Department of Economics
ISSN number 1441-5429

The Costs of Fleet Variety

Monash Economics Working Papers no. 2026-02

Filip Premik, Dan Yu

Abstract:

We study how heterogeneity in capital inputs affects firm performance. Drawing on detailed data on municipal bus fleets in Poland, we exploit plausibly exogenous variation generated by public procurement and nationally coordinated sales behavior of bus manufacturers to identify the causal effects of variety in fleet composition across brands and other technical dimensions. More heterogeneous fleets exhibit lower vehicle utilization and, for a fixed level of output, require more units of capital and generate higher costs. Our results emphasize that the pro- ductive capacity of capital depends on its internal structure, not only on its aggregate quantity or value.

Keywords: Heterogeneous capital, capital utilization, productivity, fleet composition, orga- nization of production

JEL Classification: D24, L23, L62

The Costs of Fleet Variety

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

Abstract

We study how heterogeneity in capital inputs affects firm performance. Drawing on detailed data on municipal bus fleets in Poland, we exploit plausibly exogenous variation generated by public procurement and nationally coordinated sales behavior of bus manufacturers to identify the causal effects of variety in fleet composition across brands and other technical dimensions. More heterogeneous fleets exhibit lower vehicle utilization and, for a fixed level of output, require more units of capital and generate higher costs. Our results emphasize that the productive capacity of capital depends on its internal structure, not only on its aggregate quantity or value.

JEL codes: D24, L23, L62.

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

[1]Monash University, filip.premik@monash.edu; 2. University of Alberta, dy6@ualberta.ca. We are grateful to IGKM, the urban transport chamber of commerce in Poland, for sharing their data with us. We thank Arthur Campbell, Maxime Gravoueille, Gordon Leslie, Laura Puzzello, as well as seminar and conference participants for their helpful comments and suggestions.

1 Introduction

Fleets, which are collections of durable goods used to perform repetitive tasks, constitute the backbone of modern economies. Fleets of vehicles enable the transportation of people and goods, fleets of industrial machinery underpin farming and manufacturing, and fleets of computer systems support the daily operations of nearly every organization. Accordingly, a substantial share of the production capital input in many industries takes the form of a fleet.

Fleets are rarely homogeneous. Over time, technological progress, supplier entry, and procurement decisions lead firms to acquire different makes and models that perform the same tasks but embody distinct designs and standards. When machines serving identical functions differ in construction, they are less able to share workers, spare parts, and maintenance routines. As heterogeneity increases, incompatibilities multiply, potentially requiring more units, broader skills, and larger inventories to deliver the same output. Resulting productivity losses, which we refer to as the costs of fleet variety, represent an often-overlooked source of inefficiency in the use of production capital.

We show that fleet variety has first-order negative effects on firms' productive efficiency. Using urban bus fleets as a laboratory[1] and exploiting plausibly exogenous variation from public procurement and nationally coordinated manufacturer sales strategies, we find that higher brand concentration in a fleet improves individual bus utilization, and that more homogeneous fleets require fewer buses to deliver the same service and operate at lower unit costs.

Our findings imply that heterogeneity in capital inputs carries important consequences for the efficient allocation of resources. Fleet variety generates misallocation within and across firms: technologically identical units may differ in productivity depending on the structure of a fleet in which they operate. While reducing heterogeneity can raise efficiency, variety frictions complicate reallocation, as moving capital across firms may lower its productivity

[1] Urban buses provide a canonical setting for studying fleets. 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 quite a different purpose than the rest of the fleet). Such fleet unification may have reflected an effort to limit the costs of fleet variety.

when compatibility constraints are ignored. Moreover, when fleet variety is costly, early capital choices shape future procurement decisions, limiting dynamic allocative efficiency and potentially raising barriers to entry in upstream markets.

More broadly, our results provide micro-level evidence consistent with concerns raised in the Cambridge capital controversy (Robinson, 1953; Sraffa, 1960; Samuelson, 1966). Even when capital goods are designed to perform identical tasks, heterogeneity across units within a fleet can generate systematic productivity differences independent of factor prices. As a result, measures of firm's capital based on monetary value commonly used in economic analysis of firm performance may obscure its effective productive capacity.

Although our empirical setting focuses on urban bus fleets, the mechanisms we document apply broadly wherever production capital is organized in fleets that ought to operate seamlessly with shared personnel and support infrastructure[2]. When heterogeneity limits interoperability, operational frictions arise, prompting active efforts to reduce unnecessary variety across industries. Such initiatives may emerge from within industries themselves: in postwar West Germany, bus operators imposed compatibility standards that harmonized vehicle design across manufacturers[3]. In other cases, coordination is imposed by regulators: the European Union's mandate for a common USB-C charging interface reduced technological fragmentation and improved interoperability beyond its waste-reduction objective. The widespread pursuit of compatibility standards reflects the prevalence and underlying economic importance of variety-related frictions.

Despite this prevalence, the costs of variety are notoriously difficult to measure. Firms rarely share data that link fleet composition to performance outcomes, and fleets are often only one of several inputs in production, making it difficult to isolate the effects of variety even when detailed data are available. Moreover, realized fleet composition is typically close to a firm's privately optimal choice, leaving limited identifying variation.

We overcome these challenges by exploiting the institutional and market features of municipal bus transport in Poland, where these obstacles are largely absent. This setting

[2]Among many examples, we distinguish mining companies with fleets of excavators and drilling rigs; hospitals with collections of MRI scanners, ventilators, and surgical robots; IT firms with networks of servers running different operating systems and software architectures; and military organizations with their fleets of weapons, vehicles, and communications systems.

[3]We describe this case in detail in Supplementary Online Materials, Section B.

allows us to provide comprehensive empirical evidence on the relationship between fleet variety and productivity.

Municipal bus fleets provide a particularly clean setting to study the costs of fleet variety. City buses perform a single, well-defined task: serving predetermined passenger routes, and constitute the primary capital input in the production of bus transport services. At the same time, manufacturers introduce distinct technological solutions (e.g., engine placement, drivetrain configuration, interior layout), generating meaningful differentiation across vehicles.

Polish bus operators renew fleets gradually, purchasing buses frequently but in relatively small batches. As municipally owned entities reliant on public funding, they must follow public procurement procedures that aim to foster competition among numerous manufacturers active in Poland. With limited geographic barriers and a unified legal framework, these firms compete at the national level rather than regionally. Consequently, fleet composition varies substantially across operators, largely driven by supply-side factors, i.e. manufacturers' sales strategies over time and across cities, which underpin our identification strategy.

We quantify the costs of fleet variety by estimating regressions relating measures of fleet performance to within-fleet heterogeneity. A central concern is endogeneity: operators may increase the share of preferred bus types and utilize them more efficiently for reasons unrelated to fleet variety itself. To address this concern, we instrument changes in an operator's fleet composition with averaged changes in the fleet structures of other operators serving demographically similar cities (e.g., in population size and density). This strategy exploits the national scope of bus manufacturers' sales and institutional features of public procurement (which generate correlated supply-side shocks to fleet formation across operators) to isolate variation unrelated to the operator's own demand or productivity shocks. Focusing on similar cities ensures relevance, as route structures and desired fleet compositions are comparable. Because our identification strategy relies on active market behavior on the supply-side of fleet formation, we define fleet variety primarily by bus brands (and their interaction with other characteristics). Formally, our approach corresponds to a shift–share instrumental variables design, where identification comes from changes in brand-based fleet composition among other operators.

Our estimates consistently indicate that fleet variety reduces its productivity. Buses are utilized more efficiently as the share of their brand within a fleet increases. Consequently, more homogeneous fleets require fewer vehicles to perform the same transport tasks. In turn, labor remains unresponsive to changes in the degree of fleet unification: intutively, the urban bus industry does not require driver or mechanic licensing specific to a bus type. Cost efficiency improves accordingly, as unit service costs decline with greater fleet unification.

These effects are not uniform. The utilization gains from unification increase over the bus life cycle and are strongest among smaller operators and initially underutilized vehicles. Broadening the definition of variety beyond brand shows that unification along additional dimensions, such as bus length, further amplifies performance gains. Together, the results suggest that variety frictions are shaped not only by bus producer's identity but also by how buses as capital units are matched to operational tasks.

Although our focus is on the costs of fleet variety, we acknowledge that fleet variety may also generate benefits. Firms may value variety as a way to better match heterogeneous demand conditions, to hedge against risks such as product recalls, or to reduce dependence on a single supplier with market power. As a reduced-form analysis, our paper does not take a stand on firms' optimal choices regarding the expansion or reduction of fleet variety. Instead, we study the effects of fleet variety on input utilization and the cost structure within firms. These effects are consistently negative across data sources and empirical specifications. Nevertheless, they may be offset by the benefits of variety that are unobserved to us at the time firms make their decisions, rationalizing the persistence of a positive degree of fleet heterogeneity.

Our paper contributes to a few areas of existing literature.

Heterogeneous production inputs. Heterogeneity in production inputs has long been discussed in economics, where it is typically linked to unobserved input quality (e.g., Griliches, 1957; Griliches and Mairesse, 1995; Bartelsman and Doms, 2000). Owing to data limitations, however, relatively few empirical studies directly address input quality, with recent exceptions focusing primarily on labor rather than capital (Fox and Smeets, 2011). Our study both controls for unobserved quality and extends the notion of input heterogeneity beyond quality differences: even if all buses were of identical quality, within-fleet differences in make would still generate variety-induced frictions that impair productivity. Relatedly, De Roux, Eslava, Franco, and Verhoogen (2021) recognize variety as a source of input heterogeneity but restrict attention to labor and materials and identify its effects through price variation; in contrast, we study qualitatively distinct dimensions of heterogeneity that operate independently of prices: even if all buses had the same price tag, within-fleet differences in make would still generate variety-induced frictions that impair productivity. Finally, Whelan (2002) explicitly studies capital input heterogeneity by treating computers as distinct units of capital that differ in technological vintage. Because rapid technological progress renders otherwise functional computers economically obsolete, measured depreciation depends on the distribution of vintages rather than on physical wear alone. In our setting, we control for bus vintage and evolving technology while focusing on other dimensions of heterogeneity within fleets, thereby generalizing this approach.

Empirical work typically observes aggregate capital inputs but rarely the internal structure or composition of capital within firms. Capital input heterogeneity, if unaccounted for, induces measurement error in input variables. Tybout (1992) links such measurement error to inflation distortions, bookkeeping conventions, and survey bias, and exploits variation in machinery and vehicles to correct for the resulting biases. This approach, however, is obviously not applicable to industries with costly fleet variety, as in our setting. Collard-Wexler and De Loecker (2016); Kim, Petrin, and Song (2016) circumvent input heterogeneity by proposing general-purpose structural econometric tools that are robust to heterogeneity in capital inputs, but without distinguishing between its sources or types. In contrast, our paper explicitly identifies an economically meaningful source of capital input heterogeneity—fleet variety, which we expect to be prevalent across many industries, and provides a detailed characterization of its impact on productivity. Moreover, because our empirical strategy is based on reduced-form methods, our results do not rely on functional-form assumptions.

Variety within Production Process. A long-standing literature in operations and management science studies variety as an operational friction linking greater product variety with lower productivity and higher operating costs (MacDuffie, Sethuraman, and Fisher, 1996; Fisher and Ittner, 1999; Desai, Kekre, Radhakrishnan, and Srinivasan, 2001). Variety in products or processes increases coordination burdens and degrades performance, and how organizational or design choices may mitigate, but probably not eliminate, these costs (Trattner, Hvam, Forza, and Herbert-Hansen, 2019). Our paper differs by shifting attention from output variety to heterogeneity in capital inputs. We show that even when the service task is homogeneous, differences in the composition of durable capital goods generate compatibility and coordination frictions that reduce input utilization and worsen cost structure.

Fleet commonality. Our paper also adds to a long discussion in transportation research regarding fleet commonality. This literature focuses specifically on airlines and tends to use univariate measures of fleet commonality (typically, either a variation of the Herfindahl-Hirschman index or the fleet commonality index developed by De Borges Pan and Espirito Santo Jr (2004)) along one dimension of differentiation (typically supplier's brand) to study their correlation with firm's outcomes (Brüggen and Klose, 2010; Zou, Yu, and Dresner, 2015; Agrawal, 2024), suggesting positive association between fleet homogeneity and firm's operating profitability or cost efficiency. Merkert and Hensher (2011) confirmed these results using a variant of data envelopment analysis. Merkert (2023) suggests that the effects of fleet unification reach beyond the manufacturer dimension, showing the importance of unification along the engine dimension. Our paper complements this literature by moving beyond univariate measures and correlations: we study multidimensional fleet heterogeneity, exploit plausibly exogenous variation in fleet composition, and provide causal evidence on how heterogeneity in fleets affects utilization of various production inputs within firms and their cost structure.

The remainder of our paper is structured as follows. We begin with providing motivating theoretical intuitions in section 2. Section 3 describes our empirical application. Section 4 outlines our empirical strategy, and section 5 presents the results. We conclude our analysis with section 6.

2 Production with Input Differentiation

Before turning to the empirical analysis, we outline a set of simple theoretical intuitions that clarify the mechanism we seek to quantify empirically. Consider a production function $F(\kappa, \ell)$ where output is produced using a differentiated input κ and a non-differentiated input ℓ. The differentiated input aggregates J lower-level inputs, $k = (k_1, ...k_J)$ through an aggregator $\kappa \equiv \kappa(k_1, ..., k_J)$. For simplicity, we assume that $\kappa(\cdot)$ and $F(\cdot)$ are twice continuously differentiable.

At the aggregate level, inputs are standard substitutes: marginal products are positive and diminishing, and the cross-partial derivative satisfies $\frac{\partial F^2}{\partial \kappa \partial \ell}(\kappa, \ell) > 0$. At the differentiated level, adding units of any type increases κ: $\frac{\partial \kappa}{\partial k_i}(k) > 0$ for all i, but there are gains from unification. Specifically, the marginal contribution of additional units of the same type is increasing $\frac{\partial^2 \kappa}{\partial k_i^2}(k) > 0$, and cannot be improved by adding more units of another kind: $\frac{\partial^2 \kappa}{\partial k_i \partial k_j}(k) \leq 0$.

These assumptions imply standard convex isoquants in the (κ, ℓ) space. In contrast, within the differentiated aggregate, the marginal rate of technical substitution (MRTS) between any two inputs k_i and k_j is negative and decreasing. Since MRTS is essentially a slope of the isoquant, this result implies a concave isoquant in the space of (k_i, k_j).

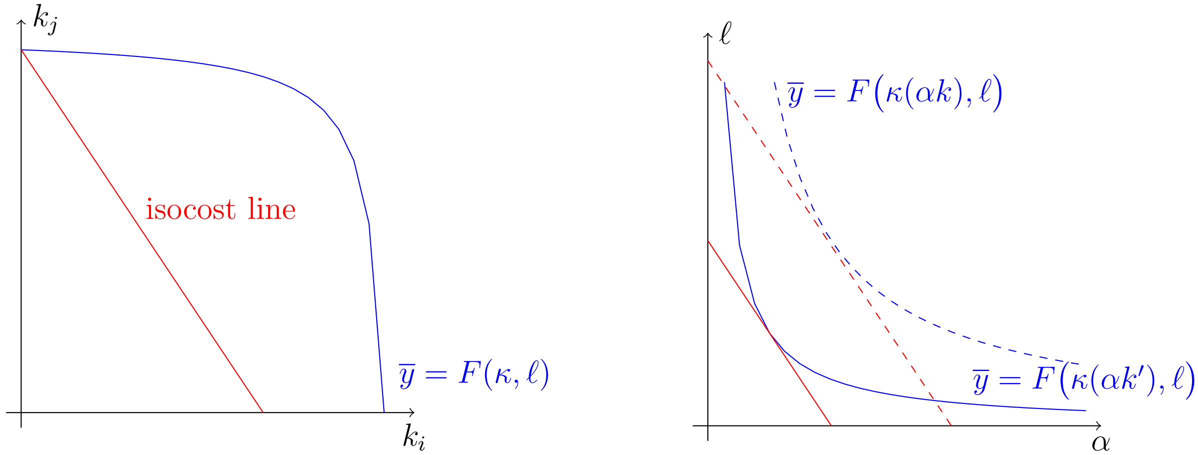
Consider a cost-minimizing firm that can freely adjust all inputs. In the (κ, ℓ) space, convex isoquants imply an interior solution: variety is productive at the aggregate level. However, concave isoquants in the (k_i, k_j) space imply a corner solution, in which the firm specializes in a single differentiated input. Increasing variety among k_i's is therefore counterproductive, even though aggregate differentiation remains valuable. We depict this relation in Panel 1a below.

In practice, firms may face short-run constraints on adjusting the structure of differentiated inputs. Consider two firms producing a common target output $\overline{y}$, each with the same total amount of differentiated units $N = \sum_{i=1}^{J} k'_i = \sum_{i=1}^{J} k_i$ but with different structures such that $\kappa' \equiv \kappa(k') > \kappa(k) \equiv \kappa$. Suppose firms can only scale existing inputs proportionally, $k_i \to \alpha k_i$, while freely adjusting ℓ. This captures settings in which new capital purchases are small relative to the installed base, as in our empirical application.

As shown in Panel 1b, the more efficiently structured firm operates on a scaled-down isocost line. Consequently, it chooses lower values of both α and ℓ to achieve the same output. Differences in the structure of differentiated inputs thus translate into persistent productivity differences when adjustment is limited.

Our paper provides empirical evidence on the relations illustrated in Figure 1.

Figure 1: optimal choice of production inputs with input differentiation



(a) isoquant and isocost in the space of differentiated inputs when all the inputs are freely adjusted.

(b) isoquants and isocosts in the space of aggregate inputs when there are rigidities in input choices. Solid lines represent the firm with a more efficient structure of differentiated inputs.

3 Empirical Setting: Urban Bus Fleets in Poland

We exploit the unique institutional setting of the market for urban bus fleets in Poland to identify and estimate the costs of fleet variety.

3.1 Urban Transportation in Poland

Urban transportation is a relatively large industry in Poland and depends heavily on buses. According to Statistics Poland (Statistics-Poland, 2024), ground public city transportation served over 3 billion passengers in 2023. Out of nearly 60 thousand kilometers within the network of services, less than three thousand kilometers were operated by vehicles other than buses.

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

We focus on the public bus operators. Public bus operators serve their own cities exclusively and do not compete with each other. They produce transportation services by operating routes in the quantity requested and paid by the city at an agreed rate. As a result, the demand for service (and to a large extent also its price) is fixed from the perspective of the operator. The overseeing authorities hold the operators accountable for providing requested services within the allocated budget. Due to this ownership and hiring structure, the set of incentives the public bus operators face is consistent with cost minimization.

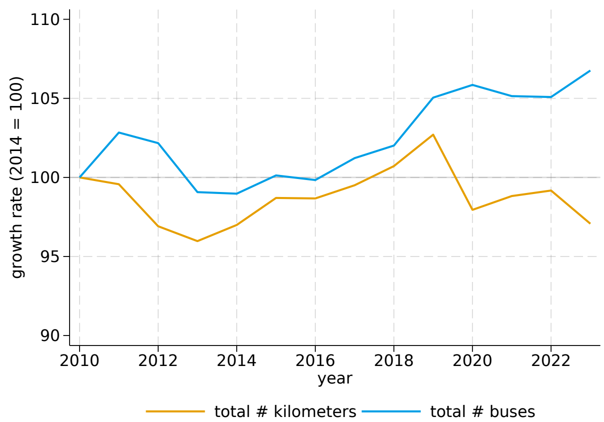
The production output in this industry is measured in the total number of kilometers driven by the buses while serving the routes and reflects the city's demand for transportation. As depicted in Figure 2, this demand remained roughly constant between 2010 and 2023[4]. The output is generated using two main production factors: a fleet of buses and a stock of labor with drivers and bus maintenance personnel playing the major role. Public bus operators typically own, use, and maintain their buses. Leasing and other forms of hire are limited. The total number of buses owned by public bus operators increased by approximately 7% between 2010 and 2023 (partially driven by an increase in the number of operators; figure 2), which accounts for slightly below 700 buses. In turn, during that time, over 10,500 new buses arrived at the depots. Therefore, most purchases of new buses are aimed at fleet renewal, not expansion.

Relying mostly or exclusively on public funds, public bus operators are required to follow the procedures of public procurement. Even though these procedures allow bus operators to affect certain suppliers' chances of winning, the outcome of a tender remains uncertain until after the bids are simultaneously open (for more details see Premik (2024)) and depends on bus providers[5] strategies concerning tender participation and pricing. Since every Polish bus operator functions in the same legal regime and Poland is geographically small enough not to generate substantial logistic challenges in delivering buses, such strategies are often designed and implemented at the national level.

[4]City's demand for serving a system of routes subject to a schedule should not be confused with passenger demand, which dropped by over a third in 2020 and has not recovered until 2023.

[5]We prefer to use the term bus providers instead of bus producers, as second-hand buses are often sold by third party intermediaries.

Figure 2: The evolution of total mileage and the number of buses in the fleets.



Note: The information on the total number of buses is sourced from the statistical yearbooks published by Statistics Poland, for example, Statistics-Poland (2024). The evolution of the total number of kilometers is based on the data collected by IGKM on 76 large public bus operators who reported their total output in each wave of the survey between 2010 and 2023.

There are two main channels from which public bus operators source their buses. Brand new buses are acquired at the primary market in which there are over half a dozen active bus producers in any given year. Their identities vary over time due to the market entry and exit of bus producers. Bus producers also tend to specialize in producing only a subset of all possible bus types. Bus operators may also buy previously used buses on the secondary market, typically imported from Western Europe. All of these lead to potentially substantial heterogeneity across acquired vehicles.

3.2 Bus Differentiation

City buses are highly differentiated products. To facilitate the analysis of such differentiation and the resulting fleet variety, we define a class of buses as a subset of buses that share a common set of characteristics. Consequently, fleet variety relates to the number of classes in the fleet.

Bus classes differ in economic interpretation, depending on the set of defining characteristics. Some of them reflect bus purpose, such as bus length or drivetrain, and are closely tied to the tasks buses perform and the routes they serve. Other dimensions reflect producer identity, such as bus brand or engine brand, capturing 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, a bus operator may prefer to increase their fleet variety by buying a few smaller buses to serve low-demand routes and save on fuel, as opposed to serving all routes with buses of a uniform size. In contrast, producer-related heterogeneity reflects technological incompatibilities across vehicles performing the same tasks, potentially generating frictions in maintenance, repairs, and operations that impair bus operators' productivity without offering any gain.

Our analysis focuses primarily on classes defined by bus brand. Brand corresponds to the identity of the bus producer. The vast majority of buses in our data are sold either directly by their manufacturers or through their exclusive dealers. As a result, producers' efforts to sell their vehicles to bus operators induce variation in bus operators' fleets that is independent from their own preferences towards brands. We exploit this variation to identify and estimate the costs of fleet variety. Moreover, brand tautologically defines classes based on producer-identity, enabling us to abstract from potential efficiency gains arising from purpose-driven fleet variety, which could otherwise confound our estimates.

Although buses differ along many dimensions other than brand, we argue that brand summarizes most of the distinctive technological solutions across vehicles. In Supplementary Online Materials, Section C, we describe other bus characteristics that may lead to the costs of variety and provide Cramér's V measures of association between each characteristic and brand. These associations are uniformly high, indicating that brand serves as a strong proxy for a broad range of underlying technological features. As a result, our reliance on brand, required by our identification strategy, is not a substantive limitation.

3.3 Data Sources

We need information on the evolution of fleet structure combined with measures of operator's productivity to identify and estimate the costs of fleet variety. We collect this data from three major sources.

Fleet data. Fleet data consists of a list of all vehicles owned by every public bus operator in Poland between 2007 and 2024. For each bus-operator pair, the indication of brand, dates of purchase and scrapping (or re-selling), a list of previous owners, and a list of technical details are available. The information comes from http://phototrans.eu/—an online photo gallery created by enthusiasts and employees of the industry that has evolved into a comprehensive database of vehicles owned by operators worldwide[6].

Using the fleet data, we construct two measures of fleet variety: one defined at the bus level and one defined at the fleet level. The bus-level measure is the share of buses belonging to a given class (in most cases, a brand) within a fleet. The fleet-level measure is the within-fleet Herfindahl–Hirschman Index (HHI) computed from these class shares.

Odometer data. Every registered vehicle in Poland is subject to periodic mandatory inspections, typically conducted at least once per 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.

We use odometer readings to construct yearly mileage for each bus, which serves as one of the main outcome variables in our analysis. Odometer readings are collected at irregular dates throughout the year, which precludes using simple differences between consecutive readings. To obtain a measure of mileage that is comparable across buses, we first compute the distance traveled between each pair of consecutive odometer readings. We then aggregate these distances to the calendar-year level, scaling them by the appropriate number of days when consecutive readings span different years. The resulting measure reflects annualized, or yearly, mileage for each bus and is the best approximation to the mileage driven in a year we can obtain.

Under the assumption that bus operators deploy their resources to the maximum feasible extent, which is consistent with cost-minimizing behavior, bus utilization defined by the yearly mileage provides a straightforward measure of bus-level productivity.

[6]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.

Fleet Performance Indicators: IGKM Data. Fleet performance indicators are provided by the Chamber of Commerce of City Transportation (IGKM) in Poland. IGKM is an association of firms operating in urban transportation services in Poland. IGKM data is collected through a yearly questionnaire distributed across firms associated with IGKM. The survey aims to help exchange data, good practices, and experiences among IGKM members. Participation in the survey is voluntary. Even though bus operators associated in IGKM provide similar services, they do it in different cities and hence do not directly compete against each other. That rules out wrong incentives to provide deceptive information in the survey. We use survey waves covering the period of 2007-2023.

The IGKM data provide our primary source of information on operator-level efficiency. They report total output (measured as the aggregate mileage of all buses in an operator's fleet within a year) as well as a range of additional indicators describing fleet size, employment, and cost structure.

3.4 Suggestive Evidence on the Costs of Fleet Variety

We use our data to provide suggestive evidence on the costs of fleet variety. We begin by describing the Polish public bus operators' fleet structure and then move to linking it with bus the operator's efficiency measures. We restrict our attention to a balanced panel of 129 bus operators who serve mainly urban routes and were present on the market every year between 2014 and 2024, that is, years for which the odometer reading data is available.

Table 1 presents a basic overview of our sample of 129 bus operators and the total of 16333 buses in their fleets between 2014 and 2024. Several important features of the market emerge. The average bus remains in service for nearly 20 years. Once a bus arrives, it typically remains in service until it is scrapped, as illustrated by very similar distributions of bus age at scrappage and leaving the fleet. Therefore, short-run adjustments in bus operators' output are achieved primarily by reallocating utilization across existing vehicles rather than by reshaping the fleet.

Over the lifecycle of a bus, new batches of buses are typically added to the fleet every 1.7 years, implying that approximately eleven procurement rounds occur during the lifetime of a single bus. Even if we assume that all buses in the new batch are the same (which is not always the case), this procurement pattern poses a substantial challenge to fleet unification, enabling desired variation in fleet structure across time and bus operators.

Table 1: Public bus operators and their fleets: average summary statistics by year (2014-2024).

	total/mean	median	standard deviation
# operators	129	—	—
# buses	16333	—	—
average fleet size (# buses)	75	35	144
average fleet age (years)	10	10	2.6
average age at scrapping (years)	20	20	3.1
average age at leaving the depot (years)	19	19	3.9
average time between introductions (years)	1.7	1.7	4
average size of an introduction (% fleet)	15	12	10
average yearly mileage per bus (km)	49560	48625	13751
average yearly mileage within fleet st. dev. (km)	17256	16916	—

Note: This table presents the summary statistics obtained on the sample of all buses.

The distributions of average fleet size, average time between deliveries of new buses, and their sizes as % of a fleet exhibit averages that are notably higher than respective medians, and large standard deviations. This reflects substantial heterogeneity in the types of cities served by the operators and resulting differences in their business operating models.

The last two rows of Table 1 reveal substantial variation in average yearly bus mileage at the bus level. In particular, the average within-fleet standard deviation of mileage (17,256 kilometers) exceeds the corresponding across-fleet standard deviation (13,751 kilometers), indicating that systematic differences in utilization cannot be explained solely by variation in the types of cities served by different operators.

Such variation may arise from multiple sources, each linked to a distinct notion of fleet variety. One possibility is heterogeneity in bus purpose. For example, buses serving suburban routes that transport passengers from outer suburbs to the city center are likely to accumulate systematically higher annual mileage than buses operating short feeder routes connecting residential areas to nearby train stations. Alternatively, the observed variation may reflect differences in bus productivity associated with producer-related fleet variety. For example, buses of brands underrepresented within a fleet may have lower utilization potential due to longer repair times or delays in the availability of compatible spare parts.

To further investigate the sources of the observed variation in bus utilization, we examine the composition of bus operators' fleets in more detail. Table 2 reports summary statistics on fleet structure based on the brand-based definition of bus classes.

Table 2: Fleet structure by brand: summary statistics (Odometer sample, 2014-2024).

	national				fleet			
	#	1st share (the highest)	2nd share (2nd highest)	HHI	#	1st share (the highest)	2nd share (2nd highest)	HHI
brand	28.7	.46	.17	.27	3.7	.62	.24	.51

Note: The table presents the bus differentiation measures related to brand, averaged over operators and years. The first four columns present statistics calculated at the national level, and the last four columns present firm-level statistics. The 1st and 2nd share denotes the average fleet shares of brands with the highest and second highest number of vehicles in the fleet, respectively.

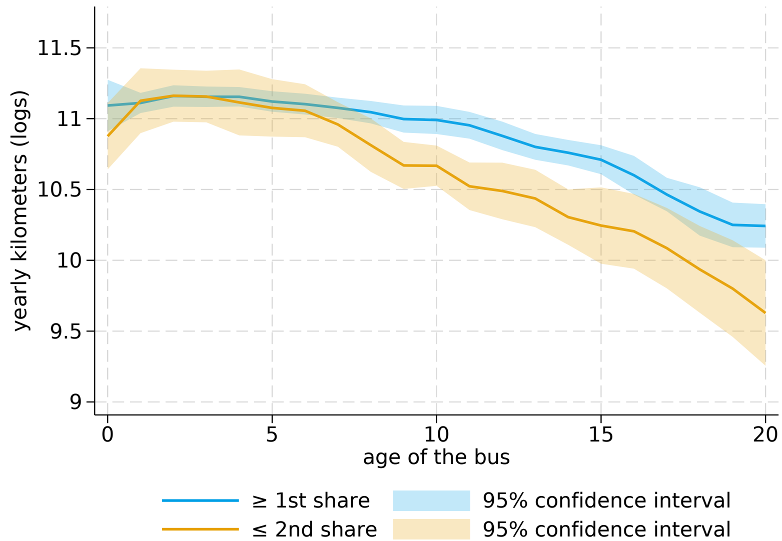
In a given year, on average nearly 29 bus brands serve urban routes in Poland. Despite frequent and relatively small procurement orders and the wide range of bus types available on the market, a typical operator maintains limited within-fleet variety. The average fleet includes buses from only 3.7 brands, i.e. less than 13% of the 28.7 brands present in the market. Consistent with this concentration, the average within-fleet Herfindahl–Hirschman Index (HHI) over brands equals 0.51, nearly twice the corresponding market-level measure. Hence, bus operators attempt fleet unification, and this attempt, to some extent, is successful.

In the next step, we connect the observed fleet unification patterns to operators' performance. Figure 3 presents the average yearly bus utilization in two groups of vehicles defined by the fraction of buses of the same brand in the fleet. We define such groups by the thresholds reflecting the average fleet shares of the first and the second most represented brands in a fleet, using the respective numbers from Table 2.

Unsurprisingly, bus age is a major factor determining yearly mileage, with newer buses being more frequently utilized than the older ones. However, this is only part of the story presented in Figure 3.

Buses of brands that are more heavily represented in the operator's fleet tend to exhibit higher utilization in the life cycle. Distinct utilization patterns emerge. In the first few years of operation, there are no visible utilization gains from having a larger number of buses of the same brand. Intuitively, newly delivered buses are less prone to malfunction and are typically covered by warranty. Beginning at approximately five years of age, however, buses

Figure 3: Bus utilization in life-cycle by brand fleet share.



Note: this graph presents estimated coefficients together with 95% confidence intervals from a linear regression where the dependent variable is log yearly mileage and the regressors are the full set of dummies for age category interacted with dummies indicating that a bus belongs to a given class within an operator's fleet defined by brand whose fleet share in addition exceeds the 1^{st} share or is below the 2^{nd} share statistics (as defined in table 2). No constant is included, but we control for year fixed effects. Standard errors are clustered on the bus operator level. Warsaw is excluded from the analysis as an outlier.

belonging to more numerous brand classes become significantly more utilized. This effect persists through the remainder of the bus life cycle and strengthens as buses age.

Intuitively, as vehicles age, they require increasingly frequent maintenance and repairs. Buses belonging to classes that are more heavily represented in the fleet are therefore likely to be repaired more quickly than rarer types, for several reasons. Spare parts are more likely to be readily available, and mechanics accumulate experience and specialized knowledge with commonly used vehicle types. As one industry representative noted in discussion, so-called *unicorn* buses, i.e., rare series within a fleet, often remain in regular service only until the first major breakdown.

The odometer reading data suggest that the costs of fleet variety—the negative relation between fleet heterogeneity and bus performance—are an important factor influencing bus utilization. Now, we show evidence on the costs of fleet variety on the fleet-level measures of bus operators' performance. Table 3 introduces basic fleet-level summary statistics based on the IGKM data.

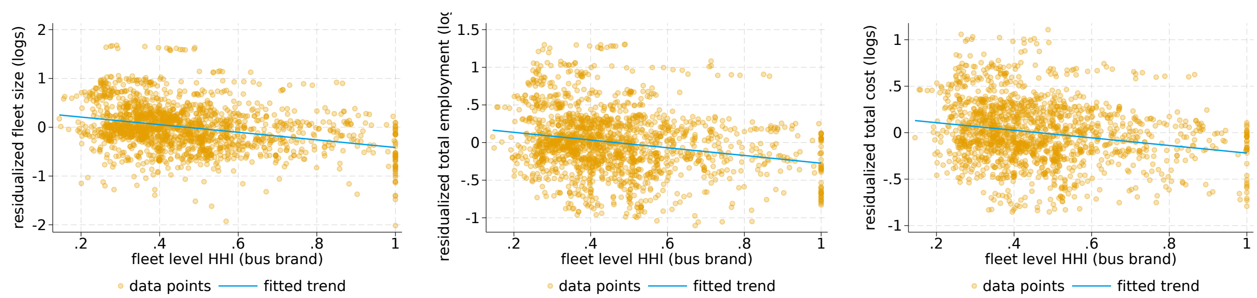
Table 3: Public bus operators and their fleets: summary statistics (IGKM sample 2007-2023).

	total/mean	median	standard deviation
# operators	117	—	—
average fleet size (# buses)	78	36	147
average labor (# employees)	245	105	475
→ including drivers (# employees)	159	64	338
→ including mechanics (# of employees)	41	19	69
average total output (thousands km)	4599	1723	9362
average total cost (thousands 2010 USD)	10146	3645	22984

Note: This table presents the summary statistics obtained on the IGKM sample.

The IGKM sample covers 117 bus operators between 2007 and 2023. The average fleet size is comparable to the one from the odometer sample, suggesting there is no significant self-selection mechanisms into the IGKM sample. These operators employ, on average, 245 employees. Drivers are the most significant part of the workforce, followed by a relatively similar number of mechanics and remaining employees. On average, these firms perform yearly over 4.5 million kilometers of service at the cost of approximately 10 million 2010 USD.

Figure 4: Fleet-level HHI (brand) and production outcomes.



(a) residualized number of buses (b) residualized labor input (c) residualized total cost

Note: this figure presents scatter plots of brand-based within-fleet HHI against the residualized logarithm of (a) the number of buses in the fleet, (b) total employment, and (c) total operating costs, together with fitted linear trends. We obtain the residualized measures by regressing the logarithm of each outcome on the logarithm of total output and fixed effects for years and operators, and retaining the residuals. By conditioning on realized output, these residualized outcomes admit a productivity-based interpretation.

Figure 4 presents the relation between brand-based within-fleet HHI and measures of firm performance based on inputs and costs. These measures are created by taking residuals from a regression of the logarithm of each outcome, respectively, on the logarithm of the total operator's output and fixed effects for years and operators. Because we condition on realized output, these residualized outcomes have an interpretation in terms of productivity.

All three graphs suggest the same result: variety is costly. More homogeneous fleets require fewer production inputs and generate smaller costs.

The descriptive evidence indicates that operators actively limit fleet heterogeneity: despite frequent procurement and a wide range of available bus types, most fleets are concentrated in a small number of brands and models. Our data also suggests that the variety is bad for the operators: it is negatively correlated with bus utilization, and positively correlated with fleet size, employment, and total costs. While the patterns documented in this section are suggestive, they are based on simple correlations observed in the raw data. In the remainder of the paper, we develop an identification strategy that allows us to establish a causal negative relationship between fleet variety and operators' productivity.

4 Empirical Strategy

Our task is to identify and estimate the effects of fleet variety on bus-level and fleet-level measures of bus operators' performance.

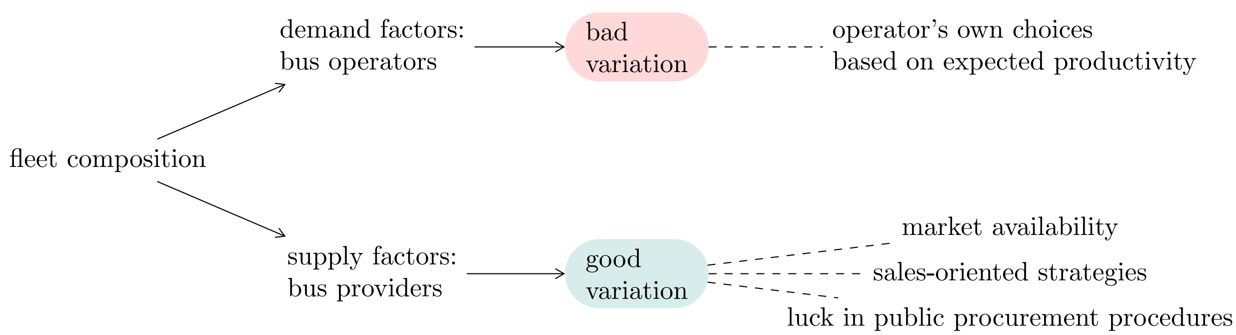
4.1 Intuition

A central empirical challenge is that fleet composition is not randomly assigned. Even under public procurement procedures, operators retain some influence over which buses they intend to purchase, basing these decisions on expectations about performance, operating conditions, and long-run costs. Any factor that makes a particular bus type well-suited to an operator, such as local terrain, route structure, or maintenance capabilities, is therefore likely to increase both its utilization (or improve other efficiency measures, such as requiring fewer buses or employees, or lowering costs) and the operator's propensity to acquire additional buses of the same type. As a result, the observed shares of brands in a fleet reflect not only technological compatibility but also latent match quality between the operator and a given brand, creating an endogeneity problem. Identifying the causal effect of fleet variety thus requires isolating changes in fleet composition that are orthogonal to operator-specific productivity considerations, or demand-side shocks.

The institutional structure of the urban bus market in Poland generates a natural source of exogenous variation in fleet composition. Most buses are purchased through public procurement procedures, which limit operators' discretion in selecting suppliers. Procurement procedures crucially require suppliers' active participation: manufacturers must decide whether to submit a compliant offer and on what terms. As a result, procurement outcomes reflect not only operators' preference for certain bus classes but also suppliers' strategic choices regarding entry, pricing, and bidding intensity. All of these contribute to what we call the supply-side shocks.

Figure 5 visualizes the sources of variation driving fleet composition. Demand-side shocks, summarizing how strongly an operator attempts to expand the fleet share of a given brand, constitute *bad* variation from the perspective of identifying the costs of variety, as they are directly related to factors that also drive an operator's performance. Supply-side shocks, by contrast, reflect brands' efforts to sell vehicles and hence affect fleet composition without directly influencing how intensively buses (and other production inputs) are utilized once in service. As a result, they contain *good* variation for the identification purposes[7].

Figure 5: Sources of variation in fleet composition.



[7] A potential concern is that supply-side shocks may also contain *bad variation* that affects the operator's productivity directly. For example, bus providers may design strategies aimed at increasing future sales by improving current bus operators' experience through enhanced after-sales service or spare-parts pricing. We argue that such channels are unlikely to be relevant in our setting. First, unless these strategies vary over time, they are easily absorbed in the empirical framework by first differencing or through fixed effects. Second, temporary improvements in after-sales service are unlikely to induce operators to adjust their fleets, given that buses are long-lived assets. Third, after-sales conditions are typically specified ex-ante in procurement contracts, leaving limited scope for discretionary supplier behavior. Finally, spare parts are often supplied by third-party intermediaries in competitive markets and hence remain outside of the bus providers' selling strategies.

On the supply side, bus manufacturers operate at a national scale. Poland is geographically compact, the regulatory regime is uniform, and transport costs do not create meaningful barriers to delivering new buses across cities. Consequently, suppliers' sales vary over time in a manner that is correlated across operators. A manufacturer that expands its market presence in one city is therefore likely to do so in other cities, particularly those with a similar route network. As a result, a bus brand's expansion in the fleets of other comparable operators reflects a supply-side shock that is correlated with the evolution of the operator's own fleet composition, and plausibly unrelated to their demand-side shocks.

Our identification strategy exploits this structure. We isolate exogenous changes in fleet structure by instrumenting our measures of variety (brand fleet shares or brand within-fleet HHI indices for bus-level and fleet-level analysis, respectively) with brand inflows (reflecting the number of buses of a given brand delivered throughout a year) among other but similar fleets. We focus on bus inflows because supply-side shocks affect which buses enter fleets, but are unlikely to directly affect which buses are withdrawn from operation. As the inflows essentially arise from changes in fleet composition, we conveniently keep our empirical analysis on the first-difference level.

4.2 Bus-Level Analysis

We formalize the above intuitions. On the bus-level, we aim to estimate parameter α of the following equation:

$$\Delta y_{ijt} = \alpha \Delta s_{c(i),jt} + \Delta x'_{ijt}\beta + \Delta\tilde{\varepsilon}_{ijt} \tag{1}$$

where Δ is the time difference operator, y_{ijt} denotes yearly mileage of bus i in fleet j and at year t, $s_{c(i)jt}$ is the fleet share of i's brand (i.e. $c(i)$; we abuse this notation for simplicity and interchangeable write just c), x_{ijt} contains any observables factors systematically affecting bus utilization (we include dummies for bus age, fixed effects at the operator-time (jt) level interacted with dummies for bus length (size) and drive, and indicators for buses being leased or on warranty at jt), and $\tilde{\varepsilon}_{ijt}$ is a composite error term.

The composite error term can be further divided into systematic (μ_{ijt}) and idiosyncratic (ε_{ijt}) components explaining bus utilization, $\tilde{\varepsilon}_{ijt} = \mu_{ijt} + \varepsilon_{ijt}$. The endogeneity issue arises when:

$$\Delta s_{c(i)jt} \not\perp \Delta\mu_{ijt}$$

that is, the fleet composition is related to unobserved systematic drivers of bus utilization. Such factors may include, for example, individual tastes of operator in buses of brand c or objective advantages of c relative to other brands (perhaps these buses are more appropriate to serve routes in mountainous terrain or navigate easier narrow urban streets).

We break the possible connection between s_{cjt} and μ_{ijt} by selecting an instrument that isolates the supply-side shocks in the following way. First, we note that changes in a fleet within a year can be decomposed as follows [8]:

$$\Delta s_{cjt} \equiv \underbrace{\Delta^+ s_{cjt}}_{\text{inflows}} + \underbrace{\Delta^- s_{cjt}}_{\text{outflows}} \tag{2}$$

where the *inflows* component $\Delta^+ s_{cjt}$ describes changes to s_{cjt} related to the arrival of new buses to the fleet and $\Delta^- s_{cjt}$ summarizes *outflows*, i.e., the proportion of buses that were scrapped or resold by j at t. We focus our attention on the inflows $\Delta^+ s_{cjt}$ as this is where the supply-shocks kick in. However, $\Delta^+ s_{cjt}$ may still contain the *bad* variation related to utilization shocks, that is, related to the μ_{ijt}. In the second step, for operator j we consider fleet inflows among other operators $j' \neq j$, i.e. $\Delta^+ s_{cj't}$, as plausibly unrelated to j's demand shocks (μ_{ijt}) but correlated with j's inflow $\Delta^+ s_{cj't}$ (and hence also with Δs_{cjt}) through the bus providers' national-level market strategies, i.e., the supply-side shocks.

Eventually, we construct our bus-level instrument by averaging inflows across operators:

$$z^B_{cjt} = \sum_{j'} \omega_{jj'} \times \Delta^+ s_{cj't} \tag{3}$$

[8] Because buses enter and exit fleets at different points within a calendar year, we construct a fleet-size measure that accounts for such timing. For each class c in fleet j and year t, we define N_{cjt} as the total number of bus-days contributed during year t. Furthermore, inflows (ΔN^+_{cjt}) are measured as the first 365 bus-days following a vehicle's entry into the fleet, and outflows (ΔN^-_{cjt}) correspond to the 365 bus-days after its exit. For example, a bus entering the fleet on December $30th$, 2018 would contribute two days to 2018 inflows and 363 days to 2019 inflows. This convention mirrors the standard annual accounting framework but ensures that measured fleet size does not mechanically depend on the exact timing of deliveries or scrappage within the year. To maintain consistency across measures, we normalize net changes by total fleet size in year t, i.e. $\Delta s_{cjt} \equiv \frac{\Delta^+ N_{cjt} - \Delta^- N_{cjt}}{N_{jt}}$. When fleet size is approximately constant, as in our setting, this normalization has a negligible impact on the analysis.

where $\omega_{jj'}$ denotes a set of exposure weights such that $\omega_{jj} = 0$ for every j[9], $\omega_{jj'} \geq 0$ for every j, j', and $\sum_{j'} \omega_{jj'} = 1$ for every j.

The exposure weights $\omega_{jj'}$ define the exposure of bus operator j to bus provider-specific shocks measured by inflows to fleets of other bus operators j'. The goal is to compare *similar* cities as they are most likely to share similar supply-side components. We achieve it by building a matrix of cosine similarity measures between city pairs jj' based on three factors: operator's total output, city population, and city population density. These factors reflect the objective demand for routes, which remains roughly constant over time, and thus are not likely to be related to individual-bus-level or fleet-level performance shocks.[10]. Having obtained the matrix of cosine similarities among cities, say $\{cs_{jj'}\}_{jj'}$, we construct our exposure weights for each operator j in the following way. We select every j' for which the cosine similarity measure between j and j' is positive and renormalize them to sum up to one, otherwise setting the exposure weight to zero:

$$\omega_{jj'} = \begin{cases} \frac{cs_{jj'}}{\sum\limits_{k': cs_{jk'} > 0, j \neq k'} cs_{jk'}} & \text{if } cs_{jj'} > 0 \text{ and } j \neq j' \\ 0 & \text{otherwise} \end{cases}$$

Using the terminology from Figure 5, our instrument captures the *good* variation—related to the supply factors that are not targeting bus utilization shocks. It may also capture what we call *bad* variation: though it is *bad* in reference to operator j''s utilization shocks $\mu_{ij't}$, but not necessarily μ_{ijt}. We argue that instruments based on $\Delta^+ s_{cj't}$ (and, more broadly, $s_{cj't}$) are valid and relevant in our setting.

Validity. Instrument validity requires that other operators j' do not purchase more buses of a given brand systematically for reasons that simultaneously increase utilization at both j and j'. We assess instrument validity by discussing potential channels through which

[9]Excluding operator j removes mechanical correlation between the instrument and operator-specific shocks. This strategy has been employed by, among others, Autor and Duggan (2003), and is commonly referred to in the literature as a leave-one-out instrument.

[10]We define the cosine similarity measures as fixed in time to simplify our analysis. Although there is some limited time variation in these factors, it is not large enough to change significantly their relative ranking and hence choices of similar cities.

$\Delta^+ s_{cj't}$ could be correlated with $\Delta\mu_{ijt}$, and by explaining how features of the market and our empirical design address these concerns.

A first potential threat arises from differences in the objective quality of bus classes. Higher-quality buses may be both more intensively utilized and more widely adopted, inducing correlation between innovations in $\Delta^+ s_{cj't}$ and $\Delta\mu_{ijt}$. We address this concern by exploiting the fact that bus quality is largely determined by technological solutions and therefore time-invariant. First differencing removes all time-invariant heterogeneity across buses. Any remaining age-related quality variation is absorbed by bus-age fixed effects.

A related concern is that a given class may be a particularly good match for both operators j and j' due to persistent local conditions, such as terrain or route characteristics. As long as these conditions are time-invariant, they are eliminated by the same differencing and fixed-effects structure. More generally, any time-invariant objective or subjective preference for a class within an operator is removed.

Time-varying demand shocks, such as the COVID-19 pandemic, affect multiple operators and could induce correlation between utilization shocks at different operators, which may be correlated with the fleet structure if an operator decides to adjust it in response to the shock. Including time-varying operator-level (jt) fixed effects to the specification removes this correlation. We additionally interact them with indicators for size and driver to account for the operator's task structure.

While corruption in public procurement could affect sales outcomes, it is unlikely that it also affects bus operators' post-sale utilization decisions. Lastly, throughout the sample, there are no vehicle recalls that would affect more than one operator and hence induce correlation between innovations in $s_{cj't}$ and μ_{ijt} through innovations in $\mu_{ij't}$.

Relevance. Relevance of our instrument requires correlation between contemporaneous inflows of buses of a given brand into fleets of comparable bus operators. It is likely to hold given the features described above of the market that encourage market-wide sales campaigns among bus providers. Moreover, individual bus operators' fleets are small relative to the scale of production on the supply side, implying that capacity constraints rarely bind[11].

[11] Capacity constraints would induce a component of negative correlation between inflows, possibly confounding our instrumental variable strategy.

City-specific characteristics may induce heterogeneity in providers' strategies, as fleet requirements differ across cities and providers specialize in particular vehicle types. As a result, innovations in $s_{cj't}$ may be weakly correlated with innovations in s_{cjt} when operators j and j' serve dissimilar cities. We address this concern by restricting comparisons in our instrument to operators located in cities that are similar along dimensions unrelated to the utilization of a single bus.

Instrument relevance restricts which class definition can be applied to define fleet variety. Our identification strategy applies naturally to brand-based classes, as bus manufacturers actively compete for market share and adjust their sales strategies over time, generating correlated supply-side variation across cities. In contrast, classes defined purely by bus purpose, such as size or drive, do not correspond to active sellers and therefore do not generate analogous supply-side shocks. However, we are still able to shed some light on the importance of such bus features in generating costs of fleet variety by applying our framework to classes that interact these features with brand.

4.3 Fleet-Level Analysis

The logic of the fleet-level analysis closely parallels that of the bus-level framework, but it requires several adjustments to account for aggregation. In particular, because class shares necessarily sum to one within each fleet, we adopt a different measure of fleet variety, a class-based Herfindahl–Hirschman Index (HHI) defined at the fleet level.

$$s_{jt}^F = \sum_c s_{cjt}^2$$

We are interested in identifying and estimating parameter α of the following equation:

$$\Delta y_{jt} = \alpha \Delta s_{jt}^F + \Delta x_{jt}' \beta + \Delta \tilde{\varepsilon}_{jt} \tag{4}$$

where y_{jt} describes a fleet-level performance indicator of fleet j and at time t, s_{jt}^F is our measure of fleet variety, x_{jt} denotes observed systematic drivers of the fleet performance measure, $\tilde{\varepsilon}_{jt} = \mu_{jt} + \varepsilon_{jt}$ is a composite error term composed of a systematic and an idiosyncratic components.

Fleet-level data provide several relevant measures of operator performance. As our primary objective is to estimate the effect of fleet variety on productivity, we focus on the (logarithms of) the total number of buses in the fleet, total employment, and total operating costs. In all specifications, we condition on total yearly mileage, which captures the operator's output and ensures both comparability across operators and a productivity interpretation of the estimates[12].

In our fleet-level specifications, we include time dummies to remove any macro-level trends. We also control for a range of covariates that are plausibly exogenous and may impact the fleet-level performance measures, including average bus age in the fleet, an indicator for extended bus repairs being done at the operator's, share of leased buses, and share of brand-new buses still on a warranty.

Our instrument now aims to isolate the supply-side shocks at the bus providers' level. We define it as:

$$z^F_{jt} = \sum_{j'} \omega_{jj'} \times \sum_{c} \Delta^+ s^2_{cj't} \tag{5}$$

The identification argument at the fleet level parallels that at the individual bus level. Validity requires that other operators j' do not purchase more buses of the same brand (or, more broadly, a class) for reasons that are correlated with fleet-level performance shocks at both j and j'. We address this concern by removing time-invariant variation and conditioning on a rich set of fleet-level covariates, thereby controlling for factors other than variety that may also affect performance. Instrument relevance, in turn, follows from the same arguments as in the bus-level analysis.

[12] This approach works because the total output is largely exogenous from the operator's perspective, as it is determined by city authorities through the design of the transport network. Alternatively, one may consider a specification in which the dependent variable would be a measure of performance per unit of output. At the cost of less interpretable coefficients, this specification is free from endogeneity concerns if the total output, in fact, is not exogenous. We present respective estimates from such specifications in table E.5 in Supplementary Online Materials, Section E. The results are qualitatively the same as from the main specification discussed in the paper, suggesting that the total output is indeed taken as given by the operators.

4.4 Estimation Remarks

Our Framework is an SSIV. The construction of our instrument induces dependence across operators that are exposed to similar shocks, making standard inference procedures potentially invalid. We address this issue by noting that our framework can be cast as a shift–share instrumental variable (SSIV) design and by applying the inference methods developed by Borusyak, Hull, and Jaravel (2022). Using their terminology, the exposure weights $\omega_{jj'}$ play the role of *shares*, while the fleet inflows $\Delta s_{cj't}$ serve as the *shifts*. Exogenous variation is introduced through these shifts, which isolate supply-side shocks in the evolution of bus operators' fleets.

The SSIV framework emphasizes the sources of identification. The orthogonality condition between our instrument z and demand-side shocks μ at both the bus and fleet levels is equivalent to requiring orthogonality between brand-level (or, more generally, class-level) fleet inflows and the exposure-weighted average of unobserved utilization shocks within the same brand (class) across similar operators. We derive this equivalence in our application and discuss its implications in detail in Supplementary Online Materials, Section D.

Interpretation of the Estimates. To facilitate interpretation, we express our estimates in terms of changes in performance associated with a typical new bus purchase that increases fleet unification. At the bus level, we report average effects corresponding to an average unifying purchase by multiplying the estimated coefficient α by the median change in the fleet share of the corresponding class following the arrival of new buses. Specifically, we use the 50th percentile of the distribution of $\Delta^+ s_{cjt}$ conditional on $\Delta^+ s_{cjt} > 0$. At the fleet level, we follow an analogous approach. We express the effects as responses to an average unifying purchase by multiplying α by the median increase in the fleet-level class HHI after new buses arrive, using the 50th percentile of the distribution of $\Delta^+ s^F_{jt}$ conditional on $\Delta^+ s^F_{jt} > 0$.

Weak IV Testing. Our instrumental variable strategy requires that supply-side variation in class fleet shares be correlated across operators. A potential concern is that this correlation may be weak, leading to weak-instrument problems. We address this issue in several ways. First, we assess instrument relevance using the Olea and Pflueger (2013) effective first-stage F-statistic, and complement this heuristic with the Kleibergen–Paap rk

LM underidentification test. Second, we conduct weak-instrument robust inference using Anderson–Rubin tests based on Lagrange Multiplier (LM) and Minimum Distance (MD) statistics. Rejecting the null in these tests provides additional evidence supporting the statistical significance of the estimated effects despite a potentially weak first stage.

Sample Selection. Constructing the estimation sample requires selecting both shift-level operators (j'), which provide the variation used to build the instrument, and unit-level operators (j), whose outcomes identify the effects of fleet variety. We restrict the shift-level sample to a balanced panel of operators active throughout the entire observation window (2014–2024 at the bus level and 2007–2023 at the fleet level) to keep exposure weights fixed over time. Unit-level operators are drawn from this sample subject to outcome availability, which reduces the sample size, particularly at the fleet level, where not all operators participate in the IGKM survey.

5 Results

This section presents the results of our empirical analysis. We begin by discussing the main findings on the effects of brand variety in fleets on the selected performance measures. We then extend the analysis to explore potential nonlinearities in the effects of variety across selected bus- and fleet-level characteristics. We conclude our results by presenting estimates based on broader definitions of variety that extend beyond bus brand alone. Additionally, Appendix A provides further evidence on the validity of our instruments, including analysis of the distribution of the exposure weights $\omega_{jj'}$ and a range of falsification tests; and Supplementary Online Materials, Section E reports a series of robustness tests that support our results.

5.1 Main Result

In our main result, we study how fleet heterogeneity defined by brand affects individual bus utilization and fleet-level measures: fleet size, number of employees, and total cost of providing the service. Table 4 presents the estimates.

Table 4: The effects of fleet variety defined by brand on operator's performance measures.

	bus-level	fleet-level		
	log yearly mileage	# buses	# employees	total cost
brand	.038 (.014)***	−.020 (.004)***	−.001 (.003)	−.021 (.006)***
N_i	11377			
N_j	128	78	78	77
$N_{j'}$	129	128	128	128
N_c	37	38	38	38
t	2014-2024	2007-2023	2007-2023	2007-2023
effective F-stat.	10.169	17.598	17.586	11.203
underid. test (χ^2)	5.528 (.019)**	3.882 (.049)**	4.030 (.045)**	4.297 (.038)**
Anderson-Rubin test (LM)	3.975 (.046)**	2.837 (.092)*	.139 (.709)	3.275 (.070)*
Anderson-Rubin test (MD)	6.232 (.013)**	18.330 (.000)***	.164 (.685)	12.472 (.000)***

The table presents the estimated percentage changes in measures of bus and fleet performance in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$). *Sample selection.* At the bus level, we consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 90 days throughout year t in operator j's fleet to include observation ijt in the sample. At the fleet level, we include in the unit-level sample only those operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand and brand×year level for fleet- and bus-level analysis, respectively. *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

Our main results consistently indicate that fleet variety matters. First, it affects the productivity of the fleet, i.e., the capital production input. A typical purchase toward fleet unification within the same brand is associated with a 3.8% increase in the utilization of buses of that brand within the fleet. This bus-level effect is mirrored at the fleet level. A typical increase in brand-based fleet concentration (HHI) resulting from a unifying purchase leads to a 2% reduction in fleet size. These effects are naturally connected: as buses become more productive in more homogeneous fleets, operators require fewer vehicles to deliver a given level of service.

Second, fleet variety also affects operators' cost structures. A median purchase toward brand-based fleet unification is associated with a 2.1% reduction in total operating costs. This cost decline likely reflects economies of scale in maintenance and logistics, including more efficient ordering, storage, and use of spare parts.

Finally, we find no systematic effect of fleet variety on operators' employment. The estimated coefficient is both statistically insignificant and economically small. This result extends to different categories of labor: drivers, mechanics, and other employees. In the Supplementary Online Materials, Table E.3, we show that none of these employment margins responds to changes in brand-based fleet variety.

This pattern is intuitive. Unlike in sectors such as aviation, different bus brands do not require separate operating licenses, so driver employment is primarily determined by route coverage and service frequency rather than fleet heterogeneity. While mechanics might be expected to adjust in response to increased technological variety, we find no evidence of such adjustment. Instead, operators appear to respond along the capital margin, reducing fleet size as productivity rises in more homogeneous fleets and reallocating maintenance effort toward more common vehicle types. Given that new bus purchases are often heavily subsidized, whereas labor costs are largely borne by operators and subject to employment protection regulations, especially in municipally owned firms, adjustment along the capital rather than the labor margin is not surprising.

Statistical tests indicate no major inference concerns related to weak instruments. In all specifications, the effective first-stage F-statistic exceeds the conventional threshold of 10 and the underidentification test is rejected at least at the 5% significance level. Weak-instrument-robust tests yield significance patterns consistent with our baseline estimates, providing additional support for the conclusion that fleet variety affects all examined performance measures except labor.

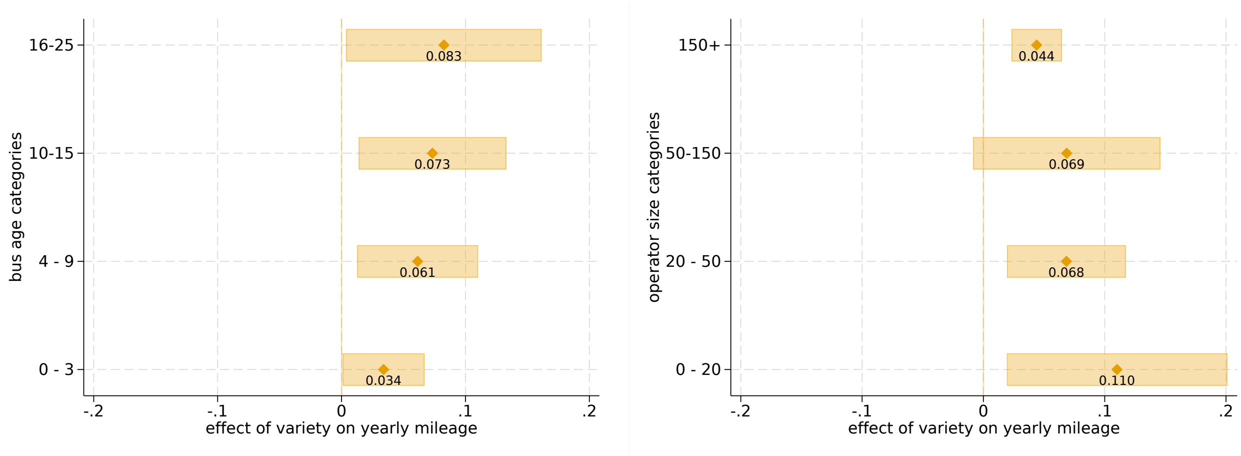
The results reported in Table 4 are obtained under a specific set of assumptions. To assess robustness, we re-estimate our models in alternative settings: varying the definition of variables, clustering levels, and sample selection. The corresponding results and discussion are presented in the Supplementary Online Materials, Section E.

5.2 When Does Variety Matter Most?

Our main results provide a linear approximation to the relationship between fleet variety and operator performance. However, this relationship may be nonlinear. For example, Figure 3 in the descriptive section suggests that the importance of variety increases as buses age. In this section, we re-estimate the bus-level regressions[13] on a set of carefully chosen subsamples to further characterize the costs of variety. Throughout, variety is defined exclusively by bus brand.

[13] The data do not provide sufficient variation to study analogous nonlinearities at the fleet level.

Figure 6: Heterogeneous effects of variety by groups of buses.



(a) by age

(b) by operator size

This figure presents the estimated percentage changes in bus yearly mileage in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^{+}|\Delta s^{+} > 0)\} - 1$) together with 90% confidence intervals. *Sample selection.* We estimate the effects on selected subsamples by considering only buses whose (a) age (b) operator size fall within a given group. We consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 270 days throughout year t in operator j's fleet to include observation ijt in the sample. While estimating the effects within groups defined by operator size, we only consider buses of age 4 and older. *Clustering.* We cluster standard errors at the brand×year level.

Figure 6 reports estimates as a function of bus age and operator size. Panel 6a confirms the intuition from the raw data: utilization gains associated with belonging to a more prevalent brand in the fleet grow over the bus life cycle. Importantly, variety already matters for new buses. For buses aged 0–3 years, the estimated effect is 3.4% and statistically significant at the 10% level. The point estimate nearly doubles for buses aged 4–9 years when warranties typically expire, and continues to increase in a concave manner, exceeding 8% for the oldest vehicles in the sample.

Panel 6b reports estimates as a function of operator size. Any operator belongs to only one of the size groups, as the group assignment is based on the minimum number of buses held by an operator over the entire sample period. In this way, we avoid sample selection concerns. We restrict attention to buses older than three years. The estimated effects are largest, exceeding 10%, among small operators and decline gradually with operator size. Nevertheless, even the largest operators experience gains from brand unification.

Taken together, these results reveal a clear and intuitive monotonicity pattern in the effects of fleet variety. This raises a natural follow-up question: are these effects driven by the extensive or intensive margin? That is, do the estimates reflect large changes driven by buses moving from near-zero utilization into active use as additional buses of the same brand enter the fleet, or do they capture more gradual increases in utilization among buses that are already in service? Relatedly, do the effects of variety differ when a class represents a small share of the fleet (e.g. 10%) versus a large share (e.g. 80%)? Figure 7 addresses these questions.

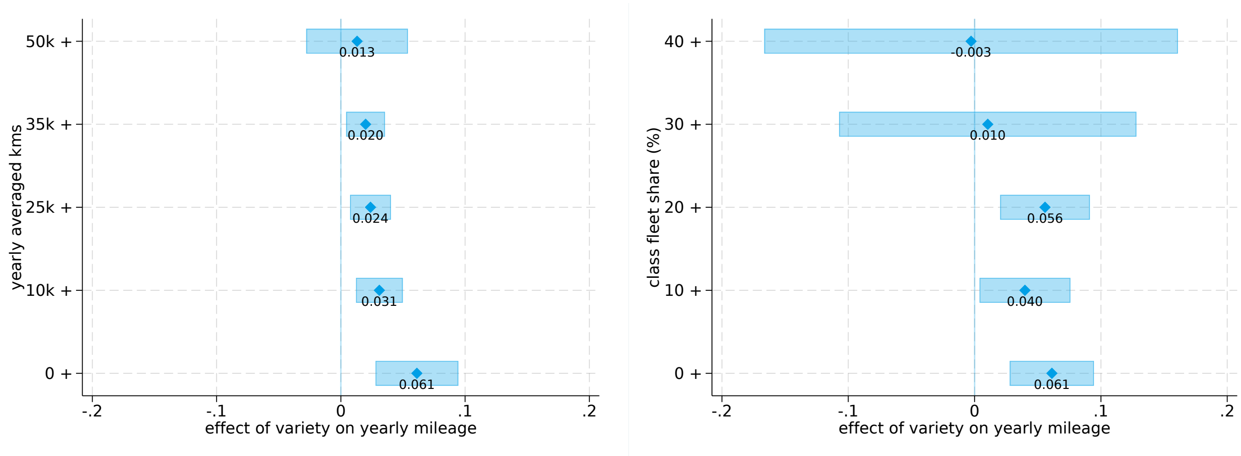
Panel 7a reports estimates obtained from subsamples defined by minimum yearly mileage thresholds. To avoid selection concerns, we require buses to exceed the threshold in every period they are observed. The estimated effect is largest in the full sample and declines monotonically as the threshold increases, reaching zero at approximately 50,000 km per year—the average (and median) yearly mileage (see Table 1). This pattern indicates that a substantial portion of the costs of variety operates through the extensive margin: buses that are nearly unused are brought into more regular service as their class becomes more prevalent in the fleet. A plausible mechanism is that greater availability of spare parts and accumulated know-how shifts repair priorities toward more common classes, reducing downtime for these buses.

At the same time, the intensive margin also plays a role. Even after excluding buses with less than 35,000 km per year,[14] we continue to find a statistically significant effect of about 2%. Finally, utilization is naturally bounded by route length and service frequency, implying limited scope for increases among already heavily used buses. Consistent with this constraint, the effects of variety are mechanically concentrated among underutilized vehicles, as reflected in Panel 7a.

Panel 7b reports estimates based on subsamples that progressively exclude buses whose brand's fleet share remains below a given threshold throughout the sample. As before, the estimated effect is strongest in the pooled sample. Raising the threshold to 10% or 20% has little impact on the magnitude of the effect, while further increases, above 30%, lead to estimates that are close to zero. At this point, it is difficult to distinguish whether the attenuation reflects a true absence of effects or a loss of statistical power due to the shrinking

[14] A yearly mileage of 35,000 km is a decent result. It corresponds to nearly 100 km per day. Given an average operating speed of about 20 km/h, this implies roughly five hours of daily service.

Figure 7: Heterogeneous effects of variety by utilization and share.



(a) by yearly mileage

(b) by fleet share of the class

This figure presents the estimated percentage changes in bus yearly mileage in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^{+}|\Delta s^{+} > 0)\} - 1$) together with 90% confidence intervals. *Sample selection.* We estimate the effects on selected subsamples by considering only buses whose (a) yearly mileage (b) class share in fleet exceeded a given threshold for every year in the sample. We consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 270 days throughout year t in operator j's fleet to include observation ijt in the sample. We only consider buses of age 4 and older. *Clustering.* We cluster standard errors at the brand×year level.

sample. Even under the conservative interpretation that the effect is genuinely zero beyond this threshold, the results imply strong incentives for operators to maintain at least a 30% fleet share for each brand. Notably, this threshold corresponds to operating with roughly three to four brands, which is consistent with the observed average of 3.7 brands per fleet reported in Table 2.

5.3 Variety Beyond Brand

So far, we have focused on fleet variety defined by bus brand. While this is an important dimension, it is clearly not the only source of heterogeneity in operators' fleets. In this section, we examine additional dimensions of bus heterogeneity by defining more granular classes that interact brand with other characteristics[15]. These include features related to the producer's identity, such as generation or engine brand, as well as features related to bus

[15] Recall that our identification strategy requires brand to remain part of the class definition.

purpose, including vehicle size defined through length categories and drive (e.g conventional, compressed natural gas, or electric).

Table 5: The effects of extended definitions of fleet variety on bus utilization.

	log yearly mileage				
	(1)	(2)	(3)	(4)	(5)
brand	.038 (.014)***				
brand × generation		.013 (.021)			
brand × engine brand			.020 (.024)		
brand × size				.049 (.019)***	
brand × drive					.035 (.018)*
N_i	11377	11377	11377	11377	11377
N_j	128	128	128	128	128
$N_{j'}$	129	129	129	129	129
N_c	37	99	90	88	63
t	2014-2024	2014-2024	2014-2024	2014-2024	2014-2024
effective F-stat.	10.169	12.491	3.342	9.757	5.542
underid. test (χ^2)	5.528 (.019)**	8.599 (.003)***	2.809 (.094)*	5.460 (.019)**	4.179 (.041)**
Anderson-Rubin test (LM)	3.975 (.046)**	.416 (.519)	.662 (.416)	3.344 (.067)*	3.024 (.082)*
Anderson-Rubin test (MD)	6.232 (.013)**	.433 (.511)	.724 (.395)	4.655 (.031)**	4.197 (.040)**

The table presents the estimated percentage changes in bus yearly mileage in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^{+}|\Delta s^{+} > 0)\} - 1$). *Sample selection.* We consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 90 days throughout year t in operator j's fleet to include observation ijt in the sample. *Clustering.* We cluster standard errors at the brand×year level *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

Table 5 reports the bus-level estimates. Column (1) replicates the baseline specification to facilitate comparison, while columns (2)–(5) assume augmented class definitions. Additional unification along dimensions related to the producer's identity beyond brand (brand×generation and brand×engine brand) does not appear to affect individual bus utilization. The corresponding estimates are small in magnitude and statistically insignificant, a conclusion that is reinforced by tests robust to weak instruments.

In contrast, variety along dimensions related to bus purpose yields different results. Expanding a brand's fleet share through the purchase of buses of the same size is associated with a larger increase in utilization, with the effect reaching 5%. Drive type also appears to matter, although the estimates are less precise. This is unsurprising given that more than 90% of observations correspond to diesel buses, which limits the variation available to identify effects related to drivetrain heterogeneity.

Table 6: The effects of extended definitions of fleet variety on fleet level outcomes.

	(1)	(2)	(3)	(4)	(5)
	# of buses				
brand	−.020 (.004)***				
brand × generation		−.037 (.008)***			
brand × engine brand			−.026 (.006)***		
brand × size				−.023 (.010)**	
brand × drive					−.024 (.005)***
effective F	17.598	3.313	4.354	4.742	11.660
underid. (χ^2)	3.882 (.049)**	2.963 (.085)*	2.780 (.095)*	4.204 (.040)**	5.734 (.017)**
AR test (LM)	2.837 (.092)*	3.518 (.061)*	2.177 (.140)	3.283 (.070)*	3.502 (.061)*
AR test (MD)	18.330 (.000)***	11.480 (.001)***	6.318 (.012)**	7.186 (.007)***	13.529 (.000)***
clustering	brand	brand	brand	brand	brand
	# of employees				
brand	−.001 (.003)				
brand × generation		−.003 (.007)			
brand × engine brand			−.002 (.011)		
brand × size				−.002 (.005)	
brand × drive					−.002 (.005)
N_j	78	78	78	78	78
$N_{j'}$	128	128	128	128	128
N_c	38	100	96	91	62
t	2007-2023	2007-2023	2007-2023	2007-2023	2007-2023
effective F	17.586	2.238	4.043	4.895	11.395
underid. (χ^2)	4.030 (.045)**	2.184 (.139)	2.668 (.102)	4.548 (.033)**	5.867 (.015)**
AR test (LM)	.139 (.709)	.181 (.671)	.045 (.832)	.120 (.729)	.098 (.755)
AR test (MD)	.164 (.685)	.203 (.652)	.050 (.823)	.125 (.723)	.100 (.752)
clustering	brand	brand	brand	brand	brand
	total cost				
brand	−.021 (.006)***				
brand × generation		−.049 (.031)			
brand × engine brand			−.029 (.013)**		
brand × size				−.025 (.008)***	
brand × drive					−.021 (.008)***
N_j	77	77	77	77	77
$N_{j'}$	128	128	128	128	128
N_c	38	100	96	91	62
t	2007-2023	2007-2023	2007-2023	2007-2023	2007-2023
effective F	11.203	1.694	4.095	5.303	8.753
underid. (χ^2)	4.297 (.038)**	2.192 (.139)	2.989 (.084)*	4.463 (.035)**	5.663 (.017)**
AR test (LM)	3.275 (.070)*	4.960 (.026)**	2.387 (.122)	5.578 (.018)**	3.435 (.064)*
AR test (MD)	12.472 (.000)***	10.236 (.001)***	4.941 (.026)**	8.507 (.004)***	10.808 (.001)***

The table presents the estimated percentage changes in fleet-level outcomes in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ \mid \Delta s^+ > 0)\} - 1)$. *Sample selection.* we include in the unit-level sample only those operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand level *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

Table 6 reports the fleet-level estimates. The results show a consistent pattern: further unification within a brand along dimensions related both to the producer's identity and to bus purpose yields larger gains in terms of reduced fleet size and total operating costs. As in the baseline analysis, we find no evidence of adjustment on the labor margin.

6 Discussion

This paper documents the costs of variety in capital inputs. We provide empirical evidence that differentiation among capital goods generates operational frictions that reduce firm performance. In our application, urban bus operators require larger fleets and incur higher costs to deliver a given level of service when their vehicles span a wider range of types. These frictions are multidimensional: for example, unifying fleets simultaneously across brand and bus size yields larger productivity gains than unification along a single dimension.

The sources of variety in our setting reflect institutional features such as public procurement rules that limit discretion in supplier choice. In many other markets, however, variety-related frictions persist even when firms are free to choose fleet suppliers. Technological change, supplier entry and exit, and deliberate product differentiation can all impede unification. In response, firms may accept supplier lock-in or diversify across suppliers to reduce dependence, both of which can entail efficiency losses.

Although our empirical setting focuses on urban buses, the underlying mechanisms are likely to be more general. Many industries rely on fleets of machines produced by different suppliers using distinct technologies. Even when capital goods are nominally similar, internal heterogeneity can complicate operation and maintenance, creating inefficiencies that are difficult to detect in aggregate data.

Taken together, our results highlight that capital is not an innocuous input: its productive value depends on internal compatibility. They also suggest that partial standardization, through organizational choices or policy interventions, can yield substantial efficiency gains even without requiring full uniformity.

Appendix

A SSIV Diagnostic Results

Here, we present a set of diagnostic results supporting our instrumental variable strategy.

A.1 Distribution of Exposure Weights

One requirement for shift-share IV designs is that the number of observed shocks grows with the sample size, or equivalently, that the Herfindahl-Hirschman index based on average shock exposure converges to zero as the sample size increases.

In our setting, the number of distinct shocks is large, as we exploit variation in fleet inflows across all operators, brands, and multiple years. However, a potential concern is that cities may cluster along their observable characteristics. In such a case, many exposure weights $\omega_{jj'}$ would be equal to zero, and the instrument value for a given operator j would effectively be driven by a small number of shocks.

To assess this concern, we examine the distribution of effective comparisons defined as the number of cities j' similar to operator j with positive exposure weights $\omega_{jj'}$, as well as the distribution of shift-level exposure weights across operators, $\tilde{\omega}_{j'} = \sum_k \omega_{kj'}$.

Table A.1: Distributions of exposure weights: order statistics.

	percentiles				
	min	25	50	75	max
# of similar cities ($\omega_{jj'} > 0$)	32	61	64	68	97
$\tilde{\omega}_{j'}$ (not-averaged)	0.64	0.975	1.004	1.052	1.133

The table presents the order statistics of the distribution of the number of similar cities for each unit-level operator j, and of the distribution of the shift-level exposure weights $\tilde{\omega}_{j'}$ (un-averaged, i.e. without taking into account the sample size).

Table A.1 reports order statistics for both measures. Every operator j in the unit-sample is matched with a large number of similar shift-level cities j'. Moreover, every shift-level city j' contributes as a comparison for some j while creating the instrument, as evidenced by the distribution of $\tilde{\omega}_{j'}$ concentrated away from zero. Its maximum is also relatively low, suggesting that there is no cities that would dominate in terms of exposure weights.

Numerically, note that the Herfindahl index based on average shock exposure would sum squares of $\tilde{\omega}_{j'}$ divided over the sample size (i.e. scaled by regression weights e_ℓ in BHJ terminology). With over 16 thousand buses, more than 100 operators, and many years of observation, the average exposure-weight-based HHI is going to be numerically very low on both bus and fleet level analysis. Hence, BHJ requirements regarding exposure weights seem to hold in our setting.

A.2 Falsification Tests

We follow the falsification approach proposed by BHJ to verify instrument validity. The tests regress proxies for unobserved confounding in the equation of interest on the instrument. Systematic correlation would raise concerns about validity.

We implement two such tests: a generic pre-trend test and an application-specific alternative. In the pre-trend test, we replace the outcome and covariates with their lagged values, which breaks the intended link between fleet composition and outcomes while preserving potential confounders. We consider lags of two and five years.

Table A.2 presents the results. Most of the pre-trend test estimated coefficients are statistically insignificant at any typically assumed confidence level. There are no systematic patterns of correlation across different lags. All of this speaks in favor of our instrumental variable strategy.

In the second test that we consider, we come back to the decomposition of changes in the fleet structure. Equation (2) shows that the evolution in the class fleet share is a sum of two elements: one related to the arrival of the new buses $\Delta^+ s_{cjt}$ and another related to the scrapping of existing buses $\Delta^- s_{cjt}$. In the main body of the paper, we argue that the latter depends purely on the operator's decisions, which makes it a perfect proxy for unobserved confounders. Intuitively, classes for some reason less preferred by an operator are likely to be scrapped more often.

To perform our test, we regress $\Delta^- s_{cjt}$ on the instrument $z_{cjt} = \sum_{j'} \omega_{jj'} s_{cj't}$ to seek a correlation between unobserved systematic drivers of performance and the instrument. Table A.3 shows that there is no such correlation. This result further supports the plausibility of our instrumental variable strategy.

Table A.2: Pre-trend falsification tests: The effects of current fleet variety on lagged bus- and fleet-level outcomes.

	bus-level	fleet-level		
	log yearly mileage	# buses	# employees	total cost
	(a) two-year lagged dependent variable			
brand	.023 (.015)	.014 (.008)*	.004 (.006)	.011 (.011)
N_i	10471			
N_j	128	78	78	77
$N_{j'}$	129	128	128	128
N_c	37	38	38	38
t	2016-2024	2009-2023	2009-2023	2009-2023
effective F	11.775	9.303	11.193	12.691
underid. (χ^2)	4.544 (.033)**	2.068 (.150)	2.229 (.135)	2.437 (.118)
AR test (LM)	2.083 (.149)	3.945 (.047)**	.821 (.365)	2.101 (.147)
AR test (MD)	2.798 (.094)*	4.992 (.025)**	.517 (.472)	1.311 (.252)
	(b) five-year lagged dependent variable			
brand	.052 (.031)*	−.004 (.018)	−.001 (.004)	.008 (.008)
N_i	7455			
N_j	128	76	76	75
$N_{j'}$	129	128	128	128
N_c	37	38	38	38
t	2019-2024	2012-2023	2012-2023	2012-2023
effective F	6.819	1.794	1.239	1.794
underid. (χ^2)	3.050 (.081)*	.926 (.336)	.756 (.384)	.948 (.330)
AR test (LM)	1.591 (.207)	.064 (.800)	.016 (.899)	1.734 (.188)
AR test (MD)	2.026 (.155)	.055 (.814)	.017 (.895)	1.647 (.199)

The table presents the estimated percentage changes in measures of lagged (2 and 5 years respectively) bus and fleet performance in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$). *Sample selection.* At the bus level, we consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 90 days throughout year t in operator j's fleet to include observation ijt in the sample. At the fleet level, we include in the unit-level sample only those operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand and brand×year level for fleet- and bus-level analysis respectively. *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

Table A.3: Falsification test based on Δs^-.

	$\Delta^- s_{cjt}$
z_{cjt} (brand)	−.363 (.498)
N_j	128
$N_{j'}$	128
N_c	38
t	2006-2023

The table presents estimated coefficient in the regression of $\Delta^- s_{cjt}$ on the instrument z_{cjt}. *Clustering.* We cluster standard errors at the brand level. *Presentation.* Standard errors obtained from the shift-level regression are given in parentheses. *P-values.* *** <.01,** <.05,* <.1.

References

Agrawal, A. (2024): "Empirical Analysis of Fleet Commonality Strategy's Impact on the Operating Performance of the Indian Aviation Sector," *Transportation Research Record*, p. 03611981241235230.

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.

Collard-Wexler, A., and J. De Loecker (2016): "Production function estimation with measurement error in inputs," .

De Borges Pan, A. G., and A. Espirito Santo Jr, Respicio (2004): "Developing a Fleet Standarization Index for Airline Pricing," *Journal of Air Transportation*, 9(2).

De Roux, N., M. Eslava, S. Franco, and E. Verhoogen (2021): "Estimating production functions in differentiated-product industries with quantity information and external instruments," Discussion paper, National Bureau of Economic Research.

Desai, P., S. Kekre, S. Radhakrishnan, and K. Srinivasan (2001): "Product differentiation and commonality in design: Balancing revenue and cost drivers," *Management science*, 47(1), 37–51.

Fisher, M. L., and C. D. Ittner (1999): "The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis," *Management science*, 45(6), 771–786.

Fox, J. T., and V. Smeets (2011): "Does input quality drive measured differences in firm productivity?," *International Economic Review*, 52(4), 961–989.

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," .

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.

MacDuffie, J. P., K. Sethuraman, and M. L. Fisher (1996): "Product variety and manufacturing performance: evidence from the international automotive assembly plant study," *Management science*, 42(3), 350–369.

Merkert, R. (2023): "The impact of engine standardization on the cost efficiency of airlines," *Research in Transportation Business & Management*, 46, 100797.

Merkert, R., and D. A. Hensher (2011): "The impact of strategic management and fleet planning on airline efficiency–A random effects Tobit model based on DEA efficiency scores," *Transportation Research Part A: Policy and Practice*, 45(7), 686–695.

Olea, J. L. M., 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*.

Robinson, J. (1953): "The Production Function and the Theory of Capital," *The Review of Economic Studies*, 21(2), 81–106.

Rust, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

Samuelson, P. A. (1966): "A Summing Up," *The Quarterly Journal of Economics*, 80(4), 568–583.

Sraffa, P. (1960): *Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory.* Cambridge University Press, Cambridge.

Statistics-Poland (2024): "Statistical Yearbook of the Republic of Poland 2024," *Warszawa.*

Trattner, A., L. Hvam, C. Forza, and Z. N. L. Herbert-Hansen (2019): "Product complexity and operational performance: A systematic literature review," *CIRP Journal of Manufacturing Science and Technology*, 25, 69–83.

Tybout, J. R. (1992): "Making noisy data sing: Estimating production technologies in developing countries," *Journal of econometrics*, 53(1-3), 25–44.

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

Zou, L., C. Yu, and M. Dresner (2015): "Fleet standardisation and airline performance," *Journal of Transport Economics and Policy (JTEP)*, 49(1), 149–166.

Supplementary Online Materials

B 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 generates operational frictions that raise life-cycle costs: depots must hold large inventories of non-interchangeable spare parts, maintenance staff require manufacturer-specific skills, and drivers face repeated retraining. Fleet variety thus reduces 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 acted as a coordinating buyer that aggregated demand across a large share of the national bus fleet. This coordination gave the association quasi-monopsonistic influence, allowing it to internalize the system-wide costs of excessive variety that individual operators could not address in isolation.

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, 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 successful coordinated standardization in fleet management. See the article by Achilles Lutz (2008), *1968–2008: Vor 40 Jahren erster Standard-Linienbus in Hamburg*, published in *HOV-Aktuell (No. 27)*, for a detailed historical account of the VÖV *Standard-Linienbus* and its role in coordinating bus design and procurement in West Germany.

C More on Bus Differentiation

This section considers alternative definitions of bus classes and the resulting measures of fleet variety. Table C.1 summarizes the bus characteristics that induce fleet heterogeneity. We also report Cramér's V statistics measuring the association between brand and alternative class definitions, verifying that brand captures a substantial share of the variation across them.

Table C.1: Bus Classes: Aspects of Bus Differentiation.

class	description	# categories	Cramér's V (×brand)
	classes related to producer's identity		
brand	the identity of bus manufacturer	37	1.000
generation	bus model reflecting a common technological and regulatory vintage (a subset of brand)	96	.984
engine brand	the identity of the engine manufacturer	29	.600
engine family	engine's equivalent of a generation	72	.656
	classes related to bus purpose		
size	categories based on the length of the bus (e.g. single bodied 12-meter, articulated 18-meter, ...)	5	.303
drive	defined by the type of fuel powering the drivetrain (e.g. diesel, compressed natural gas, electric, ...)	6	.429
floor	defined by the number and location of possible steps in the bus (e.g. low floor, low entry, ...)	3	.640
displacement	categories defined by the combined volume of air moved (displaced) by pistons in the cylinders of an internal combustion engine (typically between 6 and 12 liters)	5	.406

Note: The table presents definitions of the main aspects of bus differentiation considered in the paper. The last column presents the Cramér's V measure of association between two qualitative variables: brand and other chosen characteristics.

Classes related to producer identity capture differences in technological solutions chosen by manufacturers of buses and key components. These differences can be substantial, affecting core structural elements such as chassis design, engines, and interior layouts, with implications for operation, maintenance, and repair. The primary class in this category is bus *brand*, or in other words, the manufacturer's identity, but it is not the only one. Manufacturers periodically update their designs, allowing us to distinguish *generations* within a brand that share common technological features over a given production window. A further dimension is *engine brand*: engines are complex components often supplied by specialized manufacturers, distinct from the bus producer. Within engine brand, we also distinguish *engine families*, grouping engines that share key characteristics such as displacement or chassis integration.

A second type of classes groups buses by features that determine their use. City buses differ in *size*, captured by length categories ranging from small vehicles suited for narrow city streets to long articulated buses serving high-demand corridors. They also vary by *drive* describing propulsion technology, e.g. diesel, compressed natural gas, or electric, which is a key technical distinction and requires access to dedicated refueling infrastructure. Another important dimension is *floor* design: e.g. low-floor buses improve accessibility for passengers with disabilities. Finally, for buses with internal-combustion engines, *displacement* measures the total volume swept by the pistons and is also a relevant characteristic, as internal-combustion buses account for over 93% of our sample.

Table C.2 shows that the considered bus characteristics induce substantial heterogeneity across fleets. In a given year, on average, nearly 29 bus brands and 70 bus generations serve urban routes in Poland. Buses are equipped with engines produced by 22 different manufacturers, and the list of distinct bus features extends far beyond that. In total, if we define a super-type as the Cartesian product of all eight considered dimensions of heterogeneity, the market comprises over 280 distinct bus types.

Table C.2: Bus differentiation dimensions: summary statistics (Odometer sample, 2014-2024).

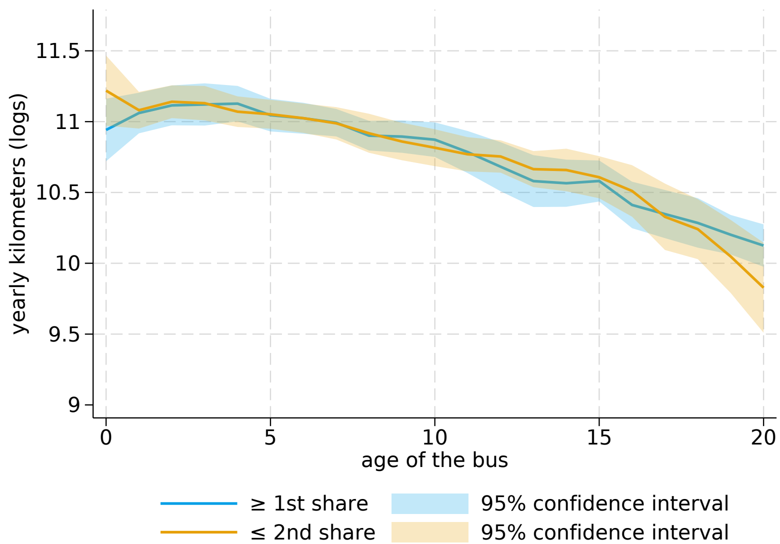
	market				fleet			
	#	1st share (the highest)	2nd share (2nd highest)	HHI	#	1st share (the highest)	2nd share (2nd highest)	HHI
	classes related to producer's identity							
brand	28.7	.46	.17	.27	3.7	.62	.24	.51
generation	69.7	.31	.13	.14	6	.48	.24	.36
engine brand	22.6	.32	.23	.2	4.2	.56	.25	.44
engine family	55.7	.21	.14	.1	6.2	.46	.23	.34
	classes related to bus purpose							
size	5	.54	.32	.4	2.8	.71	.22	.61
drive	4.6	.87	.06	.78	1.8	.9	.08	.86
displacement	5	.33	.31	.29	3.1	.62	.26	.52
floor	3	.88	.07	.8	2.2	.84	.13	.78
all types	280.9	.1	.1	.03	10.	.35	.19	.24

Note: The table presents the bus differentiation measures averaged over operators and years. We analyze the total number of bus types within each class, the fleet share of the first and second most represented group within the class, and a Herfindahl-Hirschman index of concentration within a class. The first four columns present statistics calculated at the market level, and the last four columns present firm-level statistics.

Table C.2 highlights also differences in fleet structure patterns between classes related to the producer's identity and classes related to the bus purpose. The former type of classes typically encompasses more distinct categories than the latter. Buses tend to be more standardized with respect to their purpose rather than their technological solutions. Moreover, differences between operator-level and market-level concentration measures are smaller when we consider classes related to bus purpose. This is consistent with the presumption that benefits from fleet unification with respect to bus purpose are more likely to be counterbalanced by potential gains of maintaining more classes that better suit the operator's portfolio of tasks.

Figure C.1 provides a piece of suggestive evidence indicating that classes based on bus purpose may not be the main source of the costs of variety.

Figure C.1: Bus utilization in life-cycle by brand fleet share.



Note: this graph presents estimated coefficients together with 95% confidence intervals from a linear regression where the dependent variable is log yearly mileage and the regressors are the full set of dummies for age category interacted with dummies indicating that a bus belongs to a given class within an operator's fleet (defined by size) whose fleet share in addition exceeds the 1st share or is below the 2nd share statistics (as defined in table 2). No constant is included, but we control for year fixed effects. Standard errors are clustered on the bus operator level. Warsaw is excluded from the analysis as an outlier.

Strikingly, the pattern of gains from fleet unification in the life cycle of a bus disappears when variety is defined along the bus-purpose dimension. As shown in Figure C.1, utilization rates do not differ significantly at any point in the bus life cycle between buses belonging to size categories that are common versus rare within an operator's fleet. Compare this result to the significant divergence in utilization rate between brands prevalent and rare in a fleet.

D Our Approach within the SSIV Framework

We demonstrate how our strategy fits within the BHJ framework. We begin with relaxing the standard *iid* assumption on the data-generating process. By doing so, we are able to treat shifts as random variables which create specific patterns of dependency in the instrument across *similar* cities (i.e. observations exposed to the same random shocks). Conditional on the controls described in the main body of the paper, the moment condition identifying parameters of the model is given by[16]:

$$\mathbb{E}\Big[\sum_i \sum_j \sum_t z_{c(i)jt}\mu_{ijt}\Big|\text{controls}\Big] = 0 \tag{D.1}$$

The unit-level orthogonality condition D.1 requires instruments to be orthogonal to the second-stage residuals, which capture systematic unobserved bus-level utilization shocks. Following BHJ, we can reinterpret this orthogonality conditions, changing the level of observation from the units to the shifts:

$$\mathbb{E}\Big[\sum_c \sum_{j'} \sum_t \tilde{\omega}_{j'} s_{cj't}\tilde{\mu}_{cj't}\Big|\text{controls}\Big] = 0 \tag{D.2}$$

where $\tilde{\omega}_{j'} = \sum_j \omega_{jj'}$ measures the total exposure of different bus operators j to operator j''s (at the shift-level, i.e. source of exogenous variation), and $\tilde{\mu}_{cj't} = \frac{\sum_j \omega_{jj'} \sum_{i\in c(j)} \mu_{c(i)jt}}{\sum_j \omega_{jj'}}$ is the average exposure-weighted unobserved total utilization shock within class c at time t among operators j that are *similar* to j' according to the weights $\omega_{jj'}$. Our identification assumption requires that conditional on the controls, these shocks are orthogonal to the fleet shares while weighted by the total exposure weights.

Expressing our approach in terms of the SSIV framework is also useful because it allows us to conveniently calculate standard errors from the shift level regression, i.e. at the source of exogenous variation, avoiding problems related to exposure to common shocks at the level of units.

[16] For notational simplicity, we suppress ε_{cjt} in this orthogonality condition as it is assumed to be idiosyncratic and hence by construction unrelated to the instruments.

D.1 Bus-Level Regressions

For simplicity of exposition, we suppress any covariates in our derivations. Consider bus i of class c at fleet j and period t. The fleet share of buses of the same class as i is given by $s_{ijt} \equiv s_{c(i)jt}$: share of buses of class $c(i)$ in fleet j at t, where $c(i)$ is class c of bus i. Suppressing other covariates, we are interested in the following equation:

$$y_{ijt} = \alpha s_{c(i)jt} + \underbrace{\eta_{ijt}}_{=\mu_{ijt}+\varepsilon_{ijt}}$$

where, as in the main text, y_{ijt} is a performance measure of bus i in fleet j and at time t, η_{ijt} is a composite error terms including μ_{ijt} summarizing unobserved systematic drivers of bus performance, and ε_{ijt} being an idiosyncratic component.

We consider an SSIV formulation of the instrument:

$$z_{ijt} \equiv z_{c(i)jt} = \sum_{j'} \omega_{jj'} s_{c(i)j't}$$

with $\omega_{jj'} \geq 0$, $\omega_{jj} = 0$, $\sum_{j'} \omega_{jj'} = 1$ for every j.

The identifying moment condition requires no relation between the instrument and the unobservables at the unit level. Departing from this formulation, we can also derive the identification restriction at the level of shifts:

$$\begin{aligned}
\mathbb{E}\Big[\sum_i \sum_j \sum_t z_{ijt}\mu_{ijt}\Big] &= \mathbb{E}\Big[\sum_i \sum_j \sum_t \sum_{j'} \omega_{jj'} s_{c(i)j't}\mu_{ijt}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_j \sum_{c(j)\in j} \sum_{i\in c(j)} \omega_{jj'} s_{c(j)j't}\mu_{ijt}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_j \sum_{c(j)\in j} \omega_{jj'} s_{c(j)j't} \underbrace{\sum_{i\in c(j)} \mu_{ijt}}_{=\overline{\mu}_{c(j)jt}}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_j \underbrace{\sum_c \omega_{jj'} s_{cj't}\overline{\mu}_{cjt}}_{\text{wlog assume } \overline{\mu}_{cjt}=0 \text{ for } c\neq j}\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \sum_j \omega_{jj'} s_{cj't} \overline{\mu}_{cjt} \underbrace{\frac{\sum_k \omega_{kj'}}{\sum_k \omega_{kj'}}}_{=\tilde{\omega}_{j'}}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \underbrace{\frac{\sum_j \omega_{jj'} \overline{\mu}_{cjt}}{\sum_j \omega_{jj'}}}_{=\tilde{\mu}_{cj't}}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \tilde{\mu}_{cj't}\Big] \\
&= 0
\end{aligned}
$$

In other words, the identification requires fleet shares of c to be independent from the average total measure of utilization shocks $\tilde{\mu}_{cj't}$ within class c among cities to which j' is similar, at every period. $\tilde{\mu}_{cj't}$ is average unobserved total utilization shock within class c at time t among operators j that are *similar* to j'. The identification requires these are orthogonal to the j' own class c shares.

$$
\tilde{\mu}_{cj't} = \frac{\sum_j \omega_{jj'} \overline{\mu}_{cjt}}{\sum_j \omega_{jj'}} = \frac{\sum_j \omega_{jj'} \sum_{i \in c(j)} \mu_{c(i)jt}}{\sum_j \omega_{jj'}}
$$

Note that this shift-level statement is equivalent to the unit-level statement of the instrument being unrelated to the unobservable.

In the same fashion, we can derive the shift-level regression coefficient.

$$
\begin{aligned}
\hat{\beta}^{SSIV} &= \frac{\sum_i \sum_j \sum_t z_{ijt} y_{ijt}}{\sum_i \sum_j \sum_t z_{ijt} s_{ijt}} \\
&= \frac{\sum_i \sum_j \sum_t \sum_{j'} \omega_{jj'} s_{c(i)j't} y_{ijt}}{\sum_i \sum_j \sum_t \sum_{j'} \omega_{jj'} s_{c(i)j't} s_{ijt}} \\
&= \frac{\sum_t \sum_{j'} \sum_j \sum_{c(j) \in j} \omega_{jj'} s_{c(j)j't} \sum_i y_{ijt}}{\sum_t \sum_{j'} \sum_j \sum_{c(j) \in j} \omega_{jj'} s_{c(j)j't} \sum_i s_{ijt}} \\
&= \frac{\sum_t \sum_{j'} \sum_j \sum_c \omega_{jj'} s_{cj't} \overline{y}_{cjt}}{\sum_t \sum_{j'} \sum_j \sum_c \omega_{jj'} s_{cj't} \overline{s}_{cjt}} \\
&= \frac{\sum_t \sum_{j'} \sum_j \sum_c \omega_{jj'} s_{cj't} \overline{y}_{cjt} \frac{\sum_k \omega_{kj'}}{\sum_k \omega_{kj'}}}{\sum_t \sum_{j'} \sum_j \sum_c \omega_{jj'} s_{cj't} \overline{s}_{cjt} \frac{\sum_k \omega_{kj'}}{\sum_k \omega_{kj'}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \frac{\sum_j \omega_{jj'} \overline{y}_{cjt}}{\sum_j \omega_{jj'}}}{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \frac{\sum_j \omega_{jj'} \overline{s}_{cjt}}{\sum_j \omega_{jj'}}} \\
&= \frac{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \tilde{y}_{cj't}}{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s_{cj't} \tilde{s}_{cj't}}
\end{aligned}
$$

where $\overline{y}_{cjt} = 0$ for cjt at which $c \notin j$.

The definitions of shift-level variables are:

$$
\tilde{y}_{cj't} = \frac{\sum_j \omega_{jj'} \overline{y}_{cjt}}{\sum_j \omega_{jj'}} = \frac{\sum_j \omega_{jj'} \sum_{i \in c(j)} y_{ijt}}{\sum_j \omega_{jj'}}
$$

And:

$$
\tilde{s}_{cj't} = \frac{\sum_j \omega_{jj'} \overline{s}_{cjt}}{\sum_j \omega_{jj'}} = \frac{\sum_j \omega_{jj'} \sum_{i \in c(j)} s_{c(i)jt}}{\sum_j \omega_{jj'}}
$$

D.2 Fleet-Level Regressions

Derivations for the fleet-level analysis are analogous. Abstracting of potential covariates, our working equation is:

$$
y_{jt} = \alpha \underbrace{HHI_{jt}}_{=\sum_c s_{cjt}^2} + \underbrace{\mu_{jt}}_{=\mu_{jt} + \varepsilon_{jt}}
$$

where the components are defined in the same way as for the bus-level equation.

Our instrument is given by:

$$
z_{jt} = \sum_{j'} \omega_{jj'} \sum_c s_{cj't}^2
$$

The identification restriction requires no relation between the instrument and fleet-level class HHI:

$$
\begin{aligned}
\mathbb{E}\Big[\sum_j \sum_t z_{jt} \mu_{jt}\Big] &= \mathbb{E}\Big[\sum_j \sum_t \sum_{j'} \omega_{jj'} \sum_c s_{cj't}^2 \mu_{jt}\Big] \\
&= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_j \sum_c \omega_{jj'} s_{cj't}^2 \mu_{jt}\Big]
\end{aligned}
$$

$$= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \sum_j \omega_{jj'} s^2_{cj't} \mu_{jt} \underbrace{\frac{\sum_k \omega_{kj'}}{\sum_k \omega_{kj'}}}_{=\tilde{\omega}_{j'}}\Big]$$

$$= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s^2_{cj't} \underbrace{\frac{\sum_j \omega_{jj'} \mu_{jt}}{\sum_j \omega_{jj'}}}_{=\tilde{\mu}_{j't}}\Big]$$

$$= \mathbb{E}\Big[\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s^2_{cj't} \tilde{\mu}_{j't}\Big]$$

$$= 0$$

Therefore, to ensure identification, we need (the squares of) fleet shares of c to be independent from the average fleet-level performance shock $\tilde{\mu}_{j't}$ among cities to which j' is similar, at every period. These shocks are defined by

$$\tilde{\mu}_{j't} = \frac{\sum_j \omega_{jj'} \mu_{jt}}{\sum_j \omega_{jj'}}$$

Note that this shift-level statement is equivalent to the unit-level statement of the instrument being unrelated to the unobservable.

Using the same transformation from unit to shift level but applied to the dependent and endogenous variable in our equation, we derive the shift-level regression coefficient:

$$\hat{\beta}^{SSIV} = \frac{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s^2_{cj't} \tilde{y}_{j't}}{\sum_t \sum_{j'} \sum_c \tilde{\omega}_{j'} s^2_{cj't} H\tilde{H}I_{j't}}$$

E Additional Results

In this section, we present additional results that support the results described in the main body of text.

E.1 Labor Decomposition

In the main text, we treat buses as a differentiated capital input and labor as the non-differentiated input. However, one may claim that labor is actually a differentiated input too, as we can distinguish three main types of employees of our bus operators: drivers, mechanics, and others. Table E.3 shows that the fleet variety does not influence any of these disaggregated labor inputs.

Table E.3: The effects of fleet variety defined by brand on operator's labor structure.

	fleet-level		
	# employees all	# employees drivers	# employees mechanics
effect of variety	−.001 (.003)	−.001 (.004)	.007 (.010)
N_j	78	78	78
$N_{j'}$	128	128	128
N_c	38	38	38
t	2007-2023	2007-2023	2007-2023
effective F-stat.	17.586	18.313	16.603
underid. test (χ^2)	4.030 (.045)**	3.619 (.057)*	3.585 (.058)*
Anderson-Rubin test (LM)	.139 (.709)	.039 (.843)	.441 (.507)
Anderson-Rubin test (MD)	.164 (.685)	.043 (.835)	.582 (.445)

The table presents the estimated percentage changes in fleet-level labor components in response to a median change towards unification along bus brand in fleet structure, expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$. *Sample selection.* We only consider operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand level. *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. P.vals:*** <.01,** <.05,* <.1.

E.2 Clustering

Our identification variation comes from bus providers' market strategies. Given the specifics of the market, these decisions are most typically made at the level of bus producer, or brand. Hence, the preferred level at which to cluster standard errors would be the brand level. In principle, this is possible at every equation we estimate, as brand is used in every definition of class considered.

However, the number of brands, and hence classes in our classification, is limited. As a result, we would need a relatively long t dimension to be able to get enough statistical power. This is why, at the bus-level analysis, we cluster standard errors at the brand×year level. This way of clustering adds more clusters, which enhances statistical power. In turn,

Table E.4: The effects of extended definitions of fleet variety on bus utilization with standard errors clustered at the brand level.

	(1)	(2)	(3)	(4)	(5)
	# of buses				
brand	.038 (.011)***				
brand × generation		.013 (.025)			
brand × engine brand			.020 (.018)		
brand × size				.049 (.012)***	
brand × drive					.035 (.012)***
N_i	11377	11377	11377	11377	11377
N_j	128	128	128	128	128
$N_{j'}$	129	129	129	129	129
N_c	37	99	90	88	63
t	2014-2024	2014-2024	2014-2024	2014-2024	2014-2024
effective F-stat.	1.467	7.380	.832	1.284	1.175
underid. test (χ^2)	.693 (.405)	2.243 (.134)	.784 (.376)	.885 (.347)	.733 (.392)
Anderson-Rubin test (LM)	.929 (.335)	.405 (.525)	.902 (.342)	.716 (.397)	.940 (.332)
Anderson-Rubin test (MD)	2.684 (.101)	.367 (.545)	1.206 (.272)	1.033 (.309)	1.843 (.175)

The table presents the estimated percentage changes in bus yearly mileage in response to a median change in fleet structure oriented towards unification along bus brand (expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$). *Sample selection.* We consider only buses that make at least 25km in a year to remove units that are completely not suitable for utilization. We also remove the Warsaw operator's buses from our unit-level sample as a clear outlier. We require bus i to be observed within at least 90 days throughout year t in operator j's fleet to include observation ijt in the sample. *Clustering.* We cluster standard errors at the brand level *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

as the identification argument relies predominantly on the evolution of shares but not levels themselves, the assumption of errors being independent across years is not that controversial.

As an additional robustness check, E.4 presents the estimation results of the main bus-level coefficients under the brand-level clustering. The patterns of parameter significance remain the same, leading to the same conclusions. However, we do not have enough power to conduct the weak instruments testing as depicted in the lowest panel of the table.

E.3 Fleet-Level Measures Normalized by the Output

In our fleet-level specifications, we consider a logarithm of an input (fleet size, # of employees) or a cost as the dependent variable, and condition on the logarithm of total output. This formulation is convenient as it delivers an intuitive interpretation of the coefficient of interest in terms of semi-elasticity. However, it is correct only under the assumption that the total output is exogenous in our setting, which we argue it is.

Table E.5: The effects of variety on fleet-level outcomes expressed per unit of operator's total output.

	fleet-level		
	# buses	# employees	total cost
brand	−.016 (.004)***	−.007 (.006)	−4.878 (1.474)***
N_j	78	78	77
$N_{j'}$	128	128	128
N_c	38	38	38
t	2007-2023	2007-2023	2007-2023
effective F-stat.	17.966	18.372	13.195
underid. test (χ^2)	3.499 (.061)*	3.637 (.056)*	3.940 (.047)**
Anderson-Rubin test (LM)	2.217 (.137)	.750 (.386)	2.941 (.086)*
Anderson-Rubin test (MD)	15.162 (.000)***	1.308 (.253)	12.704 (.000)***

The table presents the estimated coefficients α in equations in which fleet-level outcomes are expressed as values per unit of output. *Sample selection.* we include in the unit-level sample only those operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand level *Presentation.* Standard errors obtained from the shift-level regression and are given in parentheses under the estimated effects, and p-values under the test statistics. *P-values.* *** <.01,** <.05,* <.1.

Another way of looking at the effects of variety at the fleet level, one that is robust to the potential output endogeneity, is to redefine the left-hand side variables as inputs (or cost) per unit of output. Table E.5 presents the estimated coefficients. Even though these estimates' units are not comparable with the estimates from the main text, the patterns of statistical significance are the same. Hence, if we were using these results in place of the original specification, we would have reached the same conclusions.

E.4 2018-2019 outliers

Our fleet level analysis of the fleet size implicitly requires that the operators predominantly introduce new buses to replace others rather than expand the fleet. If the latter is true, the fleet size would mechanically increase even if the fleet brand HHI increased and more homogeneous fleets require fewer buses.

As figure 2 in the main body of text shows, our operators tend to keep the number of buses roughly constant, which implies that purchases are predominantly made towards fleet renewal, not expansion. However, two years: 2018 and 2019 seem to break this pattern. As a robustness check, we re-estimate the main result at the fleet level on data in which we remove observations in 2018 and 2019.

Table E.6: The effects of variety on fleet-level outcomes expressed per unit of operator's total output: estimates excluding fleet expansion period 2018-2019.

	fleet-level		
	# buses	# employees	total cost
brand	$-.033$ (.008)***	.000 (.006)	$-.014$ (.007)*
N_j	78	78	77
$N_{j'}$	128	128	128
N_c	38	38	38
t	2007-2017, 2020-2023	2007-2017, 2020-2023	2007-2017, 2020-2023
effective F	6.838	7.195	5.156
underid. (χ^2)	4.993 (.025)**	5.276 (.022)**	5.149 (.023)**
AR test (LM)	3.966 (.046)**	.001 (.975)	3.067 (.080)*
AR test (MD)	21.5 (.000)***	.001 (.974)	4.675 (.031)**

The table presents the estimated percentage changes in fleet-level labor components in response to a median change towards unification along bus brand in fleet structure, expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$. *Sample selection.* We exclude years 2018 and 2019 from the sample. We only consider operators who maintained at least 20 buses at every point in the sample. *Clustering.* We cluster standard errors at the brand level. *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. P.vals:*** <.01,** <.05,* <.1.

Table E.6 presents the results. As expected, the effect on fleet size becomes stronger in magnitude, reaching 3.3%.

E.5 Considering all operators at the fleet level

Our main result relies on a subsample of operators for which the observed fleet size amounts to at least 20. We focus on larger bus operators as the adjustment at the fleet level measures is more meaningful among them. In particular, small operators are less likely to adjust their fleet at all. Our effect of 2.1% would translate into a negligible fraction of a bus if the operator is very small.

To confirm these suppositions, we re-estimate the model using all available unit-level operators at the fleet-level analysis (recall that availability is restricted by voluntary participation in the IGKM survey). Table E.7 presents the results. Compared to our main result (obtained on a sample of *larger* operators), the effect on fleet size becomes small in magnitude and statistically insignificant, consistent with our intuitions. Notably, the effect on operator's costs nearly doubles. This suggests that small operators with heterogeneous fleets put significantly more effort in keeping buses they already have in service instead of

Table E.7: The effects of variety on fleet-level outcomes expressed per unit of operator's total output: estimates obtained on all available unit-level observations.

	fleet-level		
	# buses	# employees	total cost
brand	$-.006$ (.014)	$-.001$ (.005)	$-.041$ (.017)**
N_j	114	113	110
$N_{j'}$	128	128	128
N_c	38	38	38
t	2007-2023	2007-2023	2007-2023
effective F	3.367	6.740	3.326
underid. (χ^2)	1.521 (.218)	2.099 (.147)	1.535 (.215)
AR test (LM)	.132 (.717)	.050 (.824)	3.858 (.050)**
AR test (MD)	.165 (.684)	.057 (.812)	20.4 (.000)***

The table presents the estimated percentage changes in fleet-level labor components in response to a median change towards unification along bus brand in fleet structure, expressed by $\exp\{\alpha \cdot \text{median}(\Delta s^+ | \Delta s^+ > 0)\} - 1$. *Clustering.* We cluster standard errors at the brand level. *Presentation.* Standard errors obtained from the shift-level regression and based on delta method are given in parentheses under the estimated effects, and p-values under the test statistics. P.vals:*** $<.01$,** $<.05$,* $<.1$.

increasing the number of buses. This makes sense, as for small operators the relative cost of an additional vehicle is likely disproportionately higher than for a large operator.