^b	-0.741^a	-0.709^b	-0.746^b
	(0.300)	(0.270)	(0.270)	(0.273)	(0.281)	(0.347)
log distance	-0.281^a	-0.184^a	-0.191^a	-0.185^a	-0.156^a	-0.287^b
	(0.051)	(0.036)	(0.037)	(0.042)	(0.045)	(0.121)
intraplant		0.849^c	0.841^c	0.812	1.11^b	0.129
		(0.498)	(0.501)	(0.530)	(0.470)	(1.02)
cross-continent		-1.68^a	-1.70^a	-1.51^a	-1.52^a	-1.35^b
		(0.392)	(0.394)	(0.429)	(0.447)	(0.571)
log GDP per capita			0.257^a	0.425^a	0.498^a	0.594^a
			(0.069)	(0.055)	(0.064)	(0.179)
log(1+tariff)	-6.47^b	-4.48^b	-4.55^b	-3.71^c	-6.32^a	-5.45^b
	(2.58)	(2.03)	(2.04)	(2.16)	(2.33)	(2.15)
Observations	18,429	18,429	18,429	14,092	13,311	6,688
Squared Correlation	0.187	0.196	0.196	0.174	0.186	0.165

Column (4) constrains the set of choosers to be car models with positive sales (strictly defined). Column (5) adds the constraint that the choice set constitutes of plants already producing the required cell in terms of chemistry/material (NCM vs LFP). Column (6) adds the constraint that the sample only includes the set of 24 major countries, 15 major firms with their 138 major models. Clustered (dyad) standard-errors in parentheses, Signif. Codes: a: 0.01, b: 0.05, c: 0.1.

Table B2: BEV Sourcing decision

Dependent Variable:	BEV sourcing decision				
Model:	(1)	(2)	(3)	(4)	(5)
Border	-1.32^a	-0.674^b	-0.813^a	-0.813^a	-0.426^a
	(0.358)	(0.263)	(0.263)	(0.263)	(0.139)
log distance	-0.060	-0.083^b	0.084	0.084	0.099
	(0.044)	(0.042)	(0.062)	(0.062)	(0.062)
RTA	0.641^a				
	(0.144)				
EC(cells)		-0.348^a	-0.293^a	-0.293^a	-0.217^a
		(0.066)	(0.068)	(0.068)	(0.083)
log(1+tariff)		-5.25^a	-4.29^a	-4.29^a	-10.3^a
		(1.52)	(1.38)	(1.38)	(1.65)
cross-continent			-0.511^a	-0.511^a	-0.634^a
			(0.108)	(0.108)	(0.192)
log GDP per capita			0.262^a	0.262^a	0.116
			(0.095)	(0.095)	(0.079)
Observations	59,355	59,355	59,355	59,355	34,776
Squared Correlation	0.077	0.080	0.081	0.081	0.080
FE origin	ctry	ctry	cluster1	cluster2	cluster2
Sample	all	all	all	all	NRdest

Clustered (dyad) standard-errors in parentheses, Signif. Codes: a: 0.01, b: 0.05, c: 0.1.