# 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

March 18, 2025

PRELIMINARY VERSION, Download latest draft

**Abstract**

Many countries have implemented policies to promote transition from combustion engines to electric vehicles (EVs). As batteries constitute about one third of the cost of EVs and are costly to transport, 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 place 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, trade costs, and subsidies. This is a difficult combinatorial choice problem that cannot be solved using existing "squeezing" algorithms that have been used in the recent literature analyzing global supply chain location choices. Instead, we use a mixed integer linear program formulation that can computationally solve our real-scale multi-stage application in seconds. We use this method to 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 2015 to 2023. We then investigate counterfactual simulations for different types of industrial and trade policies and describe how those affect the production location choices across the global chains for EVs and the trade patterns from battery through assembly to final consumption destinations.

*We thank Rocco Machiavello and Rowan Shi for helpful conference discussions and participants at seminars for valuable suggestions. All four authors acknowledge internal grants from our respective institutions that helped to finance the data purchased in this project. There are no conflicts of interest to disclose.

# 1 Introduction

Industrial policies are increasingly being used to reshape the plant location patterns of industries characterized by multi-stage production and economies of scale. Such industries feature complex spatial interdependencies: At any given stage, different locations substitute for each other. However, upstream and downstream production sites are complementary. Costs of production differ, and final demand is spread in a highly uneven manner. These factors imply that there is no obvious way to structure the firm's value chain across space. Furthermore, industrial policies such as subsidies or trade restrictions induce firms to re-organize their global supply chains in complex ways. These challenges are relevant for industries such as semiconductors, solar panels, aluminum, and electric vehicles (EVs), all of which have been the subject of recent policy interventions, most recently when the Biden administration lifted its tariffs on China on each product.[1]

This paper develops a quantitative method to analyze those industrial polices for multi-stage production and applies it to the EV industry. The application to EVs follows from data availability and policy relevance. The 2022 US Inflation Reduction Act (IRA) awards consumer and production subsidies that do not extend to Asian and European-made cars. European governments are considering measures to counter US policies and have imposed tariffs on Chinese EV imports. Canada, meanwhile has promised to match US IRA production subsidies for factories that Volkswagen and Stellantis are constructing in Ontario. Deputy Prime Minister Freeland defended the Canadian government's decision to spend roughly $30 billion on production subsidies, 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 the upstream stages of the EV value chain. 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

[1] https://www.federalregister.gov/documents/2024/09/18/2024-21217/notice-of-modification-chinas-acts-policies-and-practices-related-to-technology-transfer

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 this 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.[2] We extend the multi-stage production cost minimization problem considered in OR to allow for endogenous demand and market entry by multi-product firms.

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 policies inspired by 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.

[2] "Squeezing" algorithms following the contribution of Jia (2008) cannot accommodate our multi-stage framework with fixed costs. We discuss this in further detail below.

The paper is organized as follows. Section 2 positions this paper within the literatures on sourcing across the value chain as well as the emerging work on EVs. Section 3 describes a set of 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. We specify the variable profits associated with each potential path in section 5. The first step towards quantification of variable profits, conducted in section 6, is to obtain the parameters governing demand for each EV model in each market. Section 7 develops and estimates an estimation framework to obtain variable costs along each potential path (cell-location, assembly-location, consumer country). The method is conditional on open facilities, so it cannot identify fixed costs. To do that, we use a simulated method of moments (SMM) on the full model in Section 8. With all parameters estimates, Section 9 reaches the end goal of simulating the effects of a menu of industrial policies drawn from recent approaches taken in North America and Europe. 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 a given stage of production.

**Global value chains with constant returns** The GVC literature starting with Yi (2003, 2010) and more recently Antràs and De Gortari (2020), Tyazhelnikov (2022), and Johnson and Moxnes (2023) 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. However, those methods work with a constant-returns assumption. Yet, the location choice typically involves opening facilities, which require fixed costs and generate increasing returns to scale. More generally, non-constant returns to scale induce inter-dependencies between locations at each stage.

**Plant location/sourcing with inter-dependencies** 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. With $\sigma$ denoting the elasticity of substitution between varieties in the demand function, and $\theta$ 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*.[3]

Our structure requires a different approach because it combines both submodularity and supermodularity forces: 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 (product entry) increase the profitability of adding production plants. Thus, there is no realistic way of imposing either global sub- or super-modularity to our application. A critical feature of our approach is that we do not need either one to hold.

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.

[3]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.

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 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. [4]

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.

**Comparison to recent work on industrial policy towards EVs** Two recent working papers also examine the application of industrial policies to the Electric Vehicle industry: Allcott et al. (2024) and Barwick et al. (2025). While there are several areas where our papers overlap with these two, the crucial difference is our focus on the endogenous location of factories.

Allcott et al. (2024) stands out for its careful treatment of the institutional features of the IRA, including the leasing loophole and the detailed treatment of subsidy eligibility over time. The policy simulations in our paper are inspired by the IRA provisions of making subsidies contingent on the value chain, but are not intended to evaluate the IRA's effects. Another difference is Allcott et al. (2024) attempt a full welfare evaluation including profit shifting and multiple externalities. Our approach is to contrast changes in two outcome variables that we see as capturing the directions through which the policies affect welfare: number of production lines in the Americas—which captures producer surplus—and expenditures on EVs—which we see as correlated with the environmental benefits of EV adoption.

Barwick et al. (2025) looks into the question of how subsidies (consumption and production related) affect learning-by-doing in the EV battery industry in a model where

[4] 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.

oligopolistic EV assemblers and battery manufacturers bargain over the final battery prices. This results in an elegant estimation equation that accounts for both vehicle and battery equilibrium markups (obtained from the bargaining equilibrium) in the formation of vehicle's price, together with marginal costs. Compared to Barwick et al. (2025), in our framework industrial policy potentially reduces marginal costs via adding plants closer to downstream purchasers, rather than through achieving cost reductions via learning by doing. Also, our framework for solving the firm's problem has maximization of profits over the whole value chain. The implicit assumption is that bargaining reaches the efficient production decision and relative bargaining power determines lump-sum transfers, rather than bargaining over prices, which causes a distortion.

# 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.

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 the elements of its core value chain (the battery pack) are sourced. All plants along the value chain have been geocoded 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 2015–2023 (with IHS projections for 2024). 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).

## 3.3 Location of plants and sourcing along the value chain

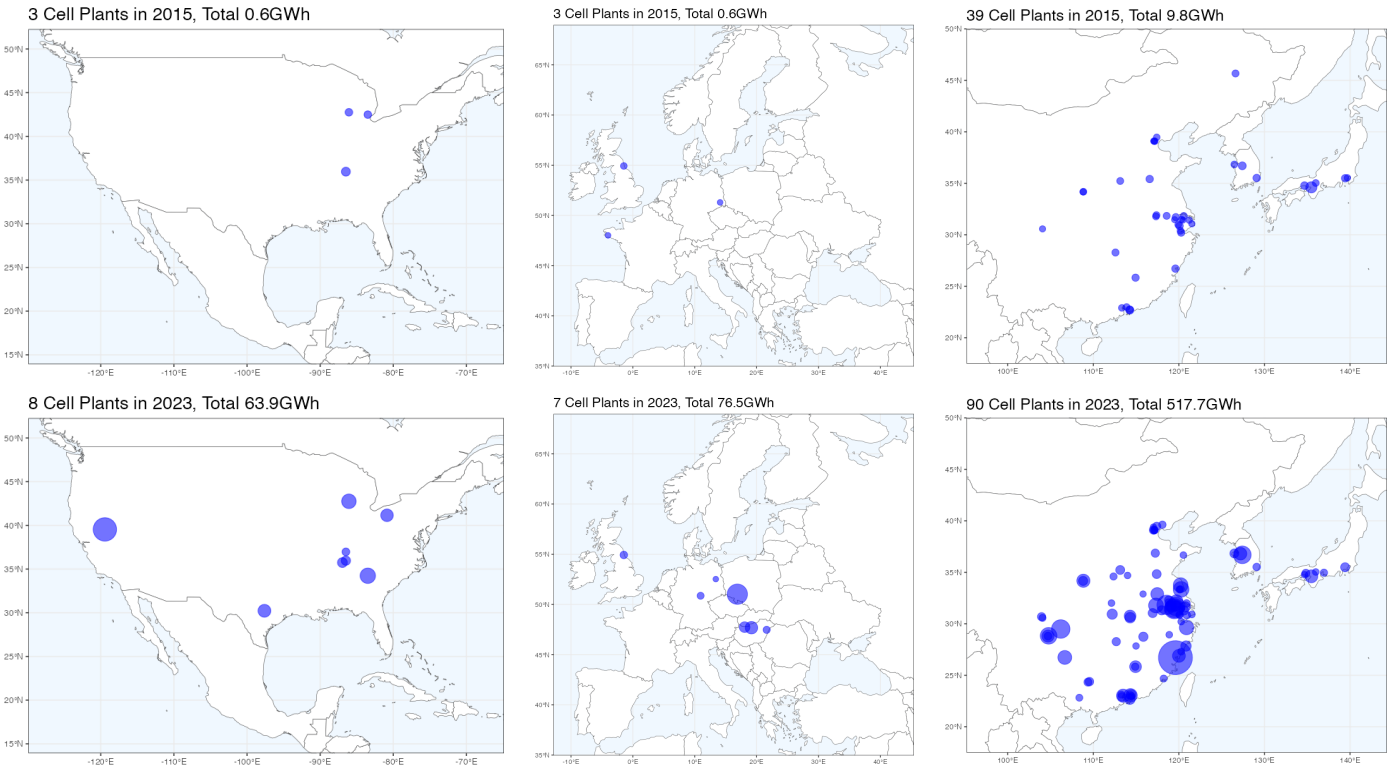
Figures 1 and 2 show the impressive growth in the numbers of plants for cells and vehicles from 2015 to 2023—in all the main regions. The size of plant symbols corresponds to total output in gigawatt-hours for cells and total sales for vehicles.

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 |

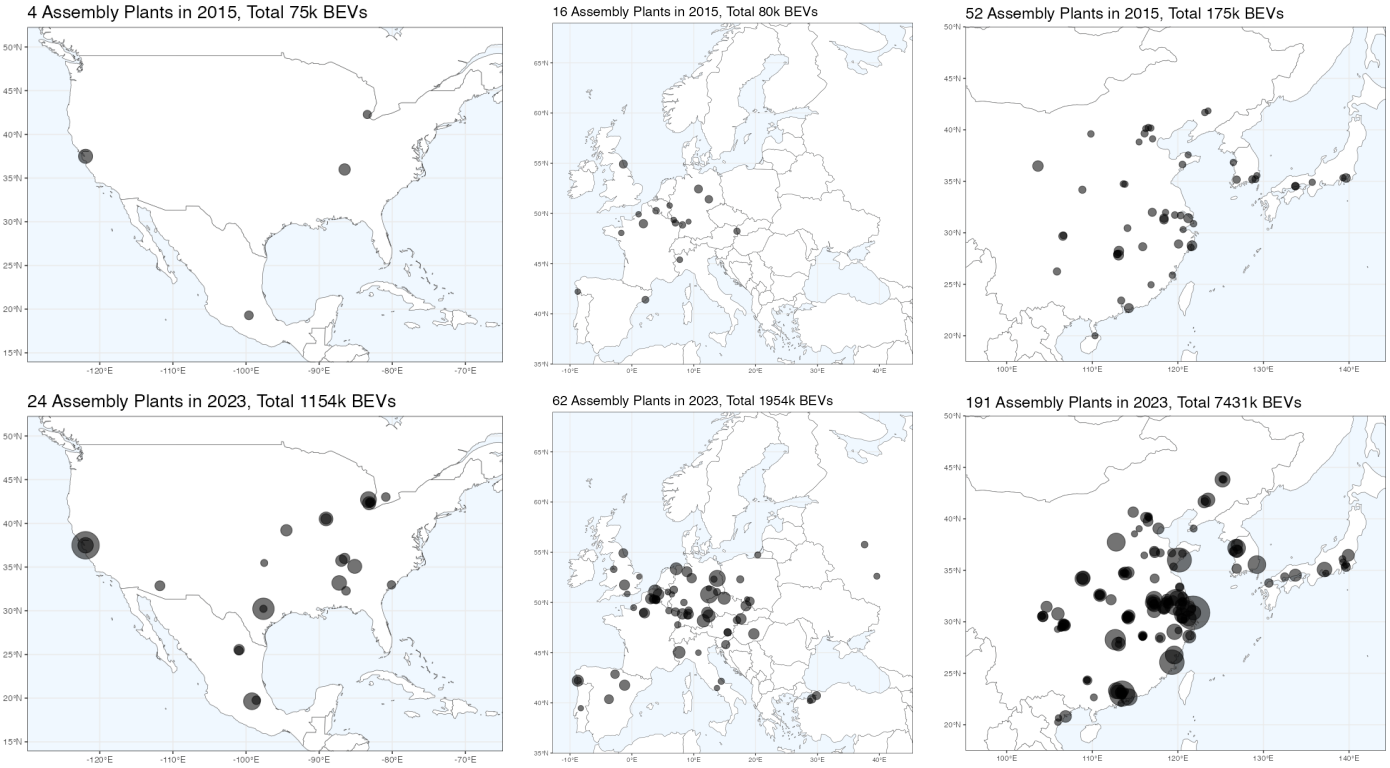
Note: Rest of world row reports averages across 59 countries for numbers of models and EV share and the sum for Sales.

Figure 1: Cell plants in the major regions



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 2 with multiple vehicle plants (Italy, Spain, Portugal, and Turkey) lack local cell production.

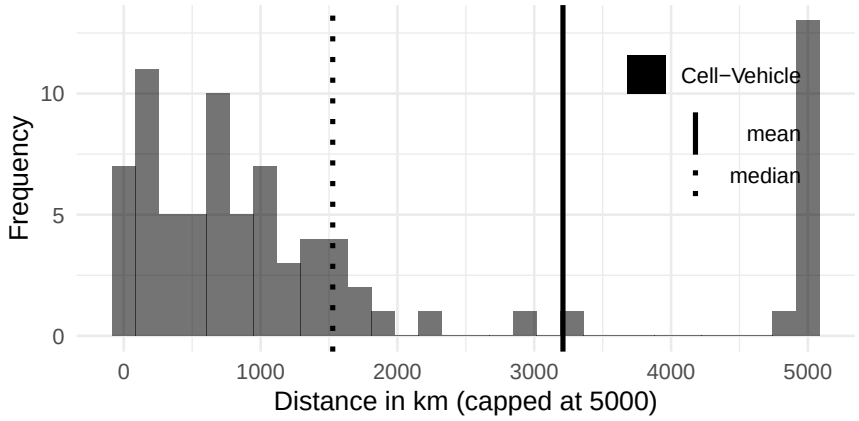
Figure 2: EV plants in the major regions



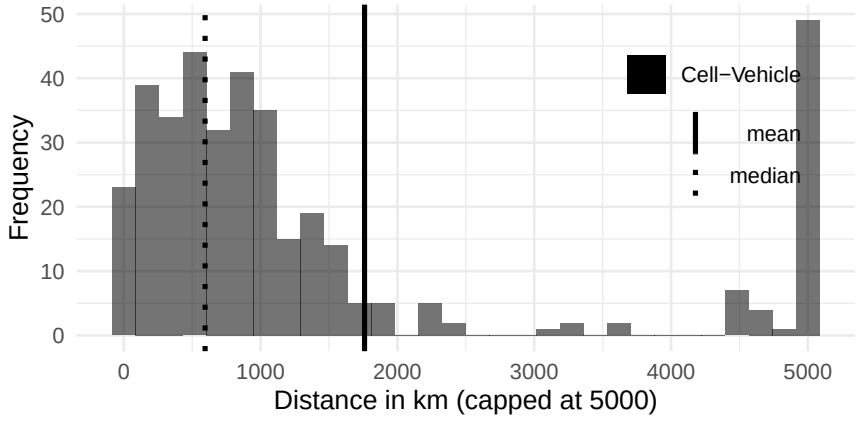

The maps show the geographic dispersion of plants but they do not tell us which plants trade with which. Figure 3 provides some insight into the distances that components travel. By showing medians and means we see the huge skewness in the distribution of distances that the output of a stage travels before becoming an input.

Figure 4(a) shows that when we disaggregate as much as the data allow, cells are almost always single-sourced. This is not the case at more aggregate levels of models, brands or 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. The share of cell multi-sourcing is about 38% for firms and falls regularly to 3% at the most detailed level, which account for the shape, material, and power of the cell. The dashed lines are accounting for multi-sourcing which are not transitory, i.e. which last more than 3 years, and are all systematically lower. Even at the final assembly stage,

Figure 3: Distances shortening between stages

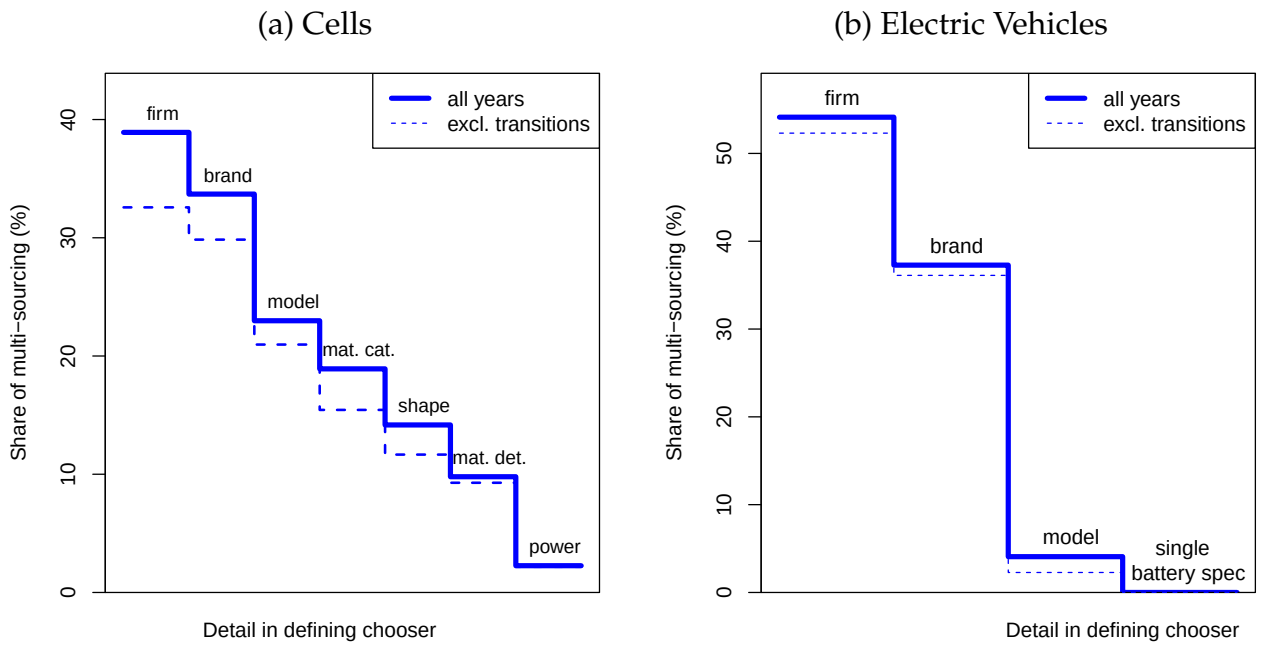



2015



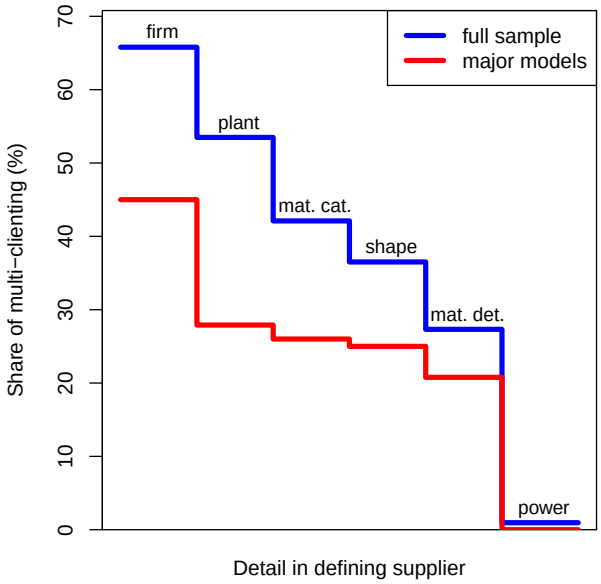

2022

Figure 4: Multi-sourcing is rare for both cells and EVs (2022)




Note: The shares represent the number of choosers which decide to source from multiple options. In panel (a), the chooser goes from a firm-assembly plant-year to a brand-model-battery specification (shape, material and power)-assembly plant-year combination. The options are the different cell plants. In panel (b), the chooser goes from a firm-destination-year to a brand-model-destination-year combination. The options are the different assembly plants.

Figure 5: Single clients served by a battery plant for a type of battery




Note: The shares represent the number of battery plants (for a particular battery material, shape and power) that supply multiple car makers in 2022.

reported in the panel (b), we see that market countries tend to single source, with a narrowly defined vehicle only coming from two different plants 4.07% 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 2016. This residual multi-sourcing for EVs certainly hides further detail that we do not observe fully in the flows. Indeed, we observe the flows detailed for a model, but not for its detailed specifications which often involve different batteries. When this battery specification is defined in terms of material and total power of the pack, we observe that car models with a single battery specification available are never multi-sourced.

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.

Figure 5 examines a related but different pattern of the data regarding the frequency of multi-supplying inside a region. For each cell plant, we count the number of clients

supplied with increasing degree of detail in the definition of this client. For instance, the first step of the blue curve shows that more than 65% of cell plants serve more than one car maker firm. This share falls by more than 10 p.p. when the client is a car assembly plant. When looking at the detailed material and shape of the cell, the share is around 30%. The largest drop is when accounting for the power of the cell: the share of multi-clienting is then negligible, either for all cell plants (blue line) of the ones that serve the set of major models that we use in our simulations later in the paper (red line). This suggests that in almost all cases battery manufacturers produce a specific cell for a single car assembler/buyer, rather than supplying a generic type of cell to several buyers.

Figure 6 shows the relationship between the size of the plant and the number of specifications (material category and shape for cells, platforms for cars) it produces (cells in panel a and vehicles in panel b). Those are computed for the year 2022.

Figure 6: Size of plant and count of groups in 2022

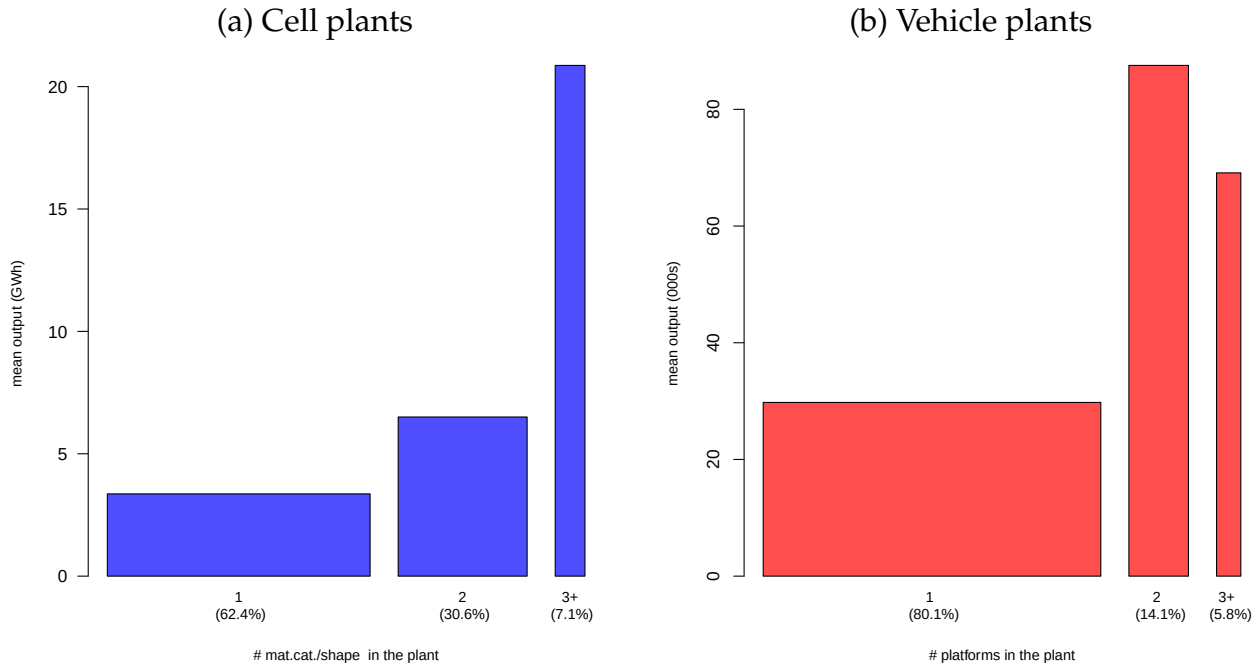



Figure 6, also shows that as a cell plant producing several specifications (different material category-shape combinations) is substantially larger on average. This suggests that adding a new type of cell for a new client entails large fixed costs (which requires a larger scale of production to be profitable).

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. In Appendix E.2, we present some examples suggesting that capacity constraints do not seem to be a major concern in our case. Between 2015 and 2023, figure E8 shows that the Tesla factory in Fremont multiplied its output by 10 (from about 50 thousand vehicles to around 500). On top of that, the Shanghai factory was able to add more than a 150 thousand vehicles on its first year in 2020, and reached 950 thousands in 2023). Continuing on the Tesla example, one can investigate cases of multi-sourcing. In figure E6 in the same appendix takes the example of the four Tesla models sold in Germany. The Model 3 is sourced both from Fremont and Shanghai in 2020, but this turns out to be a transition, since Fremont was the only source until 2019, and Shanghai becomes the sole supplier of Model 3 to Germany starting in 2021. The story is very similar with the Model Y, which starts to get imported from Shanghai in 2021. Then Germany also sources from the Berlin factory when it opens in 2022. In 2023, Berlin becomes dominant, while imports from Shanghai fall and are predicted by IHS to go to 0 in 2024. Panel (b) of the same figure shows that those are not isolated examples. The Volkswagen ID.4 SUV is sold in the United States and Canada in 2021 from the German factory of Mosel. In 2022, The American plant in Chattanooga starts to produce the car, and both sources co-exist for one year, before the imports from Germany are stopped in 2023.

In line with those data patterns, 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 a large number of news reports for those investments. 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 2: Investment costs for battery and EV plants

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

Source: Authors' calculations based on collected news articles.

Table 2 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.

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 239 EV assembly plants and 149 pack plants in the IHS EV supply chain data set. Multiplying by the reported means, total EV investment from 2015 to 2022 amounts to $464 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), $\{m : m \in M_f\}$, and a headquarters location. The models have a vector of "appeal" in each market $n$. While the firm chooses which models to offer where, they cannot add new models.

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: $\mathbf{L} \equiv \{L_1, ..., L_K\}$. The index $\ell_k$ represents a particular production location at level $k$. 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 market $n$ for model $m$ is given by $\boldsymbol{\ell}_{mn} \equiv \{\ell_1, ..., \ell_K\}_{mn}$.

## 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. A firm chooses where to open facilities at each level from $k = 1$

to $K$. Firms also 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 $N$. The payment of fixed costs activates a location for groups of models that share common characteristics. The grouping mapping at each stage is given by the matrices $\Gamma^k$. Firms can add extra models at a given location for no additional fixed cost payment, so long as they share a common grouping, that is $\Gamma^k_{mg_k} = 1$ for all $m$ that use group $g_k$ at stage $k$.[5]

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 7: Schematic of a supply chain with $K = 2$

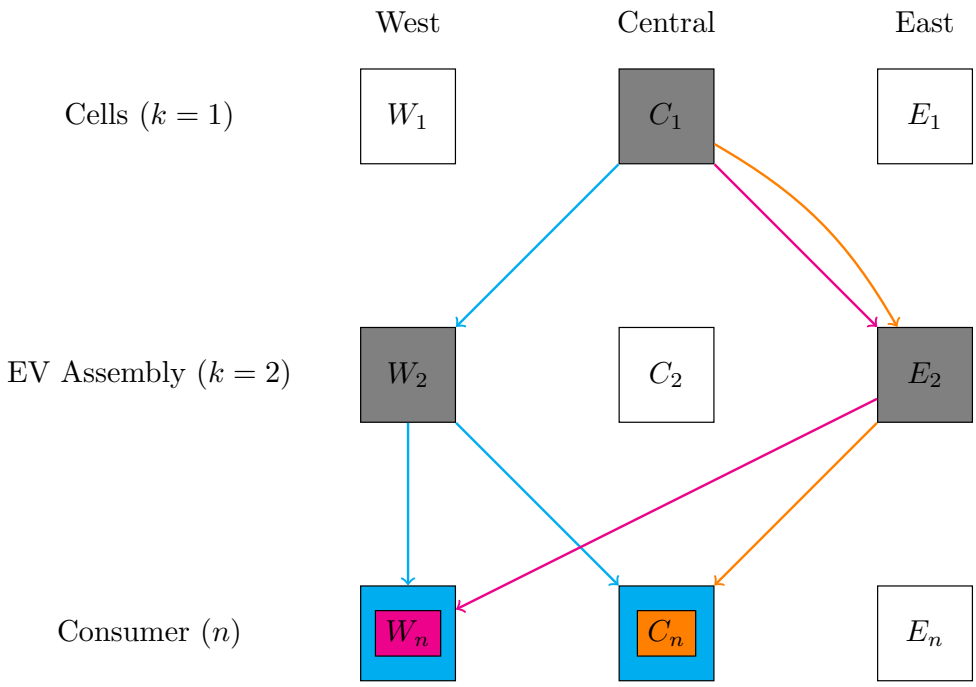



Figure 7 depicts a simplified version of our model to clarify the decisions we are modeling. As in our application, we set $K = 2$ stages of production for cell manufacture and vehicle assembly.[6] In addition, there is the decision to enter a consumer market. The

[5] In our application to BEVs, cell groups refer to *material-shape* combinations and assembly groups refer to vehicle *platforms*.

[6] We can solve MMM-UFLP for up to $K = 4$ stages. Appendix C shows the $K = 3$ formulation. We use $K = 2$ in this paper because estimation of simulated moments requires a very large number of evaluations and large amounts of memory usage.

schematic portrays a single firm that makes three models. All three models here share a common stage 1 group. At stage 2, the Magenta and Orange models have a common group, but the Cyan model is in a different group. There are three potential locations: East, Central, and West.

In figure 7, 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 assembly 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 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, $A_n$, 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 $\phi_{fg_k\ell_k}$ in location $\ell_k$ for a plant to produce type $g_k$ inputs (or final assembly for $k = K$). We denote this facility location choice with a binary variable $y_{fg_k\ell_k} \in \{0, 1\}$.

Let $\phi_{mn}$ 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.

[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 $n$.

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.

## 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] We add three features to that framework: multi-product firms, endogenous choice of market entry, and endogenous quantities (profit maximization versus cost minimization). In a typical MUFLP, there are two types of decisions involved. One is the facility location decision $y_{\ell_k}$ 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]

$$\max_{\mathbf{x},\mathbf{y},\mathbf{z}} \quad \sum_{m\in M_f}\sum_{n\in N}\sum_{\ell_1\in L_1}\sum_{\ell_2\in L_2} \pi(c(\boldsymbol{\ell}_{mn}), A_n) x_{mn\ell_1\ell_2} - \sum_{g_1\in G_1}\sum_{\ell_1\in L_1} \phi_{fg_1\ell_1} y_{fg_1\ell_1} - \sum_{g_2\in G_2}\sum_{\ell_2\in L_2} \phi_{fg_2\ell_2} y_{fg_2\ell_2} - \sum_{m\in M_f}\sum_{n} \phi_{mn} z_{mn} \tag{1}$$

[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

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

$$\sum_{\ell_1 \in L_1} x_{mn\ell_1\ell_2} \quad \leq \quad \sum_{g_2 \in G_2} \Gamma^2_{mg_2} y_{fg_2\ell_2}, \;\; n \in N, m \in \mathcal{M}_f, \ell_2 \in L_2 \tag{3}$$

$$\sum_{\ell_2 \in L_2} x_{mn\ell_1\ell_2} \quad \leq \quad \sum_{g_1 \in G_1} \Gamma^1_{mg_1} y_{fg_1\ell_1}, \;\; n \in N, m \in \mathcal{M}_f, \ell_1 \in L_1 \tag{4}$$

$$x_{mn\ell_1\ell_2} \;\; \geq 0, \qquad y_{fg_1\ell_1} \in \{0,1\}, \quad y_{fg_2\ell_2} \in \{0,1\} \quad z_{mn} \in \{0,1\}. \tag{5}$$

The $M_f \times N$ inequalities defined in (2), are the *activity constraints* that govern market entry: the left-hand-side sum will be equal to one if the model $m$ is sold in market $n$ (only a single $x_{mn\ell_1\ell_2}$ path is set to one at the optimum), or will be equal to zero otherwise (all the $x_{mn\ell_1\ell_2}$ paths are set to zero). 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. The inequality formulation improves speed, as explained in Appendix B.

The $M_f \times N \times L_2 \times G_2$ inequalities contained in (3) and the $M_f \times N \times L_1 \times G_1$ contained in(4) are *activity constraints* that govern whether a given facility is open for a model: the left-hand side sum equals one if a path through a facility in $\ell_k$ is used (only a single $x_{mn\ell_1\ell_2}$ path is set to 1 at the optimum), or equals zero otherwise (all the $x_{mn\ell_1\ell_2}$ paths are set to 0). These constraints guarantee that a facility must be active to be used in supply chains specific to groups of cell ($k = 1$) and vehicle ($k = 2$) types.

Equation (5) restricts facility activation and market entry to be binary. Some statements of the uncapacitated FLP explicitly constrain the $x_{mn\ell_1\ell_2}$ 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 sufficient in equation (5) to restrict the assignment variables (the $x$s) to be positive. While fractional assignments are not explicitly excluded, the $x$ are binary at the optimum.[11]

Appendix section A demonstrates the computational feasibility and illustrates how

[11] 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*.

policies work within a stripped down model with a single firm. We also show in section B how computation time for the MILP method varies with the number of locations and stages and how it compares to the method of Arkolakis et al. (2023).

# 5 Mapping paths to profits

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

## 5.1 Demand

On the demand side we use a constant elasticity of substitution (CES) form.[12] One convenient feature of CES is that the average own-price elasticity is widely reported in the literature. Table D2 in the appendix provides 18 different estimates, taken mainly from the Industrial Organization literature. We calibrate $-\eta = 4.0$, the median of these estimates.

Each country $n$ spends an exogenous amount $R_n$ on light vehicles of all fuel types. The battery-powered EV share $R_n^{\text{EV}}/R_n$ under CES is given by $(P_n^{\text{EV}}/P_n)^{1-\eta}$, where the $P_n^{\text{EV}}$ and $P_n$ are the price index for EVs and all light vehicles. The market share of model $m$ depends on the ratio of the model's quality-adjusted price $p_{mn}/\xi_{mn}$ to $P_n^{\text{EV}}$. We show how to calibrate the $\xi_{mn}$ in section 6. 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

$$q_{mn} = R_n^{\text{EV}} \left(P_n^{\text{EV}}\right)^{\eta-1} \xi_{mn}^{\eta-1} p_{mn}^{-\eta}, \tag{6}$$

where $P_n^{\text{EV}}$ is a CES index of firm-level quality-adjusted price index of models offered in

[12] This assumption differs from most of the literature on the car industry which uses nested logit or mixed logit demands. The main difference between the two is that while logit demand has price responsiveness $\alpha$ parameter that multiplies price, CES has an elasticity $\eta$ that multiplies log price. In the discrete choice derivation of logit demand, $\alpha$ 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 $\alpha$ as is common in mixed logit models, but we would lose the tractability needed for our facility-location problem.

$n$:

$$P_n^{\text{EV}} = \left( \sum_f \left( P_{fn}^{\text{EV}} \right)^{1-\eta} \right)^{1/(1-\eta)}, \text{ with } P_{fn}^{\text{EV}} \equiv \left( \sum_{m \in M_f} z_{mn} (p_{mn}/\xi_{mn})^{1-\eta} \right)^{1/(1-\eta)}, \quad (7)$$

where $z_{mn}$ is set to 1 for all models $m$ sold in $n$ (and 0 otherwise).

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

$$\frac{R_n^{\text{EV}}}{R_n} = \left[ \frac{P_n^{\text{EV}}}{P_n} \right]^{1-\eta} \quad \text{where} \quad P_n \equiv [(P_n^{\text{EV}})^{1-\eta} + (P_n^{\text{ICE}})^{1-\eta}]^{1/(1-\eta)}, \quad (8)$$

and $P_n^{\text{ICE}}$ is the price index for ICE vehicles, defined analogously to $P_n^{\text{EV}}$.

To obtain profits as a function of costs, the first step is a pricing equation. The latter has two pieces. The first is the path cost $c(\boldsymbol{\ell}_{mn})$, to which we return below. The second piece relates model-level costs to quality, with the underlying assumption that bigger, better cars cost more to manufacture, regardless of location.[13] To capture the idea that costs rise with the common factor of quality, we model the cost of quality as $\tilde{\xi}_m^{\alpha_c}$, where $\tilde{\xi}_m$ is the geometric mean of the $\xi_{mn}$ across all markets (for a given $m$) and $\alpha_c$ is the elasticity of marginal cost with respect to model quality.

Under monopolistic competition, each firm is sufficiently small that it neglects the impact of its pricing changes – which depend on its chosen paths – on the aggregate price $P_n$. With CES demand (6), firms apply a constant markup $\mu_{mn} = \mu = \eta/(\eta - 1)$ to their delivered marginal cost $c(\boldsymbol{\ell}_{mn})$ for all their models: $p_{mn} = \mu c(\boldsymbol{\ell}_{mn}) \tilde{\xi}_m^{\alpha_c}$. The resulting variable profit for each model is given by

$$\pi_{mn} = \frac{c(\boldsymbol{\ell}_{mn}) \tilde{\xi}_m^{\alpha_c} q_{mn}}{\eta - 1}, \quad (9)$$

where $q_{mn}$ is defined in (6) as a function of $c(\boldsymbol{\ell}_{mn})$ and $P_n^{\text{EV}}$. This completes the specification of $\pi(c(\boldsymbol{\ell}_{mn}), P_n)$ and establishes that variable profits depend on only the path and the aggregator ($A_n = R_n P_n^{1-\eta}$), thus complying with Assumption 1. The next step is to flesh out the cost function $c(\boldsymbol{\ell}_{mn})$.

[13] Table E3 shows that battery pack size, charging time, motor power, and vehicle weight explain 83% of the within-market variation in car prices and 52% of the within-market variation of appeal.

## 5.2 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).[14]

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 $\varepsilon^k_{m\ell_k\ell_{k+1}}$.[15]

In this specification with $K = 2$ stages of production, the marginal cost of cells (produced in $\ell_1$, delivered to $\ell_2$) and of vehicles (assembled in $\ell_2$, delivered to $n$) are given by

$$\text{Cells:} \quad c^1_{m\ell_1\ell_2} = w^1_{\ell_1} \tau^1_{\ell_1\ell_2} \varepsilon^1_{m\ell_1\ell_2}, \tag{10}$$

$$\text{Vehicles:} \quad c(\boldsymbol{\ell}_{mn}) = \left(w^2_{\ell_2}\right)^{\alpha_{22}} \left(c^1_{m\ell_1\ell_2}\right)^{\alpha_{12}} \tau^2_{\ell_2 n} \varepsilon^2_{m\ell_2 n}, \tag{11}$$

where $\alpha_{k2} \in (0, 1)$ denotes the cost share of stage $k = 1, 2$ factors in final assembly ($k = 2$). 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 $\tau^2_{\ell_2 n}$.

We can also extend these marginal cost determinants to include the costs of other

[14] 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.

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

inputs that come from other locations (other than cell production and vehicle assembly). A key assumption is that the trade costs of those inputs to the cell or assembly locations are negligible. In this case, those additional input costs are subsumed into an additive constant term (for log costs) in the sourcing regressions in the next section. We do not explicitly incorporate those additional input costs for notational simplicity.

Equation (9) translates an estimate of path costs—specified in (11)—into profits, which is the key requirement of the MUFLP. The drawback of this formulation of the 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 sourcing regressions and identified by differences in log costs). We substitute path costs into equation (7) to express the price index in terms of an index of the path costs of all competing models $m$ sold in $n$ ($z_{mn} = 1$):

$$P_n^{\text{EV}} = \mu C_n^{\text{EV}}, \quad \text{where} \quad C_n^{\text{EV}} \equiv \left[ \sum_m z_{mn} \left( c(\boldsymbol{\ell}_{mn}) \left( \tilde{\xi}_m^{\alpha_c} / \xi_{mn} \right) \right)^{1-\eta} \right]^{1/(1-\eta)}. \tag{12}$$

Now profits can be expressed as a function of the cost index:

$$\pi(c(\boldsymbol{\ell}_{mn}), A_n) = \left( \frac{c(\boldsymbol{\ell}_{mn}) \left( \tilde{\xi}_m^{\alpha_c} / \xi_{mn} \right)}{P_n / \mu} \right)^{1-\eta} \frac{R_n}{\eta} = \left( \frac{c(\boldsymbol{\ell}_{mn}) \left( \tilde{\xi}_m^{\alpha_c} / \xi_{mn} \right)}{C_n^{\text{EV}}} \right)^{1-\eta} \frac{R_n^{\text{EV}}}{\eta}. \tag{13}$$

The factor $R_n^{\text{EV}}/\eta$ 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.[16]

Equation (13) expresses variable profits for any potential path. It is the crucial input in the formulation of the MMM-UFLP shown in equation (1). Now we outline the estimation and calibration methods we apply to the raw data to fill in all the missing parameters.

## 5.3 Roadmap to estimation of the model

To estimate the parameters of the model we divide the problem into three main pieces. Our focus here is the cost-side of the firm's problem, so we calibrate the demand side using standard methods described in section 6. The next two sections estimate the cost parameters. In section 7, we use the assignment (or sourcing) decisions in the data (which

[16]The second equality comes from $R_n^{EV} = R_n \left( P_n^{\text{EV}} / P_n \right)^{1-\eta}$. Replacing $P_n^{\text{EV}} = \mu C_n^{\text{EV}}$ and solving for $R_n = R_n^{\text{EV}} \left( C_n^{\text{EV}} / (P_n/\mu) \right)^{\eta-1}$ gives the result that profits can be expressed in terms of the EV aggregates.

are conditional on open markets and facilities) to estimate the bilateral variable costs $\tau^k_{\ell_k \ell_{k+1}}$ for all locations at each level of production. In section 8, we then estimate all the remaining parameters using simulated method of moments matching those generated from the MMM-UFLP solution to their counterparts in our data. Those parameters include those governing the distribution of fixed costs, which we describe later in that section.

# 6 Calibration of demand parameters

We calibrate three sets of demand parameters: the own price elasticity $\eta$, the demand shifters $\xi_{mn}$, and the cost elasticity of quality $\alpha_c$. For $\eta$ we take the median from a large (and growing) literature that reports the average own price elasticity. The calibration of $\xi_{mn}$ demand shifters uses IHS model-level car prices in each market, as well as the sales of EV and ICE cars. Finally, $\alpha_c$ comes from a regression of price on mean $\xi_{mn}$ (across markets).

We invert (6) to recover $\xi_{mn}$ as a ratio of appeal relative to the geometric mean of EV models in market $n$ (we exclude plug-in hybrids throughout).[17] Letting $\tilde{p}_n$ and $\tilde{q}_n$ denote the geometric means of EV prices and quantities in market $n$, the inversion yields

$$\xi_{mn} = \left(\frac{p_{mn}}{\tilde{p}_n}\right)^{\frac{\eta}{\eta-1}} \left(\frac{q_{mn}}{\tilde{q}_n}\right)^{\frac{1}{\eta-1}}, \tag{14}$$

where we have adopted the Hottman et al. (2016) normalization that the geometric mean of $\xi_{mn}$ is one in every market $n$ (only relative appeal matters).

With an estimate of $\eta$, we can then compute the relative appeal, $\xi_{mn}$ of each EV model in each market. Appendix Table D2 provides a number of average own-price 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 D2, we see that the median over all 18 studies is 4.

Table 3 shows the top 10 and bottom 10 models in terms of appeal. Topping the list of appeal are models by Porsche, BMW, Audi and all four Tesla models. The four top models are between 2.5 and 3 times more attractive than the average offer over the set

[17] 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 3: Appeal of top 10 and bottom 10 models

| brand model | N mkts | geom. mean | best $\xi_{mn}$ country | best $\xi_{mn}$ value | worst $\xi_{mn}$ country | worst $\xi_{mn}$ value |
|---|---|---|---|---|---|---|
| Most appealing models | | | | | | |
| Porsche Taycan | 19 | 3.02 | CHN | 5.82 | CAN | 1.91 |
| Tesla Model Y (N) | 18 | 2.86 | CHN | 4.24 | JPN | 1.96 |
| Mercedes-Benz EQS | 19 | 2.76 | CHN | 6.09 | CAN | 1.76 |
| BMW iX | 18 | 2.55 | CHN | 4.12 | FRA | 1.55 |
| Audi e-tron GT | 18 | 2.06 | AUT | 3.38 | PRT | 1.41 |
| Tesla Model 3 (N) | 19 | 1.87 | CAN | 3.26 | NOR | 1.23 |
| Audi e-tron | 18 | 1.85 | NOR | 2.84 | POL | 1.10 |
| Tesla Model S | 15 | 1.85 | USA | 4.07 | PRT | 0.87 |
| Tesla Model X | 14 | 1.74 | USA | 4.59 | PRT | 1.08 |
| BMW i4 | 19 | 1.68 | PRT | 2.31 | CAN | 0.92 |
| Least appealing models | | | | | | |
| Citroen C4 | 14 | 0.70 | ESP | 1.14 | POL | 0.34 |
| Toyota bZ4X | 15 | 0.68 | CHN | 1.03 | USA | 0.41 |
| Renault Zoe | 15 | 0.65 | FRA | 1.24 | NOR | 0.37 |
| Skoda Enyaq Coupe | 11 | 0.64 | CHE | 1.07 | BEL | 0.44 |
| Kia Niro (N) | 10 | 0.55 | GBR | 1.06 | USA | 0.25 |
| Opel Corsa | 15 | 0.55 | DEU | 0.98 | FIN | 0.32 |
| MG ZS (N) | 13 | 0.54 | GBR | 0.92 | CHN | 0.17 |
| BMW iX1 | 10 | 0.53 | NOR | 0.80 | GBR | 0.20 |
| Kia Soul | 12 | 0.51 | PRT | 0.76 | ITA | 0.21 |
| Dacia Spring | 11 | 0.38 | ITA | 0.68 | DNK | 0.22 |
| MG 4 (N) | 11 | 0.36 | GBR | 0.57 | AUT | 0.19 |

Note: Models ranked by their geometric mean. Only models with data in 10 or more market included.

of countries where they are sold. It is revealing that the destination in which those are particularly appealing is China, a market in which many inexpensive models are available. Our computations reveal that, given its high price, the Mercedes-Benz EQS must be 6 times more attractive than the average model in China to achieve its actual level of sales there, against less than 2 times the average model sold in Canada. The bottom appeal list features smaller cars made by a groups of firms that has almost no intersection with the set of firms making the top 10 models. The models in that bottom group are revealed to have an appeal between 35 and 70% of the average car. Another notable pattern is that there seem to be some home brand advantage: The Renault Zoe has most appeal in France, and the two MG models have most appeal in the UK, where the brand was born (even though it is now owned and operated by SAIC).

Table 4: Estimating the cost elasticity of quality

| dep. var: | log(price) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| appeal model-market ($\xi_{mn}$) | 0.371$^a$ | 0.420$^a$ | | |
| | (0.046) | (0.052) | | |
| model C-group geom. mean ($\tilde{\xi}_m$) | | | 0.558$^a$ | |
| | | | (0.062) | |
| model geom. mean ($\tilde{\xi}_m$) | | | | 0.575$^a$ |
| | | | | (0.066) |
| *Fixed-effects* | | | | |
| V_iso_d | Yes | | | |
| V_mfr_group | Yes | | | |
| V_iso_d-V_mfr_group | | Yes | Yes | Yes |
| *Fit statistics* | | | | |
| Observations | 1,040 | 1,040 | 1,040 | 1,040 |
| $R^2$ | 0.669 | 0.729 | 0.762 | 0.763 |
| Within $R^2$ | 0.330 | 0.390 | 0.464 | 0.467 |

Note: Standard errors (in parentheses) clustered at the car model level. Signif. Codes: a: 0.01, b: 0.05, c: 0.1.

The last step is to estimate $\alpha_c$, the cost elasticity of quality, which we do using our calibrated appeal parameters together with price data. Given the constant markup $\mu$, the delivered price of $m$ in $n$ is $p_{mn} = \mu c(\boldsymbol{\ell}_{mn})\tilde{\xi}_m^{\alpha_c}$. Denoting $f(m)$ as the firm selling model $m$, we run the following regression (using data for 2022):

$$\log p_{mn} = \text{FE}_{f(m)n} + \alpha_c \log \tilde{\xi}_m + \zeta_{mn}. \tag{15}$$

The firm-destination fixed effects capture a large set of possible determinants of prices: i) Country-level taxes on cars (or subsidies on EVs) and other destination-specific costs. ii) Firm-level cost characteristics (for instance allowing Tesla to have higher costs than BYD), iii) Firm-destination shocks in costs or demand which might affect the price of a specific firm in a given country (one might think of various reputation effects, Tesla might be facing difficult times in Germany in 2025 for instance). In equation (15), the coefficient on $\log \tilde{\xi}_m$ therefore captures the additional costs of a model X compared to a model Y over all markets, accounting for each destination's characteristics such as ownership fees that are present in some markets and not in others. Table 4 reports the results of the regressions with preferred estimate of $\alpha_c = 0.56$ in column (3).

# 7 Estimation of variable costs via sourcing regressions

## 7.1 Conditional choice estimation of variable costs

Substituting upstream cost (10) into the next stage's cost equation (11), the logged cost of the complete production path is

$$\ln c(\boldsymbol{\ell}_{mn}) = \alpha_{22} \ln w^2_{\ell_2} + \ln \tau^2_{\ell_2 n} + \alpha_{12}(\ln w^1_{\ell_1} + \ln \tau^1_{\ell_1 \ell_2}) + u(\boldsymbol{\ell}_{mn}), \tag{16}$$

where $u(\boldsymbol{\ell}_{mn}) = \alpha_{12} \ln \varepsilon^1_{m\ell_1\ell_2} + \ln \varepsilon^2_{m\ell_2 n}$ 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:

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

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 a Gumbel shock (log of a Fréchet), 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 two components of $u(\boldsymbol{\ell}_{mn})$ directly. Ben-Akiva and Lerman (1985, p. 287) and Cardell (1997) 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.

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 $\ell_2$ (due to constant returns to scale in the variable costs of production). We can therefore use equation (10) to derive the conditional probability of $\ell_1$ being the cost-minimizing location of supplying to $\ell_2$ as

$$\mathbb{P}^1_{\ell_1|\ell_2}(m) = \frac{\left(w^1_{\ell_1}\tau^1_{\ell_1\ell_2}\right)^{-\theta_1}}{\Phi^1_{\ell_2}(m)}, \text{ with } \Phi^1_{\ell_2}(m) = \sum_{g_1\in G_1}\Gamma^1_{mg_1}\sum_{\ell_1\in L_1} y_{fg_1\ell_1}\left(w^1_{\ell_1}\tau^1_{\ell_1\ell_2}\right)^{-\theta_1}, \tag{17}$$

where (as in the MUFLP) $\Gamma^1_{mg_1} = 1$ if model $m$ uses group $g_1$ for cells, and $y_{fg_1\ell_1} = 1$ if location $\ell_1$ has activated this group $g_1$. Our estimating equation is

$$\mathbb{P}^1_{\ell_1|\ell_2}(m) = \exp\left[\text{FE}^1_{\ell_1} + \text{FE}^1_{\ell_2}(m) + \beta^1_D \ln D_{\ell_1\ell_2} + \beta^1_t \ln\left(1 + t^1_{\ell_1\ell_2}\right)\right], \tag{18}$$

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 $\beta^1_D$ and $\beta^1_t$ being their respective regression coefficients. The term $\text{FE}^1_{\ell_1}$ is a source $\ell_1$ fixed effect. The second fixed effect is specific to the *chooser* and is given by

$$\text{FE}^1_{\ell_2}(m) = -\ln\left\{\sum_{g_1\in G_1}\Gamma^1_{mg_1}\sum_{\ell_1\in L_1} y_{fg_1\ell_1}\exp\left[\text{FE}^1_{\ell_1} + \beta^1_D \ln D_{\ell_1\ell_2} + \beta^1_t \ln\left(1 + t^1_{\ell_1\ell_2}\right)\right]\right\}.$$

This chooser fixed effect is more specific than the assembly plant location as the chooser in our estimation corresponds to a plant, firm, model and the power and chemistry of the battery being sourced.[18]. The correspondence between estimated coefficients and structural variables is

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

Note that since tariffs are direct cost shifters, they can be used to provide a direct estimate of $\theta_1$ (as in Head and Mayer, 2019). Here, we do not restrict $\beta^1_t$ to be equal to $-\theta_1$, a flexibility allowing for unobserved bilateral trade costs. The tariff coefficient, $\beta^1_t$, incorporates variation in those unmeasured 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

[18]These narrow definitions of choosers ensure single-sourcing in almost all cases as we see in figure 4

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 assembly plant $\ell_2$ is

$$\mathbb{E}c^1_{.,\ell_2}(m) = \left(\Phi^1_{\ell_2}(m)\right)^{-1/\theta_1}. \tag{19}$$

We refer to this term as the *inclusive cost* which depends on the sourcing potential of $\ell_2$ and differs across car models due to various choice sets of the firm $\mathcal{L}_1(f)$ and the cell group $g_1$ used by $m$.

The second decision to consider is the one made by a model-market combination when sourcing the EV from an assembly plant. Using (11), the probability that $\ell_2$ is selected as the location of assembly is

$$\begin{aligned}\mathbb{P}^2_{\ell_2|n}(m) =& \frac{\left[\left(w^2_{\ell_2}\right)^{\alpha_{22}}\left(\mathbb{E}c^1_{.,\ell_2}\right)^{\alpha_{12}}\tau^2_{\ell_2 n}\right]^{-\theta_2}}{\Phi^2_n(m)}, \text{ with} \\ \Phi^2_n(m) =& \sum_{g_2\in G_2}\Gamma^2_{mg_2}\sum_{\ell_2\in L_2} y_{fg_2\ell_2}\left[\left(w^2_{\ell_2}\right)^{\alpha_{22}}\left(\mathbb{E}c^1_{.,\ell_2}(m)\right)^{\alpha_{12}}\tau^2_{\ell_2 n}\right]^{-\theta_2}.\end{aligned} \tag{20}$$

The estimating equation for destination markets when choosing among potential vehicle assembly facilities is

$$\mathbb{P}^2_{\ell_2|n}(m) = \exp\left[\text{FE}^2_{\ell_2} + \text{FE}^2_n(m) + \beta^2_D \ln D_{\ell_2 n} + \beta^2_t \ln\left(1+t^2_{\ell_2 n}\right) + \beta^2_\Phi \text{FE}^1_{\ell_2}(m)\right], \tag{21}$$

where $\text{FE}^1_{\ell_2}(m)$ was estimated in equation (18), and accounts for determinants of the upstream (cells) costs. The correspondence between the estimated coefficients and structural parameters is

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

Again, note that we do not restrict $\beta^2_t$ to be equal to $\theta_2$ 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.[19] As we estimate the two equations separately, we do not impose this coefficient restriction. It would be difficult to test in any case due to the unknown $\alpha$ parameters that go into the $\beta_\Phi$ 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 (18) and (21) as high-dimensional fixed effects PPML regressions.

Combining the two production stages, we can express the log of the variable cost of the full path as a function of our current estimates and some additional parameters:

$$\begin{aligned}\ln c(\boldsymbol{\ell}_{mn}) = &- \kappa_2 \Big\{ \text{FE}^2_{\ell_2} + \beta^2_D \ln D_{\ell_2 n} + \beta^2_t \ln\left(1 + t^2_{\ell_2 n}\right) \\ &+ \kappa_1 \left[\text{FE}^1_{\ell_1} + \beta^1_D \ln D_{\ell_1 \ell_2} + \beta^1_t \ln\left(1 + t^1_{\ell_1 \ell_2}\right)\right] \Big\} + u(\boldsymbol{\ell}_{mn}).\end{aligned} \tag{22}$$

Parameter $\kappa_2$ scales up and down the relative importance of observed determinants of marginal costs relative to fixed costs and it also determines the variance of the $u$ shocks. $\kappa_1$ provides the weights to use on the cell stage relative to vehicle assembly. Our model implies $\kappa_1 = -\beta^2_\Phi = \alpha_{12}\,(\theta_2/\theta_1)$. Thus, we obtain $\kappa_1$ from the sourcing estimation as the coefficient on the inclusive cost of cells, $\beta^2_\Phi$. For reasons outlined Ben-Akiva and Lerman (1985), our prior is that $\theta_2 \leq \theta_1$. Consequently, $-\beta^2_\Phi$ is a lower bound of the $\alpha_{12}$ cost share.

## 7.2 Filters applied before estimation

The simulated method of moments we use in section 8 to estimate fixed costs parameters is computationally intensive because each iteration of the estimation solves each firm's problem for many draws of the shocks. The global minimizer we use to avoid local minima of the SMM objective function requires multiple sets of initial parameters. Then the local solver requires many iterations. To keep the time manageable, we reduce the dimensions of the problem as follows:

1. Restrict to 24 countries that are important as cell or EV producers or as markets for EVs. These countries account for 99.9% of cells produced, 98.6% of EVs produced, and 97.1% of EV sales.

2. Restrict to 15 large multinationals that account for 77% of worldwide sales and 98% of sales outside China.

[19] 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).

Table 5: Top 15 firms in 2022

| No. | Manufacturer | # Markets | # Models | Sales Cum. Share (%) | Sales-exCHN Cum. Shr (%) |
|---|---|---|---|---|---|
| 1 | Tesla | 56 | 4 | 17.9 | 28.3 |
| 2 | Volkswagen | 65 | 23 | 25.9 | 42.2 |
| 3 | Hyundai | 53 | 15 | 30.7 | 53.6 |
| 4 | Stellantis | 20 | 19 | 34.6 | 62.5 |
| 5 | BMW | 61 | 8 | 37.9 | 68.5 |
| 6 | Renault | 38 | 6 | 40.3 | 74.2 |
| 7 | Mercedes-Benz | 57 | 9 | 42.6 | 78.9 |
| 8 | Geely | 52 | 15 | 48.0 | 83.6 |
| 9 | Ford | 36 | 4 | 49.6 | 87.3 |
| 10 | SAIC | 33 | 22 | 60.4 | 90.8 |
| 11 | Nissan-Mitsubishi | 51 | 8 | 62.6 | 94.2 |
| 12 | General Motors | 10 | 7 | 63.7 | 95.7 |
| 13 | Toyota | 40 | 8 | 64.3 | 96.8 |
| 14 | Rivian | 4 | 3 | 64.6 | 97.6 |
| 15 | BYD | 24 | 14 | 76.6 | 98.3 |

3. Restrict to 136 models offered by the 15 large firms, with price data available in 2022.[20]

None of the above filters is needed for the sourcing estimation because the $\xi_{mn}$ are not relevant for sourcing and the PPML estimator used for sourcing is very quick. However, we impose the filters to ensure the parameters correspond to the sample used later in the SMM. When estimating fixed costs we found it helpful to also aggregate potential plants from their specific latitude and longitudes to 15 locations of each stage with 5 per continent (Americas, Europe, Asia).[21] We do not impose this aggregation in the sourcing regressions because the point locations give more precise distance and country information for estimating the distance and border effects.

The filter of keeping only the top 15 firms retains a large share of the world's sales. This is because it mainly removes Chinese firms that serve the Chinese market only. In our data there are 52 firms that assemble EVs in 2022 for whom we have sales by market.

[20]Prices are used solely in the inversion to determine the demand shifters, $\xi_{mn}$. We predict $\xi_{mn}$ in countries in four countries that lack prices, but models that do not have prices in any market have to be dropped. This eliminates the electric versions of some minivans and light trucks.

[21]In the SMM stage, Americas has Canada, Mexico, and 3 regions in the United States; Europe has a German-speaking region (Germany + Netherlands + Austria), a French-speaking region (France + Belgium + Luxembourg), Eastern Europe, Southern Europe (Italy + Spain + Portugal + Greece), and UK. Asia consists has Japan, South Korea, and 3 regions in China. When measuring distances, location latitude and longitude are averages of plants' coordinates for each stage within each location.

Of those, 14 sell exclusively in China. Table 5 shows that the top 15 firms—ranked by size of sales outside China —account for 98% of all EVs sold outside China. Because the Chinese market is so large, these firms account for a smaller share (77%) of the world. Other than GM, Rivian, Stellantis and BYD, the top firms sell in 33 or more of the 73 countries that buy EVs in our dataset in 2022.

## 7.3 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.[22] These estimates show variable costs of production and delivery of cells to assembly sites, and of assembled vehicles to dealerships in the 24 markets.

Table 6: Combined sourcing regression results

| Stage: | (1)<br>Cells | (2)<br>Vehicle |
|---|---|---|
| border | -0.953$^a$ | -1.04$^a$ |
| | (0.319) | (0.254) |
| log distance | -0.382$^a$ | -0.112$^c$ |
| | (0.021) | (0.062) |
| RTA | 0.458 | 0.869$^a$ |
| | (0.320) | (0.214) |
| Inclusive cost of cells | | -0.234$^a$ |
| | | (0.084) |
| log GDP per capita | 0.213$^c$ | 0.206$^b$ |
| | (0.118) | (0.087) |
| log(1+tariff) | -8.49$^a$ | -8.56$^a$ |
| | (2.34) | (1.74) |
| Observations | 7,945 | 15,793 |
| Squared Correlation | 0.322 | 0.265 |

Note: EC(cells) are the expected costs of cells for assemblers, $\mathbb{E}c^1_{\cdot,\ell_2}$, defined in equation (19). Log GDP per capita is measured at the sub-national level.

Geographic proximity matters a great deal for sourcing of both cells and vehicles. We see this in the large border effects. Distance (measured using the great circle formula) influences cell sourcing but has a less significant effect on vehicle sourcing. This

[22] Additional specifications that vary the control variables for each stage are shown in the appendix tables E6 and E7. Those estimations also consider the whole sample available without the restriction imposed here on major countries and firms.

last result contrasts with the ones obtained (with a similar estimation method) for vehicles of all fuel types between 2000 and 2016 (and therefore mostly ICEs) from Head and Mayer (2019), where distance effects were stronger (the comparable coefficient being -0.323). BEVs therefore seem to be less sensitive to physical distance once assembled.[23]

Tariffs strongly influence sourcing decisions at both stages. The tariff elasticities are 8.49 and 8.56 for cells and vehicles, respectively. The coefficient on vehicle sourcing is remarkably similar to the one obtained by Head and Mayer (2019). Another influential policy variable is RTAs, which have a strong effect for EVs. This variable captures the removal of non-tariff barriers to trade which occurs inside those agreements, and seem more influential for cars than for batteries.

As we detailed above the coefficient on the inclusive cost of cells in the vehicle sourcing regression (Column 2) has a special meaning. It provides the weight, denoted $\kappa_1$ to use on the cell stage in equation (22). We can use the coefficient to determine the implied cost share of cells: $\alpha_{12} = -\theta_1\beta_\Phi^2/\theta_2$ As $\beta_\Phi^2 = -0.234$, if $\theta_1 = \theta_2$, the implied cost share of cells is around 23%. AlixPartners (2021) computes the share of batteries including packs and the battery management system as 8,000 EUR or 34% of the car they studied. Cells accounted for 75% of that cost, or 25.5% of the vehicle cost. This figure is remarkably close to our estimate but we need to add the caveat that the coefficient on the inclusive cost of cells varies across specifications (as seen in the appendix).

The regressions include source country fixed effects that we display and discuss in the Appendix E.5. We capture intra-national variation in costs using log GDP per capita measured at the GAUL1 administrative level (States and provinces in North America). This productivity variable is more significant for cells than for vehicles.[24] However, we do not use those country-level marginal cost estimates. There is selection bias in the choice sets of open facilities for each firm: our framework predicts that facilities are opened for a combination of low variable and fixed costs. We therefore use the SMM estimation described in the next section to recover both cost terms jointly.[25]

[23] An important distinction between the two studies is that the present one is much more granular, being able to geo-locate all plants, rather than to simply assign them to countries.

[24] Sub-national productivity has a stronger and more significant effect on sourcing decisions for vehicles in column (3) of table E7 where we use the whole universe rather than the filtered data as in columns (4) and (5).

[25] A second concern of the fixed effects for some countries is that they are identified by the choices of one or two models, leading to over-fitting. Despite these two concerns, the fixed effects retain a useful purpose in the sourcing regressions. Conditioning on them allows us to recover consistent estimates of the dyadic frictions.

# 8 Estimation of remaining paramters

The remaining parameters are the location-specific marginal cost differences $\text{FE}^k_{\ell_k}$ and $\kappa_2$ scaling parameter from the path marginal cost (22), and the location fixed costs $\phi_{fg_k\ell_k}$. For now, we keep the product entry decisions $z_{mn}$ fixed at their observed $\{0,1\}$ status. The associated fixed costs $\phi_{mn}$ are therefore irrelevant, so a choice of $\{\kappa_2, \text{FE}^k_{\ell_k}, \phi_{fg_k\ell_k}\}$, along with our previously estimated/calibrated parameters from sections 6-7 fully determines the firms' profit function in (1). We group the locations into three continents $r \in \{\text{Am}, \text{As}, \text{Eu}\}$ and estimate those remaining parameters using SMM. We parametrize the fixed cost $\phi_{fg_k\ell_k}$ as a draw from a lognormal distribution

$$\phi_{fg_k\ell_k} \sim \text{LN}\left(\log\left(\rho^k_r \times \text{WREV}\right) + \rho^{\text{HQ},k} \times \log(D_{h(f)\ell_k}), \sigma\right), \tag{23}$$

where $\rho^k_r$ denotes the expected fixed cost of a single production line in continent $r$ for stage $k$ as a share of *observed* world EV revenue, and $\rho^{\text{HQ},k}$ denotes the elasticities of fixed costs with respect to distance from firm headquarters for each stage of production. Currently, we fix $\sigma$ to 1 so as to focus on the effects of dispersion for the marginal path cost shocks. Also note that changes in $\sigma$ would affect the *mean* of the log fixed cost for the *chosen* facilities. For the rest of the paper, we define $\boldsymbol{\rho} \equiv \{\rho^k_r, \rho^{\text{HQ},k} \mid r \in \{\text{Am}, \text{As}, \text{Eu}\}, k \in \{1,2\}\}$ as the vector of 8 parameters that determine the fixed costs $\phi_{fg_k\ell_k}$ draws.

We use the same continental grouping by $r \in \{\text{Am}, \text{As}, \text{Eu}\}$ for the location-specific marginal cost differences, reducing those fixed effects to a vector $\mathbf{FE} \equiv \{\text{FE}^k_r \mid r \in \{\text{As}, \text{Eu}\}, k \in \{1,2\}\}$ of 4 parameters, normalizing the parameters for $r = \text{Am}$ to zero. The $\kappa_2$ parameter then captures the levels of the marginal costs relative to the fixed costs – as well as the dispersion of the $u(\cdot)$ shocks. There are thus 13 parameters $\{\kappa_2, \mathbf{FE}, \boldsymbol{\rho}\}$ to estimate.

We draw a $136 \times 15 \times 15 \times 24$-dimensional matrix of marginal path-level cost $u(\cdot)$ shocks and a $15 \times 15 \times 2$-dimensional matrix of fixed costs $S = 100$ times.[26] For one set of draws and given a parameter guess $\{\kappa_2, \mathbf{FE}, \boldsymbol{\rho}\}$, we solve each firm's MMM-UFLP optimal location problem (this also uses the previously estimated/calibrated parameters from sections 6-7). Because the market-level aggregator $R^{\text{EV}}_n/(C^{\text{EV}}_n)^{1-\eta}$ is fixed, based on our data for the observed EV expenditure and paths of models sold in each destination markets, each firm's MMM-UFLP can be solved independently. We then obtain a simulated moment by averaging across the $S$ simulations.

We use 60 targeted moments, matching those simulated averages to their observed

[26] 15 firms $f$, 15 locations $\ell_k$, 136 car models $m$, and 24 destinations $n$ for each stage $k$.

values in our data. Those moments are organized into 4 blocks, grouping locations into the same three continents $r \in \{\text{Am}, \text{As}, \text{Eu}\}$ corresponding to the cost parameters. The first block targets the number of production lines by continent $r$ and stage $k$. A production line is defined as a firm $f$ having an active facility to produce group $g_k$ at location $\ell_k$. The remaining three blocks target dyadic flows across continents: (i) the fraction of production lines at each destination region and stage installed by firms from their home region (where headquartered); (ii) the fraction of models traded between regions at each stage;[27] (iii) the share of import value between regions at each stage. These moments are informative about the overall magnitude of fixed costs of building production facilities, as well as relative variable costs when those facilities are activated. The production line moments in the first two blocks help to identify the mean fixed costs by continent, and how they vary with distance to firm headquarters. The latter two blocks use the sourcing decisions and bilateral flows to pin down the level of variable costs relative to fixed costs and relative variable cost differences across continents.

Formally, the vector of moment function, $v(\cdot)$, specify the differences between the observed outcomes and those predicted by the model. At the true parameter value, $\mathbb{E}\big[v\left(\kappa_2, \mathbf{FE}, \boldsymbol{\rho}\right)\big] = 0$. SMM finds an estimate such that

$$\min_{\kappa_2, \mathbf{FE}, \boldsymbol{\rho}} \hat{v}\left(\kappa_2, \mathbf{FE}, \boldsymbol{\rho}\right)' W \hat{v}\left(\kappa_2, \mathbf{FE}, \boldsymbol{\rho}\right), \tag{24}$$

where $\hat{v}(\cdot)$ is the simulated estimate of the moment function and $W$ is a weighting matrix. We weight the first set of moments by the inverse of observed production line to represent percentage difference so that the units of all moments are comparable and weighted equally. As robustness checks, we also applied the optimal weighting matrix which is the inverse of the estimated variance-covariance matrix of all moments. The first (identity) weighting matrix leads to a better fit of large flows across continents which are more relevant for counterfactuals.

One difficulty in estimating this type of problem is that there could be multiple local optima. To search for a global optimum, we implement a multi-start global optimization algorithm, TikTak, explained and benchmarked by Arnoud et al. (2019). The algorithm starts with a uniform exploration of the parameter space and uses the information it accumulates to increasingly focus the search on the most promising region.[28]

[27] At the cell stage, it is the share of BEV models assembled in each region that used cells produced from respective regions. At the car stage, it is the share of BEV models sold in each region that are assembled from respective regions.

[28] We implement TikTak with 1000 random starts. The algorithm retains the 25 parameter vectors with the best objective function values and augments this set with a "prepend" parameter vector (based on priors or a vector of previous estimates). We use BOBYQA as the local optimizer and then polish the outcome of

Table 7: SMM parameter estimates

| Param. | Region ($r$) | Description | Estimate |
|---|---|---|---|
| $\kappa_2$ | | Variable cost weight and $u(\cdot)$ dispersion | 0.80 |
| $\text{FE}_r^1$ | As | Variable cost advantage of cell plant | 1.01 |
| $\text{FE}_r^1$ | Eu | (by region, Am = 0) | 0.73 |
| $\text{FE}_r^2$ | As | Variable cost advantage of assembly plant | -0.02 |
| $\text{FE}_r^2$ | Eu | (by region, Am = 0) | -0.28 |
| $\rho_r^1$ | Am | Fixed cost of cell plant (by region) | 0.08 |
| $\rho_r^1$ | As | | 0.03 |
| $\rho_r^1$ | Eu | | 0.06 |
| $\rho_r^2$ | Am | Fixed cost of assembly plant (by region) | 0.08 |
| $\rho_r^2$ | As | | 0.03 |
| $\rho_r^2$ | Eu | | 0.05 |
| $\rho^{\text{HQ},1}$ | | Fixed Cost HQ distance elas. (cells) | 0.05 |
| $\rho^{\text{HQ},2}$ | | Fixed Cost HQ distance elas. (assembly) | 0.51 |

Table 7 shows the results of the SMM estimation. The weight on observed variable costs relative to fixed costs and also the variance of unobserved path cost shocks is $\hat{\kappa}_2 = 0.8$. This implies a $\theta_2$ of 1.25. This magnitude is less than 8.56 implied from the tariff elasticity in Table 6, column (2), suggesting a large role for omitted non-tariff barriers for vehicles.

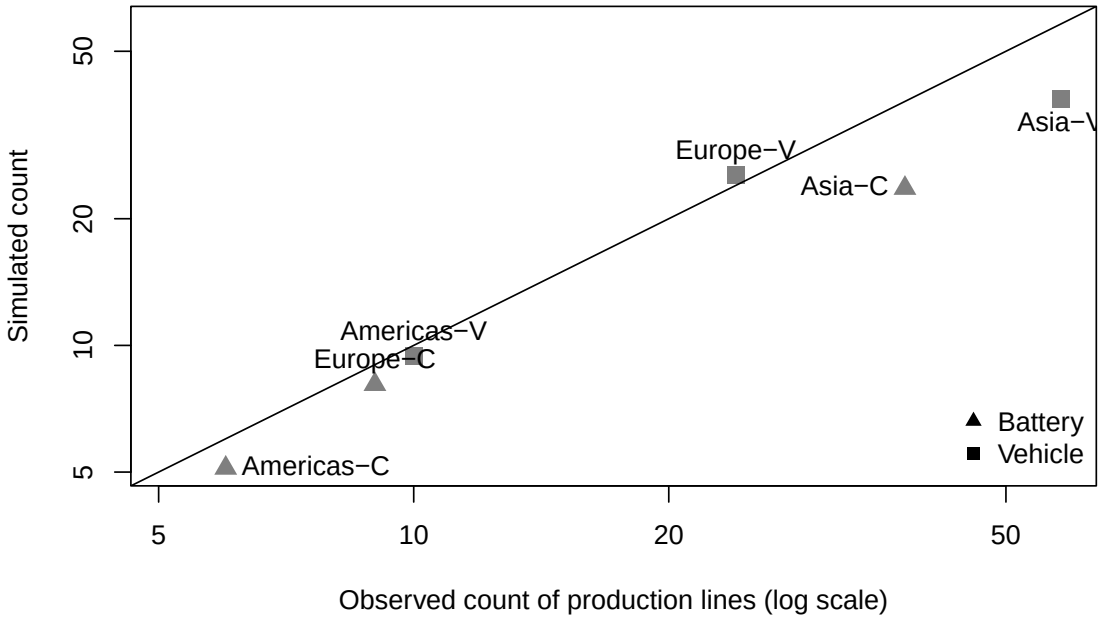
The $\boldsymbol{\rho}$ parameters determine the mean of the log fixed cost draws. We estimate the fixed costs to be highest for the Americas and lowest for Asia. Those continental differences are moderated by the distance to the firms' headquarters. Both sets of fixed costs rise as firms build plants further from their headquarters, but the effect on assembly is an order of magnitude larger than on cells.

Figures 8 and 9 display the within-sample performance of the SMM estimation for the moments we target. Figure 8 shows the count of production lines on each continent in the data (horizontal axis) and in the model (vertical axis). Recall that production lines are defined here as combinations of EV model and production location. Thus, cells for Tesla Y and cells for Tesla 3 produced in separate factories would be two production lines, as would both models produced in the same factory. The log difference between these vectors is one of our targeted moments, so it is not surprising that the model does a good job of matching the data in this figure.

The second important type of fit targeted in the SMM is between the share of models chosen from each origin continent to serve the markets in each destination continent. This

the global optimizer using BOBYQA with a tightened convergence criteria.

Figure 8: Fit of the estimated model: production lines by continent

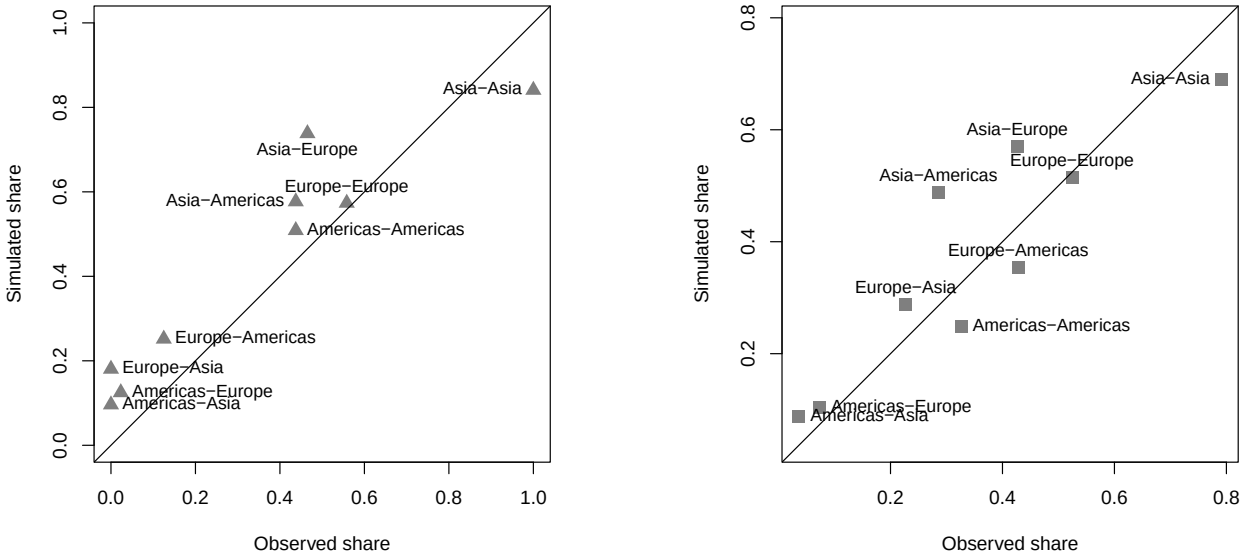



Note: A production line is defined as a combination of EV model and production location.

Figure 9: Fit of the estimated model: origin-destination model sourcing

(a) battery cells

(b) vehicle assembly

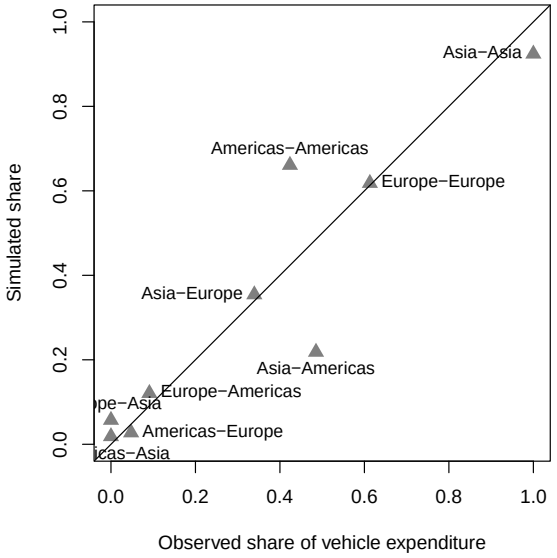



Note: The SMM estimation targets the share of models each continent sources from the others, for both cells and vehicles.

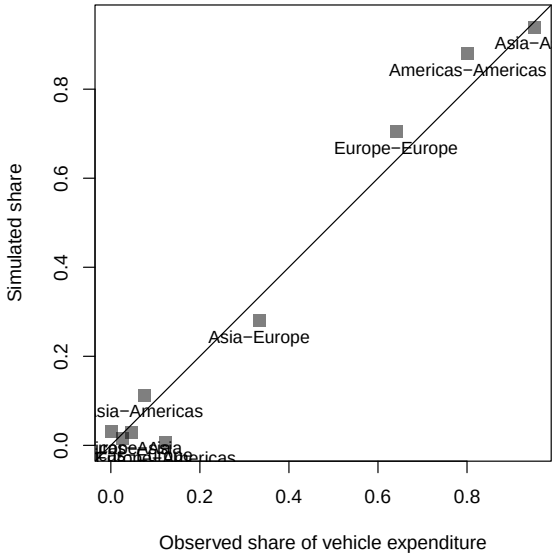
is shown in Figure 9. Panel (a) shows the share of cells sourced from each continent to serve each destination continent. Panel (b) shows the same for vehicle assembly. The simulation fits cell sourcing patterns quite closely, except for Asian producers share of their own continental market and of the European one. The fit is not as close for vehicle sourcing, with the America-Americas share underpredicted in the simulation. Asia to Americas is correspondingly overpredicted but Europe to Americas is quite close. This pattern warrants addition exploration. The model correctly predicts that Americas will be the source for few cell or vehicle models in other continents and that Asia cell and vehicle makers will dominate their own continent.

Figure 10: Fit of the estimated model: origin-destination expenditure shares

(a) Cells




(b) Vehicles




Note:

# 9 Counterfactuals

Many policies currently incentivize EV purchases and production at home or at least in friendly countries. We consider five variations around this theme:

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.

4. **Production subsidies for US-made cells** Appendix F shows the results for these 20% subsidies as a standalone policy and in combination with the value-chain conditional subsidy as "3+4".

5. **Tariffs on Chinese imported cells and vehicles**. We impose a 20% tariff on cells and vehicles imported from China as policy 5, restrict the tariff to vehicles as 5V and add it on to the value-chain conditional subsidy as "5+3" in the appendix.

All these policies are set at a level of 20% to make it easy to compare them. Other than this magnitude, they are designed to reflect different aspects of US policy towards EVs under the IRA (and also as carried out in tariffs under the Trump 1, Biden, and potentially Trump 2 administrations). Policy 1 resembles the status quo before the IRA when subsidies were not contingent on production locations.[29] Policy 2 add the "buy American" requirement at the assembly level. We impose this at the continental level, which is similar to the requirement that the IRA includes FTA partners Canada and Mexico. Policy 3 mirrors the IRA policies regulating the input sources of EVs to be eligible for the subsidy. Beginning in 2023, 40% of an EV battery's minerals and 50% of the components must come from the US or Free Trade Agreement (FTA) partners. In 2027 and 2029, this requirement will increase to 80% for minerals and 100% for components. In our simulation, the rule is that only cells produced inside the Americas are eligible for the consumer subsidy.

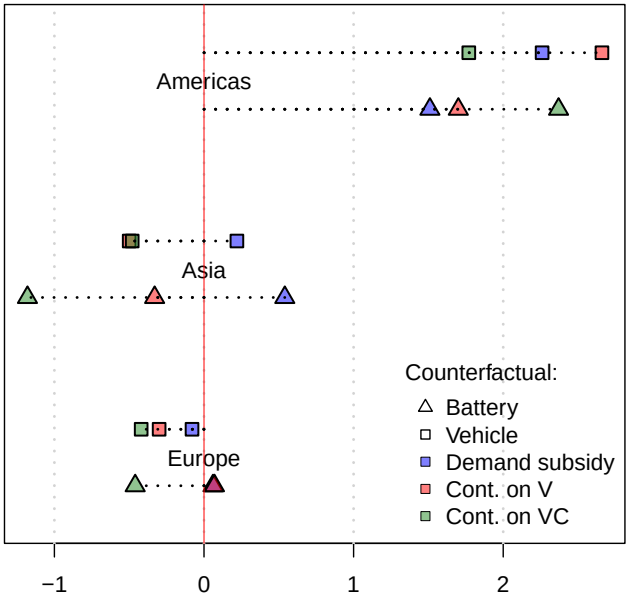
## 9.1 Policy Simulation Results

The primary outcome variables are the number of production lines activated in each continent and total expenditures on EVs. The former is our proxy for production and employment, and we treat the latter as a proxy for emissions reductions. We do not model directly how CO2 emissions from internal combustion engines change in response to our policies. The presumption is that rises in EV sales come in large part from reductions in gas engine vehicles. In the US and Europe, Figure ES.1 in Bieker (2021) shows that battery EVs emit 60–69% less CO2 over their lifetime than gas cars.

[29] There were other restrictions on eligibility that we do not build in here.

Figure 11: Counterfactual results with 20% subsidy by continent

(a) Continental production lines




(b) Expenditures on EVs

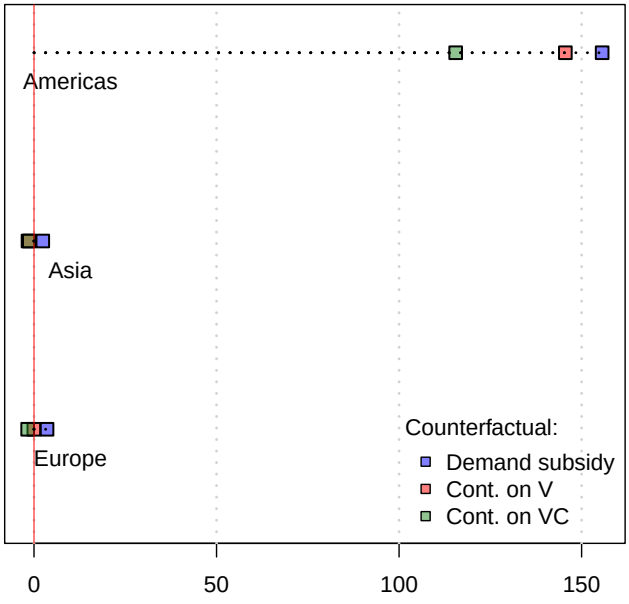



Figure 11(a) provides the continental summary of our $y$ plant activation indicators. Blue squares (EVs) triangles (cells) correspond to the change in the number of production lines for each continent in response to the 20% reduction in the purchase price to all buyers in North America. We see that this activates an average of about two production lines for assembly and slightly less for cell-making lines in the Americas. The rise in the number of Vehicles assembled in the Americas comes partly at the expense of Europe where V production lines fall. Thus, a subsidy that is ostensibly neutral in its application—because it is not contingent on the value chain locations—actually moves production around the three continents.

Insisting on domestic final assembly ("Buy American" generalized to include Canada and Mexico) raises production lines by around 2.5 compared to the baseline and little compared to the non-contingent subsidy. This hurts Asian assembly relative to the baseline and particularly compared to the non-contingent subsidy. European assembly is more adversely affected by the buy-American subsidy. The dyadic figures in Figure 12 show the assembly-contingent subsidy promotes cell exports from Asia to the Americas, while lowering vehicle exports. The complementarity between assembly and cell production appears clearly in these results.

While successful in terms of domestic protection, figure 11(b) shows that Policy 2 low-

ers expenditures on EVs and thus is not appealing as an environmental policy relative to the unconditional subsidy in Policy 1.

Table 8: Contributions to cost changes (in %) in Americas

| Policy | Elig. share | | Cost index redn. | | | Tot. Exp. |
|---|---|---|---|---|---|---|
| | # path | revenue | subsidy | costs | total | increase |
| 1: Unconditional | 100 | 100 | 20 | 4.9 | 23.9 | 155.6 |
| 2: Continental V | 30.7 | 96.6 | 19.0 | 4.4 | 22.7 | 145.5 |
| 3: Continental V+C | 18.9 | 84.1 | 15.9 | 2.8 | 18.7 | 115.5 |

Policies defined in paper. All numbers in percent. Cost changes from subsidy is $(\sum_{m\in\text{EV}}\sum_{n\in\text{Am}} r_{mn}(1-\mathbb{I}_{mn}s)^{1-\eta})^{1/(1-\eta)}$, where $r_{mn}$ is the $mn$ share of continental expenditure in the baseline simulation (no policy). $\mathbb{I}_{mn}$ indicates $m$ is eligible for the subsidy of $s = 20\%$ offered in market $n$, $\mathbb{I}_{mn} = 1$ for policy 1 for $n \in$ Am. Elig. is the average of $\mathbb{I}_{mn}$. Production cost reductions are $(\sum_{m\in\text{EV}}\sum_{n\in\text{Am}} r_{mn}(\hat{c}_{mn})^{1-\eta})^{1/(1-\eta)}$ and result from new paths chosen due to new plant activations closer to market $n$. The "total" cost reduction includes both $s\mathbb{I}$ and $\hat{c}$ effects. All changes average across simulations.

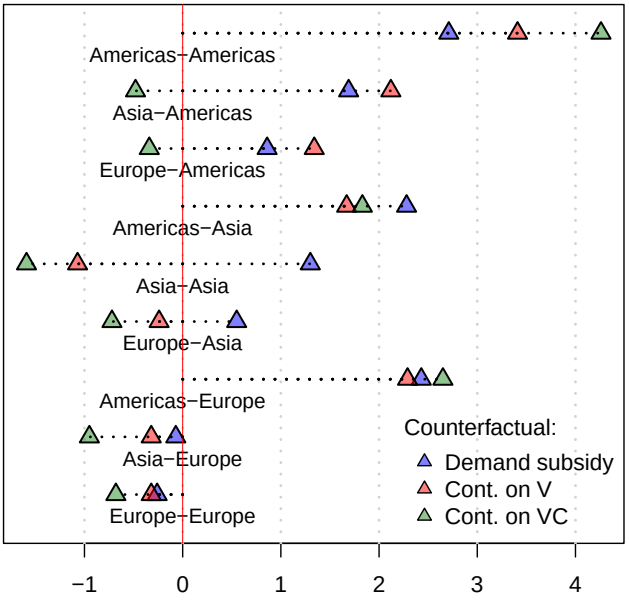
Policy 3 is worse from an environmental perspective. This is because few EVs sold in the Americas end up qualifying for the subsidy when it is contingent on American cell sourcing. Table 8 shows that only 19% of the paths are eligible. Among the small number who do draw low enough fixed costs in the Americas to source cells within the region are the largest selling vehicles since they represent 84% of sales. Overall, the cost index decrease is 19%, which is 5 p.p. lower than with the unconditional policy. This is because the cost savings coming purely from the subsidy is reduced by incomplete eligibility, which also lowers the savings related to having shorter distances in the optimized paths. Policy 3 manages to be worse than non-intervention from the point of EV production in both Asia and Europe, while increasing EV production in the Americas *less* than the other two policies. This is an example where the policy is successful at attracting production of the major input (cells), but with a detrimental effect on the production of vehicles that use those local cells.

In the appendix table F8 we see that augmenting policy 3 with a cell production subsidy would add vehicle production lines relative to policy 3 alone. Policy 3+4 does maximize the number of cell production lines in the Americas, so if that goal is sufficiently important for non-environmental reasons, it could have a justification.

Appendix section F presents the numbers underlying the visualizations in Figures 11 and 12 in Tables F8,F9, and F10.

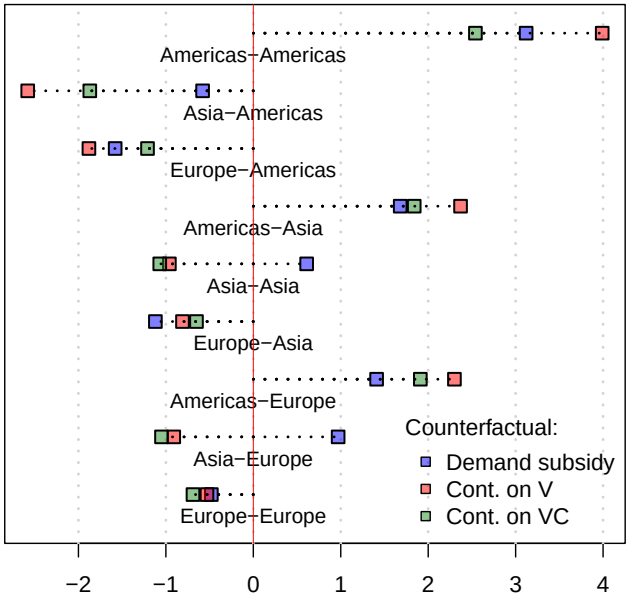
Figure 12: Counterfactual results with 20% subsidy: by origin-destination

(1) Battery Cells




Change in number of battery models wrt baseline

(2) Vehicle assembly




Change in number of vehicle models wrt baseline

# 10 Conclusion

The battery and electric vehicle industry is an important example of a value chain where optimal location conditions are determined in a complex way balancing trade costs, fixed costs, and a mix of substitution and complementarity effects. Just solving the multi-stage problem computationally has been a challenge, leading most researchers to leave out fixed costs. However, fixed costs are important because they lead to increasing returns, which themselves make policy consequences much harder to predict.

Our main contributions are 1) to apply integer programming methods to this problem and show it can solve for optimal locations with fixed costs very quickly and 2) to devise a two-step estimation procedure to estimate the parameters governing variable and fixed costs. Our current best estimates imply that contingent subsidies work as protectionist measures, especially when applied to the final stage. But they impede the environmental goals of the government relative to non-contingent subsidies.

# References


AlixPartners (2021). Accompagner la filière automobile dans la transition énergétique, connectée et partagée. Slide Deck.

Allcott, H., R. Kane, M. S. Maydanchik, J. S. Shapiro, and F. Tintelnot (2024, December). The effects of Buy American: Electric vehicles and the Inflation Reduction Act. Working Paper 33032, National Bureau of Economic Research.

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*, 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.

Arnoud, A., F. Guvenen, and T. Kleineberg (2019). Benchmarking global optimizers. Working Paper 26340, National Bureau of Economic Research.

Barwick, P. J., H.-S. Kwon, S. Li, and N. B. Zahur (2025, January). Drive down the cost: Learning by doing and government policies in the global EV battery industry. Working Paper 33378, 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.

Bieker, G. (2021). Global comparison of the life-cycle greenhouse gas emissions of passenger cars. White Paper.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*(2), 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.

Fajgelbaum, P. D., P. K. Goldberg, P. J. Kennedy, and A. K. Khandelwal (2020). The return to protectionism. *The Quarterly Journal of Economics 135*(1), 1–55.

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.

# Appendix

## A 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[30] (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 2 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.

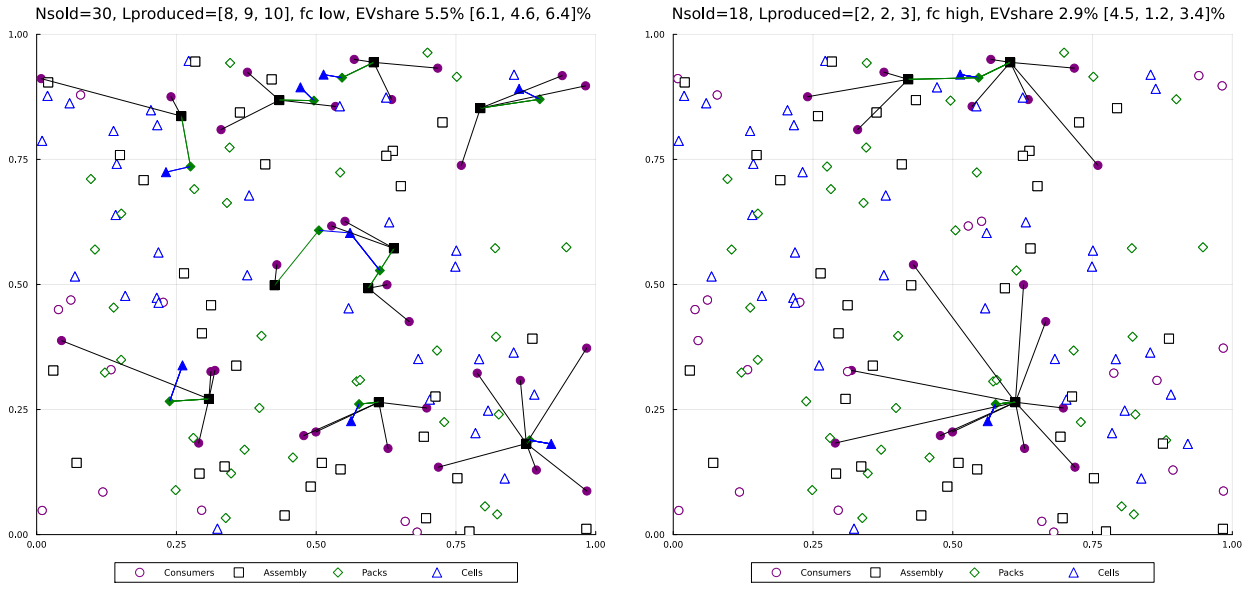
Each of the mapped simulation below takes about 20–40 seconds to solve the firm's location problem, depending on the computer and the levels of fixed costs and subsidies. This establishes that the mixed integer programming solver can handle realistically sized problems in a reasonable amount of time.[31]

We begin the investigation of the MUFLP by simulating the model without any subsidies for EVs. The left panel of Figure A1 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 demand, (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

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

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

Figure A1: Simulation: no policy interventions




(a) Low fixed costs

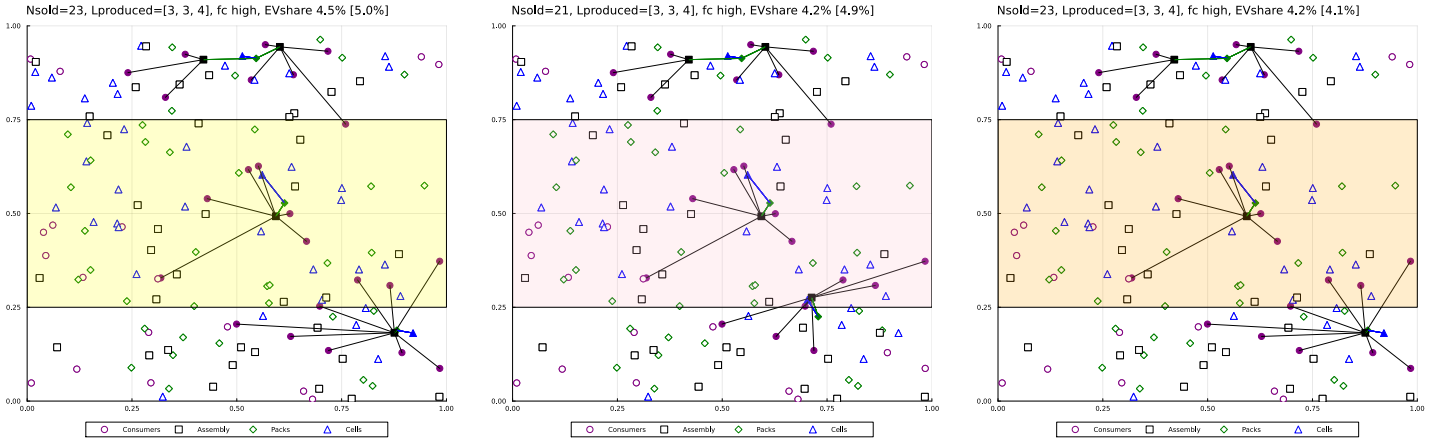
(b) High fixed costs

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 A2 brings consumer subsidies into the picture, continuing the high fixed cost case from Figure A1. 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 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%

Figure A2: Simulation: $5,000 subsidies in Middle




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

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 affordability 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. As in that paper the point where Laffer curve effects kick in depends on the size of the subsidy.

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.

# B Computational performance of the algorithm

In this section, we evaluate the computational performance of the mixed integer linear programming solution to combinatorial discrete choice problems using a set of simulations.[32] We conduct three different speed tests, each highlighting different aspects of the

[32]Computational performance tests are run on author's MacBook Pro with Apple M2 Max chip and 64GB memory via serial processing.

algorithm.

## B.1 Single-stage UFLP compared with AES squeezing algorithm

To make a fair comparison we need a problem that can be solved using both AES and MILP. AES requires $K = 1$ and MILP requires single sourcing. A simplified version of the problem considered by Arkolakis et al. (2023) is stated in Box B.1.

**Box B.1: Special case of AES single-stage UFLP formulation**

The firm's variable profit selling to market $n$ is a function of its set of potential sourcing locations, denoted $\mathcal{L}$. With iso-elastic demand proportional to $p^{-\eta}$, the variable profit in market $n$ is $\pi_n(\mathcal{L}) = kc_n(\mathcal{L})^{1-\eta}$, where $\mu = \eta/(\eta-1)$ is the optimal markup and $k = (\eta-1)^{\eta-1}\eta^{-\eta}$. The firm chooses $\mathcal{L} \subset L$ to maximize profit net of fixed costs $\phi$ incurred in each activated plant:

$$\mathcal{L}^* = \arg\max_{\mathcal{L}} \left\{ \sum_n \pi_n(\mathcal{L}) - \sum_{\ell \in \mathcal{L}} \phi_\ell \right\}.$$

With a continuum of inputs as in Tintelnot (2017), the cost function is

$$c_n(\mathcal{L}) = \left[ \sum_{\ell \in \mathcal{L}} c_{\ell n}^{-\theta} \right]^{-1/\theta}, \quad \text{where} \quad c_{\ell n} = w_\ell \tau_{\ell n} / z_\ell.$$

For finite $\theta$, the cost minimizing solution entails multi-sourcing (the supplier's optimal sourcing shares lie between 0 and 1 for all members of the choice set). The limiting case as $\theta \to \infty$ is given by $c_n(\mathcal{L}) = \min_{\mathcal{L}} c_{\ell n}$. Perfect substitution implies single-sourcing from the least-cost $\ell \in \mathcal{L}$. Randomness arises from firm-location productivity shocks $(z_\ell)$ and the locations of producers and consumers, both drawn from a uniform distribution on a unit square. The Euclidean distances between locations determine $\tau_{\ell n}$.

In the first test, we compare how long it takes to solve for the optimal set of facility locations comparing three methods:

1. formulating firm's total profit as mixed integer linear programming (MILP)
2. using the squeezing algorithm proposed by Arkolakis et al. (2023) (AES)

3. "brute-force": computing the profit associated with every possible configuration, then sorting to find the maximum profit configuration (BF)

The comparable setting between MILP and AES is a single-stage combinatorial discrete choice problem by a single-product firm whose market entry is exogenous. The AES method converges on this problem as their $\theta \to \infty$, that is, as the different location choices become perfect substitutes.

Box B.2 shows the formulation of the single-stage UFLP as a MILP.

**Box B.2: Single-stage UFLP formulation as a MILP**

$$\max \sum_{n \in N} \sum_{\ell \in L} \pi(c_{\ell n}) x_{\ell n} - \sum_{\ell \in L} \phi_\ell y_\ell \quad \text{subject to}$$

$$\sum_{\ell \in L} x_{\ell n} = 1, \quad n \in N \tag{B.1}$$

$$x_{\ell n} \leq y_\ell, \quad n \in N, \ell \in L \tag{B.2}$$

$$x_{\ell n} \geq 0, \quad n \in N, \ell \in L \tag{B.3}$$

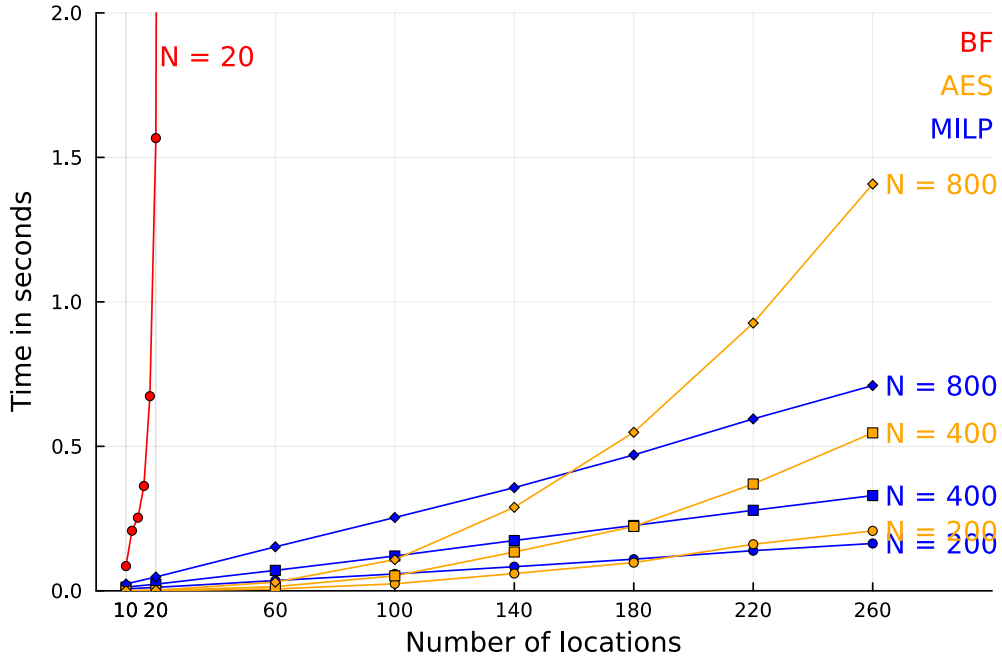
$$y_\ell \in \{0, 1\}, \quad \ell \in L \tag{B.4}$$

We generate data by randomly drawing $L$ potential production locations and $N$ markets on a unit square. The firm is assumed to sell its product in every market. The number of locations $L$ varies within $\{10, 20, 60, 100, 140, 180, 200\}$, and the number of markets varies within $\{200, 400, 800\}$. We found that brute-force demands too much memory to be applied for more than 22 locations and markets.[33] For brute-force, we restrict $L \in \{10, 12, 14, 16, 18, 20, 22\}$ and $N = 20$. The data is simulated 100 times to compute the average algorithm's run time.

Figure B3 shows that the run times (seconds) for MILP and AES are comparable at low number of potential locations, with AES being slightly faster. As the number of locations increases, the computational performance between these two algorithms start to diverge and MILP dominates. The reason is that for a single-stage uncapacitated facility location problem, the possible assignment for each market is $L$, and thus the computation time of MILP increases with locations at a linear rate (conditional on the number of markets). Nevertheless, both algorithms drastically reduce the run time compared to brute force, and extend our capability of solving combinatorial discrete choice problem in real world

[33]The memory requirement is proportional to $2^L$, so we can project the memory needed for $L = 25$ to be almost 200GB.

Figure B3: Three ways to solve a single-stage UFLP




applications. Panel (a) of Table B1 provides more details of the average run time for each parameter setting.

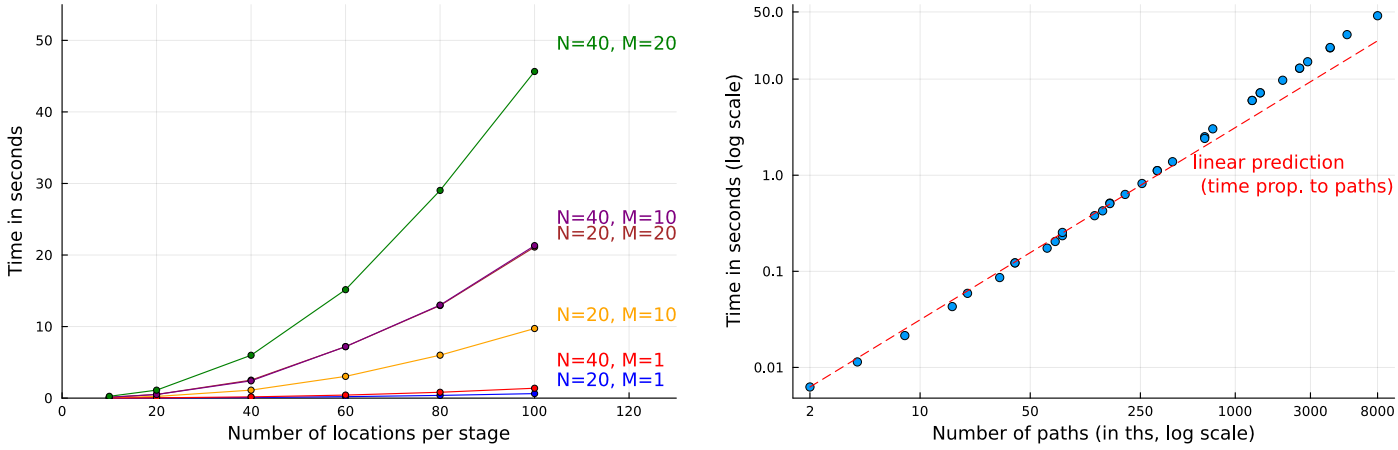
## B.2 Times for multi-stage UFLP

Here we evaluate the computational speed solving a two-stage supply chain for MILP alone because introducing more than one stage of production violates single crossing difference condition required to use AES. We also relax other restrictions in the first test by allowing for multi-product firm, represented by $M > 1$, and endogenous market entry. We vary the number of locations $L \in \{10, 20, 40, 60, 80, 100\}$, the number of markets $N \in \{20, 40\}$, and the number of products $M \in \{1, 10, 20\}$, and simulate 40 repetitions.

When $K = 2$, there are $L^2$ number of potential assignments for each product-market. For this reason time is convex in the number of locations in Panel (a) of Figure B4. Moreover, because the number of products and the number of markets are isomorphic in determining the size of the problem, only the product $N \times M$ matters and we observe almost equal computation time for $N = 40, M = 10$ and $N = 20, M = 20$.[34] Panel (b) of Table B1 shows how the time scales with the size of the problem. The "size" can be thought of as the number of variables and the number of constraints. However, it turns out that the

[34]The slight difference between column (3) and column (5) shown in Panel (a) of Table B1 is due to simulation error.

Figure B4: Computational Performance of MILP




(a) Two stages: MILP only

(b) Time as a function of $L_1 \times L_2 \times N \times M$

number of paths is what varies the most. The plot in panel (b) of Figure B4 shows that time is very close to linear in $L_1 \times L_2 \times N \times M$. The log-log regression has a coefficient of 1.1, where 1.0 would correspond to perfect linearity.

In the last test, we add one more stage in the supply chain by having $K = 3$. The three-stage problem is formulated as a MILP in section C. The number of potential assignments for each product-market increases to $L^3$. Therefore, we limit the potential locations to 30, and only use 10 repetitions. The run time increases more for each location increments in Panel (c) compared to Panel (a) in Table B1.

## B.3 Explanation for strong performance of MILP formulation

1. The role of the LP relaxation: integral solutions avoid long branch-and-bound searches.

2. Dual Simplex method

As noted in the main text, we formulate the market entry condition as an inequality even though in the solution, it holds with equality. 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 sub-problems where the results from one sub-problem are used to select columns from the original model for inclusion in the next sub-problem.

Table B1: Average computation time in seconds

Panel (a): $K = 1$

| $L_k$ | $N$ = 200 | | $N$ = 400 | | $N$ = 800 | |
|---|---|---|---|---|---|---|
| | MILP | AES | MILP | AES | MILP | AES |
| 10 | 0.01 | 0.0 | 0.01 | 0.0 | 0.02 | 0.0 |
| 20 | 0.01 | 0.0 | 0.02 | 0.0 | 0.05 | 0.0 |
| 60 | 0.04 | 0.01 | 0.07 | 0.01 | 0.15 | 0.03 |
| 100 | 0.06 | 0.02 | 0.12 | 0.05 | 0.25 | 0.11 |
| 140 | 0.08 | 0.06 | 0.17 | 0.13 | 0.36 | 0.29 |
| 180 | 0.11 | 0.1 | 0.23 | 0.22 | 0.47 | 0.55 |
| 220 | 0.14 | 0.16 | 0.28 | 0.37 | 0.59 | 0.93 |
| 260 | 0.16 | 0.21 | 0.33 | 0.55 | 0.71 | 1.41 |

Panel (b): $K = 2$

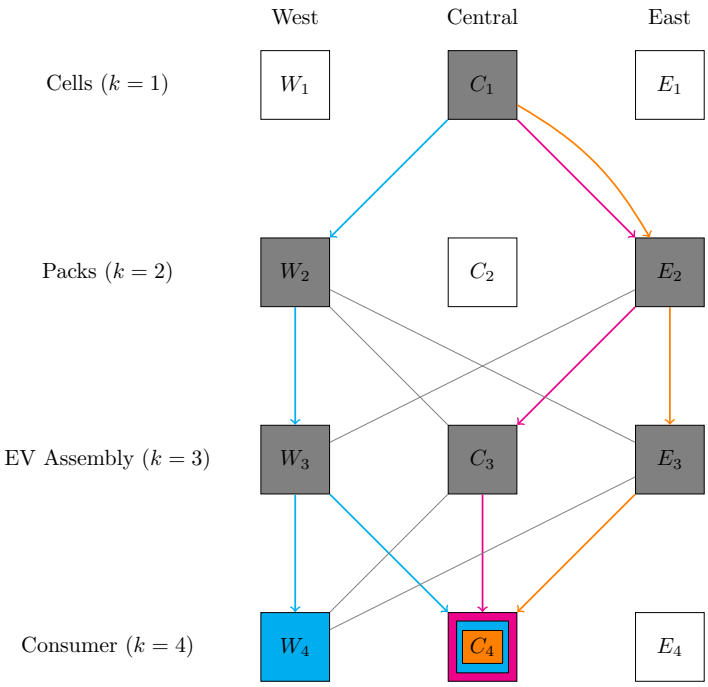
| $L_k$ | $N$ = 20 | | | $N$ = 40 | | |
|---|---|---|---|---|---|---|
| | $M$ =1 | $M$ =10 | $M$ =20 | $M$ =1 | $M$ =10 | $M$ =20 |
| 10 | 0.01 | 0.06 | 0.12 | 0.01 | 0.12 | 0.25 |
| 20 | 0.02 | 0.23 | 0.51 | 0.04 | 0.51 | 1.11 |
| 40 | 0.09 | 1.12 | 2.51 | 0.17 | 2.41 | 5.99 |
| 60 | 0.2 | 3.04 | 7.19 | 0.43 | 7.2 | 15.16 |
| 80 | 0.38 | 6.0 | 12.94 | 0.82 | 13.0 | 29.03 |
| 100 | 0.63 | 9.73 | 21.12 | 1.38 | 21.31 | 45.65 |

Panel (c): $K = 3$

| $L_k$ | $N$ = 20 | | | $N$ = 40 | | |
|---|---|---|---|---|---|---|
| | $M$ =1 | $M$ =10 | $M$ =20 | $M$ =1 | $M$ =10 | $M$ =20 |
| 10 | 0.06 | 0.67 | 1.53 | 0.12 | 1.51 | 3.51 |
| 20 | 1.86 | 26.01 | 57.32 | 4.08 | 57.32 | 128.02 |
| 30 | 6.63 | 87.0 | 187.33 | 14.66 | 186.46 | 409.75 |

# C Three-stage MMM-UFLP formulation

Figure C5: Schematic of a supply chain with $K = 2$




$$\max \sum_{m \in M_f} \sum_{n \in N} \sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} \pi(c(\boldsymbol{\ell}_{mn}), A_n) x_{mn\ell_1\ell_2\ell_3} - \sum_{k=1}^{3} \sum_{\ell_k \in L_k} \phi_{\ell_k} y_{\ell_k} - \sum_{m \in M_f} \sum_{n} \phi_{mn} z_{mn} \tag{C.5}$$

subject to

$$\sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq z_{mn}, \quad n \in N, m \in M_f \tag{C.6}$$

$$\sum_{\ell_1 \in L_1} \sum_{\ell_2 \in L_2} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_3}, \quad n \in N, m \in M_f, \ell_3 \in L_3 \tag{C.7}$$

$$\sum_{\ell_1 \in L_1} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_2}, \quad n \in N, m \in M_f, \ell_2 \in L_2 \tag{C.8}$$

$$\sum_{\ell_2 \in L_2} \sum_{\ell_3 \in L_3} x_{mn\ell_1\ell_2\ell_3} \leq y_{\ell_1}, \quad n \in N, m \in M_f, \ell_1 \in L_1 \tag{C.9}$$

$$x_{mn\ell_1\ell_2\ell_3} \geq 0, \quad n \in N, m \in M_f, \ell_k \in L_k, k = 1, 2, 3 \tag{C.10}$$

$$y_{\ell_k} \in \{0, 1\}, \quad \ell_k \in L_k, k = 1, 2, 3 \tag{C.11}$$

$$z_{mn} \in \{0, 1\}, \quad n \in N, m \in M_f. \tag{C.12}$$

The inequalities (C.7)–(C.9) 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 $\ell_k$ 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.

# D Data Appendix

The first three data sets are all proprietary data sets sold to us by IHS Markit. The other data used in the paper are publicly available.

## D.1 IHS HVBD

The High Voltage Battery Data (HVBD) module provides value chain data by light vehicle model. We use only Battery Electric Vehicles (BEVs) in our analysis, excluding hybrids, defined here as all vehicles with combustion engines. The data show configurations of brand models, showing the supplier firm and plant location for each cell, module, pack, and battery management system used in a car or light truck. The location and owner of the vehicle assembly plant is also shown, but not the destination market where the vehicle is sold. Quantities for each configuration are provided from 2015 to 2024. However, the last year is currently a projection.

## D.2 IHS Sales

The Sales module provides plant of production and destination market for each model sold. The sales quantities are available from 2015 to 2024 (again 2024 is currently a projection). This data set lacks detail on the battery. Furthermore, it does not distinguish BEVs, hybrid, and internal combustion versions with the same nameplate (model name). For that we must combine information from the New Registrations data set.

## D.3 IHS New Registrations (NRq,NRp)

The data we refer to as "newreg" comes to us in two pieces: The NRQ files has quantities and no prices. The NRP file has prices and quantities, but not for all models. NRQ does not have quantities for Japan.

We can obtain $\xi_{mn}$ 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.

## D.4 Tariff data

The primary source of tariffs is WITS, a database maintained by the World Bank. The rates reported are ad-valorem, and at the HS6 digits level of detail. For battery cells, we take the rates reported for HS 850790. For vehicles, we apply the rates of 870380 for passenger cars, and 870490 for commercial vehicles. Note that 870380 is the HS for electric-only vehicle since the major revision of 2017. We use 870390 until the country declares 870380. For both stages we use the data provided by Fajgelbaum et al. (2020) to adjust the rates following the wave of Trump tariffs against China. Regarding the frequent missing values in WITS raw data, we follow the same method as in Head and Mayer (2019): we fill the holes with linear interpolation, and when the missing data is for the latest years, we replace it with the latest available for this pair of countries and this HS.

## D.5 Geography

Distances are calculated from plant to plant for cells to assembly, using latitude and longitude data and exploiting the R **geosphere** package for great circle distances. For distance from the assembly plant to consumer market, we use the **world.cities** data set contained in the R **maps** package. It contains latitude and longitude of close to 1000 cities for France, Japan, Germany, the USA, and Italy, over 500 cities for most countries, with Sweden having the fewest cities at 104. We average the **geosphere**::distGeo distances for each plant-city pair to obtain plant-market distances. To allow discontinuities associated with intercontinental trade, we define continents based on by the R **countrycode** package, which also provides iso codes for each of the countries named in the IHS data sets.

## D.6 Sub-national per capita GDP

## D.7 Demand elasticity estimates

Table D2 provides the preferred (absolute) value of the own price elasticities estimated in 18 studies focusing on cars (ICE and EVs).

Table D2: 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 | | |

# E Additional empirical results

## E.1 Explaining price and appeal of EVs

Table E3: Regressions of price and appeal on characteristics

| dep. var: | log(price) | | | log(appeal) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| pack size | 1.24$^a$ | | 0.267 | 1.26$^a$ | | 0.076 |
| | (0.119) | | (0.175) | (0.152) | | (0.328) |
| charging time | -0.155$^a$ | | -0.032 | -0.339$^a$ | | -0.113 |
| | (0.059) | | (0.047) | (0.106) | | (0.080) |
| motor power | | 0.498$^a$ | 0.442$^a$ | | 0.741$^a$ | 0.685$^a$ |
| | | (0.074) | (0.073) | | (0.098) | (0.099) |
| vehicle weight | | 0.809$^a$ | 0.592$^a$ | | 0.523$^b$ | 0.565 |
| | | (0.165) | (0.197) | | (0.229) | (0.417) |
| Observations | 1,102 | 935 | 924 | 1,102 | 935 | 924 |
| Within $R^2$ | 0.728 | 0.830 | 0.827 | 0.392 | 0.518 | 0.516 |

Note: Standard errors (in parentheses) clustered at the car model level. Signif. Codes: a: 0.01, b: 0.05, c: 0.1.

## E.2 Evidence on multi-sourcing and capacity increases

Figures E6 and E7 are examples of multi-sourcing for vehicles found in our data. In figure E6, we focus on sales of Tesla models in Germany. The Model S and Model X are always sourced in Fremont, those are single sourcing cases all along. The Model 3 (in blue) comes from both Fremont and Shanghai in 2020. But this year is also the opening of the Chinese factory. Sales from Shanghai continue to increase thereafter, while the Californian-made Model 3 are not exported to Germany as soon as 2021. The story is very similar with the Model Y (in red). It comes in 2021 from Shanghai, and there is multi-sourcing in 2022, with the Berlin factory opening and gradually ramping up production. The made-in-China Model Y volumes fall in 2023 and go to zero in 2024. Both cases of multi-sourcing appear to be transitions. The story is the same for the Volkswagen ID.4 sold in North America as shown in figure E7(a). The model is launched in 2021 in the USA and Canada, and sourced in the German plant of Mosel. The US-based Chattanooga plant starts producing this car in 2022, with the flows from Germany going to zero in 2023. Mexican imports are negligible, but come from the US factory.

Figure E6: Transitory multi-sourcing by Tesla in Germany

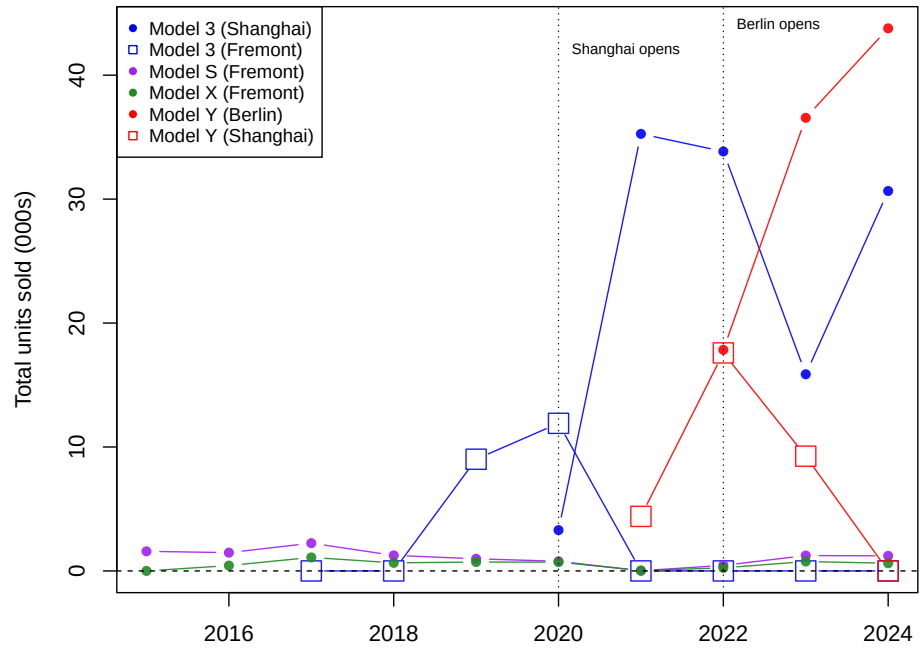


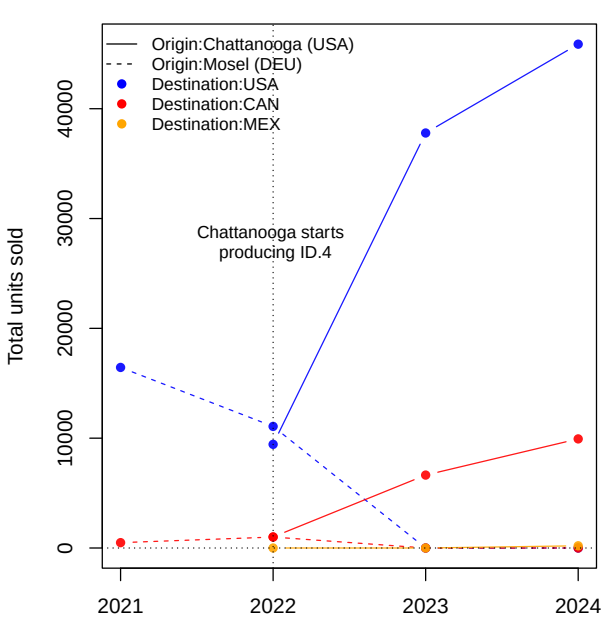

Figure E7(b) shows a different case where the same VW ID.4 is imported by Sweden from two German plants starting in 2022, and continues that way onward. It turns out that those two factories are producing different shapes of the NCM battery which powers this car. Although we do not observe it directly in our sales data, those ID.4s coming from Mosel and Emden are likely to be different trims with different battery specifications. We will treat those as such in our regressions.

Figure E8 shows capacity increases in two factories owned by Tesla: the first one in Fremont, California, and the Shanghai factory opened in 2020. Total output in 000s of cars produced is depicted in black, and specific models are in other colors. In both cases, we see that the production goes from relatively small amounts (around 50 thousands in 2015 for Fremont, and less than 200 thousands for Shanghai in 2020) to several multiples of the initial number very fast. The Fremont factory output is multiplied by around 10 in 8 years, and the factor is more than 4 in 3 years for Shanghai. Capacity constraints do not seem a dominant feature in this industry.

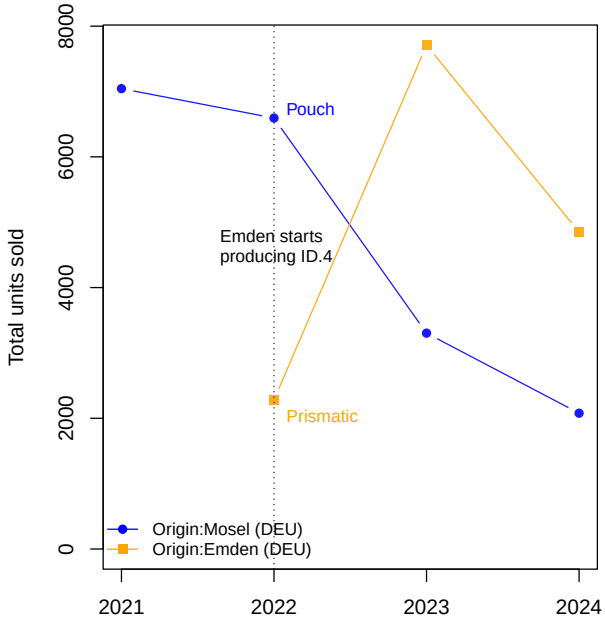
## E.3 Cell and vehicle groupings

Our model assumes that a distinct fixed cost is paid for each group $g_k$ installed in cell and assembly plants. The matrix $\Gamma^1$ with rows $m$ and column $g_1$ is composed of indicators equal to 1 if model $m$ uses group $g_1$ for cells and zero otherwise. The groups for cells

Figure E7: Transitory vs trim-based multi-sourcing

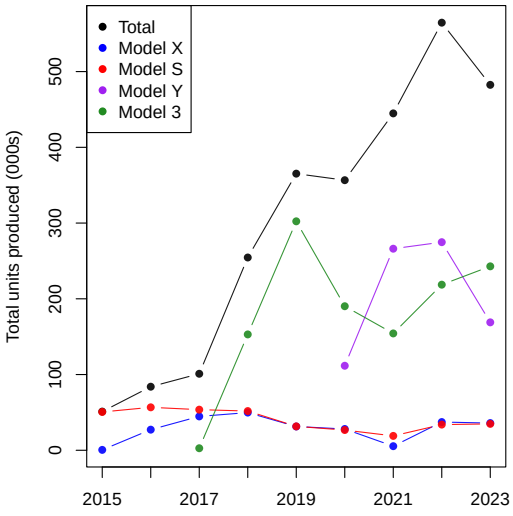



(a) VW ID.4 sold in North America

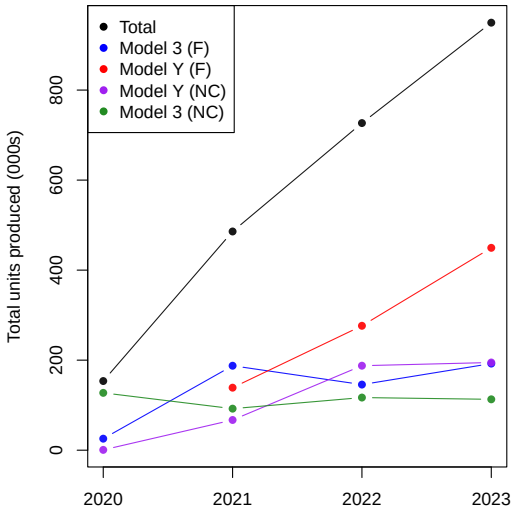



(b) VW ID.4 sold in Sweden

Figure E8: Tesla's production increases in Fremont and Shanghai




(a) Tesla-Fremont




(b) Tesla-Shanghai

distinguish combinations of material category (NCM or LFP) and shape (cylinder, pouch or prismatic) of the cell. For vehicles, the indicators in the $\Gamma^2$ matrix specify the *platform* $g_2$ for each model $m$, e.g. "GEN III" for Tesla Model 3 and Model Y, "MEB" for VWs such as ID.4, ID.3, Audi Q4, and Skoda Enyaq. Platforms establish the architecture of core vehicle systems (e.g. steering, chassis) and dictate the placement of their components. They require a specific organization of the assembly line and its machine tools.

Table E4: Cell variants per model

| # of matcat-shapes per model | count |
|---|---|
| Single group | 102 |
| Two groups | 24 |
| — one dominant by 75% | 13 |
| — none dominant | 11 |

Note: The table counts material-shape groups cells used in a model in 2022 and separates models into Single to Two groups categories (material consists of two categories, NMC or LFP).

Table E4 computes the number of cell groups (material category and shape) that can be found for each of the car models we consider in the counterfactual analysis. Out of the 128 models, there are 24 which are built with two different cell groups. In more than half of the cases (13 out of 24), one group is largely dominant (more than three quarters of the output), such that we attribute the dominant group to this car model. For the remaining 11 car models with no dominant group, we consider both versions as separate car models, which brings the overall number to 139.

Table E5 describes how many variants of the cells are produced in a typical cell factory. It turns out that the vast majority of plants (85%) produce only one shape or one material category (70%) of cells. Even for both characteristics constituting $g_1$, the most frequent case (63%) is that a plant produces only one combination. Even expressed in shares of total output (measured in GWh), a little more than a quarter of world cell production takes places in plants producing more than two groups.

Table E5: Cell plants are specialized in shape and material produced

| | Shape | | | Mat. Cat. | | | Shape & Mat. Cat. | | |
|---|---|---|---|---|---|---|---|---|---|
| # variants: | count | % | % GWh | count | % | % GWh | count | % | % GWh |
| 1 | 73 | 84.9 | 66.0 | 60 | 69.8 | 45.5 | 54 | 62.8 | 37.8 |
| 2 | 12 | 14.0 | 32.0 | 24 | 27.9 | 33.9 | 26 | 30.2 | 35.7 |
| 3 | 1 | 1.2 | 2.1 | 2 | 2.3 | 20.6 | 4 | 4.7 | 4.1 |
| 4 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 | 2 | 2.3 | 22.4 |

Note: The table gives counts and shares of shapes and material category produced in cell plants. Material category consists of two categories (NMC or LFP). Counts are number of cell plants in 2022.

## E.4 Stage-level regression results

Table E6: Cell Sourcing decision

| Dependent Variable: | Cell sourcing decision | | | | | | |
|---|---|---|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| border | -1.73$^a$ | -1.72$^a$ | -1.61$^a$ | -1.52$^a$ | -1.57$^a$ | -0.953$^a$ | -0.494 |
| | (0.331) | (0.331) | (0.331) | (0.362) | (0.345) | (0.319) | (0.354) |
| log distance | -0.258$^a$ | -0.262$^a$ | -0.236$^a$ | -0.233$^a$ | -0.235$^a$ | -0.382$^a$ | -0.358$^a$ |
| | (0.035) | (0.035) | (0.029) | (0.033) | (0.029) | (0.021) | (0.039) |
| RTA | 1.14$^a$ | 1.14$^a$ | 0.689$^b$ | 0.653$^b$ | 0.672$^b$ | 0.458 | -0.076 |
| | (0.270) | (0.271) | (0.288) | (0.303) | (0.290) | (0.320) | (0.365) |
| log GDP per capita | | 0.147$^a$ | 0.167$^a$ | 0.322$^a$ | 0.167$^a$ | 0.213$^c$ | 0.700$^a$ |
| | | (0.035) | (0.059) | (0.059) | (0.059) | (0.118) | (0.197) |
| log(1+tariff) | -4.38$^c$ | -4.40$^b$ | -7.43$^a$ | -6.87$^b$ | -7.20$^b$ | -8.49$^a$ | -15.0$^a$ |
| | (2.24) | (2.24) | (2.71) | (3.45) | (2.80) | (2.34) | (4.96) |
| Observations | 25,526 | 25,526 | 14,518 | 10,929 | 14,485 | 7,945 | 4,624 |
| Squared Correlation | 0.237 | 0.236 | 0.376 | 0.344 | 0.375 | 0.322 | 0.302 |
| Sample | all | all | all | all | all | Majors ($f$) | Majors ($m$) |

All specifications include origin country fixed effects. Columns (1) and (2) activate a cell plant if it supplies that vehicle maker in the current year. Column (3)–(7) require the cell plant to supply the vehicle maker with the material (NCM, LFP, Others) and shape (Cylinder, Prismatic, Pouch) required for the considered car model. Column (3)–(7) estimate a common fixed effect for France and Germany (since France is not in the largest connected set). Column (4) constrains the set of choosers to be car models with positive sales (strictly defined). Column (5) constrains suppliers to be on the "White List" (see text). Column (6) further constrains the sample to the set of 24 major countries and 15 major firms. Column (7) reduces the sample to the 138 major models of those 15 firms. Clustered (dyad) standard-errors in parentheses, Signif. Codes: a: 0.01, b: 0.05, c: 0.1.

Table 6 in the main text provides results for our benchmark specifications of both cell and vehicle sourcing regressions. In this appendix, we provide supplementary regressions using different specifications and samples. The first two columns of the cell sourcing Table E6 use as choice set for each chooser all the plants that supply the vehicle maker of this car model in the current year. Crossing national borders and longer distances to the assembly plant where the cells are used both reduce strongly the probability of being chosen. An active RTA raises it while tariffs have the expected negative sign. Columns (3) to (7) restricts the choice set to the plants that produce the required grouping of cells $g_1$ for this car model. The most important change is that the tariff elasticity rises while the RTA effects is reduced. Column (6) reduces the sample to the set of major firms, and column (7) to the set of major models. Unfortunately, the number of observations drops quite drastically with this last specification, making it hard to statistically assess the distinct

effects of tariffs, national borders and RTAs. Column (6) is our baseline.

Table E7: BEV Sourcing decision

| Dependent Variable: | BEV sourcing decision | | | | | |
|---|---|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| Border | -1.25$^a$ | -0.896$^a$ | -0.894$^a$ | -0.648$^a$ | -1.04$^a$ | -1.07$^a$ |
| | (0.355) | (0.266) | (0.280) | (0.172) | (0.254) | (0.261) |
| log distance | -0.123$^a$ | -0.091$^b$ | -0.092$^b$ | -0.115$^a$ | -0.112$^c$ | -0.109 |
| | (0.041) | (0.042) | (0.044) | (0.041) | (0.062) | (0.068) |
| RTA | 0.697$^a$ | 0.382$^a$ | 0.384$^a$ | 0.520$^a$ | 0.869$^a$ | 0.937$^a$ |
| | (0.147) | (0.117) | (0.120) | (0.142) | (0.214) | (0.222) |
| Inc. cost of cells | | -0.140$^a$ | -0.151$^a$ | -0.046 | -0.234$^a$ | -0.340$^a$ |
| | | (0.042) | (0.045) | (0.044) | (0.084) | (0.095) |
| log(1+tariff) | | -4.75$^a$ | -4.85$^a$ | -10.8$^a$ | -8.56$^a$ | -7.99$^a$ |
| | | (1.57) | (1.60) | (1.52) | (1.74) | (1.65) |
| log GDP per capita | | | 0.589$^a$ | 0.247$^a$ | 0.206$^b$ | 0.282$^b$ |
| | | | (0.102) | (0.072) | (0.087) | (0.131) |
| Observations | 92,332 | 92,332 | 92,332 | 63,607 | 15,793 | 14,120 |
| Squared Correlation | 0.114 | 0.117 | 0.121 | 0.117 | 0.265 | 0.273 |
| Sample | all | all | all | Majors ($f$) | Majors ($f$) | Majors ($m$) |

Clustered (dyad) standard-errors in parentheses, Signif. Codes: a: 0.01, b: 0.05, c: 0.1. All specifications include origin country fixed effects. "Majors" sample imposes the filters detailed in section 7.2. Columns (5) and (6) restrict choice sets based on vehicle platforms.

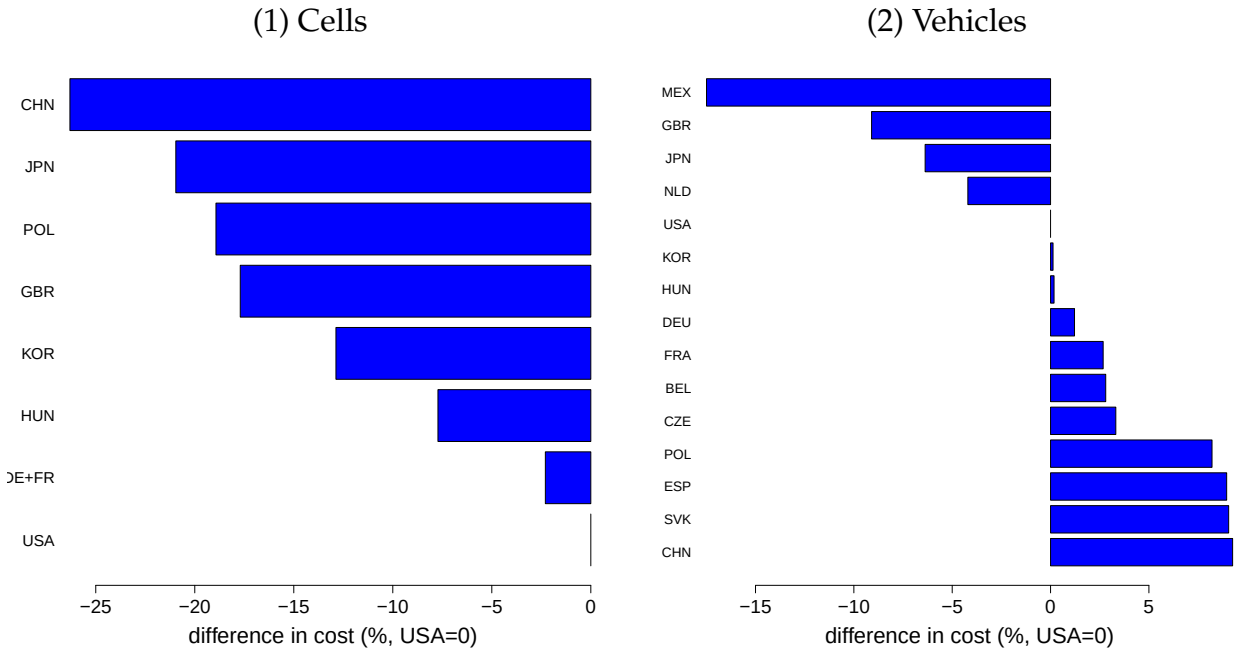
Results of vehicle sourcing are in Table E7. Column (1) reports results for the three geographical frictions with expected signs. The tariffs and inclusive costs of cells are introduced in column (2). Both substantially reduce the probability for an assembly plant to be chosen as our model predicts. The sign of income per capita (conditional on country FEs) reported in column (3) is positive as for cell sourcing. The restriction to major countries and firms reinforces strongly the effect of tariffs. Column (5), restricting choice sets to plants using the right platform, is parallel to column (6) of the cell sourcing regression table, and is therefore our baseline regression.

## E.5 Country-level fixed effects in the BEV sourcing regression

Figure E9 uses the estimates of country fixed effects from our baseline regression (column 2 of Table 6 in the main text) 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 $k$.[35] The competitive-

[35]This is a slight abuse of notation, since in practice we estimate a fixed effect for the country where plant $\ell_k$ is located. Some of those country fixed effects cannot be meaningfully identified in our sourcing

Figure E9: Relative costs in the value chain of BEVs

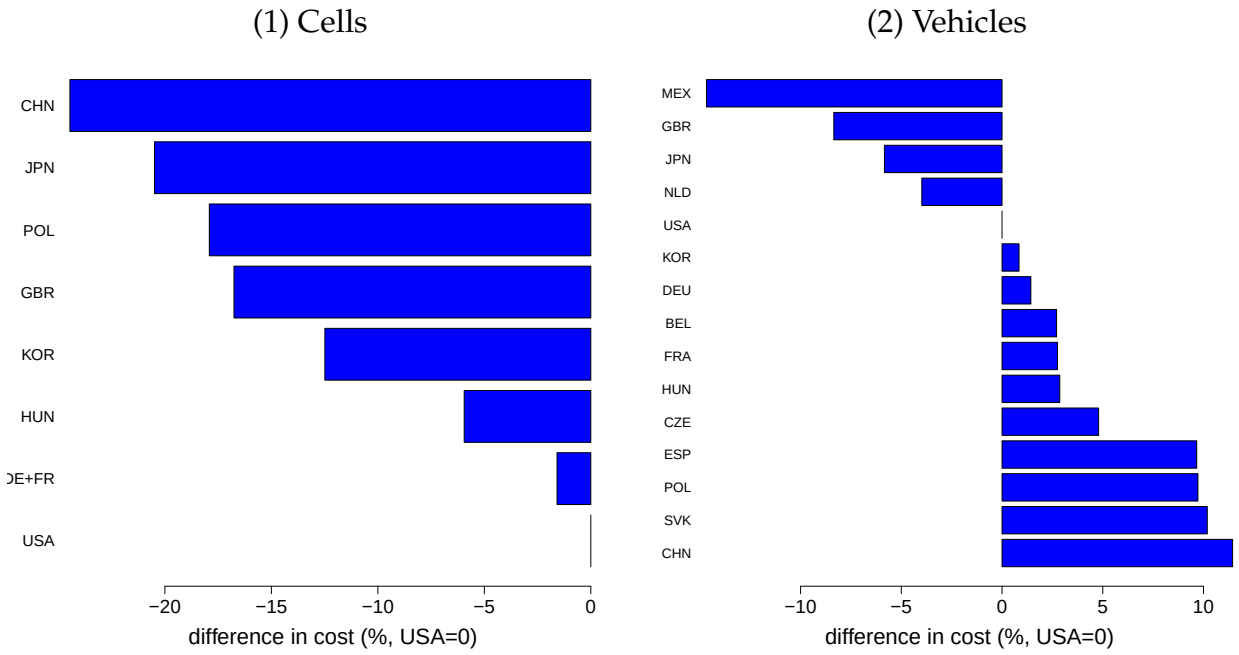



ness of China is visible for cells. The United States ranks at the bottom of the list. Polish and British-made cells are revealed to be slightly less competitive than the Japanese ones. Mexico, Great-Britain, and Japan are identified as very good places for EV assembly as opposed to China, which is revealed as the least cost competitive country for EV assembly. The low costs attributed to Mexico reflect the fact that the model (the Ford Mustang Mach-E) assembled there sell in many places in the world (not only in North America). In contrast, the vast majority of Chinese-assembled EVs are only sold locally. Regarding China, we can take the example of the Volkswagen electric SUV ID.4. In 2022, it 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 reflect those patterns: while the situation is evolving rapidly, in our sample most of the Chinese plants still serve only China: in 2023, 47% of Chinese plants served only China; but this number was as high as 83% in 2020, and even 96% in 2017. The average number of destinations in 2023 was 6.6 for Chinese plants, 55.5 for Belgian ones. Other consumers in the world prefer to source their BEVs from alternative plants, which translates into high estimated costs.

regressions: a fixed effect can be unavailable because 1) that country is never chosen (Canada in our sample), 2) none of our chosen models choose it (Italy after excluding commercial vans), or 3) the FE is in a disconnected set (France for cells produced only for the Bolloré Bluecar at the beginning of our sample). In this case we create a FE by pooling the country in the disconnected set with one that is in the connected set (Germany in this case).

Figure E10 reproduces figure E9 but accounts for the fact that our estimation includes productivity effects through income per capita.

Figure E10: Relative costs in the value chain of BEVs - including productivity effects



# F Complete counterfactual results

The main text of section 9 presents the results graphically for a subset of the policies we investigated. Here we present the complete set of results in tabular form.

Table F8: Production lines† by policy

| Receiving | Policy | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Production | 0 | 1 | 2 | 3 | 4 | 3+4 | 5 | 5V | 1+5V | 2+5V | 3+5V | 3+5 |
| | | | | | | **Vehicles** | | | | | | |
| Americas | 9.4 | 11.7 | 12.1 | 11.2 | 9.8 | 11.8 | 9.4 | 9.4 | 11.7 | 12.1 | 11.2 | 11.2 |
| Asia | 38.5 | 38.7 | 38.0 | 38.0 | 38.5 | 37.9 | 38.5 | 38.5 | 38.7 | 38.0 | 38.0 | 38.0 |
| Europe | 25.4 | 25.3 | 25.1 | 24.9 | 25.3 | 24.9 | 25.4 | 25.4 | 25.3 | 25.1 | 24.9 | 24.9 |
| | | | | | | **Cells** | | | | | | |
| Americas | 5.1 | 6.6 | 6.8 | 7.5 | 5.8 | 8.5 | 5.1 | 5.1 | 6.6 | 6.8 | 7.5 | 7.5 |
| Asia | 23.6 | 24.2 | 23.3 | 22.4 | 23.4 | 21.9 | 23.6 | 23.6 | 24.2 | 23.3 | 22.4 | 22.4 |
| Europe | 8.1 | 8.1 | 8.2 | 7.6 | 7.9 | 7.5 | 8.1 | 8.1 | 8.1 | 8.2 | 7.6 | 7.6 |

† Production lines are defined as model-plant combinations.
* Policies defined in section 9. Policy 0 is the baseline with all existing 2022 policies absorbed in the estimated path costs. 3+4 is the sum of the expenditures under policies 3 and 4. 3+5 is the sum of the expenditures under policies 3 and 5. 5V is the expenditure under policy 5 for vehicles only.

Table F9: Count of models imported by origin-destination continent

| Origin–Destination | Policy 0 | 1 | 2 | 3 | 4 | 3+4 | 5 | 5V | 1+5V | 2+5V | 3+5V | 3+5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | **Vehicles** | | | | | | |
| Americas-Americas | 12.1 | 15.3 | 16.1 | 14.7 | 12.6 | 15.6 | 12.1 | 12.1 | 15.3 | 16.1 | 14.7 | 14.7 |
| Asia-Americas | 23.9 | 23.4 | 21.4 | 22.1 | 23.6 | 21.3 | 23.9 | 23.9 | 23.3 | 21.3 | 22.0 | 22 |
| Europe-Americas | 17.4 | 15.8 | 15.5 | 16.2 | 17.2 | 15.9 | 17.4 | 17.4 | 15.8 | 15.5 | 16.2 | 16.2 |
| Americas-Asia | 9.6 | 11.3 | 12.0 | 11.5 | 9.9 | 12.2 | 9.6 | 9.6 | 11.3 | 12.0 | 11.5 | 11.4 |
| Asia-Asia | 75.8 | 76.4 | 74.9 | 74.8 | 75.8 | 74.4 | 75.8 | 75.8 | 76.4 | 74.9 | 74.8 | 74.8 |
| Europe-Asia | 31.6 | 30.5 | 30.8 | 31.0 | 31.4 | 30.9 | 31.6 | 31.6 | 30.5 | 30.8 | 31.0 | 31.0 |
| Americas-Europe | 8.5 | 9.9 | 10.8 | 10.4 | 8.8 | 11.1 | 8.4 | 8.5 | 9.9 | 10.8 | 10.4 | 10.3 |
| Asia-Europe | 46.8 | 47.8 | 45.9 | 45.8 | 46.7 | 45.4 | 46.8 | 46.8 | 47.8 | 45.9 | 45.8 | 45.8 |
| Europe-Europe | 42.2 | 41.7 | 41.7 | 41.5 | 42.1 | 41.5 | 42.2 | 42.2 | 41.7 | 41.7 | 41.5 | 41.5 |
| | | | | | | **Cells** | | | | | | |
| Americas-Americas | 7.9 | 10.6 | 11.3 | 12.2 | 8.8 | 13.4 | 7.9 | 7.9 | 10.6 | 11.3 | 12.2 | 12.2 |
| Asia-Americas | 9.1 | 10.8 | 11.2 | 8.6 | 8.9 | 8.3 | 8.9 | 9.1 | 10.8 | 11.2 | 8.6 | 8.5 |
| Europe-Americas | 4.0 | 4.9 | 5.3 | 3.7 | 3.7 | 3.4 | 4.0 | 4.0 | 4.9 | 5.3 | 3.7 | 3.7 |
| Americas-Asia | 8.5 | 10.8 | 10.1 | 10.3 | 9.8 | 11.6 | 8.5 | 8.5 | 10.8 | 10.1 | 10.3 | 10.3 |
| Asia-Asia | 73.8 | 75.0 | 72.7 | 72.2 | 73.4 | 71.5 | 73.8 | 73.8 | 75.0 | 72.7 | 72.2 | 72.2 |
| Europe-Asia | 16.0 | 16.5 | 15.7 | 15.2 | 15.7 | 14.9 | 16.0 | 16.0 | 16.5 | 15.7 | 15.2 | 15.3 |
| Americas-Europe | 6.5 | 8.9 | 8.7 | 9.1 | 8.6 | 11.6 | 6.3 | 6.5 | 8.9 | 8.7 | 9.1 | 9.1 |
| Asia-Europe | 37.9 | 37.8 | 37.5 | 36.9 | 37.5 | 36.4 | 37.9 | 37.9 | 37.8 | 37.5 | 36.9 | 36.9 |
| Europe-Europe | 29.5 | 29.2 | 29.1 | 28.8 | 29.3 | 28.6 | 29.5 | 29.5 | 29.2 | 29.1 | 28.8 | 28.8 |

* Policies defined in section 9, Policy 0 is the baseline.

Table F10: Expenditures on EVs by policy*

| Receiving continent | Policy 1 | 2 | 3 | 4 | 3+4 | 5 | 5V | 1+5V | 2+5V | 3+5V | 3+5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Americas | 155.6 | 145.5 | 115.5 | 18.8 | 158.1 | -0.5 | -0.0 | 155.6 | 145.5 | 115.5 | 115.4 |
| Asia | 2.4 | -1.6 | -1.3 | 1.5 | -0.8 | -0.0 | 0 | 2.4 | -1.6 | -1.3 | -1.5 |
| Europe | 3.6 | -0.0 | -1.7 | 1.0 | -0.6 | -0.1 | 0 | 3.6 | -0.0 | -1.7 | -1.8 |

* Policies defined in section 9, with combinations explained in the notes to table F8.

Table F11: Contributions to cost changes (in %) in Europe and Asia

| Policy | Elig. share | | Cost index redn. | | | Tot. Exp. |
|---|---|---|---|---|---|---|
| | # path | revenue | subsidy | costs | total | increase |
| | | | Europe | | | |
| 1: Unconditional | 0 | 0 | 0 | 1.1 | 1.1 | 3.6 |
| 2: Continental V | 0 | 0 | 0 | -0.1 | -0.1 | -0.0 |
| 3: Continental V+C | 0 | 0 | 0 | -0.6 | -0.6 | -1.7 |
| | | | Asia | | | |
| 1: Unconditional | 0 | 0 | 0 | 0.4 | 0.4 | 2.4 |
| 2: Continental V | 0 | 0 | 0 | -1.0 | -1.0 | -1.6 |
| 3: Continental V+C | 0 | 0 | 0 | -0.6 | -0.6 | -1.3 |

See note from previous table.