The Costs of Fleet Variety

Filip Premik[1] & Dan Yu[2]
September 18, 2026

Abstract

This paper studies how heterogeneity in capital affects firm performance. We focus on fleets: collections of capital units performing equivalent tasks, and exploit correlated fleet renewal across Polish municipal bus operators and their common exposure to bus suppliers to isolate exogenous changes in fleet composition. We use this variation to show that more brand-diverse fleets require more buses and incur higher operating costs to provide the same output, while individual-bus utilization rises with own-brand fleet share. Our results provide empirical evidence that equal physical quantities or monetary values of capital need not imply equal productive capacity when capital composition differs.

JEL codes: D24, L23, L62.

Keywords: Heterogeneous capital, capital utilization, productivity, fleet composition, organization of production.

[1]Monash University & Group for Research in Applied Economics filip.premik@monash.edu; 2. University of Alberta, dy6@ualberta.ca.

1 Introduction

Capital constitutes a fundamental input in economic models of production, where it is typically summarized by a single aggregate measure expressed in monetary terms or as a physical quantity. Behind this aggregate representation, however, often lies a collection of separate durable units performing repeated and comparable tasks—a *fleet*[1]. Units in a fleet often rely on shared supporting resources, yet different equipment types may require dedicated spare parts, distinct maintenance procedures, and specialized technical expertise. Providing equally specialized support across all types may therefore be costly, creating scope for firms to concentrate these resources around more prevalent types and leave rarer ones with less immediate or tailored support. As a result, a more varied fleet may make less effective use of the same amount of capital than a more concentrated one, giving rise to what we refer to as the *costs of fleet variety*. Does greater variety in a firm's capital input reduce productive efficiency?

Answering this question is difficult for three distinct reasons. First, firms often rely on several fleets of capital goods that perform different tasks at different stages of production. Differences in productive performance may therefore reflect specialization across tasks, differences in performance across fleets, and interactions among them, making it difficult to isolate the effects of fleet variety. Second, firms rarely disclose data linking fleet composition to productive performance. Third, firms choose the composition of their capital input in response to their production needs and expected performance, so observed changes in fleet composition generally reflect endogenous choices rather than exogenous variation. We overcome these challenges by studying the municipal bus industry in Poland and find that greater fleet variety does reduce productive efficiency.

Urban bus transportation provides a particularly clean setting for isolating the effects of fleet variety[2]. Buses constitute the main physical capital input in this industry and perform a common, well-defined task: providing scheduled passenger service along designated routes.

[1] Examples of fleets range from collections of vehicles and aircraft to construction machinery and production equipment within manufacturing plants

[2] Municipal bus fleets have long served as an empirical setting in economics. In the seminal paper by Rust (1987), Harold Zurcher—a superintendent of maintenance at the Madison Metropolitan Bus Company—managed a fleet composed of only two bus brands, excluding a small number of 10-seat minibuses serving a distinct purpose from the rest of the fleet.

At the same time, there is substantial heterogeneity across buses, driven by differences in their intended purpose and by manufacturer-specific characteristics. Greater fleet variety along the former dimensions, such as bus size, may improve efficiency by allowing operators to match vehicles more closely to the tasks they perform, even if maintaining a more varied fleet is costly. The latter instead reflects distinct designs and technological solutions embodied in buses intended to perform similar tasks, leaving less scope for such task-matching benefits. We therefore define bus types by brand, providing a cleaner basis for identifying the costs of fleet variety.

We collect novel data on fleet composition and the productive performance of Polish municipal bus operators. We assemble complete records of operators' fleet composition over time, including each bus's technical characteristics, and combine them with repeated odometer readings and detailed operator-level data on output, input use, and operating costs. This allows us to study the costs of variety at two complementary levels, from the utilization of individual capital units to aggregate resource use. At the bus level, we relate changes in annualized mileage to changes in own-brand prevalence—the fraction of the operator's fleet represented by the same brand[3]. At the fleet level, we relate changes in input use and operating costs to changes in the effective number of brands, measured by the inverse Herfindahl index of brand shares[4].

A simple aggregation framework connects the two specifications. We show that if a bus's productive capacity is linear in its own-brand share, aggregation across buses yields the inverse Herfindahl index as the corresponding fleet-level measure of variety[5].

Rich data, however, do not resolve the endogeneity of fleet composition. Although Polish bus operators acquire new buses through public procurement procedures, they choose the timing of their purchases and exercise substantial discretion in designing tenders. This allows them to influence which suppliers participate and ultimately win the contracts (Premik,

[3]Importantly, repeated measurements for the same bus over time allow us to abstract from persistent differences in bus quality, removing a potentially important source of confounding.

[4]By holding total output fixed, these estimates admit a direct productive-efficiency interpretation: they capture how fleet variety affects the resources required to produce a given level of output.

[5]More generally, a power relationship at the level of the individual capital unit yields the Hill-number family of diversity measures at the fleet level, with the inverse Herfindahl index as a special case.

2026). Observed changes in fleet composition may therefore reflect local operating needs and supplier preferences that are also related to productivity.

To identify the costs of fleet variety in our data, we exploit the fact that operators serve separate local markets but procure buses under a common procurement regime, face shared national and European Union (EU) funding programs, and purchase from a common national supplier base.[6] This structure provides two sources of identifying variation. At the fleet level, we use contemporaneous renewal among peer operators facing similar operating environments to instrument changes in fleet composition. At the bus level, we use peer purchases of new buses of the same brand to instrument changes in brand prevalence. Identification requires that these peer purchasing patterns be unrelated to operator-specific productivity shocks—an assumption that is plausible on institutional grounds and supported by a battery of validity tests.

Our estimates show that fleet variety reduces productive efficiency. At the fleet level, greater variety increases both the physical capital and the operating costs required to provide a given level of output. We find no comparable increase in labor requirements. At the bus level, existing buses are utilized more intensively as their own brand becomes more prevalent in the fleet.

The effects are economically meaningful. An annual decrease in fleet variety of the average observed magnitude reduces the required fleet size by about 0.6% and total operating costs by about 2.5%. At the bus level, an annual increase in own-brand prevalence of the average observed magnitude raises the annualized mileage of existing buses by about 2.7%.

Additional bus-level analyses clarify the nature of the costs of variety. We show that the data favor a model in which bus utilization depends on a brand's fleet share and not on the absolute number of same-brand buses. This pattern is consistent with frictions tied directly to fleet composition: for example, when inventories, maintenance routines, workshop organization, or technical expertise are geared toward more prevalent brands. The costs of variety therefore depend on how capital is distributed across types, rather than reflecting fixed type-specific costs that are spread over more units as the scale of each type increases.

[6]Selling brand-new buses entails fixed costs associated with national certification and familiarity with public procurement procedures. Once these costs are incurred, suppliers face virtually no geographic barriers to competing for contracts across the country.

We also examine heterogeneity by bus age. Point estimates are close to zero for buses in their first few years of service and positive for older buses. As maintenance, parts replacement, and accumulated technical expertise become more important over time, this pattern is consistent with the costs of variety arising from how specialized support and resources are organized across types within the fleet.

Our results imply that capital composition can matter for productive capacity independently of capital scale. Neither the physical quantity nor the monetary value of capital therefore provides a sufficient statistic for its productive capacity: two firms with the same amount of aggregate capital may differ in how much output that capital can effectively generate.

This does not imply that complete standardization is optimal. Maintaining a more varied capital stock can also generate benefits by allowing firms to match equipment more closely to different tasks, reduce dependence on a single supplier, insure against recalls, and adopt new technologies as they become available. Our estimates isolate the productive cost on the other side of this trade-off: greater heterogeneity can make a given amount of capital less productive.

Our paper contributes to a few areas of existing literature.

Capital heterogeneity and productivity. A longstanding literature recognizes that different capital assets may contribute differently to production (Jorgenson and Griliches, 1967). More recent empirical work studies how the mix of functionally distinct, potentially complementary capital inputs affects productive performance (Wilson, 2009; Coyle, Bournakis, and Mei, 2025). We study a different margin: heterogeneity within a single capital input, among units performing the same task. We show that fleet composition affects productivity and trace this effect to the utilization of individual capital units. Relatedly, Kebede and McMillan (2026) show that installed capital may overstate effective capital when external constraints limit its utilization. We show empirically that this wedge can instead arise endogenously from the composition of the capital stock itself.

Accounting for input heterogeneity in empirical studies. A separate literature treats input heterogeneity as a challenge for productivity measurement: differences hidden within

observed input quantities may otherwise be attributed to productivity. This concern has traditionally centered on input quality (e.g., Griliches, 1957; Griliches and Mairesse, 1995; Bartelsman and Doms, 2000), but a lack of suitable data limited empirical work. Fox and Smeets (2011) provide one of the first detailed empirical attempts to measure input quality directly, using rich worker-level information to construct quality-adjusted labor input. de Roux, Eslava, Franco, and Verhoogen (2021) show that observed input quantities can mismeasure effective inputs when firms differ in input quality and in the range of differentiated labor and material inputs they use, and use variation in input prices to account for this heterogeneity. Whelan (2002) is a notable exception in providing a direct empirical treatment of heterogeneity in capital inputs. He shows that observed computer stocks can mismeasure effective capital when different vintages embody different levels of technology, and constructs a vintage-based measure to account for this heterogeneity. In turn, the source of input heterogeneity we study gives rise to a friction that is not captured by any of these adjustments. Even after accounting for differences in quality, price, and vintage—as we do in our empirical analysis—the composition of otherwise comparable capital units can still affect productive efficiency.

When relevant differences within capital are not captured by the observed aggregate, measured capital may fail to reflect the firm's productive capacity. One response is to leave the source of this mismeasurement unspecified and adjust the estimation procedure instead. Kim, Petrin, and Song (2016) and Collard-Wexler and De Loecker (2021) develop production-function estimators that are robust to measurement error in observed capital. We take a different route. We identify one economically meaningful source of this gap: heterogeneity in the composition of capital. We construct measures that capture this heterogeneity directly and estimate how it affects firms' productive capacity and the utilization of individual capital units.

Fleet commonality. Transportation research relates fleet commonality, typically measured using the standardization index of de Borges Pan and Espírito Santo (2004) or a Herfindahl

index over manufacturers, to airline performance.[7] Greater commonality is associated with higher profitability and lower unit costs (Brüggen and Klose, 2010; Zou, Yu, and Dresner, 2015). Merkert (2023) further shows that commonality in engines matters beyond commonality in airframes. These correlations, however, cannot separate the effect of fleet composition from the broader business strategies with which it is jointly chosen, while firm-level data do not reveal how commonality affects performance. We address both limitations. We also derive our measure of fleet variety from a simple framework linking fleet composition to vehicle utilization. We instrument fleet composition using variation outside the bus operator's control and observe the utilization of individual vehicles directly. This allows us to show not only that fleet variety raises firm-level costs, but also that it does so by reducing the utilization of buses belonging to less prevalent brands.

The remainder of the paper is organized as follows. Section 2 develops the framework and derives measures of fleet variety used in the empirical analysis. Section 3 describes the institutional setting and data, and presents descriptive evidence on fleet composition and performance. Sections 4 and 5 present the fleet- and bus-level empirical analyses, respectively. Section 6 further investigates the nature of the bus-level utilization effect. Section 7 concludes.

2 Productive Capacity of Heterogeneous Capital: A Simple Framework

We develop a simple framework linking heterogeneity in capital inputs to productive capacity. Although we state the framework in terms of bus fleets, reflecting our empirical application, the same logic applies more generally to any setting in which capital consists of heterogeneous units.

We assume that the productive capacity of an individual unit of capital depends on the prevalence of its own *kind* in the fleet. We then aggregate this relationship across buses

[7] Airlines provide a familiar example. Southwest Airlines has long relied predominantly on the Boeing 737 family, a strategy commonly associated with savings from fleet commonality in training, maintenance, and spare parts.

to characterize how fleet composition affects overall productive capacity. This formulation allows us to derive a natural measure of fleet variety that is closely related to standard measures of concentration and variety, which in turn inform our specification choices in the empirical analysis.

Class prevalence and productive capacity. Consider a bus operator with a fleet of N buses partitioned into classes $c = 1, \ldots, C$, where a class captures a relevant dimension of technological differentiation and, in our baseline analysis, corresponds to bus brand. For simplicity, we abstract from ex-ante differences in bus quality, so any productivity differences in our framework arise solely from fleet composition. Let n_c denote the number of buses in class c and $s_c = {}^{n_c}/_{N}$ its fleet share, or class prevalence.

Suppose that a bus in class c provides effective annual service capacity $v(s_c)$, with $v'(s_c) > 0$. Thus, a bus can be utilized more intensively when its class is more prevalent in the fleet. This may be the case, for example, because spare parts are more readily available, maintenance routines are performed more frequently, or experience with the relevant technology has accumulated. Effective fleet capacity is then:

$$K(N, \mathbf{s}) = \sum_c n_c v(s_c) = N \sum_c s_c v(s_c)$$

where $\mathbf{s} = (s_1, \ldots, s_C)$. Physical fleet size is therefore not a sufficient statistic for productive capacity: two equally large fleets can provide different amounts of effective capital because they differ in composition.

Abstracting from other production inputs, let $\bar{K}(y)$ denote the amount of effective capital required to provide output y. The physical fleet size required to provide this output while fleet structure is $\mathbf{s}$ satisfies:

$$N(y, \mathbf{s}) = \frac{\bar{K}(y)}{\sum_c s_c v(s_c)}$$

The effect of fleet composition on required fleet size is summarized by the scalar $\sum_c s_c v(s_c)$, and its particular form depends on how own-class prevalence affects bus utilization.

Hill-number representation. Consider a parsimonious power specification for the individual effective service capacity:

$$v(s_c) = \gamma s_c^{q-1}, \qquad \gamma > 0, \quad q > 1$$

Substituting this expression into the fleet-size requirement for output y yields:

$$N(y, \mathbf{s}) = \frac{\bar{K}(y)}{\gamma \sum_c s_c^q} \equiv \frac{\bar{K}(y)}{\gamma} (x_q)^{q-1}$$

where $x_q = (\sum_c s_c^q)^{\frac{1}{1-q}}$ is a Hill number[8] of order q. It has a natural interpretation as an effective number of classes: x_q is the number of equally represented classes that would generate the same value of the underlying concentration measure as the observed fleet composition.

This specification delivers the central testable prediction of the framework: with $q > 1$, greater fleet variety increases the physical fleet required to provide a given level of output. Whenever maintaining additional capital is costly, this also raises operating costs[9]. Moreover, if fleet variety also directly increases per-bus operating costs through maintenance, parts, or other support frictions, its effect on total operating costs is even larger.

Choice of q. The parameter q governs how the utilization gain from a marginal increase in own-class prevalence varies with the initial prevalence of the class. For $1 < q < 2$, $v(\cdot)$ is concave, so the marginal gain is larger for relatively rare classes; for $q = 2$, $v(\cdot)$ is linear and the marginal gain is constant; and for $q > 2$, $v(\cdot)$ is convex and the marginal gain is larger for already prevalent classes. The same curvature carries over to the fleet level: q governs how the physical fleet required to provide a given level of output varies with the effective number of classes x_q.

[8] Hill numbers constitute a family of diversity measures introduced by Hill (1973) and widely used in ecology to quantify biodiversity (Chao, Chiu, and Jost, 2014).

[9] Section C of the Online Appendix extends the framework to operating costs under a simple cost structure. It also derives a first-order approximation linking the relationships implied by the framework to the logarithmic specifications used in our empirical analysis.

Distinguishing among values of q would require additional structural restrictions linking the curvature of $v(s_c)$ to observed bus utilization[10], and so q is not identified by our reduced-form empirical design[11]. Instead, for any admissible value of q, the framework defines a corresponding measure of fleet variety and relationship between class prevalence and productive capacity. Our empirical strategy estimates the effects under that specification. We set $q = 2$ in the baseline specification[12], which gives the model a simple linear-approximation interpretation while remaining ex ante agnostic about whether standardization gains are concentrated among relatively rare or prevalent classes. In addition, the corresponding Hill number becomes the inverse of the familiar HHI concentration index.

Our framework offers a simple account of how the costs of variety arise and propagate through the fleet. We now turn to the empirical setting to test these predictions.

3 Empirical Setting: Urban Bus Fleets in Poland

We examine the relationship between capital heterogeneity and productive performance in the setting of public urban bus operators in Poland. Urban bus service provides a particularly clean setting because buses constitute the main capital input, allowing us to abstract from interactions with other forms of capital. The Polish setting provides plausibly exogenous variation in fleet composition that we exploit for identification.

This section introduces our empirical setting. We first describe the data sources and the organization of urban bus transportation in Poland. We then document the dimensions along which buses differ and explain how we use these differences to measure capital heterogeneity within operators' fleets. Finally, we show how fleet composition is related to performance at

[10] As a primitive governing bus-level utilization, q depends on how individual units of capital are managed in terms of maintenance, servicing, and task assignment. Modeling these low-level managerial decisions would be interesting in its own right, but is beyond the scope of this paper.

[11] For any assumed value of q, our instrument provides exogenous variation in the corresponding change in class prevalence. Identifying q itself, however, would require additional excluded variation that identifies how the effect of prevalence changes with its level. Because the level of brand prevalence is itself endogenous, this variation cannot be obtained simply by interacting our instrument with observed prevalence.

[12] We assess the quantitative consequences of this choice. In the Online Appendix (Sections F and G), we consider alternative values of q and examine how they affect both the strength of our identifying variation and the estimated effects. Although the magnitudes of our estimates vary with q, the results consistently imply that greater fleet variety reduces productive efficiency.

both the fleet and individual-bus levels, providing descriptive evidence for the mechanisms highlighted by the framework and formally quantified in the next section.

3.1 Data Sources

Our analysis combines information on the evolution of operators' fleet composition with performance measures at both the operator and individual-bus levels. We obtain them from three main sources.

Fleet data. The fleet data consist of a list of all vehicles owned by every public bus operator in Poland between 2007 and 2023. For each bus-operator pair, we observe the bus brand, dates of purchase and departure from the fleet, previous owners, and a range of technical characteristics. The information comes from http://phototrans.eu/—an online photo gallery created by enthusiasts and industry employees that has evolved into a comprehensive database of vehicles owned by operators worldwide[13]. We use it to construct our measures of fleet composition over time.

Odometer data. Every registered vehicle in Poland is subject to periodic mandatory inspections, typically conducted at least once a year. Since 2014, odometer readings have been recorded at each inspection. These data are stored in the official vehicle registry, along with detailed technical information on each bus and its engine, and can be accessed publicly provided that three identifiers are available: the vehicle identification number (VIN), the current registration number, and the date of first registration. All three identifiers are contained in our fleet data. The odometer readings provide our main bus-level outcome, annualized bus mileage.

Fleet Performance Indicators: IGKM Data. Fleet performance indicators are provided by the Chamber of Commerce of City Transportation (IGKM) in Poland, an association of firms operating urban transportation services. IGKM collects data through an annual questionnaire distributed among its members to facilitate the exchange of data, good practices,

[13]Despite not being an official source, phototrans.eu provides an accurate description of Polish municipal operators' fleets. Premik (2024) cross-validates the fleet data with procurement data on every auction for brand-new city buses. All buses implied by the auction data are found in the fleet data.

and experience[14]. We use survey waves covering 2007–2023. The data provide our primary source of information on operator-level performance, including total output (measured as the aggregate mileage of all buses in an operator's fleet within a year), fleet size, employment, and operating costs.

3.2 Urban Transportation in Poland

Urban transportation in Poland is a relatively large industry and relies heavily on buses. According to Statistics Poland (Statistics Poland, 2024), urban public ground transportation served more than 3 billion passengers in 2023. Of nearly 60 thousand kilometers of service routes, fewer than 3 thousand were operated by vehicles other than buses.

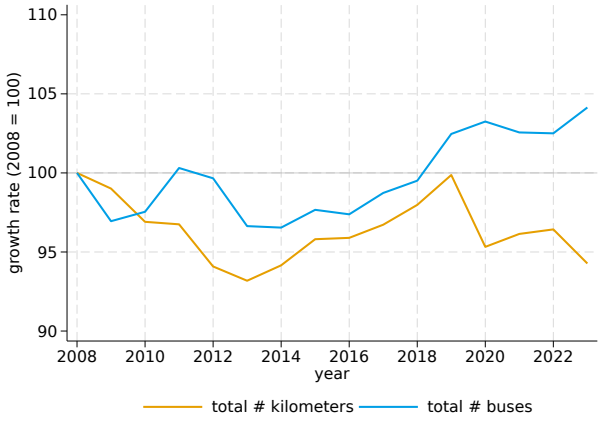
Every larger municipality organizes its own public transportation system. The city designs routes and collects fares, while services are provided either by public bus operators owned by local authorities or by private operators hired under fixed-term contracts. Public operators play the dominant role, operating nearly 90% of all buses in the market (Statistics Poland, 2024).

We focus on public bus operators. These operators serve their own cities exclusively and do not compete with one another. They provide transportation services by operating the routes in quantities requested by the city, which compensates them at an agreed rate. As a result, demand for bus services—and, to a large extent, its price—is fixed from the operator's perspective. The overseeing authorities hold operators accountable for providing the requested services within the allocated budget. Given this ownership and contracting structure, the incentives faced by public bus operators are broadly consistent with cost minimization.

Output in this industry is measured by the total number of kilometers driven by buses while serving scheduled routes and largely reflects the city's demand for transportation services. As depicted in Figure 1, this output remained roughly constant between 2010 and

[14]Participation in the survey is voluntary. Although IGKM members provide similar services, they operate in different cities and do not compete directly with one another, limiting strategic incentives to misreport information.

Figure 1: Total output and fleet size over time.



Notes. The figure shows the evolution of total output and fleet size. *Sources.* Fleet size is from Statistics Poland; total output is from the IGKM survey. *Sample.* The output series covers operators reporting in every wave from 2008–2023; the two series therefore cover overlapping but non-identical populations. *Presentation.* Both series are indexed to 2008=100.

2023[15]. The two main production inputs are buses and labor, primarily drivers and maintenance personnel. Public bus operators typically own, operate, and maintain their buses, while leasing and other forms of hire are limited. The total number of buses owned by public operators increased by approximately 5% between 2008 and 2023 (partially reflecting an increase in the number of operators; Figure 1), or slightly more than 600 buses. Over the same period, 12137 new buses arrived at operators' depots. Therefore, most new-bus purchases replace existing vehicles rather than expand fleets.

3.3 Fleet Evolution

Fleet composition changes gradually as operators introduce new batches of buses frequently—on average every 1.8 years—with each introduction accounting for only a relatively small share of the fleet (Table 1). As a result, an operator's fleet composition reflects the cumulative outcome of many acquisition decisions made over the life of its buses.

Buses are long-lived capital units: the average vehicle leaves an operator's fleet at around 19 years of age and is typically scrapped at roughly the same age. The resale market therefore plays little role in adjusting fleet composition, as buses generally remain with

[15]The city's demand for operating a scheduled system of routes should not be confused with passenger demand, which dropped by more than a third in 2020 and had not fully recovered by 2023.

Table 1: Public bus operators and their fleets: summary statistics.

	total/mean	median	standard deviation
# operators	123	—	—
# buses	21634	—	—
fleet size (# buses)	76	35	144
bus age (years)	11	11	2.4
age at scrapping (years)	19	19	2.3
age at leaving the fleet (years)	19	19	2.7
time between deliveries of buses (years)	1.8	1.5	.97
size of delivery (% fleet)	14	13	8.3
annualized mileage per bus (km; 2014-2023)	50658	50999	13245

Notes. The table summarizes public bus operators and their fleets, 2007–2023. *Statistics.* The first two rows report totals; the remaining rows report means, medians, and standard deviations across operators. *Definitions.* Age at leaving the fleet includes departures for any reason, whereas age at scrapping counts only vehicles withdrawn from service. A delivery is a batch of new buses arriving at an operator in a given year. Annualized mileage is available from 2014 onward.

the same operator until the end of their service life. Short-run adjustments in output are consequently achieved primarily by reallocating utilization across existing vehicles rather than by reshaping the fleet. At the same time, heterogeneity created by successive acquisition decisions is difficult to reverse, allowing incompatibilities in maintenance, spare parts, and operating routines associated with fleet variety to persist for many years.

Public bus operators acquire vehicles through two main channels: purchases of brand-new buses and purchases of second-hand buses. The two channels differ in how purchases are organized and timed. We exploit this distinction in our empirical strategy.

Brand-new purchases. Brand-new buses constitute the main acquisition channel and are purchased through public procurement procedures governed by EU directives (2004/18/EC and 2014/24/EU) and national procurement law. This creates a common institutional environment across operators, covering how tenders are conducted and awarded, the rules governing supplier participation, and a common appeals system through which procurement decisions can be challenged. Together with the absence of important geographic barriers to supplying buses across Poland, this helps sustain a national market in which the same manufacturers compete for contracts across operators. Procurement activity is also shaped by common funding cycles, driven primarily by access to national and EU funds, which generate waves of new-bus purchases across otherwise separate local markets. Common institutions and funding conditions therefore induce correlated variation in the timing and intensity of fleet renewal.

Within this common framework, however, contracting authorities retain considerable discretion over the requirements that determine which buses may be offered in a tender and the scoring rules used to rank submitted bids. Using data from the same market, Premik (2026) shows that these choices substantially affect which suppliers are eligible to bid and which ultimately submit bids. As a result, the manufacturer composition of an operator's purchases may reflect its own procurement choices and underlying needs. Our instrumental-variable strategy exploits the resulting common variation in the timing and intensity of bus acquisitions across operators to isolate variation in fleet composition that is not driven by an operator's own procurement decisions.

Second-hand purchases. The second-hand market is much less institutionalized. During the sample period, such buses were typically imported from Western Europe and acquired in response to operators' immediate fleet needs and the availability of used vehicles on the market. The timing of second-hand purchases is therefore more idiosyncratic across operators and less closely tied to the common procurement and funding cycles governing purchases of brand-new buses. Second-hand transactions are also often carried out through third-party intermediaries, without direct participation by the original manufacturers.

The second-hand market nevertheless provides useful information about manufacturers' broader market activity. For each brand, the market-wide balance between brand-new and second-hand acquisitions provides a measure of manufacturer activity. Reflecting brand-level market dynamics, this balance is unlikely to be determined by any individual operator's procurement choices. It therefore plays a role in our bus-level analysis, where the balance between new and second-hand arrivals determines the direction in which peer purchases predict changes in a brand's fleet share.

3.4 Bus Differentiation

City buses are highly differentiated products. To operationalize the notion of a class introduced in Section 2, we define a bus class as a set of buses sharing a common set of characteristics. Fleet composition is then described by the distribution of buses across these classes, and fleet variety summarizes heterogeneity within that distribution.

Bus classes differ in their economic interpretation depending on the characteristics used to define them. Some dimensions reflect bus purpose, such as bus length or drivetrain, and are closely tied to the tasks buses perform and the routes they serve. Others reflect producer identity, such as bus brand or engine brand, and capture differences in technological solutions developed and chosen by manufacturers.

While both dimensions contribute to overall fleet variety, they differ fundamentally in how they affect productivity. Purpose-related fleet variety reflects heterogeneity in tasks and may be efficiency-enhancing. For example, an operator may increase fleet variety by purchasing a small number of shorter buses to serve low-demand routes and save on fuel, rather than serving all routes with buses of a uniform size. In contrast, producer-related heterogeneity reflects technological incompatibilities across vehicles performing similar tasks, potentially generating frictions in maintenance, repairs, and operations without corresponding gains from better matching buses to their tasks.

Our analysis focuses primarily on classes defined by bus brand, which identifies the manufacturer. In the primary market, buses are typically sold either directly by manufacturers or through their exclusive dealers, so manufacturers' market activity generates common variation in the brands entering operators' fleets. This variation is substantial: multiple manufacturers are active in any given year, with entry and exit over time and specialization in buses serving different purposes. In the secondary market, where manufacturers are typically not directly involved, the mix of used buses available also varies across brands and over time. Together, these market-wide changes provide the brand-specific variation used in our bus-level analysis. Defining classes by brand therefore gives us a measure of variety centered on producer identity rather than on functional differences across buses that may themselves generate efficiency gains.

Although buses differ along many dimensions other than brand, brand captures a substantial share of the underlying producer-related technological differentiation. In Section B of the Online Appendix, we consider alternative characteristics related to producer identity, such as bus generation, engine brand, and engine family, as well as characteristics related to bus purpose, such as size, drive, and engine displacement. Brand explains substantially more variation in the former than in the latter. This distinction is useful for our analysis:

Table 2: Brand variety in the market and within fleets.

	market	fleet
number of brands	25	3.9
effective # brands (1/HHI)	4.9	2.3
share of the largest brand	.38	.6
share of the second largest brand	.18	.24
HHI	.22	.49

Notes. The table compares market-wide brand variety with variety within operators' fleets. *Sample.* All public bus operators in the odometer sample, 2007–2023. *Definitions.* Market statistics pool all buses in the sample; fleet statistics are computed within operator-year and then averaged. The HHI is the sum of squared brand shares, and its reciprocal is the effective number of brands, corresponding to the order-two Hill number used in the baseline analysis.

brand captures common manufacturer-specific technological solutions while remaining much less closely tied to the purposes buses are designed to serve. This pattern supports our interpretation of brand as primarily capturing manufacturer-related technological differentiation rather than differences in the purposes buses are designed to serve.

3.5 Descriptive Evidence on Fleet Variety and Performance

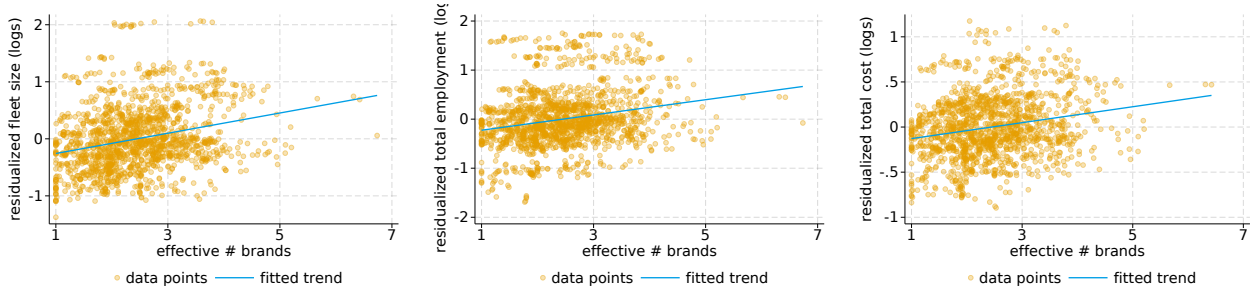
The market features described above create substantial scope for fleet variety. Operators make frequent, relatively small purchases, while buses remain in service for many years and are rarely resold. At the same time, the market includes a relatively large set of suppliers offering differentiated products, allowing heterogeneity to accumulate within fleets over time. Yet Table 2 shows that individual fleets remain relatively concentrated. In a given year, Polish operators collectively use about 25 brands, while the average operator uses buses from only 3.9 brands.

Operators distribute their fleets unevenly across these brands: the largest brand accounts for 60% of a typical fleet. For this reason, the raw number of brands alone provides an incomplete measure of fleet variety. Our framework in Section 2 motivates a measure that accounts for both the number of brands in a fleet and their prevalence. The resulting order-two Hill number gives the number of equally represented brands that would generate the same concentration as the observed fleet. Thus, although a typical operator uses 3.9 brands, its effective number of brands is only 2.3.

The corresponding market-level number is more than twice as large, showing that individual operators maintain substantially more concentrated fleets than the market as a whole.

A potential explanation, suggested by our framework, is that standardization reduces the resources required to provide service. Figure 2 provides descriptive evidence consistent with this interpretation. It plots our baseline measure of fleet variety, the effective number of brands, against fleet size, employment, and total operating cost after residualizing each outcome with respect to output and operator and year fixed effects. The empirical patterns are consistent with the framework: fleets with a larger effective number of brands tend to require more resources to provide the same service.

Figure 2: Fleet variety and operator performance.



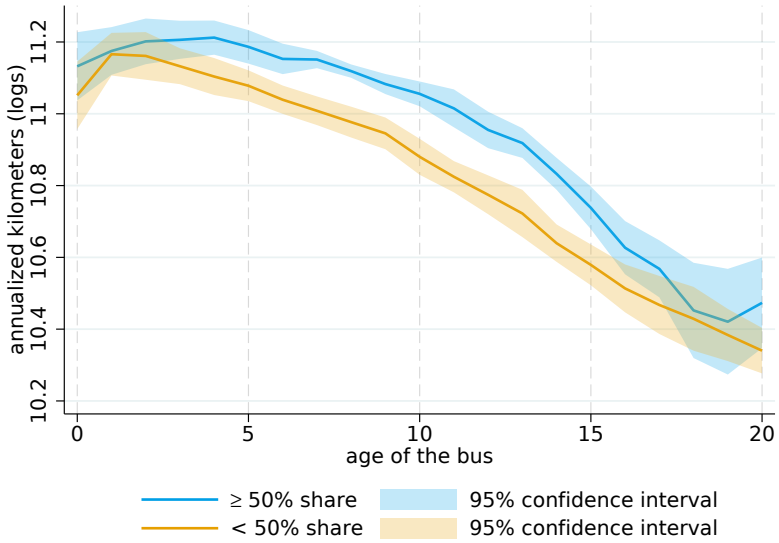
(a) # buses (b) # employees (c) total cost

Notes. Each panel relates fleet variety to an operator-performance outcome. *Specification.* Outcomes are residualized with respect to log total output and operator and year fixed effects; lines show fitted linear relationships. *Sample.* Operator-years in the IGKM sample, 2007–2023, excluding Warsaw.

We next turn to the corresponding relationship at the bus level. Figure 3 compares average utilization over the bus life cycle between vehicles belonging to brands that are more and less prevalent in an operator's fleet. Upon arrival, brand-new buses accumulate similar mileage regardless of how prevalent their brand is within the fleet. A gap emerges after the first 3–5 years, roughly when standard warranties on new buses expire. Buses of more prevalent brands are then used more intensively than those of less prevalent brands, with the gap widening with age and remaining statistically significant nearly until the end of the average bus's service life.

The timing of this divergence is consistent with maintenance-related costs of variety. While warranties may limit differences early in a bus's life, greater availability of spare parts and accumulated experience with more prevalent technologies may become increasingly

Figure 3: Bus utilization over the life cycle by the brand's fleet share.



Notes. The figure shows predicted log annual mileage by bus age separately for buses whose brand accounts for at least half of the fleet and those whose brand does not. *Specification.* Predictions come from a regression of log annual mileage on bus-age indicators interacted with the prevalence-group indicator, absorbing operator×year fixed effects. *Sample.* Bus-years for vehicles up to 20 years old with annual mileage of at least 1,000 km. *Clustering.* Standard errors are clustered by operator. *Presentation.* Shaded areas show 95% confidence intervals.

important as buses age. Buses of rarer brands may therefore spend more time out of service and accumulate less mileage[16].

Roadmap of the empirical exercise. The descriptive evidence is consistent with the framework: greater fleet variety is associated with lower productive efficiency, while buses belonging to more prevalent brands are utilized more intensively. However, these patterns do not have a causal interpretation because operators likely choose the composition of their fleets in response to their own needs and operating conditions. We therefore turn to an empirical strategy that isolates exogenous variation in fleet composition and allows us to estimate its causal effects.

The empirical analysis proceeds in three steps. We first examine whether greater fleet variety raises the resources required to provide a given level of output at the firm level. This is the natural level at which firms' resource use is measured and therefore provides our main

[16]Anecdotal evidence from an interview conducted in the course of this research points in the same direction: one industry representative noted that so-called *unicorn* buses, i.e. rare series within a fleet, often remain in regular service only until their first major breakdown.

result in a form that is directly comparable across settings. We then turn to the performance of individual buses, where variation at the level of individual capital units allows us to test the key assumption of our framework: whether utilization depends on the prevalence of a bus's own class within the fleet. We also use this variation to rule out alternative explanations for this relationship and clarify which dimensions of fleet composition constitute economically relevant variety.

4 Fleet-Level Effects of Fleet Variety

The central question of our study is whether greater capital heterogeneity raises the resources required to provide a given level of output. We begin our analysis at the fleet level.

4.1 Fleet-Level Specification

We estimate the parameter α^F from the following equation:

$$\Delta y_{jt} = \alpha^F \Delta x_{jt} + \beta^F \Delta w_{jt} + \Delta \varepsilon_{jt} \tag{1}$$

where j indexes bus operators and t indexes calendar years. The outcome y_{jt} is a measure of operator performance, x_{jt} measures fleet variety, w_{jt} collects observable determinants of performance, and ε_{jt} summarizes unobserved factors affecting the outcome. We formulate the analysis in first differences to eliminate potentially confounding time-invariant variation across operators, and instrument Δx_{jt} with z_{jt} to isolate the exogenous variation in changes of fleet composition.

Outcomes. We consider three outcomes: number of buses in the fleet, number of employees[17], and total operating cost. All three outcomes enter the specification in natural logarithms. To give these outcomes a productive-efficiency interpretation, every specifica-

[17] Although our simple framework does not explicitly model labor, we include it in the empirical analysis as another important input in the production of urban bus transportation services.

tion conditions on the operator's total output, measured by the aggregate mileage of buses in service[18].

Fleet variety. Following the framework discussed in Section 2, our baseline measure of fleet variety is the order-two Hill number defined over bus brands, i.e. the reciprocal of the HHI index obtained on brand shares[19]. Thus, x_{jt} is the effective number of brands in the fleet: the number of equally represented brands that would generate the same concentration as the observed fleet. Higher values correspond to greater fleet variety.

Controls. In addition to the total output, w_{jt} includes the average age of buses; indicators for the presence of bus types associated with vehicle purpose as defined by drive, floor, and size; an indicator for leased buses; and the shares of buses still under warranty and originally purchased as brand-new. All specifications also include year fixed effects. These controls absorb common time shocks and observable changes in fleet characteristics that may affect performance independently of fleet variety.

More details about construction of the variables used in our analysis can be found in Section E of the Online Appendix.

4.2 Peer Fleet Renewal as an Instrument

Operators actively choose how to update their fleets, so changes in fleet composition may respond to shocks that also affect performance. We therefore treat Δx_{jt} as endogenous in equation (1) and instrument it with fleet renewal at other operators. Our instrument exploits a feature of fleet formation documented in Section 3: comparable operators often renew their fleets at similar times because they face common funding opportunities, investment cycles, and procurement conditions, while the brands they purchase remain operator-specific. This common timing of renewal provides variation in operator j's fleet composition that does not

[18]This interpretation relies on total output being exogenous from the operator's perspective, which is plausible in our setting because city authorities largely determine service through the design of the public transport network. As an alternative, Table F.16 in the Online Appendix re-estimates the main specification while expressing the outcomes per unit of output. The results are qualitatively unchanged.

[19]This specification corresponds to setting the curvature parameter q in our framework to 2. Section E of the Online Appendix varies q jointly with the power p applied to the shifts underlying our instrument. The parameter p governs the curvature of the instrument, while first-stage strength varies with both p and q. For any $q > 1$ with a sufficiently strong first stage, the qualitative results remain unchanged.

originate in its own acquisition decisions and, crucially, is plausibly unrelated to operator-specific shocks to performance.

Let $\Delta^{+}_{\text{tot}} s_{jt}$ denote total inflows into operator j's fleet as a share of its current fleet size. We construct the instrument from squared contemporaneous inflows at all other operators j':

$$z_{jt} = \sum_{j'} \omega_{jj'} \left(\Delta^{+}_{\text{tot}} s_{j't}\right)^2 .$$

The weights $\omega_{jj'}$ place greater weight on operators serving *similar* cities, as their purchasing and fleet-renewal patterns are more likely to comove. We define them using cosine similarity between operators' vectors of time-invariant characteristics, based on (average) total output, city population, and population density; and normalized so that $\omega_{jj} = 0$, $\omega_{jj'} \geq 0$, and $\sum_{j'} \omega_{jj'} = 1$.[20]

Relevance. Relevance requires two links. First, renewal at peer operators must predict operator j's own renewal. Common funding cycles, investment needs, and procurement conditions make operators more likely to renew their fleets at similar times, while averaging across a broad set of peers[21] reduces the idiosyncratic component of individual operators' purchases.

Second, operator j's own fleet renewal must translate into changes in its fleet variety. Although operators acquire buses relatively frequently, most years involve little renewal, whereas large purchases that replace a substantial fraction of the fleet occur only occasionally. Large renewal episodes replace a substantial fraction of the fleet and often involve several bus types. Since manufacturers specialize in particular types of buses, large renewals are more likely to involve multiple brands and therefore to generate meaningful changes in fleet variety. This motivates squaring peer inflows in the instrument, which places greater weight on the relatively infrequent large renewal episodes that are most informative about changes

[20] The leave-one-out restriction $\omega_{jj} = 0$ removes mechanical correlation between the instrument and operator-specific shocks. Related leave-one-out constructions are used by, among others, Autor and Duggan (2003).

[21] For averaging to reduce idiosyncratic variation, each operator must have a meaningful number of sufficiently similar peers. This condition is satisfied in our data: the average operator has an effective number of about 53 peers under our similarity weights. Table F.9 in the Online Appendix reports the distribution of these weights.

in fleet composition[22]. Section F of the Online Appendix provides evidence supporting both relevance links in detail in Tables F.8–F.11.

4.3 Identification and Inference

Validity. Identification requires peer renewal to be unrelated to shocks that directly affect operator j's performance, conditional on our controls and time effects. Several features of the setting make this restriction plausible. Public operators serve separate local markets, do not share customers or production inputs, and make their own acquisition decisions. Moreover, the fleet-level instrument depends only on the total volume of peer renewal, not on which brands or technical characteristics peers purchase. It therefore does not transmit supplier-specific choices at peer operators into operator j's performance.

We examine several potential violations of this restriction[23]. Adding operator j's own inflows and outflows leaves the estimate largely unchanged, indicating that the cost effect is not driven by the volume of fleet renewal itself. Similarly, removing persistent components of purchasing variation across peers and exposed operators does not reduce the fleet-size or cost estimates and, if anything, increases them[24]. Timing tests show that peer inflows in other years generally do not predict current outcomes and that current instrumented changes in fleet variety do not predict past outcomes (except for employment, for which the pre-trend evidence is less reassuring). We also find no evidence that the instrument is systematically associated with predetermined operator characteristics. Any remaining confounder would need to be transitory and contemporaneous, affect purchasing among operators that our weights classify as similar, influence operator j's performance through a channel other than its fleet composition, and be unrelated to predetermined operator characteristics. These restrictions make a confounding explanation difficult to reconcile with the data, although they cannot mechanically rule it out.

[22]The choice of curvature in the instrument is guided primarily by first-stage strength: convex transformations generate the strongest first stages. Table F.4 in the Online Appendix reports results for alternative choices of the curvature parameter and shows that they do not affect the economic interpretation of the estimates or the substantive conclusions.

[23]Details of these exercises are reported in Tables F.5–F.7 in the Online Appendix.

[24]However, removing the operator-level variation eliminates much of the available variation in the data and sharply weakens the first stage, making this exercise less informative than an alternative baseline specification.

Shift-share inference. As operators with similar exposure weights are affected by common peer-level shocks, observations are not independent across exposed operators j. We therefore cast the design as a shift-share instrumental variable (SSIV) design and apply the inference framework of Borusyak, Hull, and Jaravel (2022). Using their terminology, $\omega_{jj'}$ are the exposure shares and $\left(\Delta^{+}_{\text{tot}} s_{j't}\right)^2$ are the shifts. We conduct inference at the shift level and cluster standard errors by peer operator. Section D of the Online Appendix derives the corresponding shift-level representation and orthogonality condition.

Our instrument is designed to capture common procurement waves across operators. Consequently, multiple observed peer-year shifts may reflect the same underlying source of variation, so the effective amount of identifying variation may be substantially smaller than the raw number of observed shifts[25]. We therefore report weak-instrument diagnostics alongside the results. These include the Montiel Olea and Pflueger (2013) effective first-stage F-statistic and the Kleibergen–Paap rk LM underidentification test, together with Anderson–Rubin LM and Minimum Distance tests that remain valid under weak identification.

Sample. Shift-level operators form a balanced panel active throughout 2008–2023[26], allowing the exposure structure to remain fixed over time. Exposed operators are drawn from this panel subject to availability of the IGKM outcomes[27]. Our baseline sample further restricts exposed operators to cities with at least 40,000 residents, as very small transit systems have small fleets and infrequent renewal, leaving little scope for meaningful changes in fleet composition and little identifying variation.

4.4 Results

Table 3 shows that greater fleet variety worsens operator performance. A one-unit increase in the effective number of brands raises fleet size by about 3% and total operating cost by about 13%, for a given level of output. Because a one-unit change is large relative to the observed level of variety—operators have, on average, just over two effective brands (Table 2)—the

[25] Although Rotemberg-weight decompositions (Goldsmith-Pinkham, Sorkin, and Swift, 2020) show that identifying variation is relatively diffuse across individual shifts and peer operators, it is quite concentrated across years. We provide more related details in Table F.12 in the Online Appendix.

[26] Note that 2007, the first year in the IGKM sample, supplies only the levels for the first difference.

[27] Table F.3 in the Online Appendix describes in detail the resulting sample.

Table 3: The effects of fleet variety on operator performance.

	fleet-level outcomes		
	# buses	# employees	total cost
effective # brands	.030 (.013)**	−.011 (.029)	.131 (.042)***
implied effect of a mean annual variety-reducing change	-.6%	.2%	-2.5%
operator-years	1121	1109	1034
exposed operators j	79	79	78
peer operators j'	123	123	123
years t	2008-2023	2008-2023	2008-2023
Effective F	11.260	10.345	12.479
Underidentification test (KP rk LM)	10.751 (.001)***	10.355 (.001)***	12.752 (.000)***
Anderson-Rubin test (LM)	4.400 (.036)**	.153 (.695)	5.503 (.019)**
Anderson-Rubin test (MD)	5.139 (.023)**	.153 (.695)	7.599 (.006)***

Notes. The table presents fleet-level shift-share IV estimates of the effect of fleet variety on operator performance. *Estimand.* The first row reports α^F, the effect of a one-unit increase in the order-two Hill number over bus brands; the second rescales it to a typical variety-reducing change, $\exp\{\alpha^F \overline{\Delta x}\} - 1$, where $\overline{\Delta x} = \text{mean}(\Delta x_{jt} \mid \Delta x_{jt} < 0)$. *Specification.* All regressions condition on the change in total output and include the controls described in the text. *Sample.* Exposed operators serve cities with at least 40,000 residents. *Clustering.* Standard errors are clustered by peer operator. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

second row reports effects for a more empirically relevant change. A reduction in variety equal to the average annual decline observed in the data lowers fleet size by 0.6% and total operating cost by 2.5%. The effective first-stage F-statistic exceeds 10 for all three outcomes, and weak-instrument-robust tests yield the same pattern of statistical significance.

Extensions. We find no systematic effect on employment. This may partly reflect the composition of labor: drivers, whose numbers are largely pinned down by the service schedule, account for over half of employment, whereas mechanics, whose workload may respond more directly to fleet variety, make up a much smaller share. However, estimates obtained for drivers, mechanics, and other employees separately are also imprecise and statistically insignificant (Tables F.3 and F.14 in the Online Appendix). Moreover, employment is the only baseline outcome for which the pre-trend tests raise some concerns. We therefore do not draw strong conclusions about the labor response.

The cost result is not driven by the minimum-city-size restriction. Including operators serving smaller cities leaves the cost effect positive and larger (Table F.15 in the Online Appendix). However, the fleet-size response becomes smaller and less precise. This pattern may reflect the more discrete nature of capital adjustment in small fleets, where each bus

represents a substantial share of the capital stock and operators therefore have less scope to accommodate greater variety by adjusting fleet size.

Finally, the strength of the first stage depends on how we define bus classes: it is strongest when classes are defined by brand, while definitions based on other characteristics generally yield much weaker first stages, regardless of whether related to producer identity or bus purpose. Interacting brand with these characteristics provides limited additional first-stage strength, with the clearest improvement coming from drive type (Table F.13 in the Online Appendix). This pattern clarifies the scope of the design: common renewal shocks generate substantial variation in brand composition, but much less in other dimensions of fleet composition.

5 Bus-Level Effects of Fleet Variety

The fleet-level analysis establishes that greater variety raises the amount of capital and operating costs required to provide a given level of output. We now turn to the bus-level analysis, in which we estimate how fleet variety affects utilization of individual capital units.

5.1 Bus-Level Specification

Let y_{ijt} denote a performance measure of bus i in fleet j and year t, and denote its brand by $c \equiv c(i)$. We estimate the parameter α^B from the following equation:

$$\Delta y_{ijt} = \alpha^B \Delta s_{cjt} + \beta^B \Delta w_{ijt} + \Delta \varepsilon_{ijt},$$

where s_{cjt} is the class-level measure of fleet variety, w_{ijt} collects the controls, and ε_{ijt} captures unobserved determinants of bus mileage. As in the fleet-level analysis, we work in first differences. Here, differencing additionally removes persistent differences in individual-bus quality and time-invariant brand–operator match quality.

Outcome. Our bus-level outcome is annualized bus mileage. Under cost minimization, operators have an incentive to use serviceable buses intensively. Annual mileage therefore reflects both how intensively a bus is assigned to service and how consistently it is available

for assignment. Conditional on the tasks to which a bus is assigned, lower availability due to maintenance or other operating frictions appears as lower annual mileage, making it the empirical counterpart of the annual effective service capacity in Section 2. Annualized mileage enters the specification in natural logarithms. We construct annualized mileage from consecutive odometer readings; Section E of the Online Appendix describes the construction in detail.

Fleet variety. At the bus level, we measure fleet variety using own-brand prevalence, defined as brand c's share of operator j's fleet, s_{cjt}. With the baseline curvature parameter $q = 2$, the framework implies a linear relationship between productive capacity and s_{cjt}[28].

Controls. We absorb bus-age, operator×year×size, operator×year×drive, and operator×year×floor fixed effects. We control for whether the bus is under warranty or leased.

More details about construction of the variables used in our analysis can be found in Section E of the Online Appendix.

5.2 Brand-Specific Instrument

Although the specification controls for a broad set of factors affecting bus utilization, changes in brand prevalence may still be endogenous. In particular, operators may acquire more buses of brands whose vehicles are more productive for reasons specific to their setting. To address this concern, we adapt the peer-based instrument developed in Section 4 to isolate brand-specific variation in fleet shares.

The bus-level instrument builds on the same source of variation as the fleet-level design: operators participate in the same national procurement market and therefore tend to renew their fleets at similar times. We again use contemporaneous inflows at peer operators, with the same peers and similarity weights $\omega_{jj'}$. At the bus level, however, these inflows must also predict changes in the prevalence of a particular brand.

[28] Section C of the Online Appendix shows how the logarithmic empirical specification provides a first-order approximation to the relationship in the framework. Table G.17 examines alternative values of the curvature parameter q and of powers p applied to the corresponding shift. Although the estimated magnitudes vary somewhat and generally increase with q, the qualitative conclusion is robust: lower own-brand prevalence is consistently associated with lower bus utilization.

For this purpose, we exploit manufacturers' activity in the national market. Although tender requirements and scoring rules are set by individual operators, manufacturers decide which tenders to contest and how aggressively to compete. A manufacturer seeking to expand its presence may therefore participate more broadly and win contracts across several operators, particularly those with similar procurement needs. Purchases of its new buses by peer operators can consequently reveal brand-specific market activity shared across otherwise separate local markets. We measure this activity using $\Delta^{+}_{\text{new}} s_{cj't}$, the inflow of brand-new buses of brand c into peer operator j''s fleet. We focus on new buses because second-hand acquisitions are more idiosyncratic: their timing often depends on the availability of particular used vehicles, and transactions frequently occur through third-party providers rather than manufacturers themselves.

Peer new-bus inflows capture the intensity of this activity, but not necessarily the direction in which they should predict changes in operator j's brand prevalence. A peer purchase of brand c should predict an increase in its prevalence at operator j when the manufacturer is actively supplying and expanding in the primary market. When the manufacturer is instead withdrawing from, or only weakly active in, the market, a peer purchase may predict a decrease in brand c's prevalence at j. We distinguish these cases using the brand's market-wide activity.

Let m^{new}_{ct} denote the number of operators receiving at least one brand-new bus of brand c in year t, and m^{used}_{ct} the number receiving at least one second-hand bus. We set $d_{ct} = 1$ when $m^{\text{new}}_{ct} \geq m^{\text{used}}_{ct}$ and $m^{\text{new}}_{ct} > 0$, and $d_{ct} = -1$ when $m^{\text{new}}_{ct} < m^{\text{used}}_{ct}$. The sign d_{ct} distinguishes brand-years in which acquisitions are more widespread across operators in the primary market from those in which they are more widespread in the secondary market [29]. Our preferred instrument is:

$$z^{\text{signed}}_{cjt} = d_{ct} \sum_{j'} \omega_{jj'} \Delta^{+}_{\text{new}} s_{cj't}$$

[29]The sign d_{ct} is defined at the market level and is therefore common across exposed operators. Operator j's own acquisition may contribute to these market-wide operator counts, but individual operators are small relative to the national market, limiting their contribution to the assigned sign. A leave-one-out definition of the sign would make d_{ct} depend on the exposed operator j, so that the signed shift would no longer be common across exposed observations and the exact shift-share representation used for estimation and inference would be lost. We therefore keep the market-wide definition.

We also compare our preferred signed instrument with the corresponding unsigned benchmark, which imposes a common predictive direction across brand-years.

Relevance. The first-stage estimates confirm our motivation for signing the shifts. In brand-years dominated by new-bus inflows, peer new-bus inflows positively predict changes in own-brand prevalence, whereas in brand-years dominated by second-hand inflows, the relationship reverses. Pooling the two regimes therefore offsets relevant variation, while signing the shifts aligns these opposing first-stage relationships and substantially strengthens the first stage (Table G.18 in the Online Appendix). Note that signing does not introduce a new source of variation: it allows the same underlying brand-level market activity to predict changes in prevalence in the appropriate direction, avoiding the cancellation that weakens the unsigned instrument

5.3 Identification and Inference

Validity. The general exclusion logic is the same as at the fleet level: peer operators are separate local monopolies and do not share customers or production inputs with operator j. At the bus level, however, there is an additional brand-specific concern. A shock that increases purchases of brand c at peer operators could also directly affect the utilization of existing buses of brand c at operator j. In that case, the instrument would be correlated with unobserved determinants of mileage.

Our specification already removes several natural sources of correlation between brand prevalence and bus utilization. First differencing absorbs time-invariant differences in bus quality, while bus-age effects account for the systematic decline in utilization over the vehicle life cycle documented in Section 3. Finally, operator-year interactions with bus size, drive type, and floor type absorb differences in utilization arising from the assignment of buses with different purposes to different routes.

We examine in several ways the remaining concern: time-varying, brand-specific productivity shocks that are correlated with changes in the prevalence of the same brand. First, if the instrument captures slow-moving productivity shocks that also drive acquisitions, instrumented changes in prevalence should predict utilization before the corresponding fleet

change occurs. We find no evidence of such pre-trends: instrumented changes in prevalence predict neither mileage changes nor mileage levels in preceding years. Second, shifts from more distant years do not predict current utilization, supporting the interpretation that identification comes from contemporaneous manufacturer activity rather than brand-specific trends. Third, excluding all peers located within 200 km[30] leaves the estimate essentially unchanged, making geographically localized shocks affecting both peer purchases and operator j's utilization an unlikely explanation. Finally, we control for peers' second-hand inflows of the same brand and for the age composition of their same-brand fleets to absorb broader brand-specific demand and replacement cycles that could jointly affect peer purchases and the utilization of existing same-brand buses at operator j. The estimate remains positive and statistically significant, with only a marginal reduction in magnitude. These exercises are reported in detail in Tables G.22–G.25 in the Online Appendix.

SSIV inference. The bus-level instrument is also an SSIV, and we follow the inference framework introduced in Section 4. The shifts are now indexed by brand c, peer operator j', and year t. We conduct shift-level inference with two-way clustering by peer operator and brand×year, allowing for dependence across shifts associated with the same peer and with common brand-year activity[31]. We report the same effective-F, underidentification, and weak-IV-robust statistics as in the fleet-level analysis[32].

Sample. Our baseline excludes buses in their final two calendar years before leaving the fleet. Utilization declines sharply as buses approach retirement, when mileage may reflect factors unrelated to fleet composition, including mechanical condition, maintenance needs,

[30] A distance of 200 km is substantial in the Polish context, corresponding to roughly one-third of the country's north–south or east–west extent.

[31] Clustering by brand alone is not informative in this setting because identifying variation is highly concentrated across manufacturers, leaving very few effective brand clusters.

[32] Many peer–brand–year shifts may reflect the same underlying manufacturer's national market activity, so the large number of observed shifts can overstate the number of distinct sources of identifying variation. The Rotemberg-weight decomposition confirms substantial concentration across manufacturers, with an effective number of brands of 1.62 in our preferred specification. This raises the concern that the estimated effect may be driven primarily by the market leader. We therefore estimate a joint specification that allows the treatment effect to differ across three groups: the market leader, other large and established manufacturers, and more fringe manufacturers. Importantly, the group of other large and established manufacturers provides an independently relevant first stage and yields an effect of the same sign and comparable magnitude to that of the market leader. A joint test cannot reject equality of the effects across the three manufacturer groups. The detailed results are reported in Tables G.19–G.20 in the Online Appendix.

Table 4: The effect of own-brand prevalence on bus utilization.

	Δ log annual mileage	
	unsigned instr.	signed instr.
brand share	.969 (.449)**	.755 (.263)***
implied effect of a mean annual prevalence-increasing change	3.5%	2.7%
bus-years	55128	55128
exposed operators j	122	122
peer operators j'	122	122
brands c	36	36
years t	2015-2023	2015-2023
Effective F	8.253	32.5
Underidentification test (KP rk LM)	3.465 (.063)*	9.182 (.002)***
Anderson-Rubin test (LM)	2.572 (.109)	4.348 (.037)**
Anderson-Rubin test (MD)	4.988 (.026)**	10.435 (.001)***

Notes. The table presents bus-level shift-share IV estimates of the effect of own-brand prevalence on annual mileage. *Estimand.* The first row reports α^B; the second rescales it to a typical prevalence-increasing change, $\exp\{\alpha^B \overline{\Delta s_c}\} - 1$, where $\overline{\Delta s_c} = \text{mean}(\Delta s_{cjt} \mid \Delta s_{cjt} > 0)$. *Instruments.* The signed instrument multiplies peers' inflows of brand-new buses of brand c by d_{ct}; the unsigned instrument omits this sign. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

and preparation for withdrawal. Excluding these observations reduces the risk that such end-of-life dynamics confound the relationship of interest. We also exclude buses operated in Warsaw[33].

5.4 Results

Table 4 shows that buses are utilized more intensively when their own brand becomes more prevalent in the fleet. Under our preferred signed instrument, a typical purchase that raises a brand's fleet share increases the annual mileage of existing buses of that brand by about 2.7%. The effective first-stage F statistic is 32.5.

Using the unsigned instrument gives a comparable estimate of 3.5%, but a much weaker first stage with an effective F statistic of 8.3. This comparison suggests that signing mainly improves instrument relevance rather than driving the estimated utilization effect. Weak-instrument-robust inference also supports a significant positive effect under the preferred specification.

[33]Warsaw has a much larger fleet than any other city in our sample and relied relatively little on external funding during the sample period. Its purchasing cycles are therefore less synchronized with those of other operators, so it contributes many bus-year observations but little of the correlated renewal variation used by our instrument. Including Warsaw leaves the estimated effect essentially unchanged but weakens the first stage (Table G.21 in Section G of the Online Appendix), consistent with this interpretation.

Extensions. We consider a few extensions that further characterize the bus-level effect. The corresponding results are reported in Section G of the Online Appendix.

First, we examine how the effect varies over the bus life cycle. Consistent with the descriptive pattern in Figure 3, the signed-IV estimate is close to zero for buses aged 0–4 and positive for buses aged 5–14 and 15 or more, with strong first stages for both older groups (Table G.27 in the Online Appendix). The point estimates increase with age, however, we do not have enough statistical power to reject equality across the three groups. This pattern is consistent with standardization becoming more valuable as maintenance, parts replacement, and accumulated experience with a manufacturer's technology become more important over the bus life cycle.

Second, we examine alternative definitions of bus classes. Our instrument generates substantial first-stage variation in manufacturer-related dimensions of fleet composition, but much less variation in characteristics related to bus purpose. Only when we interact these characteristics with brand do we recover strong first stages (Table G.26 in the Online Appendix). In particular, defining classes by brand×size yields a positive and precisely estimated effect, supporting a role for manufacturer-related standardization even among similarly sized buses.

6 Understanding the Bus-Level Utilization Effect

So far, we have established that lower own-brand prevalence reduces utilization. We now consider two alternative mechanisms that could generate the same relationship without reflecting a cost of fleet variety. First, a brand's share can fall mechanically when the rest of the fleet expands, so the estimate could reflect fleet growth rather than a change in composition. Second, utilization may depend on the absolute number of same-brand buses rather than on their relative prevalence in the fleet, in which case the relevant friction would be one of scale of operation rather than fleet composition. These two exercises provide an indirect test of our framework by assessing whether the estimated utilization effect operates through relative type prevalence, as the model assumes.

6.1 Accumulation or Dilution

Let n_{cjt} denote the number of buses of brand c and N_{jt} total fleet size, so that $s_{cjt} = {}^{n_{cjt}}/_{N_{jt}}$. A brand's fleet share can change either because the operator acquires or retires buses of brand c (a change in n_{cjt}, henceforth the accumulation channel), or because the fleet as a whole grows or shrinks (a change in N_{jt}, henceforth the dilution channel). The latter creates a potential alternative explanation for our estimates. Panel (a) of Table 5 shows that fleet growth is associated with lower mileage per bus, while it also mechanically reduces the shares of brands whose bus counts do not increase proportionately. A positive relationship between own-brand prevalence and utilization could therefore arise even if fleet composition had no effect on utilization, simply because both respond to changes in total fleet size.

To separate the two channels, we formally decompose changes in the fleet share of brand c into accumulation and dilution:

$$\Delta s_{cjt} = \underbrace{\frac{1}{\bar{N}_{jt}} \Delta n_{cjt}}_{\text{accumulation}} - \underbrace{\frac{\bar{s}_{cjt}}{\bar{N}_{jt}} \Delta N_{jt}}_{\text{dilution}},$$

where $\Delta a_t = a_t - a_{t-1}$ denotes the first-difference operator and $\bar{a}_t = \frac{1}{2}(a_t + a_{t-1})$ is the average across the two periods. We derive this decomposition in Section H of the Online Appendix.

Panel (b) of Table 5 shows that the instrument-induced variation in brand prevalence comes almost entirely through the accumulation channel. Changes in total fleet size, ΔN_{jt}, occur only at the operator-year level and are therefore absorbed by operator-year fixed effects. The remaining cross-brand variation in dilution comes through $\bar{s}_{cjt}$, which the instrument predicts very weakly, as indicated by the low effective F statistic. Consistent with this, replacing Δs_{cjt} with the accumulation component alone leaves both the first stage and the estimated effect essentially unchanged. As a result, instrument-induced changes in brand prevalence therefore satisfy approximately:

$$\Delta s_{cjt} \approx \frac{\Delta n_{cjt}}{\bar{N}_{jt}} \tag{2}$$

Table 5: Own-brand prevalence and bus utilization: accumulation and dilution.

	panel (a): fleet growth and bus utilization		
	Δ log annual mileage		
fleet growth ΔN_{jt} (difference of 90th–10th percentile, OLS)	$-.033$ $(.007)^{***}$		
	panel (b): accumulation and dilution components		
	Δ log annual mileage		
	total change	accumulation	dilution
brand share	.755 $(.263)^{***}$	.725 $(.252)^{***}$	-17.740 (11.736)
bus-years	55128	55128	55128
exposed operators j	122	122	122
peer operators j'	122	122	122
brands c	36	36	36
years t	2015-2023	2015-2023	2015-2023
Effective F	32.5	28.8	2.318
Underidentification test (KP rk LM)	9.182 $(.002)^{***}$	9.148 $(.002)^{***}$	2.338 (.126)

Notes. Panel (a) reports the OLS change in log annual mileage associated with moving fleet growth, ΔN_{jt}, from its 10th to its 90th percentile. Panel (b) reports bus-level shift-share IV estimates using the total change in own-brand prevalence, its accumulation component, or its dilution component as the endogenous variable. *Instrument.* Panel (b) uses the signed instrument. *Sample.* Baseline bus-level estimation sample. *Clustering.* Panel (b) standard errors are two-way clustered by peer operator and brand × year. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

We use this result below to distinguish whether relative shares or absolute counts provide the more appropriate measure of fleet composition.

6.2 Shares or Counts

Our main bus-level specification measures brand prevalence by the share of brand c in operator j's fleet, s_{cjt}. An alternative is to measure it by the absolute number of same-brand buses, n_{cjt}. These two specifications correspond to different economic mechanisms. Under the *shares* specification, utilization depends on a brand's prevalence relative to the rest of the fleet. This is consistent with operational systems adapting to fleet composition: for example, inventories, maintenance routines, or workshop organization being geared toward the brands that are most prevalent. Two fleets with the same brand shares therefore imply the same prevalence-related utilization effect, even if one fleet is larger. Under the *counts* specification, utilization depends instead on the absolute number of same-brand buses. This could arise from fixed investments in spare parts, tooling, or brand-specific expertise, whose cost per bus falls as the number of buses of that brand increases.

The distinction matters for interpreting our results. If shares are the relevant measure, the utilization effect is tied to relative fleet composition and hence directly to fleet variety. If counts are the relevant measure, the friction is less about the cost of operating a varied fleet and more about the cost of being small.

Equation (2) provides a direct way to distinguish between these mechanisms. Because instrument-induced changes in brand prevalence satisfy approximately $\Delta s_{cjt} \approx \Delta n_{cjt}/\bar{N}_{jt}$, each specification can be expressed in terms of either measure:

$$\begin{aligned} &\text{(shares)} \quad \Delta \log y_{ijt} = \alpha\, \Delta s_{cjt} + \ldots \quad \approx \frac{\alpha}{\bar{N}_{jt}}\, \Delta n_{cjt} + \ldots \\ &\text{(counts)} \quad \Delta \log y_{ijt} = \alpha\, \Delta n_{cjt} + \ldots \quad \approx \alpha \bar{N}_{jt}\, \Delta s_{cjt} + \ldots . \end{aligned}$$

The two mechanisms imply opposite approximate scaling patterns across operators of different sizes. Under the shares specification, the coefficient on shares should remain stable with fleet size, while the coefficient on counts should decline approximately in proportion to inverse fleet size. Under the counts specification, the coefficient on counts should remain stable, while the coefficient on shares should increase approximately in proportion to fleet size.

We evaluate these scaling implications by comparing the effects estimated on subsamples of smaller and larger operators. Because contemporaneous fleet size may itself respond to fleet variety, we use city population as a predetermined proxy for operator scale and split operators at 90,000 residents. We use the average fleet size in each group to derive the predicted relationship between the share- and count-based estimates under the two specifications. Table 6 presents the results.

The results favor the relative-prevalence model. When the relationship is estimated in shares, the coefficients are very similar across smaller and larger operators despite a more than fivefold difference in average fleet size, and we cannot reject their equality. When estimated in counts, the coefficients instead scale approximately inversely with average fleet size, as predicted by the relative-prevalence model. By contrast, the absolute-count model predicts similar coefficients when prevalence is measured in counts and coefficients that increase with fleet size when it is measured in shares, both of which we reject.

Table 6: The effect of own-brand prevalence on bus utilization: shares versus counts.

	Δ log annual mileage	
	shares	counts
small fleets (raw estimate)	.791 (.203)***	.027 (.007)***
large fleets (raw estimate)	.742 (.326)**	.003 (.001)**
exposed operators j (small operators)	86	86
exposed operators j (large operators)	36	36
average fleet size (small operators)	29.5	29.5
average fleet size (large operators)	158	158
effective F (small operators)	18.710	9.017
effective F (large operators)	17.076	8.032
shares. H_0: share coefficients equal across groups	.022 (.881)	
shares. H_0: count coefficients scale with inverse mean fleet size		1.243 (.265)
counts. H_0: count coefficients equal across groups		11.065 (.001)***
counts. H_0: share coefficients scale with mean fleet size	11.494 (.001)***	

Notes. The table compares specifications measuring own-brand prevalence by fleet share and by the number of same-brand buses. *Design.* Operators are split at 90,000 residents, using city population as a predetermined proxy for scale; both specifications are estimated jointly across the two groups. The shift is the signed peer inflow of brand c, measured as a share of peer fleets and common to both columns. *Tests.* Since $\Delta s_{cjt} \simeq \Delta n_{cjt}/\bar{N}_{jt}$, the shares model implies equal share coefficients and count coefficients that scale inversely with fleet size; the counts model implies equal count coefficients and share coefficients that scale with fleet size. Restrictions use mean fleet size within each group. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand $\times$ year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below Wald tests. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Importantly, the exercise retains meaningful identifying power in both operator groups: the first stages remain informative and the estimates sufficiently precise to distinguish between the two scaling patterns. We therefore conclude that the data favor relative shares, rather than absolute counts, as the relevant measure of brand prevalence. Hence, the bus-level evidence supports the mechanism embedded in our framework: productive efficiency depends on the relative prevalence of a capital type within the fleet.

This result clarifies where the costs of fleet variety come from. The utilization penalty associated with operating a minority brand depends on its relative presence in the fleet, rather than simply on having too few buses of that brand. Increasing operator scale therefore does not eliminate the penalty if fleet composition remains unchanged. This evidence points to a cost of operating a relatively fragmented fleet, rather than to a fixed brand-specific cost that larger operators can amortize over more same-brand buses. It is also consistent with the fleet-level evidence, where variety is measured by a scale-free index and greater variety is associated with a larger fleet and higher operating costs for a given level of output.

7 Conclusion

This paper documents the productivity costs of operating heterogeneous capital. Using variation in fleet composition generated by common procurement conditions and manufacturers' national market activity, we show that operators with more heterogeneous fleets require more buses and incur higher operating costs to provide a given level of output. A typical observed reduction in fleet variety lowers total operating costs by about 2.5% and the fleet required to provide the same output by about 0.6%. Consistent with these aggregate effects, buses are utilized more intensively as their own brand becomes more prevalent in the fleet: a typical increase in own-brand prevalence raises the annual mileage of existing buses of that brand by about 2.7%.

We further examine what drives this utilization effect. It is not explained by mechanical changes in brand shares as the overall fleet expands: the identifying variation comes almost entirely from changes in same-brand holdings. The data also favor relative brand shares over absolute counts. Buses of a minority brand are used less intensively because that brand is relatively uncommon in the fleet, not simply because the operator owns few buses of that brand. The result therefore points to a cost of fleet composition rather than a fixed brand-specific cost that larger operators can spread over more buses. This dependence on relative prevalence is consistent with frictions arising from how shared resources, such as maintenance capacity, spare parts, or technical expertise, are organized across brands within the fleet.

We focus on urban bus fleets because they provide a particularly clean setting in which to identify the costs of heterogeneous capital. Buses perform a common and well-defined task, constitute the main capital input, and can be tracked individually over long periods, while repeated procurement generates substantial variation in fleet composition. Although the setting is specific, the underlying friction is more general. Durable capital is often acquired gradually from differentiated suppliers and remains in use for many years. When equipment from different suppliers requires different parts, expertise, infrastructure, or operating routines, units designed to perform similar tasks may nevertheless be operationally incompatible. Such compatibility frictions can persist for years and are difficult to detect in aggregate measures of capital. Our results therefore suggest that the productive value

of capital depends not only on its quantity or value, but also on the composition of its constituent units and how effectively they can operate together.

Why does variety arise at all if it is costly? Complete standardization may be difficult or undesirable. In our setting, purchases occur intermittently, often in funding-driven waves, and tender outcomes depend partly on which manufacturers are active and compete at the time. Operators can influence these outcomes—and indeed design tender requirements that favor brands already present in their fleets (Premik, 2026)—but cannot fully control them. More generally, technological change, supplier entry and exit, capacity constraints, and delivery deadlines can make standardization difficult to sustain. Variety may also provide benefits: diversifying across suppliers can insure against manufacturer-specific recalls and limit exposure to supplier market power and lock-in, while using different types of capital can reduce costs when they are better suited to different tasks. Capital input composition therefore reflects a trade-off between the productive costs of fleet variety documented in this paper and the constraints and benefits that make complete standardization undesirable or infeasible.

Acknowledgements

We thank Arthur Campbell, Maxime Gravoueille, Gordon Leslie, Laura Puzzello, and seminar and conference participants for helpful comments and suggestions. We are grateful to IGKM, the urban transport chamber of commerce in Poland, for sharing their data.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflicts of interest.

References

Achilles, L. (2008): "1968–2008: Vor 40 Jahren erster Standard-Linienbus in Hamburg," *HOV-Aktuell*, (27).

Autor, D. H., and M. G. Duggan (2003): "The Rise in the Disability Rolls and the Decline in Unemployment," *The Quarterly Journal of Economics*, 118(1), 157–206.

Bartelsman, E. J., and M. Doms (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-Experimental Shift-Share Research Designs," *The Review of Economic Studies*, 89(1), 181–213.

Brüggen, A., and L. Klose (2010): "How Fleet Commonality Influences Low-Cost Airline Operating Performance: Empirical Evidence," *Journal of Air Transport Management*, 16(6), 299–303.

Chao, A., C.-H. Chiu, and L. Jost (2014): "Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers," *Annual Review of Ecology, Evolution, and Systematics*, 45, 297–324.

Collard-Wexler, A., and J. De Loecker (2021): "Productivity and Capital Measurement Error," Working paper, Duke University and KU Leuven. Revised version of NBER Working Paper 22437 (2016).

Coyle, D., I. Bournakis, and J.-C. Mei (2025): "Heterogeneous Capital and the Productivity of UK Manufacturing Firms," SSRN Working Paper 5954397.

de Borges Pan, A. G., and R. A. Espírito Santo, Jr. (2004): "Developing a Fleet Standardization Index for Airline Pricing," *Journal of Air Transportation*, 9(2), 97–110.

de Roux, N., M. Eslava, S. Franco, and E. Verhoogen (2021): "Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments," Working Paper 28323, National Bureau of Economic Research.

Fox, J. T., and V. Smeets (2011): "Does Input Quality Drive Measured Differences in Firm Productivity?," *International Economic Review*, 52(4), 961–989.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110(8), 2586–2624.

Griliches, Z. (1957): "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25(4), 501–522.

Griliches, Z., and J. Mairesse (1995): "Production Functions: The Search for Identification," Working Paper 5067, National Bureau of Economic Research.

Hill, M. O. (1973): "Diversity and Evenness: A Unifying Notation and Its Consequences," *Ecology*, 54(2), 427–432.

Jorgenson, D. W., and Z. Griliches (1967): "The Explanation of Productivity Change," *The Review of Economic Studies*, 34(3), 249–283.

Kebede, H. A., and M. S. McMillan (2026): "Firm Capacity Underutilization and the Measurement of Productivity," *Journal of Development Economics*, 182, 103786.

Kim, K. I., A. Petrin, and S. Song (2016): "Estimating Production Functions with Control Functions When Capital Is Measured with Error," *Journal of Econometrics*, 190(2), 267–279.

Merkert, R. (2023): "The Impact of Engine Standardization on the Cost Efficiency of Airlines," *Research in Transportation Business & Management*, 46, 100797.

Montiel Olea, J. L., and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31(3), 358–369.

Premik, F. (2024): "Procurement with Bid Preference and Buyer's Switching Costs: The Case of Municipal Buses," Unpublished manuscript, Monash University.

——— (2026): "Identifying Buyer's Conduct in Procurement Auctions," Unpublished manuscript, Monash University.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

Statistics Poland (2024): *Statistical Yearbook of the Republic of Poland 2024*. Statistics Poland, Warsaw.

Whelan, K. (2002): "Computers, Obsolescence, and Productivity," *Review of Economics and Statistics*, 84(3), 445–461.

Wilson, D. J. (2009): "IT and Beyond: The Contribution of Heterogeneous Capital to Productivity," *Journal of Business & Economic Statistics*, 27(1), 52–70.

Zou, L., C. Yu, and M. Dresner (2015): "Fleet Standardisation and Airline Performance," *Journal of Transport Economics and Policy*, 49(1), 149–166.

Online Appendix

A Details of Bus Differentiation

This section describes alternative dimensions of bus differentiation and assesses how much of this broader heterogeneity is captured by our baseline definition based on bus brand.

A.1 Dimensions of Bus Differentiation

Classes related to producer identity. These classes capture technological differences associated with manufacturers of buses and key components, including chassis design, engines, and interior layouts, with potentially substantial implications for operation, maintenance, and repair. The primary class is bus *brand*, defined by the identity of the bus manufacturer. We also distinguish *generations* within a brand, which group buses sharing common technological features over a production period. Because engines are often supplied by specialized manufacturers distinct from the bus producer, we separately identify *engine brand* and, within each engine brand, *engine families* that share key characteristics such as displacement or chassis integration.

Classes related to bus purpose. These classes capture characteristics that shape how buses are used. We distinguish *size*, based on length categories ranging from small vehicles suited to narrow streets to articulated buses serving high-demand routes; and *drive*, which describes propulsion technologies such as diesel, compressed natural gas, or electric and may require dedicated refueling or charging infrastructure. For buses with internal-combustion engines, we also distinguish *displacement*, the total volume swept by the pistons, as an additional technical characteristic relevant to vehicle use. Internal-combustion buses account for over 93% of our sample.

Bus differentiation in the data. Table B.1 shows substantial differentiation across all dimensions, particularly those related to producer identity. Market-wide variety substantially exceeds variety within individual fleets, and this gap remains large when raw counts are replaced by effective counts that account for unequal class shares. Differentiation in

Table A.1: Variety in the market and within fleets, by definition of the class, 2007–2023.

	market					fleet				
	#	eff. #	largest	second	HHI	#	eff. #	largest	second	HHI
classes related to producer's identity										
brand	25	4.9	.38	.18	.22	3.9	2.3	.6	.24	.49
generation	64.4	10.	.23	.13	.1	6.4	3.6	.45	.23	.33
engine brand	20.7	5.4	.31	.2	.18	4.4	2.6	.55	.26	.43
engine family	52.1	10.	.2	.15	.1	6.3	3.4	.46	.23	.34
classes related to bus purpose										
size	5	2.2	.61	.25	.44	2.8	1.7	.72	.22	.62
drive	4.3	1.2	.91	.05	.84	1.6	1.1	.93	.06	.9
displacement	5	3.2	.37	.32	.31	3.1	2.1	.62	.26	.52
all types	276.6	40.	.09	.07	.02	10.	5.7	.34	.19	.23

Notes. The table summarizes market-wide and within-fleet differentiation across the class definitions discussed above. *Sample.* All public bus operators in the odometer sample, 2007–2023. *Definitions.* Market statistics pool all buses in the sample; fleet statistics are computed within operator-year and then averaged. Shares are based on bus-days. The effective number of classes is the reciprocal of the HHI. `All types` defines classes by the full combination of characteristics listed above.

characteristics related to bus purpose is more limited, both across the market and within fleets. Overall, fleet composition is multidimensional, but individual operators use only a concentrated subset of the available types.

Our identifying variation operates primarily through bus brand. We therefore next assess how much of this broader differentiation is captured by brand.

A.2 How Much Differentiation Does Brand Capture?

We use two complementary measures of how much of the variation in the alternative bus characteristics is captured by brand.

Regression-Based Measure. For each categorical characteristic X_k, we regress its complete set of category indicators on brand fixed effects. Let $\mathrm{SS}_{kj}^{\mathrm{expl}}$ and $\mathrm{SS}_{kj}^{\mathrm{tot}}$ denote the explained and total sums of squares for category j of characteristic k. We aggregate across categories and characteristics as

$$\rho^2 = \frac{\sum_{k=1}^{K} \sum_{j=1}^{C_k} \mathrm{SS}_{kj}^{\mathrm{expl}}}{\sum_{k=1}^{K} \sum_{j=1}^{C_k} \mathrm{SS}_{kj}^{\mathrm{tot}}}.$$

The statistic lies between zero and one and measures the share of variation in the complete set of category indicators that is explained by brand. We compute the same statistic separately for characteristics related to producer identity and bus purpose.

Table A.2: Variation in Bus Characteristics Captured by Brand

	# categories	ρ^2	Greenacre-weighted R^2
all characteristics	226	0.323	0.497
→ of which: manufacturer's identity	209	0.391	0.553
→ of which: bus purpose	17	0.201	0.179

Notes. The table reports how much of the joint variation in bus characteristics is captured by brand. *Measures.* ρ^2 is the pooled share of variation in the complete set of category indicators explained by brand indicators. The Greenacre-weighted R^2 averages the R^2s from regressions of retained MCA dimensions on brand indicators, using Greenacre-adjusted inertias as weights. *Rows.* The first row pools all characteristics; the remaining rows consider characteristics related to producer identity and bus purpose separately. *Sample.* All public bus operators in the odometer sample, 2014–2023.

MCA-Based Measure. Our second approach reverses the order of aggregation and regression. We first summarize the joint variation in the characteristics other than brand using Multiple Correspondence Analysis (MCA), the analogue of principal component analysis for categorical variables, and then measure how much of the resulting variation is explained by brand. We retain dimensions whose inertia exceeds the 95th percentile of the corresponding inertia obtained when the categories of each characteristic are independently permuted across buses, using 200 permutations. This retains 79, 73, and 6 dimensions in the three respective applications.

We regress each retained MCA row score on brand fixed effects and combine the resulting R^2s using Greenacre-adjusted inertias as weights. The resulting inertia-weighted R^2 lies between zero and one and provides an alternative measure of the share of joint differentiation captured by brand.

Results. Both approaches lead to the same conclusion: brand captures a substantial but incomplete share of the broader differentiation across buses. Across all characteristics, the regression-based measure gives $\rho^2 = 0.323$, while the MCA-based measure gives an inertia-weighted R^2 of 0.497. Thus, depending on whether we aggregate the original category indicators directly or first summarize them through MCA, brand captures roughly a third to a half of the broader differentiation observed across buses.

Brand is considerably more informative about characteristics related to producer identity, for which the corresponding measures are 0.391 and 0.553, than about characteristics related to bus purpose, for which they are 0.201 and 0.179. Thus, brand primarily captures technological differentiation associated with producer identity rather than differences in the purposes buses are designed to serve.

This distinction is useful for our empirical analysis. Brand does not capture all dimensions of bus differentiation, and we consider alternative class definitions where the data provide sufficient identifying variation. At the same time, its relatively weak relationship with purpose-related characteristics suggests that brand is not simply a proxy for the tasks buses perform, reducing the concern that brand-specific variation reflects correlated changes in demand for particular bus types across similar cities.

B Additional Model Derivations

This section provides three additional derivations for the framework developed in the main text. We first derive the comparative statics of fleet variety for required fleet size, then make explicit the implications for operating costs, and finally relate the model to the empirical specifications which use logarithms of bus utilization and fleet-level outcomes.

B.1 Comparative Statics of Fleet Variety

Fleet Size. Recall that, under the power specification for bus-level performance, the fleet size required to provide output y is:

$$N(y, x_q) = \frac{\bar{K}(y)}{\gamma}(x_q)^{q-1}, \qquad q > 1.$$

It follows that:

$$\frac{\partial N(y, x_q)}{\partial x_q} = \frac{\bar{K}(y)}{\gamma}(q-1)(x_q)^{q-2} > 0$$

Thus, greater fleet variety requires a larger physical fleet to provide a given level of output.

Total cost. Suppose that operating and maintaining a bus entails a baseline cost $\kappa > 0$, and let $h(x_q) \geq 0$ denote additional per-bus operating costs associated with fleet variety, with $h'(x_q) \geq 0$. Total operating cost is:

$$TC(y, x_q) = [\kappa + h(x_q)]\, N(y, x_q) + C(y)$$

where $C(y)$ collects costs determined directly by output. It follows that:

$$\frac{\partial TC(y, x_q)}{\partial x_q} = h'(x_q)N(y, x_q) + [\kappa + h(x_q)] \frac{\partial N(y, x_q)}{\partial x_q} > 0$$

Greater fleet variety can therefore raise operating costs through two channels: directly, by increasing the cost of operating and supporting a given bus, and indirectly, by increasing the physical fleet required to provide a given level of output. Setting $h(x_q) = 0$ recovers the capacity-only benchmark in which the cost effect operates entirely through fleet size.

B.2 Relation to the Empirical Specifications

The framework is stated in levels, whereas our empirical specifications use logarithmic outcomes. We use logarithms because the outcomes are positive, differ substantially in scale across buses and operators, and proportional changes are more comparable than changes in levels. We now show that the empirical equations can be interpreted as first-order approximations to the relationships implied by the model.

Bus-level utilization. Under the power specification in Section 2, $v(s_c) = \gamma s_c^{q-1}$. For a given value of q, define $h_q(s_c) \equiv s_c^{q-1}$, so the model can then be written as $v(s_c) = \gamma h_q(s_c)$ and hence $\log v(s_c) = \log \gamma + \log h_q(s_c)$. A first-order expansion around a reference value $\bar{h}_q$ gives

$$\Delta \log v(s_c) \approx \frac{1}{\bar{h}_q} \Delta h_q(s_c) = \frac{1}{\bar{s}^{q-1}} \Delta s_c^{q-1}$$

Thus, for any value of q, the model implies a relationship that is locally linear in changes in s_c^{q-1}. In our baseline case, $q = 2$, so $h_q(s_c) = s_c$ and the approximation becomes:

$$\Delta \log v(s_c) \approx \frac{1}{\bar{s}} \Delta s_c$$

motivating our baseline bus-level specification.

Fleet-level outcomes. At the fleet level, the model implies:

$$N(y, x_q) = \frac{\bar{K}(y)}{\gamma} (x_q)^{q-1}$$

Taking logarithms yields:

$$\log N(y, x_q) = \log \bar{K}(y) - \log \gamma + (q-1) \log x_q$$

For a given q, a first-order expansion around a reference level $\bar{x}_q$ gives:

$$\Delta \log N \approx \frac{q-1}{\bar{x}_q} \Delta x_q$$

motivating our fleet-level specification, which can therefore be interpreted as a local linear approximation to the relationship between fleet variety and required fleet size implied by the model. Analogous steps derive the approximation for the total cost equation.

C Shift-Share Representation and Inference

Following Borusyak, Hull, and Jaravel (2022), both empirical designs admit a common shift-level representation. Let ℓ index exposed observations and n index shifts. Let r_ℓ denote the regression weight and $s_{\ell n}$ the exposure of observation ℓ to shift n, so that

$$z_\ell = \sum_n s_{\ell n} g_n.$$

After partialling out the controls in the corresponding specification, let $\widetilde{y}_\ell$ and $\widetilde{x}_\ell$ denote the outcome and endogenous variable. Define

$$W_n = \sum_\ell r_\ell s_{\ell n}, \qquad \widetilde{y}_n^S = \frac{\sum_\ell r_\ell s_{\ell n} \widetilde{y}_\ell}{W_n}, \qquad \widetilde{x}_n^S = \frac{\sum_\ell r_\ell s_{\ell n} \widetilde{x}_\ell}{W_n}.$$

The IV estimator can then be written as

$$\widehat{\alpha} = \frac{\sum_n W_n g_n \widetilde{y}_n^S}{\sum_n W_n g_n \widetilde{x}_n^S},$$

with orthogonality condition

$$\mathbb{E}\left[\sum_n W_n g_n \widetilde{\epsilon}_n^S\right] = 0,$$

where $\widetilde{\epsilon}_n^S$ is defined analogously. The fleet- and bus-level designs differ in their regression weights, exposures, and shifts.

C.1 Fleet-Level Design

At the fleet level, $r_{jt} = 1$, $\ell = (j, t)$, and $n = (j', t)$, with

$$s_{(j,t),(j',t)} = \omega_{jj'}, \qquad W_{j't} = \sum_j \omega_{jj'}.$$

The baseline instrument is

$$z_{jt} = \sum_{j'} \omega_{jj'} g_{j't}, \qquad g_{j't} = \left(\Delta^+_{\text{tot}} s_{j't}\right)^2,$$

where $\omega_{jj'} \geq 0$ measures similarity between exposed operator j and peer j', $\omega_{jj} = 0$, and $\sum_{j'} \omega_{jj'} = 1$. Specifications varying instrument curvature use

$$g_{j't} = \left(\Delta^+_{\text{tot}} s_{j't}\right)^p,$$

with $p = 2$ in the baseline.

The exclusion restriction therefore requires peer purchasing shifts to be orthogonal to the exposure-weighted performance shocks of the operators exposed to them.

C.2 Bus-Level Design

At the bus level, all buses of brand c at operator j in year t share the same endogenous variable and instrument, so we represent the regression at the brand–operator–year level. Let n_{cjt} denote the number of bus observations in cell (c, j, t) and set $r_{cjt} = n_{cjt}$. Then $\ell = (c, j, t)$ and $n = (c, j', t)$, with

$$s_{(c,j,t),(c,j',t)} = \omega_{jj'}, \qquad W_{cj't} = \sum_j n_{cjt} \omega_{jj'}.$$

The baseline signed instrument is

$$z_{cjt} = \sum_{j'} \omega_{jj'} g_{cj't}, \qquad g_{cj't} = d_{ct} \Delta^{+}_{\text{new}} s_{cj't},$$

where $d_{ct} = 1$ if

$$\#\{j : \Delta^{+}_{\text{new}} s_{cjt} > 0\} \geq \#\{j : \Delta^{+}_{\text{second}} s_{cjt} > 0\},$$

and $d_{ct} = -1$ otherwise. The unsigned specification sets $d_{ct} = 1$. Because d_{ct} is common across operators within a brand-year, signing aligns shifts according to whether new-bus market activity predicts an increase or decrease in the brand's prevalence.

Validity requires the signed purchasing shift generated by peer j' for brand c to be orthogonal to exposure-weighted residual utilization shocks for buses of that brand.

C.3 Inference

At the fleet level, shifts vary by peer operator × year, and we cluster standard errors by peer operator. At the bus level, shifts vary by brand × peer operator × year, and we two-way cluster by peer operator and brand × year, allowing dependence across peer shifts within a brand-year, including that induced by d_{ct}.

C.4 Rotemberg Weights

We use Rotemberg weights to describe the distribution of identifying variation across shifts (Goldsmith-Pinkham, Sorkin, and Swift, 2020; Borusyak, Hull, and Jaravel, 2022). For a generic shift n,

$$\lambda_n = \frac{W_n g_n \widetilde{x}^S_n}{\sum_m W_m g_m \widetilde{x}^S_m}, \qquad \widehat{\alpha} = \sum_n \lambda_n \widehat{\alpha}_n, \qquad \widehat{\alpha}_n = \frac{\widetilde{y}^S_n}{\widetilde{x}^S_n}.$$

Because λ_n may be negative, we measure concentration using normalized absolute weights,

$$\pi_n = \frac{|\lambda_n|}{\sum_m |\lambda_m|}, \qquad N_{\text{eff}} = \frac{1}{\sum_n \pi_n^2}.$$

We report cumulative absolute weights and effective numbers of shifts, including after aggregation by peer operator and year at the fleet level, and additionally by brand and brand $\times$ year at the bus level.

D Variable Construction

This section describes the construction of the main variables on the fleet- and bus-level.

D.1 Fleet Size and Composition

Fleet size and composition. Our bus-level data record the periods during which each bus belongs to an operator's fleet. Because buses may enter or leave during a calendar year, we measure fleet presence using bus-days rather than point-in-time counts. We define operator j's fleet size in year t, N_{jt}, as total bus-days divided by 365. A bus present throughout the year therefore contributes one unit, while buses present for only part of the year contribute proportionately less. Similarly, n_{cjt} denotes the annualized number of buses of class c, constructed from their bus-days divided by 365. Thus:

$$N_{jt} = \sum_{c} n_{cjt}, \qquad s_{cjt} = \frac{n_{cjt}}{N_{jt}}$$

Both N_{jt} and n_{cjt} therefore measure annualized fleet sizes rather than point-in-time vehicle counts.

Inflows and fleet renewal. We measure inflows using the first 365 bus-days after each bus enters an operator's fleet, so that the contribution does not depend mechanically on the calendar date of entry. For example, a bus entering on December 30 contributes two bus-days in the entry year and 363 in the following year.

Let $\Delta^{+}_{\text{tot}} s_{cjt}$ denote total inflows of brand c into operator j's fleet, expressed as a share of current fleet size. We decompose them into new and second-hand inflows:

$$\Delta^{+}_{\text{tot}} s_{cjt} = \Delta^{+}_{\text{new}} s_{cjt} + \Delta^{+}_{\text{second}} s_{cjt}$$

Summing across brands gives the fleet-level measure:

$$\Delta^{+}_{\text{tot}} s_{jt} = \sum_{c} \Delta^{+}_{\text{tot}} s_{cjt}$$

Fleet variety. Informed by our simple framework, we measure fleet variety using the Hill number:

$$x_{jt,q} = \left(\sum_{c} s^{q}_{cjt} \right)^{\frac{1}{1-q}}$$

Our baseline sets $q = 2$, yielding the inverse Herfindahl index:

$$x_{jt,2} = \frac{1}{\sum_{c} s^{2}_{cjt}}$$

This measure gives the effective number of equally represented classes in the fleet. We use bus brand as the baseline class, but also consider alternative definitions based on, among others, engine brand, size, drive type, and their interactions with brand.

D.2 Outcome Variables

Bus utilization. Our bus-level outcome y_{ijt} is annualized mileage. Mandatory vehicle inspections record odometer readings at irregular dates, typically at least once per year. For each pair of consecutive readings, we compute average daily mileage as the odometer change divided by the number of days between inspections. We allocate this mileage across calendar years according to the number of interval days falling in each year, aggregate within bus-year, and annualize to 365 days. This makes mileage comparable across buses observed for different fractions of a calendar year, which occurs because of patterns of acquiring and withdrawing buses, as well as due to some data errors in the scraped registry.

Fleet-level outcomes. We consider fleet size, employment, and total operating cost. Fleet size is total bus-days divided by 365, employment is total operator employment, and operating cost is the annual cost of providing services reported in the IGKM survey. We express all three outcomes in natural logarithms.

Our baseline specifications condition on log total operator output, measured using the service-output variable described in Section 3. The coefficients on fleet size and operating cost therefore capture changes in the resources required to provide a given level of output.

D.3 Peer Similarity and Exposure Weights

For each operator j, all other operators in the national market form the set of potential peers. We assign greater weight to operators serving similar markets, measured by total output, city population, and population density.

We construct total output from the odometer data and average it over each operator's observation window. We take 2020 city population and population density from Statistics Poland. Because all three characteristics are relatively persistent over our sample period, we use time-invariant measures to capture cross-sectional similarity between operators. Importantly, this keeps the exposure structure fixed over time, which substantially simplifies the analysis and prevents the weights themselves from responding to contemporaneous changes in operators' circumstances.

We standardize each characteristic to mean zero and unit variance and compute cosine similarity between each pair of operators. We set negative similarities and the own-operator weight to zero, then normalize the remaining positive similarities to sum to one:

$$\omega_{jj} = 0, \qquad \omega_{jj'} \geq 0, \qquad \sum_{j'} \omega_{jj'} = 1$$

Operators serving more similar markets therefore receive larger exposure weights, while sufficiently dissimilar operators receive zero weight.

E Fleet-Level Design: Supporting Evidence

This section presents additional evidence supporting the fleet-level empirical design. We examine the fleet-level sample, functional-form choices for fleet variety and the instrument, identification validity, first-stage relevance, alternative variety definitions, and extensions of the main results.

Table E.3: Operators in the Fleet-Level (IGKM) Sample: Summary Statistics, 2007–2023

	total/mean	median	standard deviation
# operators	112	—	—
# operators-years	1656	—	—
fleet size (# buses)	81	38	149
employment (# employees)	324	126	640
→ including drivers	166	69	346
→ including mechanics	43	19	71
output (thousands bus-km)	4767	1797	9540
total operating cost (thousands 2010 USD)	10475	3681	23349

Notes. The table summarizes the fleet-level analysis sample. *Source.* Annual IGKM survey, 2007–2023. *Statistics.* The first two rows report totals; all remaining rows report the mean, median, and standard deviation across operator-years. *Definitions.* Output is the distance traveled by an operator's buses in service during the year. Total operating cost is deflated and expressed in 2010 USD. Drivers and mechanics do not exhaust total employment; the remainder consists of administrative and other staff.

The estimation sample. Fleet-level outcomes come from the annual IGKM survey described in Section 3. Table F.3 summarizes the main variables. Because outcome availability varies across operators and years, the samples for fleet size, employment, and total cost differ. Baseline estimates further restrict exposed operators to cities with at least 40,000 residents, as described in the main text.

E.1 Curvature of Fleet Variety and the Instrument

Our baseline specification makes two curvature choices. The Hill-number order q determines how rare and prevalent brands contribute to measured fleet variety, while the power p on peer inflows determines how strongly the instrument weights large purchasing episodes. In the paper, we set $q = 2$ and $p = 2$: the former gives the inverse HHI, while the latter emphasizes large renewal episodes. Values $q \leq 1$ are inconsistent with the model's mechanism: at $q = 1$, composition does not affect productive capacity, while $q < 1$ reverses its prediction. We nevertheless include these values as sensitivity checks. For $q = 1$, we use the limiting Hill number $x_1 = \exp\{-\sum_c s_c \log s_c\}$.

We assess sensitivity by varying q and p jointly, reconstructing both the endogenous variable and instrument, and reestimating the fleet-level specification. Because the scale of the Hill number varies with q, we report the effect of a typical observed variety-reducing change in each measure. Table F.4 presents the results.

Strong first stages concentrate on the convex side of the grid, where large renewal episodes receive greater weight. Where identification is sufficiently strong, a typical variety-reducing

Table E.4: Effects of Fleet Variety on Operator Performance: Varying Curvature Parameters

q – Hill-number order	p – power applied to peer inflows: 1/4	1/2	1	2	3
			panel (a): fleet size		
1/4	.001 (.007) [.685]	−.015 (.029) [.149]	−.007 (.004)* [3.399]	−.006 (.003)** [4.389]	−.007 (.005) [2.224]
1/2	.001 (.007) [.739]	−.016 (.029) [.143]	−.006 (.003)** [4.118]	−.006 (.003)** [7.425]	−.006 (.003)** [4.610]
1	.001 (.006) [1.111]	−.035 (.139) [.030]	−.007 (.004)* [3.783]	−.006 (.002)*** [12.863]	−.005 (.002)** [11.780]
2	.001 (.005) [1.298]	.061 (.490) [.011]	−.009 (.006) [1.960]	−.006 (.003)** [11.260]	−.005 (.002)** [19.963]
3	.001 (.006) [1.135]	.033 (.149) [.036]	−.011 (.008) [1.205]	−.007 (.003)** [8.018]	−.006 (.002)*** [16.779]
			panel (b): employment		
1/4	.043 (.036) [.891]	−.087 (.244) [.075]	−.004 (.009) [3.144]	.002 (.006) [4.340]	.003 (.006) [2.228]
1/2	.042 (.034) [1.005]	−.100 (.303) [.062]	−.004 (.009) [3.883]	.002 (.005) [7.666]	.003 (.005) [4.872]
1	.035 (.024) [1.514]	−.953 (12.079) [0]	−.004 (.010) [3.441]	.002 (.005) [13.325]	.002 (.005) [12.933]
2	.032 (.021) [1.750]	.100 (.279) [.078]	−.006 (.014) [1.597]	.002 (.006) [10.345]	.002 (.004) [18.978]
3	.035 (.023) [1.541]	.077 (.166) [.126]	−.007 (.018) [.933]	.002 (.006) [7.201]	.002 (.005) [15.248]
			panel (c): total operating cost		
1/4	−.002 (.066) [.183]	−.018 (.042) [.565]	−.019 (.015) [4.524]	−.027 (.013)** [4.829]	−.038 (.020)* [2.434]
1/2	−.002 (.056) [.249]	−.020 (.044) [.506]	−.018 (.014) [5.400]	−.024 (.011)** [8.132]	−.031 (.012)*** [5.028]
1	−.001 (.038) [.534]	−.029 (.067) [.229]	−.020 (.014) [5.036]	−.023 (.009)** [13.096]	−.027 (.008)*** [11.293]
2	−.001 (.033) [.720]	−.068 (.214) [.042]	−.024 (.016) [3.160]	−.025 (.008)*** [12.479]	−.027 (.006)*** [16.589]
3	−.001 (.036) [.647]	−.131 (.737) [.011]	−.029 (.017)* [2.229]	−.027 (.008)*** [9.843]	−.028 (.006)*** [14.738]

Notes. The table reports fleet-level shift-share IV estimates across values of the Hill-number order q and the power p applied to peer inflows. *Estimand.* We rescale each estimate to the effect of a typical variety-reducing change, $\exp\{\alpha^F \overline{\Delta x}\} - 1$, where $\overline{\Delta x} = \text{mean}(\Delta x_{jt} \mid \Delta x_{jt} < 0)$. The baseline specification uses $q = 2$ and $p = 2$. All specifications condition on the change in exposed-operator output and exclude exposed operators serving cities with fewer than 40,000 residents. We cluster standard errors by peer operator. Standard errors from the shift-level regression, obtained by the delta method, appear in parentheses; effective F-statistics appear in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

change produces similar declines in fleet size and total cost, while employment remains close to zero. By contrast, specifications with $q \leq 1$ have little identifying power. Thus, the procurement variation exploited by the instrument primarily shifts measures with $q > 1$, the same region in which the model predicts that greater variety raises capital requirements.

E.2 Validity Tests

The fleet-level design would fail if shocks that move purchasing at operator j's peers also affect j's performance through channels other than fleet composition. We examine three concerns: direct effects of fleet renewal, broader shocks that jointly affect purchasing and performance, and pre-existing performance dynamics correlated with the instrument.

Purchase volume and common purchasing variation. Table F.5 addresses the first two concerns. We first control for the exposed operator's own inflows and outflows to account for direct effects of renewal volume. Because these variables enter fleet size mechanically, attenuation of the fleet-size coefficient is difficult to interpret. The cost effect, however, remains essentially unchanged, suggesting that renewal volume does not drive the baseline result.

Second, we examine whether common or persistent purchasing patterns across operators drive the results. Year effects already remove purchasing variation common to all operators in a given year. This matters quantitatively: Table F.8 shows that only about one third of the instrument's variation remains after removing year effects, and this residual variation identifies α^F. We then remove additional persistent components across peer operators and exposed operators (Table F.5). The fleet-size and cost estimates remain similar or increase, although the first stage weakens sharply as these exercises remove much of the remaining systematic purchasing variation.

Persistence. We use the timing of the identifying variation to test for slow-moving confounding. If a persistent common factor drives both fleet composition and performance, associations should also appear when we shift either the instrument or the outcome away from year t.

Table F.6 reports two complementary tests. Panel (a) replaces z_{jt} with $z_{j,t+k}$ while keeping the outcome, endogenous variable, and controls at t; future peer purchases should not predict current performance. Panel (b) keeps the contemporaneous instrument and fleet-variety change but replaces Δy_{jt} with $\Delta y_{j,t+k}$; negative horizons therefore test for pre-trends. Neither exercise indicates persistent confounding for fleet size or total cost. Employment exhibits some pre-trends, so we interpret its estimates more cautiously.

Table E.5: Fleet-Level Results: Purchase Volume and Persistent Purchasing Components

	Fleet-level outcomes		
	# buses	# employees	total cost
main	.030 (.013)** [11.260]	−.011 (.029) [10.345]	.131 (.042)*** [12.479]
own purchase volume (control for inflow and outflow at the unit level)	.004 (.007) [10.231]	−.023 (.031) [9.288]	.129 (.047)*** [11.125]
peer heterogeneity (shifts net of peer-operator effects)	.064 (.039) [1.925]	.094 (.086) [1.616]	.212 (.101)** [2.547]
operator heterogeneity (include exposed-operator FEs)	.053 (.030)* [2.523]	.101 (.079) [2.080]	.179 (.086)** [3.125]

Notes. The table reports fleet-level shift-share IV estimates of the effect of brand variety on operator performance under alternative controls for purchasing volume and persistent purchasing components. *Estimand.* Entries report α^F, the effect of the order-two Hill number over bus brands. *Specifications.* `Main` reproduces the baseline result. `Own purchase volume` adds the exposed operator's own inflow and outflow; these variables are consequences of the procurement wave exploited by the instrument and enter fleet size directly. `Peer heterogeneity` residualizes each shift on peer-operator fixed effects. `Operator heterogeneity` adds exposed-operator fixed effects to the first-difference regression. All specifications condition on the change in exposed-operator output and exclude exposed operators serving cities with fewer than 40,000 residents. We cluster standard errors by peer operator. Standard errors are reported in parentheses and effective F-statistics in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table E.6: Fleet-Level Timing Falsification Tests

	Fleet-level outcomes		
Horizon k	# buses	# employees	total cost
	panel (a): leads of the shift		
0	.030 (.013)** [11.260]	−.011 (.029) [10.345]	.131 (.042)*** [12.479]
1	.019 (.016) [8.700]	−.081 (.056) [6.866]	.043 (.064) [6.982]
2	−.016 (.012) [9.686]	−.047 (.031) [9.230]	−.032 (.064) [6.628]
3	.012 (.040) [.674]	−.570 (.654) [.536]	−.115 (.246) [.619]
	panel (b): lags of the outcome		
-3	.013 (.016) [11.727]	−.153 (.042)*** [10.574]	−.046 (.066) [9.869]
-2	−.008 (.014) [8.989]	−.087 (.037)** [8.557]	−.046 (.067) [7.338]
-1	.022 (.017) [12.391]	−.070 (.041)* [12.623]	.002 (.063) [12.088]
0	.030 (.013)** [11.260]	−.011 (.029) [10.345]	.131 (.042)*** [12.479]

Notes. The table reports timing falsification tests for the fleet-level shift-share IV specification. *Panel A.* Row k replaces the instrument z_{jt} with $z_{j,t+k}$, while keeping the outcome, endogenous variable, and controls at t. Future peer purchases should not predict current operator performance. *Panel B.* Row k replaces the outcome Δy_{jt} with $\Delta y_{j,t+k}$, while keeping the endogenous variable, instrument, and controls at t. Current changes in fleet composition should not predict past performance. In both panels, row 0 reproduces the baseline estimate. All specifications use the order-two Hill number over bus brands, condition on the change in exposed-operator output, and exclude exposed operators serving cities with fewer than 40,000 residents. We cluster standard errors by peer operator. Standard errors from the shift-level regression are reported in parentheses and effective F-statistics in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table E.7: Fleet-Level Balance Test: Predetermined Operator Characteristics

	Predetermined operator characteristics			
Lag k	# buses	# employees	total cost	mean age
1	.016 (.128)	−.785 (.630)	−1.153 (.698)*	17.266 (13.086)
2	−.135 (.108)	−1.030 (.588)*	−.918 (.830)	13.044 (13.383)
3	.128 (.114)	−.101 (.552)	−.045 (.604)	−1.926 (13.025)

Notes. The table tests whether the instrument predicts predetermined operator characteristics. Row k uses the characteristic measured at $t-k$ and the contemporaneous instrument. We estimate the regressions in the shift-level representation, using exposure-weighted aggregates and operator fixed effects. All specifications exclude exposed operators serving cities with fewer than 40,000 residents. We cluster standard errors by peer operator and report them in parentheses. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Selection on levels. We next test whether the identifying variation is systematically directed toward operators with different predetermined characteristics. Table F.7 regresses lagged operator characteristics on the contemporaneous instrument, absorbing operator fixed effects so that permanent differences across operators do not drive the test. We implement the regressions at the shift level using the exposure-weighted aggregates implied by the shift-share representation. We find no systematic relationship between the instrument and operators' predetermined characteristics.

The validity tests imply that any remaining confounder would need to be transitory and contemporaneous, affect similar operators, influence performance through a channel other than fleet composition, and be unrelated to predetermined operator scale. These conditions substantially narrow, but do not eliminate, the scope for confounding.

E.3 Relevance and Sources of Identifying Variation

Relevance requires two links: peer inflows must predict an operator's own inflows, and own inflows must change fleet variety. We examine both links and then assess how concentrated the resulting identifying variation is.

The mechanism (I): common timing in fleet renewal. The first link requires purchasing to comove across operators, for example because of shared funding cycles or procurement conditions. At the same time, individual purchases contain substantial idiosyncratic variation. Averaging activity across similar peers filters out part of this idiosyncratic component and strengthens the common timing signal.

Table E.8: Variation Net of Year Effects

	shift	instrument
% of variance net of year effects	94.0%	35.4%

Notes. The table reports the share of variation remaining after absorbing year effects. *Shift* refers to $(\Delta^{+}_{\text{tot}} s_{j't})^2$ for peer operator j'; *instrument* refers to the exposure-weighted average z_{jt} for exposed operator j. The sample matches the baseline fleet-level specification.

Table E.9: Concentration in Exposure Weights

	Mean	
peers with positive weight (out of 123)	62.8	
effective number of peers $\left(1/\sum'_j \omega^2_{jj'}\right)$	52.6	
	similarity weights	uniform weights
cumulative share of top-5 peers	.130	.087
cumulative share of top-10 peers	.254	.173
cumulative share of top-20 peers	.485	.347

Notes. The table summarizes how broadly exposure is distributed across peers, averaged over exposed operators. Cumulative-share rows rank peers by $\omega_{jj'}$ and sum the largest n weights. *Uniform weights* assigns equal weight to all peers with positive $\omega_{jj'}$. The sample matches the baseline fleet-level specification.

Table F.8 supports this interpretation. After absorbing year effects, about 94% of the variation in an individual peer shift remains, compared with only about one third for the exposure-weighted instrument. Peer averaging therefore greatly increases the relative importance of common purchasing movements. The remaining variation, after removing year effects, identifies α^F.

For this averaging to filter idiosyncratic purchasing effectively, the instrument must draw on a broad set of peers rather than being dominated by the purchases of a few operators. Table F.9 shows that the average operator assigns positive weight to roughly half of all peers, with an effective number of about 53. The 20 largest weights account for about 49% of exposure, compared with less than 35% under uniform weighting. Similarity weighting therefore favors comparable operators while retaining substantial averaging across peers.

The mechanism (II): own inflows change fleet variety. The second link in the first stage requires an operator's own purchases to generate meaningful changes in fleet variety. This link is stronger when renewal occurs in large, discrete batches: a large purchase can materially change the relative prevalence of different classes in the fleet, whereas small additions have little effect on fleet shares. Table F.10 shows that renewal is indeed concentrated in relatively few large purchasing episodes. For a typical operator, a major purchasing year

Table E.10: Lumpy inflows: size and timing of new fleet arrivals.

	p25	p50	p75	mean
largest annual inflow $\left(max_t\ I_{jt}\right)$	.178	.229	.296	.245
largest share of the shift in that year $\left(max_t\ I_{jt}^2 / \sum_t I_{jt}^2\right)$	.234	.319	.499	.369
effective number of inflow years, out of 16 $\left((\sum_t I_{jt}^2)^2 / \sum_t I_{jt}^4\right)$	3.273	5.107	6.691	5.282

Notes. The table reports quantiles and means across operators for three measures of the size and concentration of fleet renewal. *Definitions.* $I_{jt} = \Delta^{+}_{\text{tot}} s_{jt}$ is annual inflow as a share of the fleet; the shift is I_{jt}^2. The effective number of inflow years is the order-two Hill number of an operator's shares of the shift across the sixteen years. *Sample.* Main fleet-level estimation sample.

Table E.11: Composition of new fleet arrivals by size of the inflow.

	inflow size	brand	generation	engine brand	size	drive
tercile 1	.025	1.324	1.435	1.414	1.443	1.173
tercile 2	.069	1.600	1.892	1.703	1.746	1.224
tercile 3	.173	1.797	2.087	1.911	1.875	1.290

Notes. The table relates the size of annual inflows to the composition of arriving buses. Operator-years are split into terciles of annual inflow $I_{jt} = \Delta^{+}_{\text{tot}} s_{jt}$. `inflow size` reports the mean inflow within each tercile; the remaining columns report the mean number of classes among arriving buses by characteristic. *Sample.* Main fleet-level estimation sample.

accounts for a substantial fraction of the fleet, while only a limited number of years account for most renewal activity.

For large renewal episodes to change fleet variety, the inflows must themselves be heterogeneous. Table F.11 shows that larger inflows contain more brands, generations, engine brands, and bus sizes. Together with the lumpiness of renewal, this helps explain why large purchases generate larger changes in fleet variety and why convex transformations of inflows strengthen the first stage (Table F.4).

Concentration of identifying variation. The number of observed peer-year shifts may overstate the effective identifying variation if some shifts contribute disproportionately to the IV estimate. We therefore use the Rotemberg-weight decomposition described in Section D to assess concentration across individual shifts, peer operators, and years, using absolute weights because Rotemberg weights can be negative. Table F.12 shows that identifying variation is diffuse across shifts and peer operators but more concentrated across years. The three largest-weight years coincide with major waves of EU-funded procurement.

Table E.12: Concentration in the Rotemberg weights.

	top-5	top-20	top-50
shifts (out of 1968): % of absolute Rotemberg weights (cumulative)	8.596	17.713	27.0
operators (out of 123): % of absolute Rotemberg weights (cumulative)	11.733	31.0	59.0
	1-st	2-nd	3-rd
years (out of 16)	2011	2018	2012
% of absolute Rotemberg weights (absolute)	17.479	11.810	11.321
effective # of shifts (based on absolute Rotemberg weights)		322	

Notes. The table summarizes the concentration of identifying variation across individual shifts, peer operators, and years using absolute Rotemberg weights. *Definitions.* Absolute weights are $|\lambda_n| / \sum_n |\lambda_n|$. Shift and operator rows report cumulative shares for the n largest weights; year rows report the three largest years individually. The effective number of shifts is the reciprocal HHI of the absolute-weight shares. *Sample.* Main fleet-level estimation sample.

E.4 Alternative Measures of Fleet Variety

Our baseline defines classes by bus brand, a dimension well suited to the variation generated by our instrument: common renewal episodes can change suppliers while leaving characteristics tied to bus purpose largely unchanged. Table F.13, panel (a), reconstructs the order-two Hill number using generation, engine brand, size, and drive. First stages are generally weaker than for brand, indicating that common renewal generates more variation in supplier identity than in other dimensions of fleet composition.

Panel (b) instead interacts these characteristics with brand, retaining manufacturer differences within narrower bus types. First stages strengthen, particularly for size and drive, once brand remains part of the class definition. Where identification is sufficiently strong, the estimates are broadly consistent with the baseline, showing that the estimated cost of variety persists when manufacturers are distinguished among buses serving similar purposes.

E.5 Additional outcomes, samples, and output normalization

We conclude with three extensions of the fleet-level results.

Employment by staff category. Total employment may mask heterogeneous responses across worker types. Drivers account for more than half of employment, while mechanics—whose workload fleet variety would most plausibly affect—represent less than 15% (Table F.3). Table F.14 therefore estimates effects separately for drivers, mechanics, and other staff. We find no systematic effect in any category, although the estimates are imprecise.

Table E.13: Effects of fleet variety on operator performance: alternative class definitions.

	panel (a): classes defined by characteristic				
	brand	generation	engine brand	size	drive
	# buses				
estimate	−.006 (.003)**	−.013 (.005)**	−.021 (.020)	−.006 (.003)*	−.277 (2.358)
Effective F	11.260	1.816	.477	6.614	.005
	# employees				
estimate	.002 (.006)	.005 (.013)	.008 (.022)	.002 (.005)	−.175 (2.827)
Effective F	10.345	1.855	.426	6.533	.002
	total cost				
estimate	−.025 (.008)***	−.057 (.032)*	−.063 (.038)*	−.024 (.008)***	584 (164152)
Effective F	12.479	2.053	1.074	10.384	.000
	panel (b): classes defined by brand × characteristic				
	brand	brand × generation	brand × engine brand	brand × size	brand × drive
	# buses				
estimate	−.006 (.003)**	−.013 (.005)**	−.010 (.004)**	−.007 (.003)**	−.006 (.003)**
Effective F	11.260	1.801	2.820	7.147	11.070
	# employees				
estimate	.002 (.006)	.005 (.013)	.004 (.010)	.003 (.007)	.002 (.006)
Effective F	10.345	1.837	2.378	6.891	10.785
	total cost				
estimate	−.025 (.008)***	−.057 (.032)*	−.037 (.014)***	−.026 (.009)***	−.027 (.009)***
Effective F	12.479	2.038	4.061	11.873	11.191

Notes. The table presents fleet-level shift-share IV estimates under alternative definitions of the classes over which fleet variety is measured. Panel (a) defines classes by each characteristic separately; panel (b) interacts each characteristic with brand. *Estimand.* Estimates are rescaled to the effect of a typical variety-reducing change, $\exp\{\alpha^F \overline{\Delta x}\} - 1$, where $\overline{\Delta x} = \text{mean}(\Delta x_{jt} \mid \Delta x_{jt} < 0)$ for the corresponding Hill number. *Specification.* All regressions condition on the change in exposed-operator output. *Sample.* Exposed operators serve cities with at least 40,000 residents. *Clustering.* Standard errors are clustered by peer operator. *Presentation.* Delta-method standard errors are reported in parentheses below the estimates and effective F-statistics in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table E.14: Effects of fleet variety on employment by staff category.

	employment outcomes			
	total employment	drivers	mechanics	white collar
brand	−.011 (.029)	.000 (.028)	−.142 (.092)	.047 (.085)
Effective F	10.345	10.554	10.737	10.266

Notes. The table presents fleet-level shift-share IV estimates of the effect of brand variety on employment by staff category. *Estimand.* Estimates report α^F, the effect of a one-unit increase in the order-two Hill number over bus brands. *Specification.* All regressions condition on the change in exposed-operator output. *Sample.* Exposed operators serve cities with at least 40,000 residents. *Clustering.* Standard errors are clustered by peer operator. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Operators below the size threshold. The baseline excludes operators serving cities with fewer than 40,000 residents. Table F.15 includes smaller operators, whose fleet size adjusts

Table E.15: Effects of fleet variety without the minimum city-size restriction.

	fleet-level outcomes		
	# buses	# employees	total cost
brand	.027 (.022)	−.016 (.069)	.176 (.071)**
Effective F	8.104	5.996	9.987

Notes. The table presents fleet-level shift-share IV estimates of the effect of brand variety on operator performance without the minimum city-size restriction. *Estimand.* Estimates report α^F, the effect of a one-unit increase in the order-two Hill number over bus brands. *Specification.* All regressions condition on the change in exposed-operator output. *Clustering.* Standard errors are clustered by peer operator. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table E.16: Effects of fleet variety with alternative treatments of output.

	outcomes per unit of output			total output
	# buses	# employees	total cost	
brand	.103 (.064)	.044 (.059)	.154 (.044)***	−.077 (.064)
Effective F	11.031	10.111	12.334	11.031

Notes. The table presents fleet-level shift-share IV estimates under alternative treatments of output. The first three columns express fleet size, employment, and total cost per unit of output; the final column treats total output as the outcome. *Estimand.* Estimates report α^F, the effect of a one-unit increase in the order-two Hill number over bus brands. *Sample.* Exposed operators serve cities with at least 40,000 residents. *Clustering.* Standard errors are clustered by peer operator. *Presentation.* Standard errors from the shift-level regression are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

in coarser increments because each bus represents a larger share of the capital stock. In the broader sample, the fleet-size effect is smaller and less precise, while the cost effect is larger. This pattern is consistent with smaller operators having less scope to adjust fleet size and absorbing more of the effect through operating costs.

Alternative treatment of output. The baseline conditions on total output so that coefficients capture the resources required to provide a given level of service. Table F.16 instead expresses fleet size, employment, and total cost per unit of output and also treats output itself as an outcome. The cost effect remains similar, while the fleet-size effect is less precisely estimated and employment remains close to zero. We also find no evidence that instrumented fleet variety affects output. Together, these results support the baseline interpretation that fleet variety affects the resources required to provide a given level of service rather than the amount of service provided.

Table F.17: Bus-level estimates under alternative curvature parameters.

r: power applied to own-brand prevalence	p: power applied to peer inflows				
	1/4	1/2	1	2	3
1/4	.021 (.008)*** [32.7]	.021 (.008)*** [33]	.021 (.007)*** [27.5]	.024 (.008)*** [13.158]	.031 (.010)*** [8.214]
1/2	.023 (.009)** [33.7]	.023 (.009)** [34.6]	.023 (.008)*** [29.4]	.027 (.009)*** [14.087]	.034 (.011)*** [8.765]
1	.027 (.011)** [34]	.027 (.010)*** [36.1]	.027 (.009)*** [32.5]	.031 (.010)*** [15.573]	.040 (.014)*** [9.532]
2	.037 (.015)** [31.3]	.036 (.014)** [35.4]	.036 (.013)*** [35.8]	.041 (.014)*** [17.585]	.053 (.019)*** [10.238]
3	.048 (.020)** [26.4]	.047 (.018)*** [31.3]	.046 (.016)*** [35.5]	.052 (.018)*** [18.560]	.068 (.023)*** [10.264]

Notes. The table presents bus-level shift-share IV estimates across curvature parameters $r \equiv q-1$ and p. *Estimand.* Estimates are rescaled to the effect of a typical purchase that raises a brand's share, $\exp\{\alpha^B \overline{\Delta s_c}\}-1$, where $\overline{\Delta s_c} = \text{mean}(\Delta s_{cjt} \mid \Delta s_{cjt} > 0)$. The baseline is $r = p = 1$. *Instrument.* All specifications use the signed instrument. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Delta-method standard errors are reported in parentheses and effective F-statistics in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

F Bus-level design: supporting evidence

This section provides supporting evidence for the bus-level design, covering first-stage relevance, validity, alternative class definitions, and heterogeneity in the utilization effect.

F.1 Curvature of brand prevalence and the instrument

Our bus-level specification involves two curvature choices. Let $r \equiv q-1$, so productive capacity depends on class prevalence through s_c^r. The parameter r governs the curvature of brand prevalence, while p governs the power applied to brand-specific peer inflows. The baseline sets $r = p = 1$, leaving both the endogenous variable and shifts linear.

Table G.17 varies r and p jointly and reports effects rescaled to a typical purchase that increases a brand's share. First stages are strongest for p between 1/2 and 1, supporting the linear baseline. First-stage strength does not discipline r: estimated effects generally increase with curvature but remain positive throughout the grid. Thus, magnitudes depend on the curvature imposed on brand prevalence, while the conclusion that greater own-brand prevalence raises utilization is robust.

F.2 Relevance and Sources of Identifying Variation

Signed shifts. At the bus level, relevance requires peer purchases of brand c to predict changes in its prevalence at the exposed operator. The direction of this relationship can differ across brand-years, however. We classify a brand-year as a *new-bus regime* when acquisitions are at least as widespread across operators in the primary as in the secondary market, and as a *second-hand regime* otherwise. Peer new-bus inflows should predict higher own-brand prevalence in the former regime but can predict lower prevalence in the latter. An unsigned shift therefore combines first stages of opposite signs.

Table G.18 supports this mechanism. Panel (a) shows substantial variation across brands in acquisition regimes, while panel (b) shows that peer new-bus inflows positively predict own-brand prevalence in new-bus regimes and negatively predict it in second-hand regimes. Signing the shifts by d_{ct} aligns these relationships and substantially strengthens the first stage. Panel (c) further shows that signing nearly eliminates negative Rotemberg weight. The unsigned instrument yields the same qualitative utilization effect but a substantially weaker first stage. We therefore use the signed instrument in the baseline specification.

Concentration of identifying variation. The large number of peer–brand–year shifts may overstate the effective identifying variation if some shifts contribute disproportionately to the IV estimate. Table G.19 uses the Rotemberg-weight decomposition from Section D to summarize concentration across individual shifts, brand–years, peer operators, years, and brands.

Identifying variation is relatively diffuse across individual shifts and peer operators but concentrated across brand–years and, especially, brands. This pattern is consistent with manufacturer activity generating common brand-specific movements across many operators: numerous peer–brand–year shifts can therefore reflect the same underlying change in a manufacturer's national market presence.

Is the effect driven by one manufacturer? The concentration of identifying variation across brands raises the possibility that the baseline result is driven by a single manufacturer. Table G.20 addresses this by estimating separate effects for the dominant manufacturer, two large legacy manufacturers, and the remaining brands, with group-specific first stages.

Table F.18: Why the shift is signed: acquisition regimes and opposite first stages.

	all	Solaris	MAN	Mercedes	others
panel (a): brand possession and expansion					
N/share of bus-year observations	55128	.470	.164	.159	.208
share of brand-years experiencing brand-new expansion	.235	1	.111	.111	.219
brand's share in brand-new inflows	1	.476	.124	.140	.261
brand's share in second-hand inflows	1	.170	.437	.208	.185

	unsigned	signed	N shifts $\neq 0$
panel (b): first-stage estimates			
all observations	.370 (.135)***	.619 (.110)***	761
only brand-years with more brand-new inflows	.578 (.121)***		624
only brand-years with more second-hand inflows	−.854 (.177)***		137

	unsigned					signed				
panel (c): Rotemberg weights	all	Solaris	MAN	Mercedes	others	all	Solaris	MAN	Mercedes	others
$\sum$ positive	1.391	1.371	.001	.012	.007	1.033	.820	.128	.080	.005
$\sum$ negative	-.391	-.002	-.223	-.141	-.025	-.033	-.001	-.006	-.012	-.014
$\sum$ all	1	1.369	-.223	-.128	-.018	1	.819	.121	.068	-.008

Notes. The table provides evidence motivating the signed instrument. Panel (a) summarizes acquisition patterns for the three largest brands in new- and second-hand-bus markets; panel (b) reports first-stage estimates for the unsigned and signed instruments; panel (c) summarizes the corresponding Rotemberg weights. *Regimes.* A brand-year is classified as a new-bus regime when acquisitions of brand c are at least as widespread across operators in the primary as in the secondary market, and as a second-hand regime otherwise. *Sample.* Buses operated by the Warsaw operator are excluded. *Clustering.* First-stage standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table F.19: Concentration of identifying variation in the bus-level design.

	nonzero shifts	effective N	cumulative share of $\lvert\lambda\rvert$, %		
			top 5	top 20	top 50
peer × brand × year	761	185	7.8	21.7	41.6
brand × year	103	11.679	58.9	95.7	100
peer operator j'	119	55.7	18.3	50.0	79.7
year t	9	7.205	75.9	100	100
brand c	19	1.620	99.6	100	100

Notes. The table summarizes concentration in the absolute Rotemberg weights of the signed bus-level specification. *Definitions.* Rows aggregate peer operator × brand × year shifts along the indicated dimension. Effective N is the reciprocal HHI of the corresponding $\lvert\lambda_n\rvert$ shares; the remaining columns report the cumulative share of absolute weight held by the largest units. *Sample.* Baseline bus-level estimation sample.

The dominant manufacturer and the two legacy manufacturers are separately identified and yield effects of the same sign and similar magnitude; we cannot reject equality across groups. The remaining brands have a weak first stage because their peer inflows are too sporadic to predict prevalence elsewhere. The baseline effect therefore does not appear to be driven solely by the dominant manufacturer.

Table F.20: The effect of own-brand prevalence on bus utilization by brand group.

	Δ log annual mileage	
	unsigned instr.	signed instr.
Solaris	.816 (.309)***	.816 (.309)***
MAN & Mercedes	.342 (.242)	.590 (.311)*
other brands	1.536 (1.283)	3.055 (1.549)**
H_0: coefficients equal across groups	1.982 (.371)	2.472 (.291)
effective F (Solaris)	78.8	78.8
effective F (MAN & Mercedes)	23.8	21.1
effective F (other brands)	5.483	4.422

Notes. The table presents bus-level shift-share IV estimates separately for the dominant manufacturer, the two large legacy manufacturers, and the remaining brands. *Design.* The three effects are estimated jointly, interacting both own-brand prevalence and the instrument with time-invariant brand-group indicators. *Instruments.* The unsigned instrument uses peer inflows of brand-new buses of brand c; the signed instrument multiplies these inflows by d_{ct}. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table F.21: The effect of own-brand prevalence on bus utilization with Warsaw included.

	Δ log annual mileage	
	unsigned instr.	signed instr.
brand	1.026 (.582)*	.846 (.359)**
bus-years	65077	65077
exposed operators j	123	123
peer operators j'	123	123
brands c	36	36
years t	2015-2023	2015-2023
Effective F	6.752	18.270

Notes. The table presents bus-level shift-share IV estimates with Warsaw included. *Estimand.* Estimates report α^B, the effect of own-brand prevalence on annual mileage. *Instruments.* The unsigned instrument uses peer inflows of brand-new buses of brand c; the signed instrument multiplies these inflows by d_{ct}. *Sample.* Buses in their final two calendar years before leaving the fleet are excluded. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Which operators supply the variation. Warsaw provides a useful illustration of the relevance mechanism. It is by far the largest operator in the sample, with about 1,400 buses on average compared with just over 500 for the second largest, and purchases buses almost every year while relying relatively little on episodic external funding. Its procurement is therefore less concentrated in the common renewal waves that underpin our first stage. The baseline consequently excludes Warsaw despite its large number of bus-year observations.

Table G.21 includes Warsaw as a check. The estimated effect is essentially unchanged, while the first stage weakens. This is consistent with identification coming from operators whose purchases comove with common procurement waves rather than from those contributing the most observations.

Table F.22: Timing falsification tests for the bus-level design.

horizon k	unsigned instr.	signed instr.
	panel (a): lagged outcomes	
-3	−.123	−.239
	(.367)	(.188)
	[3.518]	[24.5]
-2	.073	.083
	(.530)	(.278)
	[4.742]	[35.6]
-1	.606	.261
	(.443)	(.256)
	[6.806]	[37.3]
0	.969	.755
	(.449)**	(.263)***
	[8.253]	[32.5]
	panel (b): leads of the shift	
0	.969	.755
	(.449)**	(.263)***
	[8.253]	[32.5]
1	.654	.472
	(.412)	(.280)*
	[7.678]	[20.5]
2	.274	.457
	(.432)	(.309)
	[11.212]	[36.5]
3	.024	.197
	(.463)	(.276)
	[4.806]	[14.921]

Notes. The table presents timing falsification tests for the bus-level specification. *Horizons.* In panel (a), row k replaces the outcome at t with the outcome at $t+k$, leaving the endogenous variable, instrument, and controls at t. In panel (b), row k replaces the shift at t with the shift at $t+k$, leaving the outcome, endogenous variable, and controls at t. *Instruments.* Unsigned and signed instruments are defined as in the baseline specification. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses and effective F-statistics in brackets. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

F.3 Validity

Timing. Table G.22 reports two complementary timing tests. Panel (a) replaces the outcome with changes in mileage from preceding years, leaving the endogenous variable, instrument, and controls at t. We find no relationship at any lag under either instrument, providing no evidence that the instrument selects buses already on different utilization paths.

Panel (b) instead replaces the contemporaneous shift with peer purchases from subsequent years. The estimated effect declines as the shift is moved further from t: only the adjacent lead remains marginally significant, while longer leads are smaller and insignificant despite strong first stages. Some one-year persistence is natural because we observe deliveries rather than contract dates, and deliveries from a single procurement may span adjacent calendar years. Overall, the timing is consistent with variation generated by procurement episodes rather than persistent brand–operator factors.

Selection on past utilization. The timing tests above use changes in mileage and therefore do not rule out selection on pre-existing utilization levels. Table G.23 instead relates the

Table F.23: Balance test: past utilization and the instrument.

	log annual mileage, lagged k years		
	1	2	3
unsigned instr.: coefficient	−1.420	−1.321	−.158
	(1.776)	(1.475)	(1.131)
signed instr.: coefficient	−1.375	−.640	−.128
	(.959)	(.722)	(.421)
bus-years	55128	44337	34537

Notes. The table reports balance tests relating the contemporaneous instrument to lagged levels of bus utilization. Column k uses log annual mileage at $t-k$ as the outcome. Rows correspond to the unsigned and signed instruments. *Specification.* Regressions use the shift-level representation and include bus fixed effects because the outcome is measured in levels. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

contemporaneous instrument to lagged mileage levels. Neither instrument predicts past utilization at any horizon, providing no evidence that identifying variation is concentrated on buses that were already used more or less intensively.

Geography. Because exposure weights are based on output, city population, and density, similar operators may also be geographically close. The instrument could therefore partly reflect shared regional conditions rather than national manufacturer activity. Table G.24 addresses this concern by progressively excluding nearby peers and renormalizing the remaining exposure weights.

Excluding peers within 200 km removes roughly one third of baseline exposure mass but leaves both the estimated effect and effective first stage essentially unchanged. Nearby operators therefore do not contribute disproportionately to identification. This is consistent with the first stage being driven by brand-specific movements in the national bus market rather than regional shocks shared by nearby operators.

Other brand-specific channels. Peer new-bus inflows may be correlated with other same-brand developments that affect utilization directly. Table G.25 controls for two such channels: peers' second-hand inflows of brand c, which may reflect buyer-specific conditions, and the age of peers' existing stock of that brand, which may proxy for common replacement cycles. Following Borusyak, Hull, and Jaravel (2022), we construct unit-level representations of these shift-level controls and include them when residualizing the unit-level variables.

Adding these controls modestly reduces the estimate but leaves it positive and statistically significant, with only a moderate weakening of the first stage. The result is therefore not

Table F.24: The effect of own-brand prevalence on bus utilization excluding nearby peers.

minimum peer distance	coefficient	effective F	exposure mass dropped
		unsigned instr.	
0 km	.969 (.449)**	8.253	—
100 km	.902 (.493)*	7.221	.113
200 km	.992 (.568)*	5.883	.327
		signed instr.	
0 km	.755 (.263)***	32.5	—
100 km	.669 (.276)**	33.4	.113
200 km	.670 (.292)**	31.2	.327

Notes. The table presents bus-level shift-share IV estimates after progressively excluding geographically close peers. *Specification.* Each row sets exposure weights to zero for peers within the stated distance and renormalizes the remaining weights to sum to one; operators without remaining peers are dropped. The 0 km row reproduces the baseline exposure set. *Instruments.* Unsigned and signed instruments are defined as in the baseline specification. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table F.25: The effect of own-brand prevalence on bus utilization with shift-level controls.

	Δ log annual mileage	
	unsigned instr.	signed instr.
brand	.862 (.503)*	.645 (.293)**
Effective F	7.128	28.7

Notes. The table presents bus-level shift-share IV estimates controlling for peers' second-hand inflows of brand c and the age composition of peers' existing same-brand stock. *Design.* Following Borusyak, Hull, and Jaravel (2022), unit-level representations of the shift-level controls are included when residualizing the unit-level variables. *Instruments.* Unsigned and signed instruments are defined as in the baseline specification. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

explained by correlated second-hand purchasing or common same-brand replacement cycles, although other brand-specific confounding cannot be ruled out.

F.4 Extensions

Alternative definitions of bus classes. Table G.26 examines alternative definitions of the classes over which prevalence is measured. Panel (a) defines classes separately by generation, engine brand, size, and drive. First stages are generally weaker than for brand and essentially absent for size and drive, indicating that the instrument generates substantially more variation in manufacturer-related composition than in characteristics tied to bus purpose. The generation estimate is negative, but its first stage is too weak to interpret.

Table F.26: The effect of own-class prevalence on bus utilization under alternative class definitions.

	panel (a): classes defined by characteristic				
	brand	generation	engine brand	size	drive
estimate	.027 (.009)***	−.064 (.032)**	.009 (.009)	1.048 (173)	−.161 (.681)
Effective F	32.5	3.619	12.324	.000	.047
	panel (b): classes defined by brand × characteristic				
	brand	brand × generation	brand × engine brand	brand × size	brand × drive
estimate	.027 (.009)***	−.064 (.032)**	.017 (.012)	.032 (.010)***	.020 (.012)
Effective F	32.5	3.620	9.286	16.641	13.861

Notes. The table presents bus-level shift-share IV estimates under alternative definitions of the class over which prevalence is measured. Panel (a) defines classes by each characteristic separately; panel (b) interacts each characteristic with brand. *Estimand.* Estimates are rescaled to the effect of a typical purchase that raises class c's share, $\exp\{\alpha^B \overline{\Delta s_c}\} - 1$, where $\overline{\Delta s_c} = \text{mean}(\Delta s_{cjt} \mid \Delta s_{cjt} > 0)$. *Instrument.* All specifications use the signed instrument. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and class × year. *Presentation.* Delta-method standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Panel (b) interacts these characteristics with brand. Once manufacturer identity is retained, first stages strengthen substantially, particularly for brand × size and brand × drive. Where identification is sufficiently strong, the estimates are broadly consistent with the baseline; in particular, buses in more prevalent brand × size classes are utilized more intensively. Thus, as at the fleet level, the design primarily identifies variation in manufacturer composition, and the utilization effect persists when manufacturers are distinguished within narrower groups of buses serving similar purposes.

Table F.27: The effect of own-brand prevalence on bus utilization by bus age.

	Δ log annual mileage	
	unsigned instr.	signed instr.
0-4	−.237 (1.950)	.143 (1.141)
5-14	1.413 (.555)**	.891 (.372)**
15+	1.221 (1.282)	1.288 (.278)***
H_0: coefficients equal across groups	.664 (.718)	2.233 (.327)
effective F (0-4)	3.512	6.866
effective F (5-14)	7.540	32.5
effective F (15+)	8.393	32.0

Notes. The table presents bus-level shift-share IV estimates separately for buses aged 0–4, 5–14, and 15 years or more. *Design.* Effects are estimated jointly, interacting own-brand prevalence and the instrument with age-group indicators. *Instruments.* Unsigned and signed instruments are defined as in the baseline specification. *Sample.* Baseline bus-level estimation sample. *Clustering.* Standard errors are two-way clustered by peer operator and brand × year. *Presentation.* Standard errors are reported in parentheses below estimates and p-values in parentheses below test statistics. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Heterogeneity by bus age. Figure 3 suggests that the utilization gap between more and less prevalent brands emerges only after the first few years of service, consistent with maintenance, parts replacement, and accumulated manufacturer-specific experience becoming more important as buses age. Table G.27 estimates the effect separately for buses aged 0–4, 5–14, and 15 years or more. The effect is close to zero for the youngest buses, for which the first stage is also weaker, and positive and significant for both older groups. Point estimates increase with age, although we do not have enough power to reject equality across groups.

G Decomposition of Changes in Brand Shares

We derive the decomposition used for *accumulation or dilution* and *shares or counts* exercises:

$$\Delta s_{cjt} = \underbrace{\frac{1}{\bar{N}_{jt}}\Delta n_{cjt}}_{\text{accumulation}} - \underbrace{\frac{\bar{s}_{cjt}}{\bar{N}_{jt}}\Delta N_{jt}}_{\text{dilution}},$$

We begin deriving the midpoint product rule for two sequences. For any two sequences a_t, b_t:

$$\Delta(a_t b_t) \;=\; \bar{a}\,\Delta b \;+\; \bar{b}\,\Delta a, \qquad \bar{a} \equiv \tfrac{1}{2}(a_t + a_{t-1}), \quad \bar{b} \equiv \tfrac{1}{2}(b_t + b_{t-1}).$$

To see that this holds, expand both terms and collect:

$$\begin{aligned}
\bar{a}\,\Delta b + \bar{b}\,\Delta a &= \tfrac{1}{2}(a_t + a_{t-1})(b_t - b_{t-1}) + \tfrac{1}{2}(b_t + b_{t-1})(a_t - a_{t-1}) \\
&= \tfrac{1}{2}\big[a_t b_t - a_t b_{t-1} + a_{t-1} b_t - a_{t-1} b_{t-1}\big] + \tfrac{1}{2}\big[a_t b_t + a_t b_{t-1} - a_{t-1} b_t - a_{t-1} b_{t-1}\big] \\
&= a_t b_t - a_{t-1} b_{t-1} \;=\; \Delta(a_t b_t).
\end{aligned}$$

Now, since $s_{cjt} = \frac{n_{cjt}}{N_{jt}}$, we can derive $n_{cjt} = s_{cjt} N_{jt}$ and apply the lemma above:

$$\Delta n_{cjt} = \bar{N}_{jt} \Delta s_{cjt} + \bar{s}_{cjt} \Delta N_{jt}$$

where $\bar{N}_{jt} = \frac{N_{jt} + N_{j,t-1}}{2}$ and $\bar{s}_{cjt} = \frac{s_{cjt} + s_{cj,t-1}}{2}$. Rearranging yields:

$$\Delta s_{cjt} = \frac{\Delta n_{cjt}}{\bar{N}_{jt}} - \bar{s}_{cjt} \frac{\Delta N_{jt}}{\bar{N}_{jt}}$$

H The VÖV Standardbus: Germany's Experiment with Standardizing City Buses

From the late 1960s through the 1990s, West Germany implemented one of the most extensive real-world initiatives in publicly coordinated product standardization: the *Standard-Linienbus*, that is, the standard city bus. Implemented by the Verband öffentlicher Verkehrsbetriebe (VÖV), a West German industry association representing public bus operators, the initiative sought to reduce operating costs and improve system reliability by limiting unnecessary variety in urban bus fleets, while preserving manufacturer competition.

In the post-war decades, German municipal bus operators expanded rapidly and typically procured buses in small batches from multiple manufacturers. Although these vehicles performed essentially identical tasks, they differed in construction, interior layouts, driver controls, and other components. As emphasized in this paper, such capital heterogeneity can generate operational frictions that raise life-cycle costs: depots must hold larger inventories of non-interchangeable spare parts, maintenance staff require manufacturer-specific skills, and drivers may require additional training. Fleet variety thus may reduce effective capital utilization even when technological differences across vehicles are modest.

By the late 1950s, large bus operators, including most prominently Hamburger Hochbahn, began advocating for a unified concept of a city bus as a way to mitigate these inefficiencies. The idea gained national traction when the VÖV formed a dedicated working group in the mid-1960s to define a common technical specification for city buses. Importantly, the objective was not to create a single centrally produced vehicle, but to standardize interfaces and layouts that mattered most for operations, maintenance, and labor allocation.

The VÖV pursued a distinctive governance model best described as *standardization by recommendation*. Rather than mandating a single design, the VÖV issued a detailed type recommendation that manufacturers could adopt voluntarily. In economic terms, the VÖV coordinated demand for new buses across a large share of the national bus fleet. This coordination gave operators greater collective influence over vehicle design and allowed them to address compatibility problems that would have been difficult for individual operators to resolve independently.

The standardization effort focused on dimensions of variety with high organizational costs. Standardized elements included vehicle dimensions and door placement, window modules, lighting, electrical interfaces, interior layout (handrails, steps, and passenger flow), and a common driver workplace designed to facilitate training and vehicle switching. At the same time, elements left to competition included engines and drivetrains, exterior styling, and certain performance characteristics, preserving scope for technological differentiation among manufacturers.

As a result, the *Standard-Linienbus* delivered substantial operational uniformity while preserving industrial competition. Multiple manufacturers produced VÖV-compatible buses that behaved similarly in service and maintenance, even though they remained distinct products. Introduced in 1968, the standard diffused rapidly across West German cities and, through incremental revisions in the early 1980s, remained the dominant city-bus architecture for several decades.

Over time, however, the conditions that had sustained the VÖV standard changed. New requirements for low-floor bus designs and tighter emissions and noise regulations, together with the increasing use of modular vehicle platforms, raised the returns to manufacturer-specific architectures and reduced the feasibility of maintaining cross-brand compatibility. Institutional and procurement changes reinforced this shift: from around 1990, decentralized and competitive tendering weakened coordination across operators and eroded demand-side market power, reducing incentives to adhere to a common specification and allowing fleet variety to re-emerge.

The VÖV standard bus stands as a rare, large-scale example of coordinated standardization in fleet management. See Achilles (2008) for a detailed historical account of the VÖV *Standard-Linienbus* and its role in coordinating bus design and procurement in West Germany.