Where Did the Money Go? Transit Project Selection Under ARRA

Arseniy Braslavskiy, University of Maryland

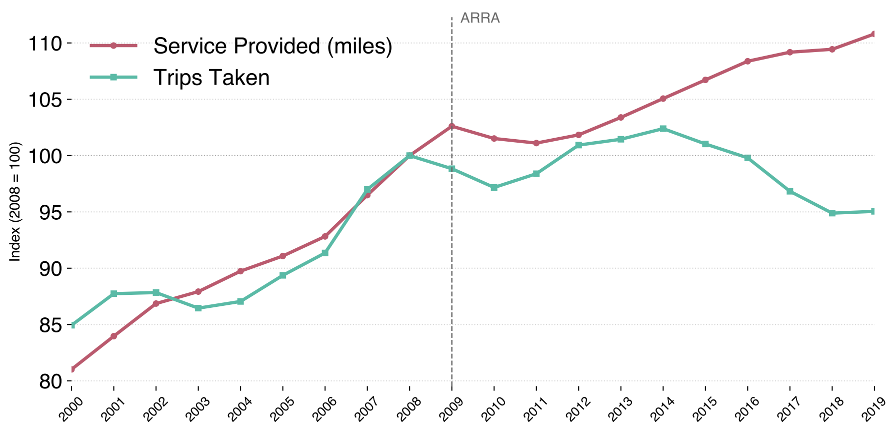
July 4, 2026

Abstract. I study how local transit agencies spent $8.4 billion of one-time capital from the 2009 stimulus, distributed by a formula and left to local discretion. Using a large language model to extract project-level data from grant texts, I find that 22 percent of projects added service capacity, 30 percent improved the quality of existing service, and 48 percent maintained it, while agencies' pre-existing characteristics explain almost none of this split. On the other hand, the type of spending mattered for the outcomes: areas that spent on expansion fared no differently from those that spent on maintenance, while areas that spent on quality improvement provided more service afterward. A simple model explains why. A system with excess capacity already carries the riders its quality attracts, so added vehicles sit unused and only an improvement in quality can attract new passengers. Empirical results thus suggest that the average American transit system overinvests in capacity and underinvests in quality. A back-of-the-envelope reallocation toward quality implies roughly 3 to 4 percent more national ridership.

1 Introduction

Public transit in the United States is financed largely by the federal government but operated locally. Federal formula grants, tied to an area's population, density, and existing service, fund a large share of transit capital and leave the choice of projects to local agencies. The formula responds to the scale of a system but not to the condition of its assets or the unmet demand it faces. Behind this design lies a common premise: that transit is underprovided, so that funding more service will raise ridership. The aggregate record sits awkwardly with that premise, as agencies have expanded service for decades while ridership has been flat or falling (Figure 1). I study how agencies used one large and plausibly exogenous infusion of federal capital, and I find that they directed it mostly toward maintenance and quality rather than capacity, that their choices were almost unrelated to the conditions they faced, and that the spending which raised service was quality improvement rather than expansion.

Figure 1: Transit service has grown while ridership has stagnated.



Note: Indexed to 2008 = 100. *Source:* National Transit Database (NTD) Time Series, all modes, non-rural reporters.

The infusion is the American Recovery and Reinvestment Act of 2009 (ARRA), which provided transit with $8.4 billion in capital in a single year and raised federal transit appropriations by about 83 percent relative to their pre-recession level (Figure 2). ARRA transit funds were distributed through the existing formula programs, required no local match, and gave agencies wide discretion over project choice. The episode is therefore a clean case of capital allocated by scale and spent at local discretion. In a companion paper, I show that these funds raised investment, with each ARRA dollar increasing total transit capital spending by roughly two dollars over the following decade (Braslavskiy, 2026). This paper addresses a different question: not how much agencies spent, which the formula fixes, but what they bought and whether it affected service and ridership. The question concerns the allocation of capital rather than its amount.

I measure what agencies bought from the text of their grant awards. Each award includes a free-form description of the funded projects, filed through FTA Form 1512; these filings are the primary data source and cover 954 awards. Because the descriptions are unstructured and vary widely in length and terminology, I use a large language model with structured outputs to convert each award into data in two steps, first splitting it into individual projects and then classifying each project by the assets it involves and the intent of the spending. This yields 3,551 classified projects, which I aggregate to urbanized areas and link to a long panel of National Transit Database outcomes. My identifying variation is the orientation of an agency's spending, the share allocated to expansion, improvement, or maintenance. As I show below, this orientation is largely unrelated to the structural characteristics of a system, so conditional on those characteristics the variation in spending type is plausibly as good as randomly assigned.

The funds arrived as an open-ended block of capital that agencies had not requested. Of the 3,551 classified projects, 22 percent add service capacity, 30 percent improve the quality of existing service, and 48 percent sustain it through replacement and routine maintenance. The assets purchased in each category match these labels: expansion concentrates in new buses and passenger facilities, improvement in software and passenger-facing upgrades, and maintenance in like-for-like vehicle replacement and facility rehabilitation.

What agencies chose to spend on bears little relation to the system or the city it serves. I assemble 59 pre-ARRA variables covering demographics, agency finances, system operations, and infrastructure, and reduce them to 14 principal components that retain about 80 percent of the variance. Together these components explain between 4 and 10 percent of the variation in how an area allocated its ARRA funds. The two leading components, capturing system scale and affluence and jointly accounting for a third of the variance in the underlying variables, have no predictive power for spending orientation. This is not driven by areas with only one or two awards: among areas with at least three, where the measure is well identified, the components still leave more than three-quarters of the variation unexplained. If agencies were allocating capital toward its highest return, the variables that proxy for need and demand would carry most of the predictive weight; instead they carry almost none, leaving spending orientation close to idiosyncratic. This idiosyncrasy serves two purposes. It supports the identification strategy by making the residual variation plausibly exogenous, and it provides evidence that a formula based on scale produces spending only loosely tied to need.

Idiosyncratic though these choices were, the type of spending mattered for outcomes. I regress the post-2009 change in an area's transit outcomes on the share of ARRA funds in each category, controlling for the principal components and the award amount. Areas that allocated ARRA to expansion show no differential change in service or ridership relative to areas that allocated it to maintenance, and the result holds when expansion is restricted to projects with an explicitly stated capacity goal. Areas that allocated ARRA to quality improvement provide significantly more service afterward, measured in vehicle revenue miles, with operating cost and peak vehicles in service rising alongside it. Pre-2009 coefficients are flat and the effect appears at 2009, which I read as evidence that it is causal. The estimate comes from about 350 areas and is significant at the 10 percent level.

A simple model rationalizes why improvement raises service and expansion does not. A transit system already carries the riders that its quality attracts, so at its current fares and quality it operates with excess capacity rather than turning riders away. Expansion adds to this unused capacity and leaves equilibrium service unchanged. Improvement, by making the service more reliable and comfortable, attracts new riders, and service rises to meet them. The binding constraint is demand at the offered quality, not the supply of vehicles. The pattern suggests that the average U.S. transit system overinvests in capacity and underinvests in quality. The model has a policy implication, which I quantify with a back-of-the-envelope reallocation: had ARRA been directed entirely toward improvement, the estimates imply about 13 percent more transit service and 3 to 4 percent more ridership nationally than the allocation agencies chose. A formula that funds scale, on the assumption that supply creates its own demand, does not direct capital to where it would do the most good.

This paper contributes to four literatures. The first studies intergovernmental grants and the discretion they grant to recipient governments. A large body of work estimates how much a transferred dollar raises spending in the targeted category, with estimates ranging from full crowd-out to the more-than-proportional pass-through known as the flypaper effect (Courant et al., 1979; Hines and Thaler, 1995; Inman, 2009; Knight, 2002; Gordon, 2004; Baicker, 2001). I shift the focus from how much grant money is spent to how it is spent: given open-ended formula funds, which projects do local agents select, and is that selection aligned with the program's objective? My companion paper on these same ARRA funds takes up the first question (Braslavskiy, 2026); this paper takes up the second, showing that the composition of the spending was largely detached from need and that its returns ran through a single margin.

Second, I contribute to the literature on optimal transit funding and the returns to transit capital. This work documents the external benefits of transit, including reduced congestion and emissions (Anderson, 2014; Parry and Small, 2009) and improved mobility for low-income riders (Glaeser et al., 2008), but says little about how capital should be allocated across systems or project types. My estimates speak to one case: where a system already meets demand at its current quality, an additional capital dollar raises ridership only if it improves service quality, not if it adds capacity the system already holds in excess. This connects to the induced-demand literature, in which usage is governed by demand rather than by the capacity supplied (Duranton and Turner, 2011; Downs, 1962).

Third, the paper makes a methodological contribution to measurement. The composition of capital spending is recorded only in unstructured grant narratives, which have not been usable for quantitative analysis at scale. I use a large language model to convert several thousand free-text project descriptions into structured classifications. The approach applies to the many settings in which administrative text contains economically meaningful detail absent from conventional data.

Fourth, I add to the evidence on the design of fiscal stimulus. Research on ARRA has focused primarily on employment and aggregate demand (Chodorow-Reich et al., 2012; Wilson, 2012; Leduc and Wilson, 2017) and has said much less about whether the projects it financed were well chosen. By examining one program's project selection, I show that a stimulus can succeed in disbursing funds quickly while still misallocating them, a distinction relevant to the design of future countercyclical infrastructure programs.

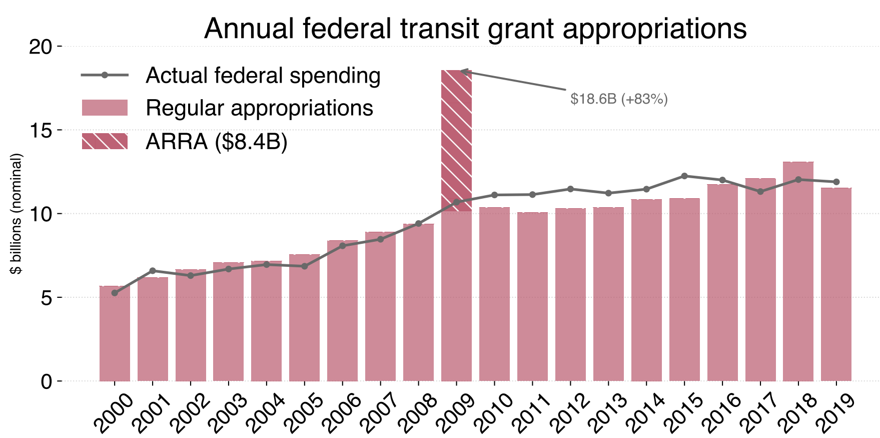
The rest of the paper is organized as follows. Section 2 describes the award text, the language model pipeline that turns it into project-level data, and the validation of that pipeline. Section 3 profiles the transit systems and the places they serve and reduces the 59 pre-ARRA variables to the principal components used in the rest of the paper. Section 4 asks how well those variables predict the orientation of an area's ARRA spending and finds that they predict little. Section 5 estimates the effect of spending type on service and ridership. Section 6 develops the model that explains the results and uses it for a reallocation counterfactual. Section 7 concludes.

2 Text Analysis of Transit Awards

I measure how transit agencies spent a large block of federal capital funds by reading the text of their grant awards. Each award carries a free-form description of the projects it funded, the only project-level record of what the money bought. The text does not lend itself to analysis: descriptions run from two words ("Fleet Replacements") to several paragraphs, with terminology that varies across agencies and regions. I use GPT-5.2, a large language model, with structured outputs to turn each award into data in two steps. First I split the description into individual projects. Then I classify each project by the assets it involves and by the likely intent of the spending.

That block of funds is the American Recovery and Reinvestment Act, which delivered an $8.4 billion tranche of transit capital funding in a single year and raised federal transit appropriations by roughly 83 percent above their pre-recession level (Figure 2). Recipients reported how they used the money through FTA Form 1512 filings, the primary data source for this paper, which record 954 awards.

Figure 2: ARRA produced a one-year spike in federal transit capital funding.



Note: Shaded bar isolates the $8.4 B ARRA tranche; solid line shows actual federal transit spending. Nominal dollars. *Source:* FTA apportionments; spending line from the National Transit Database. All agencies.

2.1 Project Identification

A single award filing frequently bundles several unrelated activities under one description: a bus purchase, a shelter renovation, and a software upgrade may appear in a single paragraph.

To recover distinct projects, I prompt the model to parse each award description and list the individually funded actions.

The extraction prompt instructs the model to produce one entry per distinct funded action in the form *verb + object* (e.g., "purchase 12 buses," "rehabilitate downtown terminal"), excluding motivations, compliance language, status or timing details, and capability lists that describe context rather than the funded activity itself. Qualifiers that change the nature of an action, such as replacement or addition status and explicit quantities, are kept when the source states them. Place names are omitted. Where an award states a dollar amount for a specific project (typically in bracketed form within the text), that amount is attached to the corresponding project entry. Awards that do not itemize spending are allocated equally across the identified projects.

Applied to the full sample, the extraction yields 3,763 distinct projects across the 954 awards.[1]

2.2 Project Classification

Each extracted project is classified in two complementary stages, one capturing what assets were involved and another capturing the likely intent or effect of the spending.

Capital component characterization (M1). The first stage extracts 38 binary variables describing the factual content of each project. They cover rolling stock (coded both in aggregate and separately for buses and railcars), guideway infrastructure, passenger-facing facilities, operations and maintenance facilities, and software or communications systems. Within each category, the model codes whether the asset appears at all, whether new units are purchased or constructed, whether the purchase is a replacement for an existing asset, whether the purchase is *exclusively* replacement, whether the asset is maintained for state-

[1] A subsequent review of the extracted project descriptions using Claude identified three categories requiring correction. First, 32 descriptions conflated a replacement and an expansion activity in a single entry (e.g., "purchase 5 buses (4 replacement, 1 for expansion)"); these were split into two separate rows and reclassified by the M1/M2 pipeline. Second, 213 descriptions corresponded to non-capital spending (primarily operating assistance, training, or federal oversight activities), accounting for 3.4% of total project funding; these are excluded from the classification and aggregation steps. Third, 24 descriptions were too generic to classify (e.g., "Transit Capital Assistance," "Invest in public transportation"), representing 2.2% of project funding; these are also excluded. Seven additional rows were deleted as duplicates or extraction artifacts. The cleaned dataset used in the analysis contains 3,788 project rows, larger than the initial 3,763 because each of the 32 split descriptions became two rows, partially offset by the seven deletions. After excluding the non-capital and generic descriptions, 3,551 projects remain for classification.

of-good-repair, and whether capabilities are upgraded. The replacement versus addition distinction is central: a bus purchase coded as replacement-only leaves fleet size unchanged, while a purchase without that qualifier expands it.

Service-intent classification (M2). The second stage assigns each project to exactly one of three mutually exclusive categories. The classifier receives both the full award description (which situates the project within the broader capital program) and the extracted project description, with explicit instructions to use award context only to resolve ambiguity about the specific project and not to infer expansion intent from co-listed projects in the same award.

The first dimension separates *expansionary* from *non-expansionary* projects. Expansionary projects add new service capacity: new lines, routes, stops, or vehicles net of replacement. Within the expansionary category, the model also records whether intent is *explicit* (the project or award description uses language such as "expansion," "new service," or "new route," or involves construction of passenger infrastructure that did not previously exist) or *implicit*, meaning a vehicle or asset acquisition with no mention of replacement, where the capacity addition is inferred from that silence. When the award context shows the purchase replaces an existing asset, the project is non-expansionary; when neither the project nor the award mentions replacement, it is implicitly expansionary. The explicit subcategory serves as a stricter alternative treatment measure.

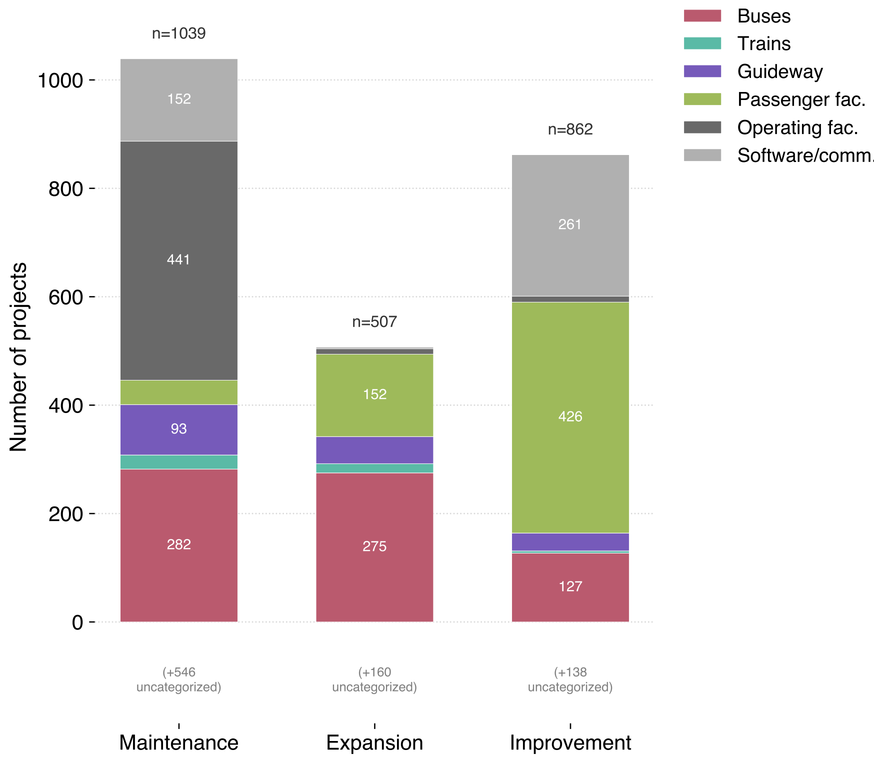
The second dimension, applied to non-expansionary projects, separates *better-service* from *sustained-service* spending. The organizing principle is whether the project represents a discretionary agency choice to improve the passenger experience or an operational necessity to maintain current service capability. *Better-service* projects are investments the agency could forgo without immediate service deterioration but chooses to make to improve what riders experience: accessibility upgrades, new passenger amenities, information and safety technology experienced by riders, and replacement of vehicles with versions whose improvements passengers can feel (low-floor boarding, accessible configurations). *Sustained-service* projects address operational necessities: without the investment, current service capability would degrade. Canonical examples are like-for-like vehicle replacement at end of useful life, preventive maintenance, rehabilitation of maintenance and operational facilities, back-office software, and maintenance shop equipment. These two dimensions sort each project into one of three labels that the rest of the paper uses, *Expansion* (expansionary projects, the model's *more-service* class), *Improvement* (better-service projects), and *Maintenance* (sustained-service projects). The paper uses these three labels throughout; the *more-service*, *better-service*, and *sustained-service* terms appear only where the classification itself is discussed. Table A4 in the Appendix lists a random sample of project descriptions in each category.

In addition to the service-intent category, the model assigns eight non-exclusive flags that capture ancillary characteristics regardless of service intent: zero-emission or hybrid propulsion, broader environmental benefit, accessibility improvements (ADA compliance, low-floor boarding, level platforms), safety enhancements, technology additions, operating cost reduction, staff working conditions, and symbolic or prestige value. These flags are used in robustness checks and heterogeneity analysis.

Complementarity of the two stages. M1 describes the factual capital transaction (what asset class was involved and in what capacity) while M2 infers the likely service impact. Because the two stages read the same project description through entirely different lenses, their coherence serves as internal validation. Figure 3 shows the M1 asset composition within each M2 group. Expansion projects, explicit and implicit alike, concentrate in bus purchases and passenger facility construction, the acquisitions that directly add capacity. Improvement projects are dominated by passenger facility work and software and communications upgrades, consistent with improving existing systems rather than adding capacity. Maintenance projects are concentrated in operating facility rehabilitation and replacement buses, with the remainder spread across the physical plant.

Appendix Figure A2 shows the same breakdown weighted by dollars. The category profiles are broadly similar, though guideway, which is a small share of projects, takes a much larger share of the spending.

Figure 3: Asset composition differs systematically across service-intent categories.



Note: Each bar shows project counts by M1 asset type within each M2 category; each project is counted once, under its primary asset category. Projects that fall in no single asset category ("Other") are noted below each bar. *Source:* LLM classification (M1 and M2) of FTA Form 1512 award descriptions; 3,252 projects across 798 awards matched to 365 UZAs.

2.3 Manual Verification

To assess classification reliability, I drew a random sample of 100 from the 3,763 extracted projects and had a research assistant code the same 46 binary fields (38 M1 capital-component variables plus 8 ancillary characteristic flags) and the service-intent category independently, using a standardized instruction sheet with worked examples.[2] Project descriptions were presented without the model's answers.

Across all binary fields, the RA and the model agree on 97.8 percent of project-field cells, with an overall Cohen's κ of 0.85. Table A2 in the Appendix reports agreement rates and Cohen's κ by variable group. Agreement is highest for rolling-stock fields (vehicles, $\kappa = 0.97$, and buses, $\kappa = 0.92$), where descriptions typically make asset type and purchase intent

[2] I thank Rayyan Salih for outstanding research assistance in completing this manual coding exercise.

unambiguous, and somewhat lower for passenger facility and flag fields ($\kappa = 0.65$ and 0.60, respectively), where coding requires inferring intent rather than reading off explicit terms. Railcar fields achieve 99 percent agreement but low average κ because rail projects are near-absent in this predominantly bus-focused sample; when nearly all projects are negative on a variable, κ is mechanically deflated regardless of actual coder concordance.

For the service-intent classification, the RA independently coded the same 100 projects using a three-category scheme (expansionary, better-service, sustained-service) consistent with the classification structure used in this paper. Two were uninformative and dropped, and the model agrees with the RA on 71 of the remaining 98 (72%; $\kappa = 0.58$). Table A3 in the Appendix reports the cross-tabulation. The expansionary category (86% agreement) and the sustained-service category (83% agreement) are coded with high fidelity; the main source of disagreement is the better-service/sustained-service boundary, which accounts for 20 of the 27 discrepant cases.

The better-service disagreements follow a systematic pattern. Of the 40 projects the RA coded as better-service, 14 were classified by the model as sustained-service. These cases (maintenance tools, staff parking, radios, computer equipment, facility energy upgrades) reflect that the RA was coding under an earlier, broader definition of better-service that encompassed any upgrade or improvement to existing systems. Under the current definition, better-service requires a discretionary choice to improve what passengers experience; investments in operational capability that do not change the rider experience are sustained-service. The six cases running the other way, which the RA coded as sustained-service and the model as better-service, are a smaller group with no comparable pattern. The disagreement rate at this boundary therefore reflects the definitional revision rather than model error.

2.4 Aggregation

Of the 3,551 classified projects, 22 percent are Expansion, split roughly evenly between explicit and implicit expansion, 30 percent are Improvement, and 48 percent are Maintenance. The analysis needs this variation at the urbanized-area (UZA) level, the geographic unit of the NTD panel, so I match each award's recipient city to its UZA using Census place-to-UZA crosswalk files. The match places 798 awards in 365 urbanized areas, carrying 3,252 of the projects into the analysis.

To turn these classifications into treatment variables, within each award and for each category I compute the share of projects that fall in the category and an indicator for whether any does. I then average these across the awards a UZA received, weighting by award dollars so that larger awards count proportionally and capital-intensive work is not understated by a simple project count.

The dollar-weighted project share, which I label *wmean2*, is the main treatment measure. Because each project falls in exactly one category, a UZA's three shares sum to one, so the measure describes the composition of its ARRA spending. The dollar-weighted any-project version, *wmeanmax*, is a coarser alternative used in robustness checks. The primary contrast sets Expansion spending against the Maintenance baseline. A secondary check uses only the explicitly expansionary share, which drops vehicle purchases that are expansionary by omission rather than by stated intent.

Appendix Figure A3 plots the joint distribution of these dollar-weighted shares across the 365 UZAs; the categories trade off against one another, most sharply between Maintenance and Expansion.

3 Transit Systems and the Places They Serve

The United States has hundreds of local transit systems, and they differ enormously in scale and finances, and they serve very different cities: a dense rail network carrying millions of riders has little in common with a handful of dial-a-ride vans in a small Sun Belt town. Before asking why agencies spent their ARRA funds the way they did, I want to know along which dimensions they differ. To this end I assemble a cross-sectional covariate dataset of 59 pre-ARRA variables covering socioeconomic composition, transit system baseline, road infrastructure, governance structure, and political environment. This profile is useful on its own. It maps how the U.S. transit landscape varies along these lines, and it gives me the terms in which to read both the formula's allocation and the agencies' spending choices.

The 23 socioeconomic variables come from the ACS 2009 five-year estimates (ACS5), with disability rate falling back to Census 2000 where ACS5 coverage is too thin. They span income and poverty, age structure, the labor market, housing stock and density, commute mode, and transit dependence by income (the transit share and car access of high- and low-earning workers).

The remaining 36 variables describe the systems and their setting. Twenty-eight transit system variables come from NTD annual time-series reports (Federal Transit Administration, 2020), aggregated from the agency to the primary UZA. Level variables are 2005–2008 pre-stimulus means; flow variables (VRM, ridership, fleet size) are per capita on 2000 Census UZA population. They cover system scale, fleet characteristics, mode mix, funding composition, capital allocation, and service efficiency. Three trend variables capture pre-ARRA dynamics (VRM growth, ridership growth, and fleet aging, each the 2000–2008 OLS slope), three FHWA 2008 variables measure auto travel intensity, road density, and freeway orientation, and two more capture governance (the VRM-weighted share of agencies that are independent public authorities) and political environment (the 2008 Democratic presidential vote share). Table A1 in the appendix lists every variable with its source.

3.1 Principal Component Analysis

These covariates are far from independent. Systems carrying many riders tend to have built more infrastructure; poor cities run tight budgets while serving large transit-dependent populations; dense places combine low car ownership with older housing. Much of the apparent variety across the 59 covariates therefore reflects a smaller set of underlying dimensions along which transit systems and their cities genuinely differ. I recover those dimensions with Principal Component Analysis.

Of the 365 UZAs in the analysis sample, I exclude the nine whose systems were also receiving Capital Investment Grants (the FTA's competitive program for major fixed-guideway expansions), since those large, separately awarded projects fall outside the formula-driven discretion this paper studies. Requiring complete data on all 59 covariates then reduces the remaining 356 to 347 by listwise deletion. I standardize every covariate before extraction, so the components reflect correlation rather than scale. The Kaiser criterion (eigenvalue above one) retains the first $K = 14$ components, which together account for 79.5% of the variance. Each component is a weighted combination of the original covariates, uncorrelated with the others by construction; the resulting scores serve as predictors in Section 4 and as controls in Section 5. Table 1 reports the cumulative variance share, the five highest-weight variables, and the highest- and lowest-scoring UZAs for each component.

Table 1: Principal components of pre-ARRA UZA characteristics

	Label	Top five loadings	Exemplar UZAs	Cum. var. (%)
1	Transit intensity	avg. seats (+0.23), cost/VRM (+0.23), peak utilization (+0.22), load factor (+0.22), total spending (+0.21)	↑ San Francisco-Oakland, CA ↑ Chicago, IL-IN ↓ Gadsden, AL ↓ Kingsport, TN-VA	22.4
2	Affluence and self-sufficiency	fed. oper. sh. (−0.25), fed. capital sh. (−0.23), transit commute sh. (+0.23), median hh income (+0.22), ADA fleet sh. (−0.22)	↑ San Francisco-Oakland, CA ↑ Chicago, IL-IN ↓ North Port-Punta Gorda, FL ↓ Florence, SC	33.7
3	College town	poverty rate (+0.34), renter sh. (+0.34), median rooms (−0.28), median hh income (−0.26), commute time (−0.26)	↑ State College, PA ↑ Ithaca, NY ↓ Simi Valley, CA ↓ South Lyon-Howell-Brighton, MI	43.0
4	Old city	65+ sh. (+0.33), median age (+0.32), zero-vehicle hh sh. (+0.31), one-vehicle hh sh. (+0.29), NILF rate (+0.25)	↑ Chicago, IL-IN ↑ San Francisco-Oakland, CA ↓ Hanford, CA ↓ Denton-Lewisville, TX	49.7
5	Demand-response growth	demand-resp. ridership sh. (+0.35), demand-resp. VRM sh. (+0.31), VRM trend (+0.31), ridership trend (+0.30), fleet Δ (+0.27)	↑ Hanford, CA ↑ Leesburg-Eustis, FL ↓ New Orleans, LA ↓ Brownsville, TX	55.3
6	Working-family poverty	unemployment (+0.44), under-18 sh. (+0.32), per capita income (−0.27), disability sh. (+0.26), DVMT (−0.23)	↑ Hanford, CA ↑ Porterville, CA ↓ Myrtle Beach, SC ↓ Bonita Springs-Naples, FL	59.8
7	New large city	pre-1940 housing sh. (−0.47), vacancy rate (+0.32), median rooms (−0.29), zero-vehicle hh sh. (−0.26), local oper. sh. (+0.24)	↑ Miami, FL ↑ Bonita Springs-Naples, FL ↓ Ithaca, NY ↓ New Bedford, MA	63.6
8	Service contraction	ridership trend (−0.46), VRM trend (−0.44), demand-resp. ridership sh. (+0.29), demand-resp. VRM sh. (+0.28), fare/trip (+0.23)	↑ Florence, SC ↑ Newark, OH ↓ Coeur d'Alene, ID ↓ Bristol, TN-Bristol, VA	66.8

(continued)

	Label	Top five loadings	Exemplar UZAs	Cum. var. (%)
9	Resort community	road mi. per 1k pop. (−0.36), DVMT (−0.31), vacancy rate (+0.28), commute time (+0.25), NILF rate (+0.22)	↑ Bonita Springs-Naples, FL ↑ Indio-Cathedral City-Palm Springs, CA ↓ Jefferson City, MO ↓ Houston, TX	69.6
10	Locally funded transit	local oper. sh. (+0.34), fleet size (+0.30), fed. oper. sh. (−0.30), VRM (+0.26), local capital sh. (+0.26)	↑ Kennewick-Richland, WA ↑ Olympia-Lacey, WA ↓ Coeur d'Alene, ID ↓ Chicago, IL-IN	72.0
11	Dense walkable city	freeway DVMT sh. (−0.41), DVMT (−0.39), fleet age trend (−0.37), authority governance sh. (−0.30), pop. density (+0.25)	↑ Chicago, IL-IN ↑ Miami, FL ↓ Texas City, TX ↓ Sherman, TX	74.0
12	Facility inefficiency	facilities sh. (+0.46), cost/trip (+0.32), rolling stock sh. (−0.31), fleet age trend (+0.24), fleet Δ (−0.22)	↑ Coeur d'Alene, ID ↑ Santa Fe, NM ↓ Hanford, CA ↓ Chicago, IL-IN	76.0
13	Fleet renewal	fleet age trend (+0.53), out-of-county work sh. (−0.41), capital intensity (−0.28), facilities sh. (−0.22), freeway DVMT sh. (−0.21)	↑ Coeur d'Alene, ID ↑ Chicago, IL-IN ↓ Montgomery, AL ↓ Charlottesville, VA	77.7
14	Operational inefficiency	fleet age trend (−0.44), cost/trip (+0.44), high-inc. transit sh. (+0.33), authority governance sh. (−0.24), facilities sh. (−0.23)	↑ Coeur d'Alene, ID ↑ Bremerton, WA ↓ Hanford, CA ↓ Miami, FL	79.5

Notes: $N = 347$ UZAs; $K = 14$ components retained by Kaiser criterion (eigenvalue > 1), accounting for 79.5% of variance. 59 pre-ARRA covariates: NTD operational variables (2005–2008 means), ACS demographic variables, FHWA road infrastructure, NTD governance structure, and 2008 county presidential vote share. All variables standardized before extraction. NTD flow variables (VRM, ridership, fleet size, DVMT) are expressed per capita. Loadings shown for the five highest-weight variables per component (sign indicates direction of correlation with the score). ↑ = two highest-scoring UZAs; ↓ = two lowest-scoring UZAs. "sh." = share; DVMT = daily vehicle miles traveled; NILF = not in labor force.

The first three components, together capturing 43% of the variance, arrange UZAs along the broadest axes of difference. **PC1** (22.4%), *transit intensity*, is the dominant dimension:

average vehicle seating, cost per VRM, peak utilization, load factor, and total spending all load positively, separating large, dense, heavily used networks such as San Francisco and Chicago from small systems running near-empty vehicles. **PC2** (11.3%), *affluence and self-sufficiency*, separates affluent metros, where many residents commute by transit and the agency funds itself locally, from poorer places that lean on federal operating and capital grants. **PC3** (9.3%), *college town*, isolates university towns such as Ithaca and State College, where renters dominate, measured poverty runs high (students inflate it), housing units are small, and commutes are short.

The next four components describe finer demographic and structural variation. **PC4** (6.7%), *old city*, captures aging, car-light populations, with large shares aged 65 and over and high zero- and one-vehicle household rates. **PC5** (5.6%), *demand-response growth*, marks small paratransit-heavy systems that were adding service and vehicles in the years before ARRA. **PC6** (4.6%), *working-family poverty*, loads on unemployment, child population share, and disability against per-capita income, identifying economically distressed working-age communities. **PC7** (3.8%), *new large city*, picks out newer, auto-oriented Sun Belt metros such as Miami and Naples, with little pre-1940 housing and high vehicle ownership.

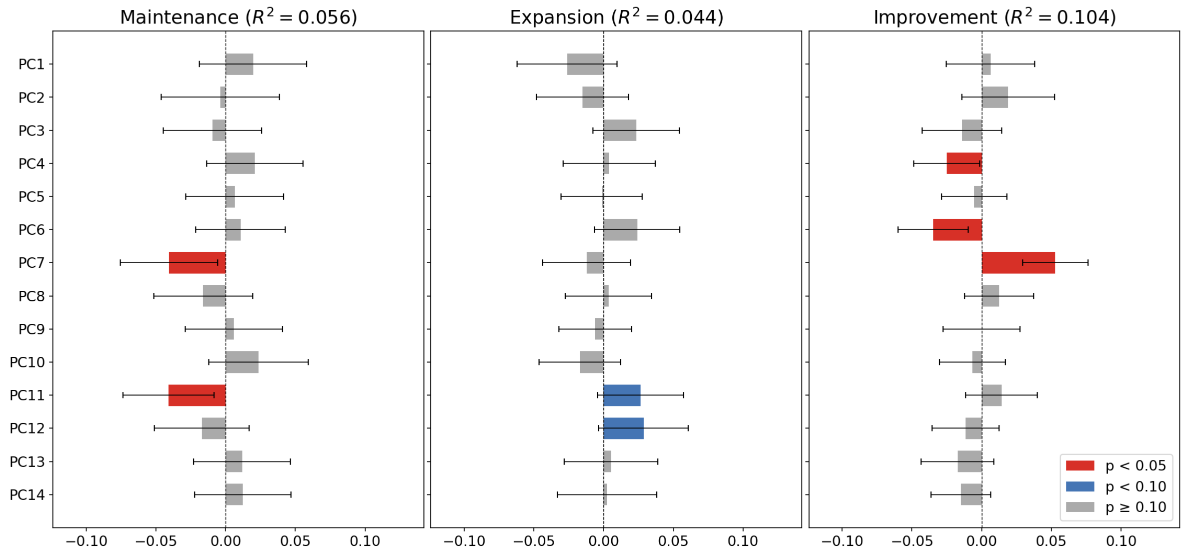
Two of the smaller components, though they explain little variance, carry much of the predictive weight in the main spending regressions. (PC8 through PC10, PC13, and PC14, which the table lists, do not.) **PC11** (2.0%), *dense walkable city*, combines high population density with low driving and limited freeway use. **PC12** (1.9%), *facility inefficiency*, flags systems that direct capital toward buildings rather than vehicles and carry high cost per trip. How these dimensions shaped what agencies bought is the subject of the next section.

4 Predicting Spending Decisions

Figure 4 reports coefficients from regressing each of the three project-type shares on all 14 principal components jointly, controlling for the total ARRA award amount. Each share is measured as dollar-weighted intensity (*wmean2*, introduced in Section 2): a project's category score is averaged within its award, and awards are then weighted by dollar amount up to the UZA level. Controlling for the award amount lets the coefficients reflect how a city's pre-existing profile shaped *how* it spent rather than how much it received; the Appendix repeats the analysis without this control (Figure A6).

Overall, the fit is modest: only 4–10% of the variation in each outcome is explained by

Figure 4: City characteristics explain little of how agencies spent their ARRA funds.



Note: Multivariate OLS; all 14 PCs plus ARRA award amount included jointly. $N = 347$ urbanized areas, HC1 standard errors, 95% CI. Color: red ($p < 0.05$), blue ($p < 0.10$), gray ($p \geq 0.10$). Outcome: *wmean2*. PC definitions in Table 1.

transit system characteristics. This is a surprising finding. The covariates entering the PCA cover transit provision and use, agency finances, and the demographics that drive demand. If spending tracked any of these, the components should capture it. That so little is explained suggests that other, more arbitrary factors played a larger role in deciding how agencies spent their ARRA awards.

Neither of the two strongest dimensions, transit intensity (PC1) and system affluence (PC2), predicts spending orientation, despite together capturing 34% of the variance in the covariate set. More striking still, the most salient structural features of a transit system do not determine how ARRA funds were directed. This is not an artifact of holding the award amount fixed. Although the formula ties that amount to scale, PC1 and PC2 are just as flat in the uncontrolled regression (Appendix Figure A6), so scale and affluence genuinely fail to predict how agencies spent.

The low predictive power does not stem from noisy measurement, however. Table 2 shows how predictive power for the expansion share increases with the number of ARRA awards per city. For cities receiving only one award, the outcome measure reduces to the expansion share within a single filing, averaging over just a handful of projects, which may not reflect the city's underlying capital priorities. As the award count rises, the within-city average stabilizes and the principal components explain an increasing share of the variance. Even among cities with at least three awards (where the spending measure is well identified), the components account for under a quarter of the variation, leaving more than three-quarters unexplained. The idiosyncrasy is real, not a byproduct of cities whose spending rests on only one or two awards.

The reading I favor is that spending was only weakly *return-targeted*. If agencies were allocating capital to maximize returns, the fundamentals that signal need and demand would carry most of the weight, and the dominant axes of scale and affluence most of all. Instead they carry almost none. This idiosyncrasy is what lets the next section treat the residual variation as good as random and recover causal effects. It is also, in itself, evidence that a formula based on scale produces spending only loosely tied to need.

Table 2: Predictive power of city characteristics increases with ARRA award count

Min. awards	N	R^2
≥1	347	0.044
≥2	148	0.089
≥3	68	0.247

Notes: OLS regression of expansion share (*wmean2*) on 14 principal components plus ARRA award amount. R^2 estimated on subsamples meeting each minimum-awards threshold.

Only a handful of the 14 components explain meaningful variation in expansion, improvement, or maintenance spending. **Older cities (PC4) and places with high working-family poverty (PC6)** spend significantly less on improvement of transit services. The two components capture different demographic profiles but share a common structural feature: transit-dependent populations with limited alternatives. PC4 is dominated by aging demographics (65-plus share and median age) and high car-free household shares, identifying communities where the elderly and the carless rely on transit by necessity. PC6 instead reflects high unemployment, large shares of children under 18, low per capita income, and high disability rates, with low driving intensity confirming that car access is constrained by economics and mobility limitations rather than age. In both cases, the marginal rider has few alternatives regardless of service quality, reducing the incentive for agencies to direct discretionary funds toward quality improvements rather than service maintenance.

New large cities (PC7), in contrast, spend more on improvement. These cities were built around cars and have a lower share of carless households. Their transit riders mostly choose transit rather than depend on it, and so care more about ride quality. Most of the increase in improvement for these systems comes at the expense of maintenance, perhaps because regular funding is large enough compared to their modest size to keep them well-maintained.

Dense walkable cities (PC11) and systems with facility-heavy capital programs (PC12) instead predict more expansion relative to maintenance. Walkable cities have robust demand for new service capacity, especially given that PC11 is orthogonal to PC1, PC2, and PC3, so this is demand above and beyond what existing transit provision already reflects. Systems that over-invest in facilities (high PC12) may prefer expansionary projects for budgeting reasons that are not clear from the data.

Appendix Figure A7 repeats the analysis using the *wmeanmax* measure, which credits an award for a category if any project within it qualifies. The improvement predictors are broadly stable: PC6 and PC7 remain significant, though PC4 (old city) loses significance and PC3 (college towns) enters. The expansion predictors change more substantially. The walkability and facility-inefficiency effects from the *wmean2* regressions do not survive; instead, cities with pre-ARRA service contraction (PC8) appear more expansionary, consistent with agencies using ARRA to reverse declining ridership trends, while locally-funded systems (PC10) appear less expansionary, perhaps because sustaining new service without federal operational support is costly.

The next section considers whether these differences in spending orientation translate into different transit outcomes. Because ARRA capital funds are fungible with other capital sources, observed differences in spending type may partly reflect project labeling choices rather than genuine differences in aggregate capital programs.

5 Consequences of Different Spending

Since ARRA spending decisions reflect pre-existing conditions, estimating the causal effect of spending type on outcomes is challenging. An agency that already runs an old fleet is the one most likely to spend on replacement, so maintenance spending and high fleet age go together for reasons unrelated to the effect of the spending itself. Similarly, underlying demand trends may drive both the type of ARRA spending chosen and subsequent ridership outcomes. However, as shown in the previous section, all 14 principal components together explain only about 10% of the variation in spending type. The remaining variation is essentially idiosyncratic, reflecting local circumstances such as the composition and priorities of the transit agency board, the project pipeline at the time of the ARRA announcement, or the personalities of key decision-makers. Controlling for all 14 PCs isolates the component of spending type not attributable to structural pre-existing characteristics. Identification rests on that residual variation being unrelated to pre-existing outcome trends, which the flat pre-treatment coefficients in the event studies below support. I therefore interpret the resulting estimates as causal.

Because the maintenance, expansion, and improvement shares of ARRA spending sum to one in each UZA, the three shares are perfectly collinear, so I estimate the effect of expansion and improvement *relative* to maintenance. I use a continuous difference-in-differences design that relates the change in UZA-level outcomes to the share of ARRA spent on each project type. To trace the dynamic response and check for parallel pre-trends, I use an event-study specification that estimates yearly effects relative to 2008. Exact functional forms for both are in Section B in the Appendix. In each case the coefficients give the effect of shifting a UZA's entire ARRA portfolio from maintenance to either expansion or improvement.

Table 3 reports the estimated coefficients on four primary outcomes: vehicle revenue miles (VRM) per capita, fleet size (vehicles per 1,000 people), fleet age (years), and unlinked passenger trips (UPT) per capita. Fleet size and fleet age are direct capital outcomes. They measure whether ARRA capital purchases translated into a larger or younger fleet. VRM and UPT capture whether those purchases fed through to service provision and ridership. Panel A is standard DiD with PCA scores interacted with the post-treatment indicator; Panel B adds unit-specific linear pre-treatment trends estimated on 2000–2008 and subtracted from the outcome before estimation, following Borusyak et al. (2024), to absorb heterogeneous pre-existing trajectories.[3] Both panels control for ARRA funding amount interacted with Post throughout, since each dollar of ARRA funding raised total capital spending by roughly two dollars (Braslavskiy, 2026). Dropping the PC controls (Table A5 in the appendix) leaves the point estimates close to these. Expansion is null either way, and the improvement coefficient is similar in size, though it clears the 10% threshold only once the controls absorb residual variance.

The striking finding in Table 3 is that expansion produces outcomes no different from maintenance spending. All estimates are insignificant, and the two fleet outcomes are small in magnitude. Trips in Panel A and miles in Panel B carry large negative point estimates, but the signs reverse across panels, so neither is reliable. The event studies in Figures 5 and 6 show no clear separation between expansion and maintenance either. Among the secondary outcomes in Appendix Figures A4 and A5, expansionary UZAs show faint patterns that

[3]The importance of correctly adjusting for pre-trends with the imputation method was demonstrated in Braslavskiy (2025).

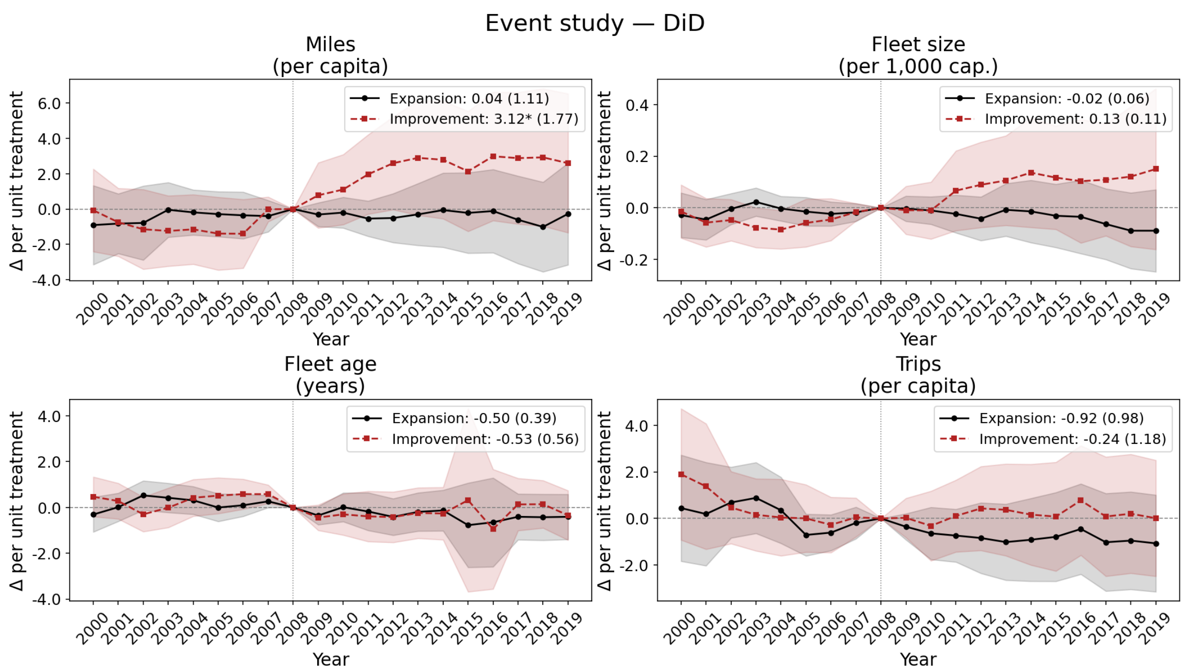
Table 3: Consequences of ARRA Spending Type on Transit Outcomes

	Miles (per capita)	*Fleet size* (per 1,000 cap.)	*Fleet age* (years)	*Trips* (per capita)
Post-2009 mean	13.28	0.57	5.76	14.40
Panel A: DiD				
Expansion × Post	0.04	-0.02	-0.50	-0.92
	(1.08)	(0.06)	(0.38)	(0.96)
Improvement × Post	3.12*	0.13	-0.53	-0.24
	(1.73)	(0.10)	(0.55)	(1.16)
Panel B: DiD with Pre-Treatment Trends				
Expansion × Post	-0.88	-0.04	-0.62	0.27
	(1.11)	(0.06)	(0.49)	(1.15)
Improvement × Post	2.82*	0.10	-0.75	1.96
	(1.49)	(0.10)	(0.65)	(1.36)
N	6,940	6,940	6,940	6,940

Notes: $\text{Post}_t = \mathbf{1}[t \geq 2009]$. Reference category is maintenance. Treatments are award-amount-weighted UZA-level shares of ARRA projects classified as expansion or improvement by the M2 LLM classifier. Outcomes from the National Transit Database panel (2000–2019): vehicle revenue miles per capita, vehicles per 1,000 capita (fleet size), fleet-average age in years, and unlinked trips per capita. All specifications include unit and year fixed effects and ARRA award amount per capita × Post. Both panels include PCA scores (PC1–PC14) interacted with the post-treatment indicator as controls. Panel B subtracts unit-specific pre-treatment linear trends from the outcome (estimated on 2000–2008) before applying two-way demeaning, following Borusyak et al. (2024). Standard errors clustered at the UZA level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

almost never reach significance. The clearest is capital intensity, the capital share of total spending. It falls, significantly so in the trend-adjusted panel, and capital cost drifts down alongside it. The other secondary outcomes show no reliable pattern. The expansion null is not an artifact of LLM misclassification. Table A6 reassigns expansion only to projects that state the goal explicitly (such as "buy X buses to expand a corridor"), and the result holds. Miles and trips rise slightly, but the qualitative pattern is unchanged. The classifier separates expansion from the rest well, so the null reflects what the spending did, not how it was labeled.

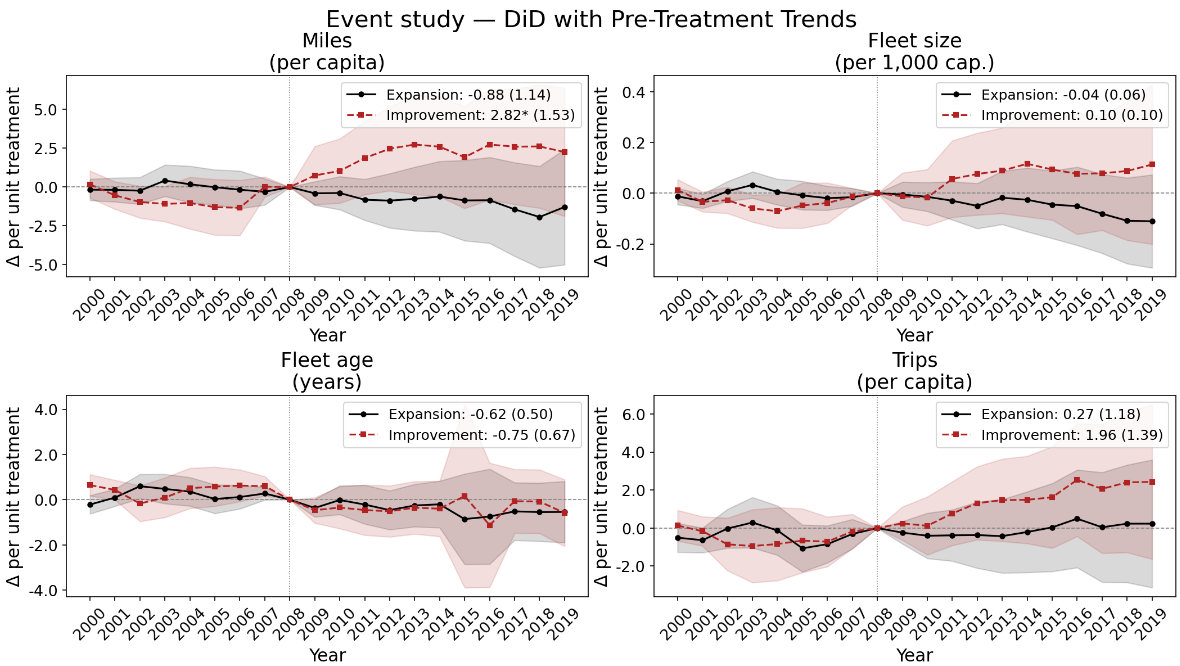
Figure 5: Event-study estimates of ARRA spending type on transit outcomes (Panel A: no detrending).



Note: Year-by-year DiD estimates with PCA scores interacted with year indicators as controls (Panel A specification). Each coefficient is the difference in outcome at year t relative to 2008 between a UZA spending entirely on expansion or improvement versus entirely on maintenance. Shaded bands are 95% confidence intervals. Legend reports the aggregated DiD (average post-period minus average pre-period coefficient) and its standard error. Outcomes per-capita from the NTD panel (2000–2019); ARRA award per capita $\times$ year controlled throughout.

Improvement spending, by contrast, raises service provision (VRM) significantly relative to maintenance. I read this as causal rather than a pre-existing trend. The pre-2009 coefficients in Figure 5 are flat, and Panel B, which removes unit-specific trends, gives a nearly identical estimate. Fleet size rises and average fleet age falls, and trips rise in Panel B once the negative pre-trend visible in the trips panel of Figure 5 is removed, though these remain insignificant. The changes share a clear post-2009 onset and large magnitudes, which leads me to read the rise in VRM as new vehicles entering service and carrying more passengers.

Figure 6: Event-study estimates of ARRA spending type on transit outcomes (Panel B: pre-treatment trend adjustment).



Note: Year-by-year DiD estimates with PCA scores interacted with year indicators as controls (Panel B specification). Unit-specific linear trends estimated on pre-treatment years (2000–2008) are subtracted from the outcome before two-way demeaning, following Borusyak et al. (2024). Each coefficient is the difference in outcome at year t relative to 2008 between a UZA spending entirely on expansion or improvement versus entirely on maintenance. Shaded bands are 95% confidence intervals. Legend reports the aggregated DiD and its standard error. Outcomes per-capita from the NTD panel (2000–2019); ARRA award per capita $\times$ year controlled throughout.

The additional outcomes in Appendix Figures A4 and A5 move the same way: operating cost and peak utilization (peak vehicles in service as a share of the fleet) both rise, and reach significance once pre-trends are removed. Because expansion and maintenance are statistically indistinguishable across nearly all outcomes, Table A7 reports improvement against the two combined, and the coefficients are similar. The improvement effect is significant at the 10% level; the wide confidence bands reflect a sample of only about 350 UZAs with errors clustered at the UZA level, not a weak or noisy pattern. Spending ARRA on improvement therefore did more to grow service than spending it on expansion, an unexpected result that the model in the next section explains.

6 The Role of Improvement Spending

This section explains the surprising finding that improvement spending expands service, while expansion spending is indistinguishable from spending on maintenance. The core intuition is simple. A transit system with more capacity than riders demand at its current quality cannot raise ridership by buying vehicles or building facilities. Raising quality, by contrast, attracts riders and grows the system. This intuition is elaborated in the following model, and welfare implications follow.

6.1 A Simple Model of Transit Spending

This model characterizes how a transit agency allocates its budget across competing uses and how that allocation determines the supply of service. For simplicity, assume that riders demand miles of service (VRM), which are translated into the number of trips at a constant rate. Riders care about fare-adjusted quality q, and demand $D(q)$, in miles of service, is upward-sloping.

The agency holds a capital stock $\bar{K}$ (vehicles and infrastructure) that determines the fare-adjusted quality of service $\bar{q}$ (fare is fixed in the short term). Running v VRM requires capital $K \geq kv$, where $k > 0$ is capital intensity. Supply is determined by the agency's spending decisions given $\bar{K}$, $\bar{q}$, and funding G.

The agency's total funding G is allocated among four uses,

$$G = O + M + E + I,$$

where O is operations (driver hours, fuel), M is maintenance capital, E is expansion capital, and I is improvement capital.

Running v VRM costs O in direct operating expenditure after recovering the fare. Maintenance M keeps the capital stock in good repair and reduces those costs: operating cost is $O(v, M) = cv + h(M)$, where $c > 0$ is a constant marginal cost per VRM and $h(M)$ is decreasing and convex. The agency therefore minimizes $O + M$ over M for any given v. Since h is convex, the first-order condition $h'(M^*) = -1$ pins a unique optimal maintenance level that is independent of v, yielding a linear minimized running cost:

$$C(v) \equiv \min_{M} \big[cv + h(M) + M\big] = C_{\min} + cv,$$

where $C_{\min} \equiv h(M^*) + M^* > 0$ is the maintenance floor, the irreducible cost of keeping the capital stock deployable even at $v = 0$.

Expansion capital E adds to $\bar{K}$, relaxing the capital constraint $K = \bar{K} + E \geq kv$. Improvement capital I upgrades the quality riders experience, $q = \bar{q} + I$. The maximum feasible VRM (setting $I = 0$) depends on whether the initial capital constraint binds. If $\bar{K} \geq k(G - C_{\min})/c$, capital is non-binding and

$$v_{\max} = \frac{G - C_{\min}}{c}.$$

If $\bar{K} < k(G - C_{\min})/c$, the agency must purchase expansion capital $E = kv_{\max} - \bar{K}$, giving

$$v_{\max} = \frac{G - C_{\min} + \bar{K}}{c + k}.$$

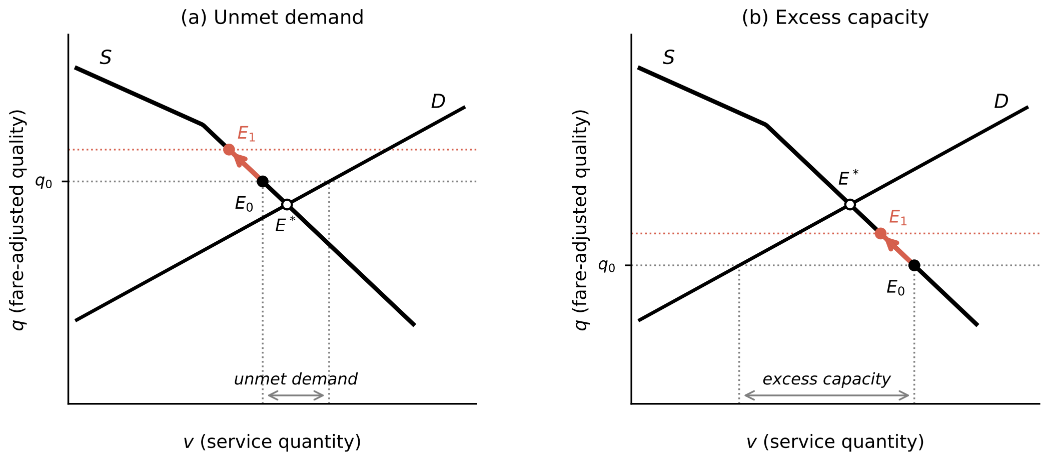
In either case the maximum-service point is $(v_{\max}, \bar{q})$. At the other extreme, at $v = 0$, the budget leaves $I = G - C_{\min}$ for improvement, so

$$q_{\max} = \bar{q} + G - C_{\min}.$$

The maximum-quality point is $(0, q_{\max})$. For any interior allocation, $I = G - C_{\min} - cv - \max(0, kv - \bar{q})$, and the supply curve is piecewise linear from $(0, q_{\max})$ to $(v_{\max}, \bar{q})$, with slope $-c$ where capital is non-binding ($v \leq \bar{K}/k$) and slope $-(c + k)$ where capital must also be purchased.

The welfare-maximizing outcome is at the intersection of supply and demand, E^*, with welfare-optimal quality q^*. Transit agencies, however, may not maximize rider welfare. They

Figure 7: Transit supply and demand: unmet demand versus excess capacity.



Note: Conceptual diagrams; axes not calibrated to data. q = fare-adjusted quality; v = service quantity. S and D are identical across panels, crossing at E^*. In panel (a), $q_0 > q^*$: riders demand more service than the budget allows, leaving unmet demand. In panel (b), $q_0 < q^*$: the agency can supply more than riders demand, leaving excess capacity. Red arrows trace a reallocation toward quality improvement, raising q_0 to q_{imp}.

may deviate either because of their incentives, such as political pressure or career concerns, or because of constraints, such as limited administrative capacity and an inability to estimate demand. Write q_0 for the quality the agency actually operates at, and E_0 for the resulting operating point. A reallocation toward improvement moves the agency from E_0 to E_1. Two principally different situations arise depending on q_0: unmet demand (Figure 7a) and excess capacity (Figure 7b). In the first case, capacity is binding at a lower level of service than riders demand. In the second case, capacity has slack, and service is limited by demand at the relatively low chosen quality. In these two cases, an increase in improvement spending along the budget constraint has drastically different consequences for the equilibrium level of service.

Unmet demand case. If a portion of maintenance and expansion spending is rerouted to improvement, quality increases and capacity decreases moving along the supply curve up and to the left: $E_0 \to E_1$. Higher quality attracts more riders, while lower capacity further limits how much service can be provided. The unmet demand gap widens because of both. The equilibrium v goes down.

Excess capacity case. If a similar change happens with slack capacity, lower maintenance and expansion spending does not limit how much service is provided. Higher quality, on the other hand, attracts more riders, who now use more service. As a result, the scale of the system v increases.[4]

6.2 Welfare Implications

The key empirical result is that UZAs that spent a higher share of their ARRA on improvement, for idiosyncratic reasons unrelated to system characteristics, saw their service grow. This places the average agency in the second regime, with excess capacity at its current quality. Improvements attract new riders, and service rises. If some demand was unmet, spending on improvement would be counterproductive and reduce the number of VRM and trips. The equivalence of maintenance and expansion spending also fits the excess capacity case: when capacity is not binding, investing less in expansion doesn't have an immediate effect on VRM.

Why the average transit system in the U.S. undersupplies quality and oversupplies capacity can be explained in several ways. The belief that more service brings more riders may simply be the default assumption. Expansion can also appeal for reasons unrelated to ridership. New routes and vehicles are easier to plan and to point to than diffuse quality upgrades, and they are more politically popular. Federal formula funding rewards prior service and population, and tracks ridership only loosely ($R^2 = 0.12$ in Figure A1). Finally, buying more buses is more straightforward than working out how to deliver better service. All of this is the cost of leaving spending to local discretion. The undersupply of quality suggests that raising the average UZA's improvement share above its current 29.7% would lift service, ridership, and welfare together.

In the back-of-the-envelope calculations in Table 4, I consider several scenarios that increase the share of ARRA spent on improvement. For estimates of the increase in VRM from increasing the improvement share, I use the values from Panel B of Table 3: 3.69 for decreasing expansion and 2.82 for decreasing maintenance share. These coefficients are applied at the UZA level and aggregated to a national total. To translate this into an increase in the number of trips, I use two methods. First, I regress the UZA-level post-2009 increase in trips on the growth in VRM, using the same controls and pre-trend adjustment as the Panel B specifications, to get a conversion rate of 0.37. Second, I run an Instrumental Variable estimation using the improvement share to instrument for the growth in VRM in a similar fashion to get a rate of 0.59. Finally, I translate the increase in the number of trips to a dollar gain

[4]If the agency initially achieves the welfare-maximizing equilibrium, $E_0 = E^*$, any change in the spending mix will lead to lower v.

using two estimates from the literature. Parry and Small (2009) put the marginal *external* benefit of bus travel, including congestion relief and smaller accident and pollution savings, at roughly 15 cents per passenger mile averaged over the day, which at an average unlinked trip of five miles is on the order of \$0.75 per added trip. To this external benefit I add the private benefit to riders. Litman (2023) places the convenience and travel-time-savings value of service-quality improvements on the order of \$1 to \$3 per trip net of the fare, but this figure values the benefit to an inframarginal rider rather than the average induced trip; applying the standard rule-of-half treatment for demand generated by a cost or quality change, I conservatively take \$0.50 to \$1.50 per added trip, bringing total welfare to roughly \$1.50 per added trip.[5]

Scenarios A raise every UZA to a target improvement-share floor given by the label (10, 25, 50, or 100 percent). UZAs already above the floor are left unchanged; for the rest, expansion spending is converted to improvement first, then maintenance once expansion is exhausted. Scenarios B hold each UZA's spending mix fixed and redirect the labeled share of ARRA funding (10 to 100 percent) in proportion to improvement shares. Scenarios C redirect the same share by the inverse of principal components 4 and 6, which predicted lower improvement shares and could plausibly have entered the formula used to allocate ARRA.

The welfare gains presented in Table 4 are sizeable but bounded. Scenarios A change the composition of spending and produce the largest gains. Even the most aggressive case, in which every area's award is moved entirely onto improvement, raises national service by roughly 13% and ridership by 3 to 4%, a noticeable but not transformative shift. Redistribution of funds holding each area's spending mix fixed has limited impacts. Scenarios C are the weakest of the three, and also the most realistic, since they reweight ARRA by characteristics the formula could plausibly have used. This shows the limits of federal funding policy when the composition of spending is fixed by local conditions. Reaching higher service requires funding that explicitly rewards ridership, to counteract local limitations and align agencies' incentives with welfare maximization. In the setting of this paper, that means spending more on quality improvements.

[5] Anderson (2014) bounds peak congestion relief alone at \$1.20 to \$4.10 per passenger mile.

Table 4: Reallocation counterfactual: implied national gains

Scenario	Improvement share	ΔVRM (M mi)	OLS ΔTrips (M)	OLS ΔWelfare ($M)	IV ΔTrips (M)	IV ΔWelfare ($M)
A: Raise improvement share						
A10	31.1%	7.6	2.8	4.2	4.4	6.6
A25	35.7%	28.4	10.4	15.6	16.6	24.9
A50	51.7%	96.1	35.2	52.8	56.2	84.3
A100	100.0%	285.6	104.8	157.2	167.2	250.8
B: Redirect ARRA by improvement share						
B10	31.3%	12.4	4.6	6.9	7.3	10.9
B25	33.7%	31.0	11.4	17.1	18.2	27.3
B50	37.7%	62.0	22.8	34.2	36.3	54.4
B100	45.7%	124.1	45.5	68.2	72.6	108.9
C: Redirect ARRA by inverse PC4 & PC6						
C10	29.9%	9.4	3.5	5.2	5.5	8.2
C25	30.2%	23.6	8.7	13.0	13.8	20.7
C50	30.8%	47.2	17.3	26.0	27.6	41.4
C100	31.9%	94.4	34.6	51.9	55.3	82.9
	Improvement share	VRM (M mi)	Trips (M)			
Baseline (levels)	29.7%	2246	4003			

Notes: Improvement share is the ARRA-dollar-weighted mean across UZAs. VRM effects use the Panel B coefficients of Table 3 (A) and the improvement-vs-rest coefficient $\beta = 3.17$ of Table A7 (B, C); trips and welfare use the conversions and $1.50 per-trip value given in the text.

7 Conclusion

ARRA gave local transit agencies $8.4 billion in capital funds and wide discretion over which projects to fund. Classified by type, the funded projects divide into three groups: 22% added service capacity, 30% improved the quality of existing service, and 48% replaced or maintained current assets. The consequences of these choices did not track the amount of capacity added. Areas that spent more on expansion saw no more service or ridership than areas that spent on maintenance, while areas that spent more on improvement provided more service afterward. This is the pattern of a system that already carries the riders its quality attracts. When capacity is not the binding constraint, added vehicles and miles go unused, and only more reliable and comfortable service draws new riders.

This allocation was only weakly tied to the conditions agencies faced. A broad set of pre-existing demographic, financial, and operational characteristics explains little of how an area spent, so the composition of ARRA spending was largely idiosyncratic. The consequences then carry the point further. The return to transit capital is conditional: it appears where spending improves the quality of service and where local demand can respond, so a capital infusion allocated without regard to either does little for ridership. This raises questions the present study cannot settle, such as whether the same mismatch holds for the much larger capital programs enacted since ARRA, and whether funding can be directed toward returns without losing the speed that makes formula allocation attractive.

How far this generalizes depends on a system's starting point. The argument holds where a system already meets the demand its current quality generates, which describes the average U.S. system in this period; where demand exceeds capacity, the conclusion reverses, and expansion would raise service while a shift toward quality would lower it. The result is specific to the first case and is not a claim that capacity never matters. The episode limits it further. Because the treatment is the composition of a single, fungible infusion, the absence of an expansion effect is better read as no detectable effect than as evidence that capacity does not matter. ARRA also arrived in a deep recession with a short spending deadline, which may have pushed agencies toward maintenance projects that were ready to start, and the reallocation exercise applies one average effect across systems that differ widely. These qualifications bound the result without overturning it: where capacity already exceeds demand, the return to transit capital comes through quality.

As fiscal stimulus, the transit portion of ARRA worked. It moved money quickly into real projects, without the delay that competitive review imposes. As a transfer meant to improve transit, it did less, because most of the money went to projects that did not raise ridership. The formula ties capital to the service a system already runs. It does not direct capital to where an added dollar would do the most good, and the discretion attached to it let agencies act on the common assumption that more service brings more riders, which the evidence does not support. What limited ARRA's effect on transit was not the amount of money but how it was allocated, across project types and across systems. A rule that rewarded ridership and the quality that supports it would have produced more service from the same funds than the scale-based formula did. If future programs are to improve transit, their allocation rules will need to reward demonstrated returns, not system scale alone.

References

Anderson, Michael L (2014). "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion". In: *American Economic Review* 104.9, pp. 2763–2796.

Baicker, Katherine (Nov. 2001). "Government decision-making and the incidence of federal mandates". In: *Journal of Public Economics* 82.2, pp. 147–194. ISSN: 0047-2727. DOI: 10.1016/S0047-2727(00)00149-3. URL: https://www.sciencedirect.com/science/article/pii/S0047272700001493 (visited on 08/24/2025).

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024). "Revisiting event-study designs: Robust and efficient estimation". In: *Review of Economic Studies* 91.6, pp. 3253–3285.

Braslavskiy, Arseniy (2025). "In the Weeds of Traffic Fatalities: Revisiting the Effect of Medical Marijuana Laws". Working paper. URL: https://arseniy.braslavskiy.com/files/In_the_Weeds_of_Traffic_Fatalities.pdf.

– (2026). "Beyond the Flypaper Effect: Crowding-In from Federal Investment in Public Transit". Working paper. URL: https://arseniy.braslavskiy.com/files/Beyond_the_Flypaper_Effect.pdf.

Chodorow-Reich, Gabriel et al. (Apr. 2012). "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act". en. In: *American Economic Journal: Economic Policy* 4.3, pp. 118–145. ISSN: 1945-7731. DOI: 10.1257/pol.4.3.118. URL: https://www.aeaweb.org/articles?id=10.1257/pol.4.3.118 (visited on 09/11/2025).

Courant, Paul N, Edward M Gramlich, and Daniel L Rubinfeld (1979). "The stimulative effects of intergovernmental grants: Or why money sticks where it hits". In: *Fiscal Federalism and Grants-in-Aid*. Ed. by Peter Mieszkowski and William H. Oakland. Washington, DC: The Urban Institute, pp. 5–21.

Downs, Anthony (1962). "The law of peak-hour expressway congestion". In: *Traffic Quarterly* 16.3, pp. 393–409.

Duranton, Gilles and Matthew A Turner (2011). "The fundamental law of road congestion: Evidence from US cities". In: *American Economic Review* 101.6, pp. 2616–2652.

Federal Transit Administration (2009). *National Transit Database 2008 Annual Database*. U.S. Department of Transportation. URL: https://www.transit.dot.gov/ntd/ntd-data.

– (2020). *2019 National Transit Database Annual Time-Series*. U.S. Department of Transportation. URL: https://www.transit.dot.gov/ntd/ntd-data.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport (Jan. 2008). "Why do the poor live in cities? The role of public transportation". In: *Journal of Urban Economics*

63.1, pp. 1–24. ISSN: 0094-1190. DOI: 10.1016/j.jue.2006.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0094119007000046 (visited on 08/16/2024).

Gordon, Nora (Aug. 2004). "Do federal grants boost school spending? Evidence from Title I". In: *Journal of Public Economics* 88.9, pp. 1771–1792. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2003.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0047272703001373 (visited on 08/22/2025).

Hines, James R. and Richard H. Thaler (Dec. 1995). "The Flypaper Effect". en. In: *Journal of Economic Perspectives* 9.4, pp. 217–226. ISSN: 0895-3309. DOI: 10.1257/jep.9.4.217. URL: https://www.aeaweb.org/articles?id=10.1257/jep.9.4.217 (visited on 08/18/2025).

Inman, Robert P. (2009). "Flypaper Effect". en. In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London, pp. 1–6. ISBN: 978-1-349-95121-5. DOI: 10.1057/978-1-349-95121-5_2956-1. URL: https://link.springer.com/rwe/10.1057/978-1-349-95121-5_2956-1 (visited on 08/20/2025).

Knight, Brian (Mar. 2002). "Endogenous Federal Grants and Crowd-out of State Government Spending: Theory and Evidence from the Federal Highway Aid Program". en. In: *American Economic Review* 92.1, pp. 71–92. ISSN: 0002-8282. DOI: 10.1257/000282802760015612. URL: https://www.aeaweb.org/articles?id=10.1257/000282802760015612 (visited on 01/11/2024).

Leduc, Sylvain and Daniel Wilson (May 2017). "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act". en. In: *American Economic Journal: Economic Policy* 9.2, pp. 253–292. ISSN: 1945-7731. DOI: 10.1257/pol.20140371. URL: https://www.aeaweb.org/articles?id=10.1257/pol.20140371 (visited on 01/11/2024).

Litman, Todd (2023). *Evaluating Public Transit Benefits and Costs: Best Practices Guidebook*. Tech. rep. Victoria, BC: Victoria Transport Policy Institute.

MIT Election Data and Science Lab (2024). *County Presidential Election Returns 2000–2024*. V20, https://doi.org/10.7910/DVN/VOQCHQ. DOI: 10.7910/DVN/VOQCHQ.

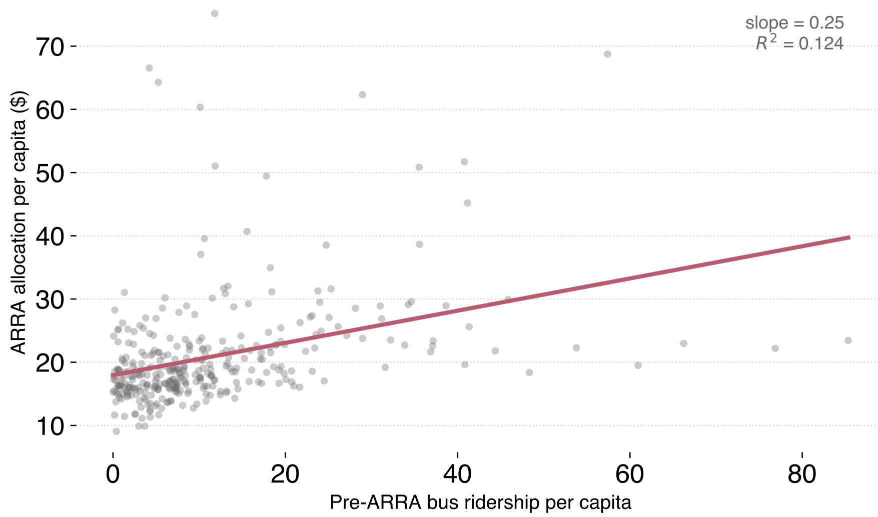
Parry, Ian W H and Kenneth A Small (2009). "Should urban transit subsidies be reduced?" In: *American Economic Review* 99.3, pp. 700–724.

Wilson, Daniel (Apr. 2012). "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act". en. In: *American Economic Journal: Economic Policy* 4.3, pp. 251–282. ISSN: 1945-7731. DOI: 10.1257/pol.4.3.251. URL: https://www.aeaweb.org/articles?id=10.1257/pol.4.3.251 (visited on 09/11/2025).

Appendix

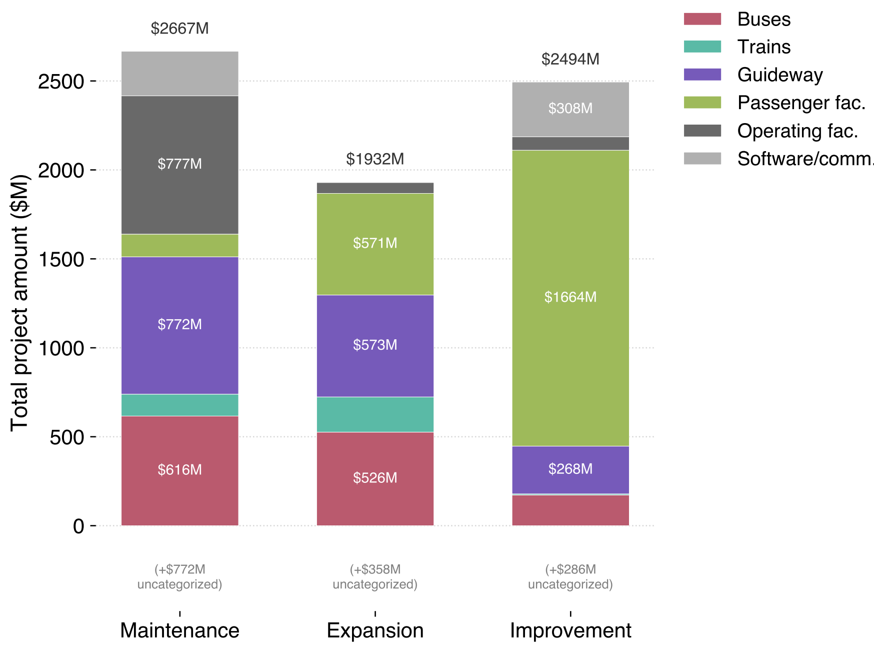
A Data Description

Figure A1: ARRA allocation only weakly tracks pre-ARRA transit use.



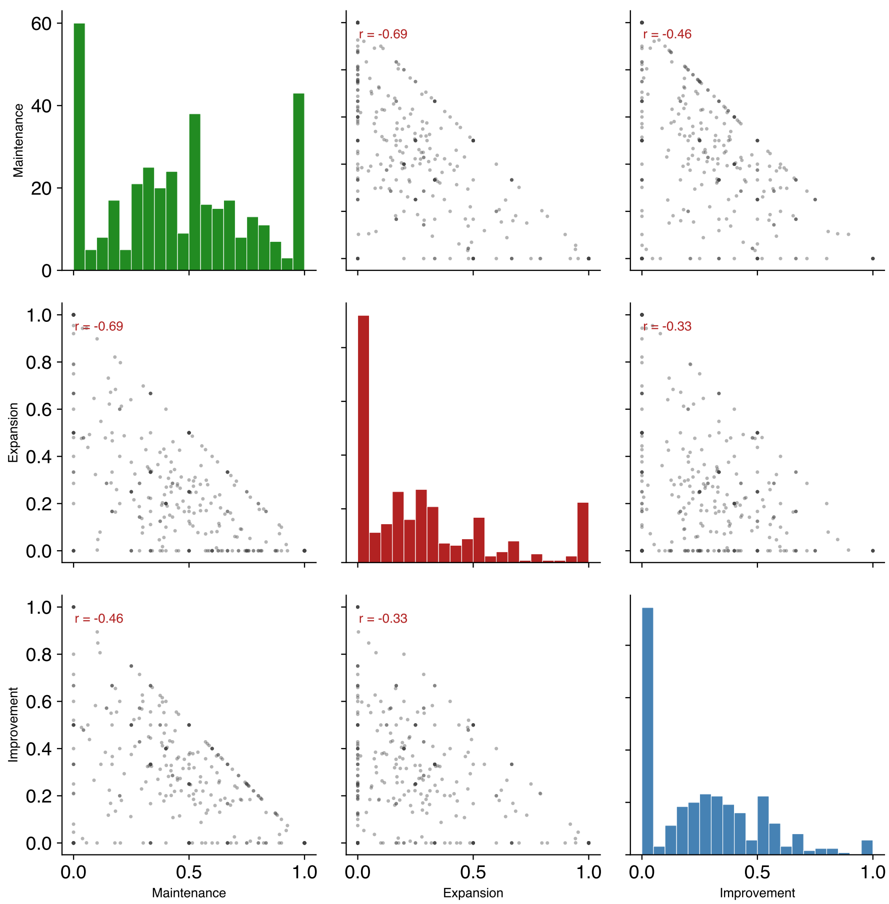
Note: Pre-2009 mean bus ridership per capita against ARRA award per capita, by urbanized area. The formula is tied to population, density, and prior service rather than to capital condition, so it tracks scale only loosely and unmet need not at all. *Source:* National Transit Database panel, excluding Capital Investment Grants (CIG) recipients.

Figure A2: Asset composition by service-intent category: dollar-weighted.



Note: Each bar shows total project dollars ($M) by M1 asset type within each M2 category; each project's dollars are assigned to a single primary asset category. Dollars for projects that fall in no single asset category ("Other") are noted below each bar. *Source:* LLM classification (M1 and M2) of FTA Form 1512 award descriptions; 3,252 projects across 798 awards matched to 365 UZAs.

Figure A3: Most ARRA spending went to maintenance and improvement projects.



Note: Dollar-weighted M2 intensity shares at the UZA level ($N = 365$). Diagonal panels show marginal histograms; off-diagonal panels show pairwise scatter plots. The three shares sum to one for each UZA. *Source:* LLM classification of FTA Form 1512 award descriptions, aggregated to UZA level.

Table A1: Pre-ARRA UZA covariates ($N = 347$)

Variable	Source	Notes
Panel A: Sociodemographic (ACS 2009[a])		
Median age	ACS 2009	
Median household income	ACS 2009	
Per capita income	ACS 2009	
Median commute time	ACS 2009	Minutes
Median rooms per unit	ACS 2009	
Poverty rate	ACS 2009	
Unemployment rate	ACS 2009	Share of civilian labor force
Not-in-labor-force rate	ACS 2009	Pop. 16+
Transit commute share	ACS 2009	All workers
Drive-alone commute share	ACS 2009	
Transit share, income ≥$75k	ACS 2009	
Car access rate, income <$35k	ACS 2009	Drive-alone + carpool
Zero-vehicle household share	ACS 2009	Occupied units
One-vehicle household share	ACS 2009	Occupied units
Renter-occupied housing share	ACS 2009	
High-density housing share	ACS 2009	Units in 20+ unit structures
Pre-1940 housing share	ACS 2009	
Single-family detached share	ACS 2009	
Vacancy rate	ACS 2009	
Share aged 65+	ACS 2009	
Share aged under 18	ACS 2009	
Disability rate	Census 2000	Pop. 18–64; ACS5 coverage insufficient
Out-of-county work share	ACS 2009	Workers employed outside home county
Panel B: Transit system — levels (Federal Transit Administration, 2020), 2005–2008 mean, per capita		
Population	Census 2000	2000 Census UZA population
Population density	Census 2000	2000 Census; persons per sq. mi.
Total spending	NTD	Operating + capital; 2009$
Vehicle revenue miles (VRM)	NTD	
Ridership (UPT)	NTD	Unlinked passenger trips
Fleet size	NTD	Active vehicles
Fleet growth	NTD	Annual change in fleet size
Mean fleet age	NTD	Years; weighted by fleet

Variable	Source	Notes
Average vehicle seats	NTD	Weighted by fleet
Capital intensity	NTD	Capital / (capital + operating)
Federal capital funding share	NTD	
Local capital funding share	NTD	
Federal operating funding share	NTD	
Local operating funding share	NTD	
Rolling stock share of capital	NTD	Share of categorized capital exp.
Facilities share of capital	NTD	Share of categorized capital exp.
Bus share of VRM	NTD	
Bus share of ridership	NTD	
Demand-response share of VRM	NTD	Dial-a-ride, paratransit
Demand-response share of ridership	NTD	
Load factor	NTD	Ridership per VRM
Average speed	NTD	Revenue miles / revenue hours
ADA-accessible fleet share	NTD	
Cost per VRM	NTD	Operating cost / VRM
Cost per trip	NTD	Operating cost / ridership
Fare per trip	NTD	Fare revenue / ridership
VRM per vehicle	NTD	Revenue miles / fleet size
Peak utilization	NTD	Peak vehicles / fleet size
Panel C: Transit system — trends (Federal Transit Administration, 2020), OLS 2000–2008		
VRM growth rate	NTD	OLS slope / period mean
Ridership growth rate	NTD	OLS slope / period mean
Fleet aging rate	NTD	OLS slope in years/year
Panel D: Road infrastructure (FHWA 2008)		
Daily VMT per capita	FHWA	Annual report HM-72
Road miles per 1,000 pop.	FHWA	Annual report HM-71
Freeway DVMT share	FHWA	Share of daily VMT on freeways
Panel E: Governance and political environment		
Authority governance share	Federal Transit Administration (2009)	VRM-weighted share of agencies classified as independent public authorities
Democratic presidential vote share	MIT Election Data and Science Lab (2024)	2008 election; county returns weighted by UZA–county population overlap

Table A2: LLM classifications show high agreement with manual coding

Group	Description	Fields	Agree (%)	Mean κ
Panel A: M1 capital component characterization				
Vehicles	Rolling stock (all types)	6	99.3	0.97
Buses	Bus fleet indicators	6	99.0	0.92
Railcars	Rail fleet indicators	6	99.0	0.35†
Guideway	Infrastructure and track	6	98.8	0.75
Fac. (pax)	Passenger-facing facilities	6	96.2	0.65
Fac. (ops)	Operations and maintenance facilities	6	97.2	0.65
Other	Software, communications, other	2	92.5	0.61
Panel B: M2 service-intent classification				
Flags	Ancillary characteristic flags	8	97.4	0.60
Category	Service-intent category	1	72.4	0.58
All binary fields			97.8	0.85

Notes: $n = 100$ randomly sampled projects. Agreement is the share of project-field cells where coder and model agree. Cohen's κ adjusts for chance agreement; values ≥ 0.80 are conventionally "almost perfect," 0.60–0.79 "substantial." Mean κ for each group is computed over fields where κ is defined. †For Railcars, three of six fields had no positive cases in the sample, leaving κ undefined for those fields and mechanically deflating the group mean.

Table A4: Random sample of project descriptions by M2 service-intent category

Sustained service	**More service**	**Better service**
Replace roof	Purchase new paratransit buses	Modify pedestrian facilities
replace vehicles	Purchase 4 buses	Install an on-board multi-camera video surveillance system including digital DVR on twelve buses
Provide preventative maintenance for the 4 new vehicles	Construct a small bus terminal (Mini Terminal)	Implement an Intelligent Transportation Systems (ITS) program for the entire WRTA fleet
Purchase 2 hybrid buses to replace 2 high-mileage buses	Construct Brownsville Multimodal Terminal]	repaint the Transfer Center building
perform necessary preventive maintenance	purchase buses	install a Bar Signal Overrun System
purchase scheduling software	Purchase 2 new bio-diesel transit buses	Rehabilitate station building windows and doors
Purchase and install up to two stationary fuel cells	Develop a Bus Rapid Transit (BRT) system	Purchase new graphics for fixed-route, paratransit and service vehicles

(continued)

Sustained service	More service	Better service
Purchase passenger shelter parts to maintain, repair and replace current and new shelters and bike racks	Purchase 5 new low-floor 40-foot hybrid/electric buses	Purchase (200) Passenger Rules Signs and (200) Non-Discrimination signs
Purchase Speedaire air compressor	purchase 3 buses for night service	Install additional bicycle racks on trains
Purchase office equipment	expand existing transit services	Purchase 130 fareboxes for use on fixed route buses (replacement)
Purchase related tools for repair	Purchase 1 new 24+2 passenger bus	Replace six 1997 diesel buses with six Hybrid Gasoline-Electric buses
Ensure ongoing preventative maintenance for vehicles and facilities	Purchase 3 new 30' hybrid-electric buses	Procure commuter train horn noise reduction equipment
Overhaul buses	Purchase new buses	rehabilitate two rail stations
purchase Hybrid Maintenance tools	Purchase 1 transit bus	Implement transit enhancement
Perform preventive maintenance on existing revenue fleet	Purchase 6 expansion paratransit vehicles	Identify existing bus stops that require upgrades to bring stops into compliance with Americans with Disabilities Act (ADA) guidance
Fund preventative maintenance expenses	purchase new vehicles for rural public transportation	purchase and install new security surveillance system
Perform preventative maintenance on the remaining fixed route fleet, maintenance facilities, and operations facilities	Purchase 1 hybrid-electric bus	Purchase trash cans, covered shelters and benches
Purchase 2 30 ft low floor coaches (replacements)	Construct new 12-bay transit center	Purchase bus shelters and new signage
Build a fleet maintenance facility	Purchase 2 new hybrid electric 35ft transit buses	Purchase four 40 ft heavy duty low-floor hybrid diesel-electric transit buses to replace aging buses
Provide support for preventive maintenance for vehicles and facilities	Add 1 22' Ford Senator E-450 vehicle to fleet	rehabilitate 1 pedestrian bridge
Purchase 10 upgraded computer systems	Continue to employ 12 builders, heavy equipment operators and electricians	Purchase 15 shelters

(continued)

Sustained service	More service	Better service
Rehabilitate existing administrative and maintenance facility	Acquire one 40 foot bus	Add bike storage units
Replace approximately five (5) 35-foot coaches with coaches of like kind and character	Purchase ten 30-foot hybrid gasoline buses	Add capabilities for dealing with flex service via the current computerized scheduling system
Replace power poles and lines	Engineering/design and construct a bus transfer center and park-and-ride facility	Install passenger waiting shelters
Install a vertical ground closed loop system	Reimburse construction costs for first segment LRT	Procure GIS software/hardware

Notes: 25 project descriptions drawn at random (seed 42) from each M2 service-intent category. Descriptions are LLM-extracted sub-project narratives from FTA Form 1512 cumulative award reports.

Table A3: M2 service-intent categories: LLM vs. RA cross-tabulation

	Expansionary	Better service	Sustained service	Total
Expansionary	**19**	1	2	22
Better service	4	**22**	14	40
Sustained service	0	6	**30**	36
Total	23	29	46	98

Notes: $n = 98$ matched projects (2 filtered as uninformative); RA coding in rows, LLM in columns. Diagonal entries (bold) indicate agreement; off-diagonal entries are disagreements. Overall accuracy: 72%; Cohen's $\kappa = 0.58$. The expansionary and sustained-service categories show high agreement (86% and 83%, respectively). The main source of disagreement is the better-service/sustained-service boundary (14 cases), reflecting that the RA coded under an earlier, broader definition of better-service that included operational facility improvements; the current definition restricts better-service to investments that directly improve what passengers experience.

B Regression Specifications

Let Y_{it} be the outcome for UZA i in year t, α_i UZA fixed effects, γ_t year fixed effects, and $\text{Post}_t = \mathbf{1}[t \geq 2009]$. Treatment variables Exp_i and Imp_i are the award-amount-weighted shares of ARRA projects classified as expansion- and improvement-oriented (maintenance is the omitted category, $\text{Exp}_i + \text{Imp}_i + \text{Sus}_i = 1$). Controls are ARRA_i (per-capita award amount) and PC_{ki} (k-th principal component score). Fixed effects are absorbed by alternating-projections demeaning; standard errors are clustered by UZA throughout.

Difference-in-Differences.

$$\begin{aligned} Y_{it} &= \alpha_i + \gamma_t \\ &\quad + \beta^{\text{exp}}\big(\text{Exp}_i \times \text{Post}_t\big) + \beta^{\text{imp}}\big(\text{Imp}_i \times \text{Post}_t\big) \\ &\quad + \delta\big(\text{ARRA}_i \times \text{Post}_t\big) + \sum_{k=1}^{K} \theta_k\big(\text{PC}_{ki} \times \text{Post}_t\big) + u_{it}. \end{aligned} \tag{1}$$

Panel B — pre-treatment trend adjustment. Before estimating (1), let $t = \text{year} - 2008$. For each UZA i, estimate a linear trend on the pre-treatment outcome and subtract it:

$$\hat{\rho}_i = \frac{\sum_{t \leq 0} (t - \bar{t}^{\text{pre}})(Y_{it} - \bar{Y}_i^{\text{pre}})}{\sum_{t \leq 0} (t - \bar{t}^{\text{pre}})^2}, \tag{2}$$

where sums run over $t \leq 0$ (years 2000–2008) and overbars denote within-unit pre-period means. The detrended outcome $\tilde{Y}_{it} = Y_{it} - \hat{\rho}_i t$ replaces Y_{it} in (1), following Borusyak et al. (2024).

Event Study (ES). Let $\mathcal{T} = \{2000, \ldots, 2007, 2009, \ldots, 2019\}$, with 2008 as the omitted reference year.

$$
\begin{aligned}
Y_{it} = {} & \alpha_i + \gamma_t \\
& + \sum_{s \in \mathcal{T}} \beta_s^{\text{exp}} \big(\text{Exp}_i \times \mathbf{1}[t = s]\big) + \sum_{s \in \mathcal{T}} \beta_s^{\text{imp}} \big(\text{Imp}_i \times \mathbf{1}[t = s]\big) \\
& + \sum_{s \in \mathcal{T}} \delta_s \big(\text{ARRA}_i \times \mathbf{1}[t = s]\big) + \sum_{s \in \mathcal{T}} \sum_{k=1}^{K} \theta_{sk} \big(\text{PC}_{ki} \times \mathbf{1}[t = s]\big) \\
& + u_{it}.
\end{aligned} \tag{3}
$$

For Panel B, only Y_{it} is detrended via (2); covariate interactions enter in levels.

Relationship between DiD and ES coefficients. For a balanced panel with time-invariant treatment, $\hat{\beta}^{\text{exp}}$ from (1) is related to $\{\hat{\beta}_s^{\text{exp}}\}$ from (3) by

$$
\hat{\beta}^{\text{exp}} = \frac{1}{n_+} \sum_{s \in \mathcal{T}^+} \hat{\beta}_s^{\text{exp}} - \frac{1}{n_- + 1} \sum_{s \in \mathcal{T}^-} \hat{\beta}_s^{\text{exp}}, \tag{4}
$$

where $\mathcal{T}^+ = \{s \in \mathcal{T} : s \geq 2009\}$, $\mathcal{T}^- = \{s \in \mathcal{T} : s < 2008\}$, $n_+ = |\mathcal{T}^+|$, $n_- = |\mathcal{T}^-|$, and the denominator $n_- + 1$ counts the reference year 2008 (whose coefficient is zero). The same identity holds for $\hat{\beta}^{\text{imp}}$. In the estimates, the difference-in-differences coefficients clustered by UZA are close to the aggregated yearly event-study coefficients, with standard errors that differ only in the second decimal.

C Additional Results

Table A5: Consequences of ARRA Spending Type on Transit Outcomes (No PC Controls)

	Miles (per capita)	*Fleet size* (per 1,000 cap.)	*Fleet age* (years)	*Trips* (per capita)
Post-2009 mean	13.28	0.57	5.76	14.40
Panel A: DiD				
Expansion × Post	0.05	-0.04	-0.22	-1.06
	(1.42)	(0.07)	(0.49)	(1.30)
Improvement × Post	2.43	0.08	-0.60	-0.20
	(1.99)	(0.10)	(0.64)	(1.42)
Panel B: DiD with Pre-Treatment Trends				
Expansion × Post	-0.20	-0.02	-0.34	0.92
	(1.43)	(0.07)	(0.78)	(1.39)
Improvement × Post	2.47	0.11	0.85	1.31
	(1.68)	(0.09)	(1.00)	(1.32)
N	6,940	6,940	6,940	6,940

Notes: $\text{Post}_t = \mathbf{1}[t \geq 2009]$. Reference category is maintenance. Treatments are award-amount-weighted UZA-level shares of ARRA projects classified as expansion or improvement by the M2 LLM classifier. Outcomes from the National Transit Database panel (2000–2019): vehicle revenue miles per capita, vehicles per 1,000 capita (fleet size), fleet-average age in years, and unlinked trips per capita. All specifications include unit and year fixed effects and ARRA award amount per capita × Post. Panel B subtracts unit-specific pre-treatment linear trends from the outcome (estimated on 2000–2008) before applying two-way demeaning, following Borusyak et al. (2024). Standard errors clustered at the UZA level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Consequences of ARRA Spending Type on Transit Outcomes (Restrictive Expansion Classification)

	Miles (per capita)	*Fleet size* (per 1,000 cap.)	*Fleet age* (years)	*Trips* (per capita)
Post-2009 mean	13.28	0.57	5.76	14.40
Panel A: DiD				
Expansion × Post	0.42	-0.03	-0.47	0.79
	(1.30)	(0.07)	(0.39)	(1.31)
Improvement × Post	3.21*	0.14	-0.45	0.34
	(1.82)	(0.10)	(0.51)	(1.05)
Panel B: DiD with Pre-Treatment Trends				
Expansion × Post	-0.90	-0.12	-1.03	-0.27
	(1.45)	(0.07)	(0.63)	(1.99)
Improvement × Post	2.94*	0.09	-0.77	1.79
	(1.57)	(0.10)	(0.61)	(1.29)
N	6,940	6,940	6,940	6,940

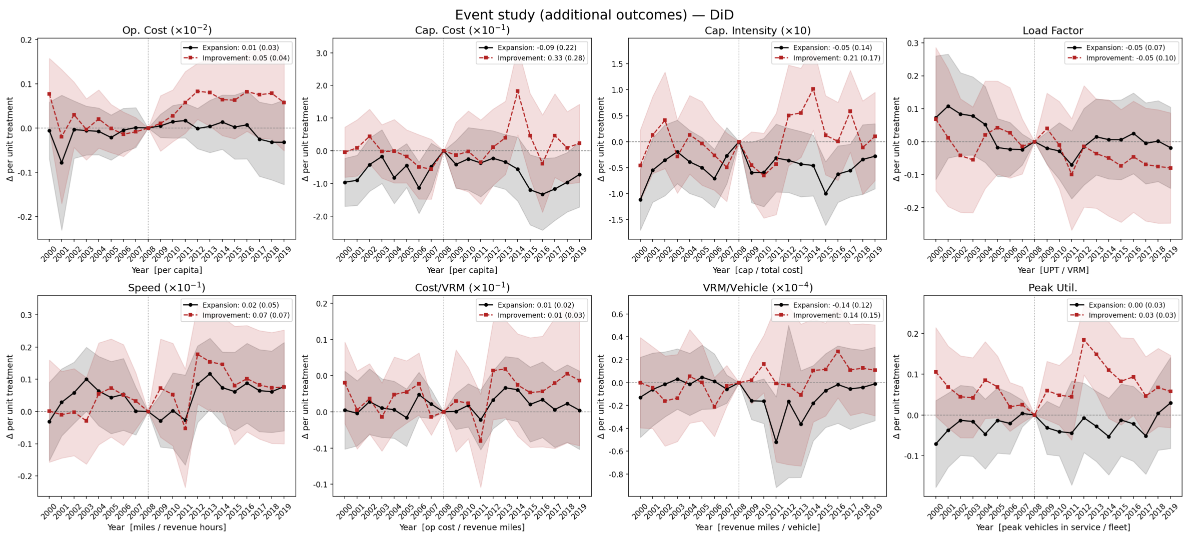
Notes: $\text{Post}_t = \mathbf{1}[t \geq 2009]$. Reference category is maintenance. Identical to Table 3, except expansion is measured with a restrictive classifier that counts a project as expansion only when it explicitly states an expansion goal (e.g. "purchase buses to extend a corridor"); projects with merely implicit expansion are reclassified into the sustained-spending reference category. Treatments are the award-amount-weighted UZA-level shares of ARRA projects classified as explicit expansion or improvement by the M2 LLM classifier. Outcomes from the National Transit Database panel (2000–2019): vehicle revenue miles per capita, vehicles per 1,000 capita (fleet size), fleet-average age in years, and unlinked trips per capita. All specifications include unit and year fixed effects and ARRA award amount per capita × Post. Both panels include PCA scores (PC1–PC14) interacted with the post-treatment indicator as controls. Panel B subtracts unit-specific pre-treatment linear trends from the outcome (estimated on 2000–2008) before applying two-way demeaning, following Borusyak et al. (2024). Standard errors clustered at the UZA level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A7: Consequences of ARRA Improvement Spending (Improvement vs. Expansion and Maintenance)

	Miles (per capita)	*Fleet size* (per 1,000 cap.)	*Fleet age* (years)	*Trips* (per capita)
Post-2009 mean	13.28	0.57	5.76	14.40
Panel A: DiD				
Improvement × Post	3.10*	0.14	-0.33	0.13
	(1.80)	(0.11)	(0.49)	(1.11)
Panel B: DiD with Pre-Treatment Trends				
Improvement × Post	3.17**	0.12	-0.50	1.86
	(1.48)	(0.10)	(0.61)	(1.40)
N	6,940	6,940	6,940	6,940

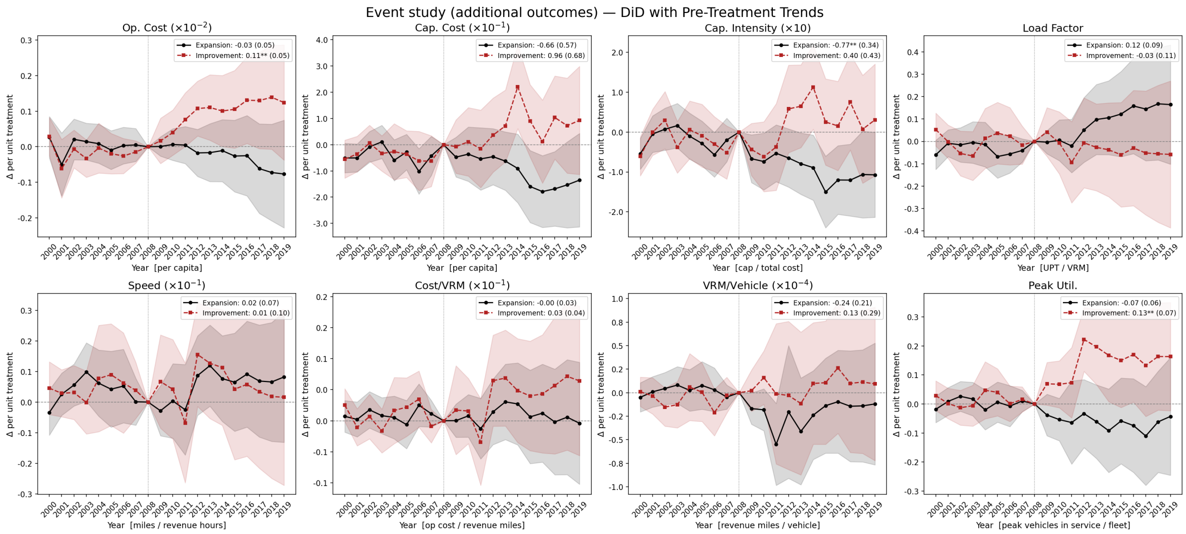
Notes: $\text{Post}_t = \mathbf{1}[t \geq 2009]$. Reference category is expansion and maintenance. Treatment is the award-amount-weighted UZA-level share of ARRA projects classified as improvement by the M2 LLM classifier; expansion and maintenance projects are the omitted reference category. Outcomes from the National Transit Database panel (2000–2019): vehicle revenue miles per capita, vehicles per 1,000 capita (fleet size), fleet-average age in years, and unlinked trips per capita. All specifications include unit and year fixed effects and ARRA award amount per capita × Post. Both panels include PCA scores (PC1–PC14) interacted with the post-treatment indicator as controls. Panel B subtracts unit-specific pre-treatment linear trends from the outcome (estimated on 2000–2008) before applying two-way demeaning, following Borusyak et al. (2024). Standard errors clustered at the UZA level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Figure A4: Event-study estimates of ARRA spending type on additional outcomes (Panel A: no detrending).



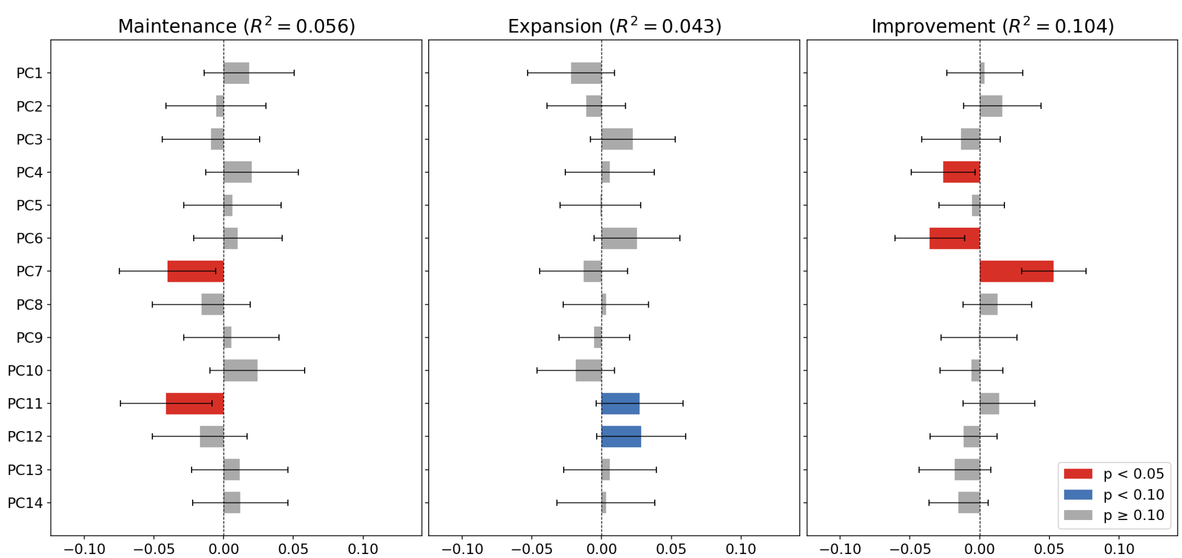
Note: Year-by-year DiD estimates with PCA scores interacted with year indicators (Panel A specification). Outcomes: operating and capital costs per capita, capital intensity (capital share of total cost), load factor (unlinked trips divided by vehicle revenue miles), average speed (vehicle revenue miles per revenue hour), cost per vehicle revenue mile, vehicle revenue miles per vehicle, and peak utilization (vehicles in maximum service divided by fleet size). Each coefficient is the difference in outcome at year t relative to 2008 between a UZA spending entirely on expansion or improvement versus entirely on maintenance. Shaded bands are 95% confidence intervals. Legend reports the aggregated DiD and its standard error. ARRA award per capita × year controlled throughout.

Figure A5: Event-study estimates of ARRA spending type on additional outcomes (Panel B: pre-treatment trend adjustment).



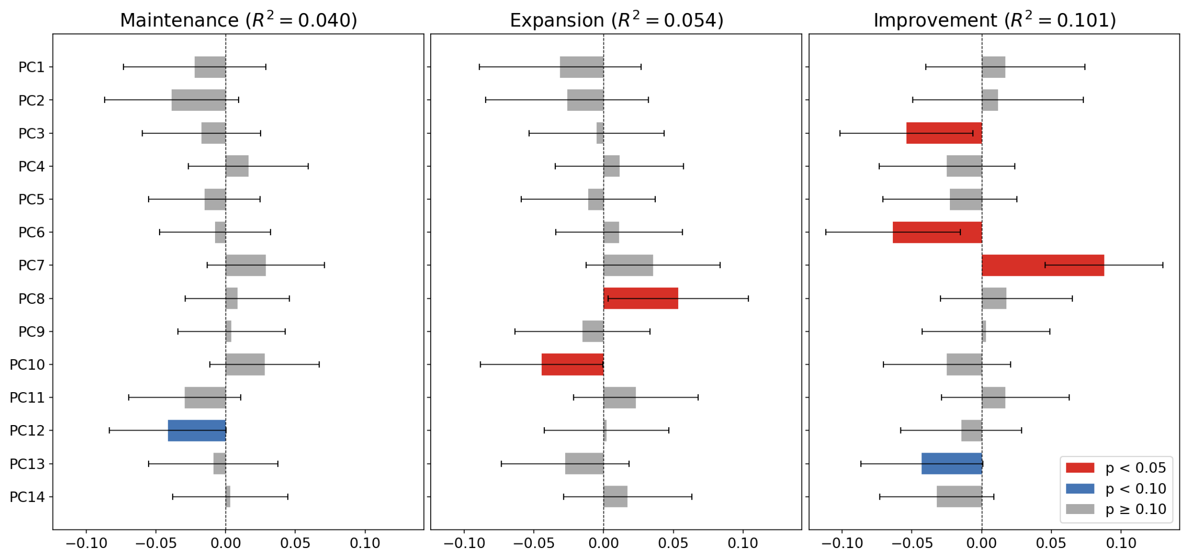
Note: Year-by-year DiD estimates with PCA scores interacted with year indicators and unit-specific pre-treatment linear trends subtracted from the outcome (Panel B specification, following Borusyak et al. (2024)). Outcomes: operating and capital costs per capita, capital intensity (capital share of total cost), load factor (unlinked trips divided by vehicle revenue miles), average speed (vehicle revenue miles per revenue hour), cost per vehicle revenue mile, vehicle revenue miles per vehicle, and peak utilization (vehicles in maximum service divided by fleet size). Each coefficient is the difference in outcome at year t relative to 2008 between a UZA spending entirely on expansion or improvement versus entirely on maintenance. Shaded bands are 95% confidence intervals. Legend reports the aggregated DiD and its standard error. ARRA award per capita $\times$ year controlled throughout.

Figure A6: Spending-orientation predictors: *wmean2* outcome, without award-amount control.



Note: Multivariate OLS; all 14 PCs included jointly (no ARRA award amount control). $N = 347$ urbanized areas, HC1 standard errors, 95% CI. Color: red ($p < 0.05$), blue ($p < 0.10$), gray ($p \geq 0.10$). Outcome: *wmean2*. PC definitions in Table 1.

Figure A7: Spending-orientation predictors: *wmeanmax* outcome, with award-amount control.



Note: Multivariate OLS; all 14 PCs plus ARRA award amount included jointly. $N = 347$ urbanized areas, HC1 standard errors, 95% CI. Color: red ($p < 0.05$), blue ($p < 0.10$), gray ($p \geq 0.10$). Outcome: *wmeanmax*. PC definitions in Table 1.