Transport Infrastructure and Policy Evaluation

Dave Donaldson*

Abstract

This chapter develops a framework whose goal is to survey and synthesize the answers that regional and urban economists have given to the question: how beneficial are transport infrastructure investments and other transport policies? Emphasis is placed on theoretical results about sufficient statistics that capture approximate impacts in both distorted and undistorted economies, as well as how new advances in data collection and causal inference are poised to leverage these theoretical results and thereby modernize and extend standard templates for infrastructure and policy evaluation. The chapter concludes with discussions of optimal policy, political economy, and practical matters of infrastructure provision.

*MIT and NBER. ddonald@mit.edu. Chapter to appear in the forthcoming *Handbook of Urban and Regional Economics* (2025). I am grateful to Nathaniel Baum-Snow, Arnaud Costinot, Victor Couture, Pablo Fajgelbaum, Cecile Gaubert, Gabriel Kreindler, Stephen Redding, and Vincent Rollet for comments.

1 Introduction

Interactions across space define the field of spatial economics, and a key mechanism that mediates those interactions is transport. But how much would an economy benefit from making infrastructure improvements or other policy interventions that make transportation cheaper? This chapter develops a framework whose goal is to survey and synthesize the answers that regional and urban economists have given to this question.

Our starting point in Sections 2-4 draws on a long tradition in Public Finance and Macroeconomics (particularly, the recent presentations in Baqaee and Farhi, 2019 and Baqaee and Farhi, 2020) to highlight the conditions under which the answers that analysts offer to this question hinge on the values that they assign to a small set of sufficient statistics. The setting is a general (closed economy) economic model in which goods and factors of production are exchanged among households and an abstract notion of firms. Many such exchanges take place across space, so they leave in their wake the spatial flow of goods, people, and capital. For example, in this framework, when firms use a truck to ship merchandise to a distant consumer this obviously involves the flow of goods along the roads involved. But equally, when individuals use the subway to commute from their residences to their workplaces this amounts to a flow of people mediated by pseudo-firms that carry out such commuting exchanges.

Against this backdrop, a core question is: what would be the aggregate welfare benefit of a government investment in one particular segment of the nation's road or subway systems? We model such improvements as an increase in the productivity of the technology that facilitates flows over the transport network segment of interest. A well-known result (due to Hulten, 1978) states that, absent any distortions in this economy, a first-order approximation to the total welfare impact of this shock is given simply by the product of two ingredients: (i) the size of the productivity improvement on the segment in question; and (ii) the share of total income (i.e. the Domar weight) that was being "spent" on flows of goods, people, and capital along that segment prior to its improvement. Perhaps surprisingly, even in full general equilibrium, no other ingredients are needed—including all other aspects of the economy, or even any understanding of why the Domar weight in question takes the value it does (that is, why the shipments or commutes are taking place the way they are). This is because in efficient economies the value-adjusted marginal product of any factor is equalized across all uses—goods, locations, points in time, etc.—so aggregate benefits do not depend (to first order) on the way that reallocations take place across such uses.

This observation, or variants of it, underpins the long-standing interest in measuring both of these ingredients. A seminal application was provided by Fogel (1964), whose evaluation of the US railroad system in the nineteenth century began with attempts to measure these two inputs. But a core theme of this chapter is that recent advances have allowed researchers to make substantial progress on how we quantify both of them. Section 5 reviews progress on new methods for measuring transport costs (of goods, people, and capital), which relates inversely to transport productivity. Building on this, we also review the important question of how infrastructure investments actually appear to have a causal effect on transport costs (i.e. improve transport productivity), so as to provide the first ingredient of any Hulten-inspired approach.

Similarly, Section 6 turns to the second Hulten input: the measurement of Domar weights on any given transport network segment throughout the economy. Modern advances in tools for data creation, storage, and analysis have endowed researchers with a blossoming ability to measure spatial flows in ways that may have seemed unthinkable even ten years ago. Combined with the advances in Section 5, these tools offer chances to bring twentieth century theory into twenty-first century implementation. For example, as pioneering as Fogel's analysis was, his approaches to measuring the causal impact of rail technology on the productivity of the transport sector, and his attempts to measure trade flows (even in goods alone, never mind in people and capital), were hampered by enormous data constraints relative to what is possible today.

However, one limitation of the Hulten-based approach is that it only provides a first-order approximation. Going to the second-order approximation requires more knowledge, but perhaps less additional knowledge than is widely appreciated. Second-order terms reflect only the change in flows that take place along the affected segment—that is, on the change in the Domar weight referred to above. This is again an immensely useful sufficient statistic result. Out of all the endogenous responses that we can expect a given infrastructure investment to have on allocations throughout a spatially-interlinked economy, only one such response needs to be measured or predicted to get second-order effects right. Indeed, even at higher than second-order approximations, the only adjustment that matters for aggregate welfare is the total change in flow along the directly affected (i.e. productivity-improving) segment. Fogel (1964) was entirely aware of this point, but his analysis rested on the strong assumption that all of the actual change in transport that took place, relative to the pre-rail era, was the causal result of the rail network, rather than some mix of this impact and that of other changes in the economy. Modern approaches to causal inference seek ways to relax strong assumptions like this one.

Section 7 therefore reviews modern approaches to the question of how infrastructure investments appear to causally impact the amount of spatial flows (again, of goods, people, and capital) that take place on any given segment. One challenge is that such impacts are unlikely to correspond to simple theoretical objects—for example, the second-order term does not depend on the slope of a normal demand curve for travel (which is defined while holding constant the prices of everything else), but instead on the total endogenous response of travel to the full set of price changes that result from the intervention under study. It is therefore tempting to trust (at least in the case of ex-post evaluations of projects that have already happened) estimates based on reduced-form empirical contrasts between relatively exposed and un-exposed spatial units, since such effects would ideally capture the entire endogenous response that is desired. However, the literature reviewed here has raised important discussions about the plausibility of finding pure control (completely un-exposed) units to which treatment units can be compared, and these discussions therefore motivate the need for approaches that combine both theoretical assumptions and empirical estimation.

A final deviation from Hulten-based approaches begins with the observation that economic distortions surely exist in most contexts of interest. For example, transport itself may be distorted due to the presence of congestion or pollution externalities. Or aspects of the wider economy may be distorted by the presence of knowledge spillover externalities, market power, or taxation. All of these distortions result in activities on which there is a non-zero distortion wedge that corresponds to the gap between the buyer's willingness to pay and the seller's marginal cost.

In environments that feature such distortions, a well-known result highlights that the first-order welfare impact needs to be augmented (that is, to go beyond the Hulten effect which continues to hold) with an additional term that measures the change in allocative efficiency caused by the infrastructure improvement. This effect is given by the interaction between pre-existing wedges on each activity and the change in the size of each activity. Again, this sufficient statistic result is useful in several ways. First, it highlights how knowledge of pre-existing wedges is essential, but knowledge of changes in any such wedges is not needed per se (since the direct effect of changing wedges is merely a transfer across agents that does not alter the aggregate notion of welfare that we use here). And second, the only allocative changes that matter are those that are distorted initially, so these responses deserve focus while responses on undistorted activities do not. Naturally, the undistorted environments discussed above are a simple case in which all activities are in this category and there is no potential for infrastructure improvements to affect allocative efficiency.

Helpfully, this augmented formula for distorted economies is true regardless of whether the underlying exogenous change is a transport infrastructure improvement or some other change in transport policy (such as a toll or a congestion charge). We therefore bundle approaches to transport infrastructure and policy evaluation in our discussion of all such environments with distortions.

With this result in mind, we turn to a wide range of examples of economic distortions that have been covered in the transport (infrastructure and policy) evaluation literature. Section 8 focuses on the case of distortions in the transport sector itself, focusing on cases such as congestion, transport policies, pollution, safety, imperfect competition, search, and corruption. Section 9 then broadens the scope of this analysis to cover distortions in the wider economy as a whole, where researchers have emphasized connections between transport evaluation and the pre-existing extent of such distortions. Examples covered include agglomeration and congestion externalities, innovation and knowledge flows, pollution and carbon emissions, imperfect competition, taxation, regulation, credit and insurance market constraints, conflict, crime, labor search and matching, inefficient unemployment, and child labor.

While Fogel (1964) was also aware of the potentially important role of distortions in the correct evaluation of infrastructure projects (see especially his discussion in Fogel, 1979), these examples highlight the scope for modern tools of measurement and causal inference to dramatically change the way these that transport infrastructure and policy evaluations are done.

At this point a skeptical reader may rightly ask how the sufficient statistic results discussed thus far can interact with quantitative models (in the service of either ex-ante or ex-post evaluation). After all, given that such modeling exercises fully commit to a particular model that aims to arrive at a fully accurate (zero-approximation) answer, pausing to compare to first- or second-order approximations may seem unnecessary.

One answer appeals to scientific communication. For example, consider a quantitative modeling exercise that arrives at a relatively large impact of a given investment. If the model is undistorted then this model's output can always be decomposed into the first- and second-order approximation terms stressed here, plus an additional effect collecting all higher-order terms. So any large total impact must arise from the fact that the model includes a large productivity shock hitting a large Domar weight, or that its structural features predict a particularly large flow response on the improved segment, or that first- and second-order approximation are inaccurate. Likewise, if the model contains distortions (yet a small Hulten effect) then it must predict that high-wedge activities will be induced to grow substantially by the investment in question. Either way, the sufficient statistic terms here provide an easy way to explain such effects to the reader. They also make it easier for researchers to compare their estimates across models and contexts.

A second answer appeals to scientific focus. All quantitative modeling exercises force the developer to make choices about which mechanisms are likely to be most important to get right if an accurate answer is to be delivered. The sufficient statistics emphasized in this chapter aim to offer guidance on such choices. For example, in undistorted cases, the key question is whether increases in transport flows across improved segments will be large or small (or even negative). Model features that have little bearing on these magnitudes seem less worthy of the modeler's (or readers') scarce attention. The same is true, in distorted cases, for distorted activities where we have little reason to expect that the program in question will cause such activities to grow or shrink.

Having exposited some core components of infrastructure and policy evaluation, the remainder of this chapter turns to two themes that take such evaluations in new directions. In particular, Section 10 moves from the foregoing discussion of *given* investments and policies to one of *optimal* investments and policies. And Section 11 covers work on the topic of infrastructure provision—such as the political economy of infrastructure location, and the question of how given investments be procured and at what cost—which deserves to be a part of any discussion of real-world comparisons of benefits with costs.

Sections 12 and 13 then conclude with comments about gaps in the literature, open questions, and directions for future work.

A distinct approach to policy and infrastructure evaluations that is not emphasized here is what is often called the pure program evaluation approach. If proxies for economic welfare can be measured at the regional level, and if regions are exposed to as-good-as random variation in transport infrastructure or policy programs (or measures of the intensity of such programs), then the average treatment effects of such programs (or their intensity) can be estimated via standard tools (Imbens and Rubin, 2015). Early examples of such an approach include Chandra and Thompson (2000), Duranton and Turner (2012), Faber (2014), Donaldson and Hornbeck (2016), and Donaldson (2018). And a vast body of work has subsequently been contributed in this tradition.

The obvious reason to allocate this chapter's space constraints away from such work is that it was the primary focus of Redding and Turner (2015) in the previous volume of this *Handbook*. But there are two additional reasons for the emphasis pursued here. One is a desire to build bridges between the program evaluation approach and the Hulten-like approach that has been applied widely since at least Fogel (1964). Hulten's first-order approximation calls for an estimate of a particular weighted average treatment effect (the Domar-weighted treatment effect of the program on productivity). The second-order approximation (one half times the productivity change-weighted treatment effect of the program on transport flows) and the allocative efficiency correction for distorted economies (the wedge-weighted treatment effect of the program on all distorted allocations) are no different. Such connections may allow for researchers to combine the benefits of program evaluation approaches with the benefits of the sufficient statistics that non-parametric theory suggests should be most important. This arbitrage seems to be only very rarely exploited in work to date.

A second point about the comparison with program evaluation approaches is more pragmatic. When two program evaluation exercises arrive at different estimates for their two separate contexts, it is natural to wonder why—and it is just as natural to wonder which estimate is more likely to be informative of some ex-ante program under consideration elsewhere. There are many reasons why estimates will differ across contexts, but the sufficient statistics approaches described here serve to narrow that potential discrepancy. For example, Hulten's theorem says that, in any undistorted economy, regardless of its underlying characteristics and features, we should expect that a good approximation to the impact of a productivity shock to certain transport segments is given by the size of that shock and the share of total income in the economy that flows over the relevant segments (i.e. their Domar weights). Put differently, two estimates should no longer disagree once they are scaled by the size of the productivity shock and the relevant Domar weights. As simple as this point is, to the best of my knowledge, program evaluation papers rarely attempt to measure these magnitudes—even though it would be fascinating to learn that the estimated impacts of a given infrastructure program deliver benefits that are substantially different from what conventional economic theory tells us could happen in an undistorted economy, and hence that infrastructure and distortions interact in rich ways.

Beyond the admirable survey in Redding and Turner (2015), this chapter draws on many other antecedents. These include surveys and summaries of work on infrastructure evaluation in general—such as those by Mohring (1993), Baldwin et al. (2011), Kockelman et al. (2013), Laird and Venables (2017), Eliasson and Fosgerau (2019), Duranton and Venables (2019), Glaeser and Poterba (2020), Donald et al. (2024), Gonzalez-Navarro et al. (2023), Laird and Tveter (2023), and Redding (2025)—as well as surveys of approaches to the historical railroad context in particular—such as those by Crafts (2004), Leunig (2010), and Atack (2024).

2 Theoretical Framework

This section lays out a framework for interpreting how we might expect transport investments and policies to impact the spatial economy. A key goal is to provide a means for discussing how recent work has contributed to this endeavor, as done in the remaining sections of this chapter.

2.1 Model setup

We consider a general model economy whose presentation follows closely that in Baqaee and Farhi (2019) and Baqaee and Farhi (2020).[1]

2.1.1 Tastes and technologies

This economy produces a set of N goods (i.e. both merchandise and services). The set of such goods considered should be expansive. In particular, it should include not only conventional notions of contributions to measured GDP but also wider phenomena that are equally important

[1] This builds on an earlier tradition exposited in, for example, Dixit and Norman (1980) and Starrett (1988).

to the study of transportation such as travel time, residential amenities, and pollution. We follow the convention that when goods are transported across any pair of locations the original good (indexed by the origin location of the transport in question) is transformed into a downstream good (indexed by the destination location). In all cases goods are produced by a "producer", even though in cases of transport there may not be a traditional firm doing the producing.

Each producer can produce its good using a constant returns, but otherwise arbitrary, technology that uses goods and factors as inputs. Primary factors of production (indexed by f) are endowed in arbitrary amounts denoted by L_f. This setup, while abstract, allows for common features of production such as non-tradable goods (e.g. a firm uses capital and labor to produce haircuts), supply chains (e.g. a firm uses other goods as inputs), and inter-regional trade. We focus on a fixed set of producers since, under CRS production, entry and exit have no first-order welfare consequences. Finally, these producers are assumed to maximize profits.

The economy also contains a representative household that owns all producers and factors of production.[2] This household uses its total income Y—derived from profits Π and earnings w_f per unit factor f—to fund consumption of goods in a way that maximizes a homothetic but otherwise arbitrary utility function. We take $Y = 1$ by choice of the numeraire, so hence Π, for example, denotes profits as a share of total income. To incorporate phenomena such as migration we imagine that one factor of production owned by the household is a fixed amount of time, and that any unit of this resource can be allocated to production in only one region. Similarly, residential amenities are incorporated because of the fact that tradable goods are indexed by location. Finally, commuting occurs whenever the time spent producing in a location differs from that consuming in it. The household takes prices as given and maximizes an arbitrary (though homothetic) utility function over the consumption of goods.

2.1.2 Flows

We now introduce notation for connecting this model to available data on the value of flows of goods. Specifically, let i (or j) denote any one of the $F + N + 1$ "entities" (factors, producers, or the household) in this economy. Then let X_{ij} denote the value that i pays for the exchange of a good

[2]See Donald et al. (2024) for a discussion of conditions under which a multi-agent spatial economy can be presented in this manner.

obtained directly from any j—for example, this could be the purchase made by the household from a producer, the payment of a producer for use of factor services, or the payment of a producer for its purchase of a good. Finally, let $\widetilde{\Omega}_{ij} \equiv X_{ij}/\sum_k X_{ik}$ denote the share of i's total expenditures that it devotes to direct purchases from j.[3] The matrix $\widetilde{\Omega}$ can be thought of as an economy-wide input-output shares matrix among all economic entities. We define the Leontief inverse of this matrix as $\widetilde{\Psi} \equiv (I - \widetilde{\Omega})^{-1}$. As usual, whereas $\widetilde{\Omega}_{ij}$ captures only the value of direct exchange from j to i, $\widetilde{\Psi}_{ij}$ includes all of the value of direct plus indirect (e.g where j sells to l, and l sells to i) exchange from j to i.

2.1.3 Distortions

The next key element of this model economy is an arbitrary set of distortions that may potentially prevail on any exchange. Such distortions could derive from phenomena such as market power, government actions such as taxation/subsidies and regulation, or externalities such as knowledge flows, pollution, or congestion. Notably, such distortions can exist both within transportation activities themselves (e.g. traffic congestion) and in the wider economy, though we shall return to this distinction below. The essence of such market failures is that they create a difference (or "wedge") between the price that the buyer pays in an exchange and the seller's marginal cost of conducting the exchange. As the well known first welfare theorem illuminates, in the face of such distortions the equilibrium in this economy is unlikely to be efficient (in the sense of maximizing the welfare of its representative household).

Letting $\mu_{ij} \geq 0$ denote the ratio of the price paid by the buyer to marginal cost of the seller on the j-to-i exchange, the case of $\mu_{ij} = 1$ occurs if there is no distortion at all, whereas cases such as a tax or markup can be captured with $\mu_{ij} > 1$ and cases such as a subsidy can be captured by $\mu_{ij} < 1$. Whenever $\mu_{ij} > 1$ on an exchange this will generate "profits"—regardless of whether the wedge is due to policy, market power, or otherwise, these wedges create revenues that are collected by some actor. We let Π denote the sum of all such profits and, as described above, rebate this to the representative household.

Finally, it proves useful below to define $\Omega_{ij} \equiv \widetilde{\Omega}_{ij}/\mu_{ij}$ and hence $\Psi \equiv (I - \Omega)^{-1}$. The crucial

[3]When an entity has no expenditures (as happens only in the case that the entity is a factor of production) we set $\widetilde{\Omega}_{ij} = 0$.

distinction between $\widetilde{\Omega}$ and Ω (and by analogy between $\widetilde{\Psi}$ and Ψ) is that $\widetilde{\Omega}$ measures exchange using X_{ij}, where the quantity changing hands is reported in terms of the value paid by the buyer. On the other hand, Ω measures exchange using X_{ij}/μ_{ij}, where the quantity is valued at the seller's marginal cost. A second useful definition is the Domar weight of any entity i (its total sales as a share of total economy-wide income), denoted by $\lambda_i \equiv \Psi_{Hi}$, i.e the row of the Ψ matrix corresponding to the household.

2.1.4 Transportation questions of interest

Two key questions that motivate much of the work summarized in this chapter are: (i) what is the impact on economic welfare of making improvements in the economy's transportation infrastructure? and (ii), what is the impact on economic welfare of making changes in the economy's transportation policies? In the framework above, transportation occurs whenever the flow of a good from entity i to entity j moves across space—a notion that can capture both the flow of goods and services as well as the flow of people (commuting, leisure travel, and migration). We model improvements in transportation infrastructure as simply productivity enhancements in the technology that carries out any such flow. Focusing on Hicks-neutral productivity changes for simplicity, we can generically denote an improvement in the technology that produces the ij exchange as ΔA_{ij}. We are particularly interested in the case of improvements, so will refer to the case with $\Delta A_{ij} > 0$, but there is nothing asymmetric in the discussion.[4]

Naturally, the productivity improvement ΔA_{ij} is unlikely to come for free. However, in line with most of the literature, for simplicity, we will ignore the resource cost of infrastructure construction. The important topic of infrastructure provision is discussed in Section 11.

While a major focus of the discussion below concerns changes in transportation infrastructure, a second type of important application derives from changes in transport policy. All policies, transportation-related and otherwise, that affect marginal decisions should be included in the set of distortions captured in μ. We can therefore examine the effects of exogenous policy changes on any given ij exchange by introducing the shock $\Delta\mu_{ij}$.

[4] See Gibbons et al. (2024) and Santamaria (2020) for studies of infrastructure removals.

3 Impacts of Infrastructure Improvements in Undistorted Economies

We begin with the special case in which all goods are produced and consumed competitively and there are no externalities or other sources of market failure. In the notation above this means that $\mu_{ij} = 1$ for all i, j. The equilibrium in such an undistorted setting will therefore maximize the utility of the representative household, a feature that considerably simplifies the quantification of comparative statics concerning the representative household's welfare. Throughout this section we focus on impacts of transport infrastructure, rather than transport policies, since our focus here on undistorted settings curtails the interest in policy impacts—for example, they are zero to first order. We return to policy evaluation in the next section.

3.1 A first-order approach

As discussed, a core question in the study of transport is: what is the economic value of improvements in transportation infrastructure? That is, we seek to understand the comparative static exercise that concerns the size of the welfare effect ΔW caused by the change ΔA_{ij} in the productivity of transport over a particular "segment" of the transport system denoted ij.[5] Such a segment may conduct one exchange (e.g. from one firm to one consumer), or an aggregation of exchanges within some set (e.g. all flows from the firms in location A to the consumers in location B), or literally only one road segment (e.g. all of the flow, regardless of origin and destination, that must pass over one stretch of roadway).

This exercise holds constant all exogenous aspects of the economy (technologies, preferences, endowments) apart from A_{ij}, but it allows all of the economy's endogenous characteristics (production, consumption, flows of goods and people, prices, wages, etc.) to change in full general equilibrium as a result of the shock ΔA_{ij}.

A foundational result for studying such productivity improvements is the so-called Hulten's theorem (Hulten, 1978). This result states that in an undistorted economy of the form considered here, we have

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log A_{ij}} = \Psi_{H,ij} \tag{1}$$

[5] For expositional purposes we focus on productivity shocks, but amenity shocks—as considered by Brinkman and Lin (2024) for example—can be incorporated analogously.

where we define $\Psi_{H,ij} \equiv \lambda_i \Omega_{ij}$ as the household's total exposure to the ij segment. This is a remarkable result, since it says that—as long as one has data on the pre-shock exposure $\Psi_{H,ij}$—no details of the model need to be specified (beyond the very important assumption that there are no distortions) for us to evaluate the first-order effects of the productivity change.

Two observations are useful for appreciating why this result is true. The first is to observe that equation (1) would be the welfare effect of the technology shock in a hypothetical setting in which the allocation of primary factors to the production of goods were held completely fixed. In such a hypothetical scenario, the productivity improvement ΔA_{ij} would enable the factors used, directly or indirectly, in the production of the ij flow to produce more of their goods. Whether those goods are consumed immediately or used in downstream production for other goods that are ultimately consumed, the value of that increase to the representative agent would be given, proportionally, by $\Psi_{H,ij}$ per proportional unit increase.

The second observation is to recall that we have considered an arbitrary general equilibrium environment in which the allocation of primary factors to the production of goods is very much not fixed. The productivity change ΔA_{ij} will cause all prices and quantities in the economy to change endogenously in response to the productivity improvement. For example, when roads get faster, we can expect more travel; with faster travel, we can expect more commuting; with more commuting, we can expect more production; with more production, we can expect more downstream production, and ultimately more final consumption; and with more consumption, we might expect even more commuting. One could go on and on. But Hulten's result states that even though all such price and quantity adjustments do very much take place, the combined effect of all such adjustments on economy-wide welfare is zero to first order. This follows from an envelope-like result. As mentioned, this undistorted economy is always acting as if it is maximizing welfare. Maximizing behavior means that small changes in all allocative choices have no impact on the objective function. Hence, the only effect of the technological improvement is the direct effect—that which holds *as if* no such adjustments took place—in (1).

To further underscore the power of Hulten's result, it is useful to note all the objects that do *not* need to be measured to learn $\frac{d\log W}{d\log A_{ij}}$. For example, even though improving one road segment (that improves the flow ij) is sure to cause changes in other flows in the system (e.g. some flow ik), neither the initial level of that other flow $\Psi_{H,ik}$ nor its change $\Delta\Psi_{H,ik}$ enter equation (1). The same is true for increases in the production of other goods near the affected road segment, changes in business practices nearby, increases in wages, elevated house prices, etc. Even though such effects surely arise and may appear important for understanding the impact of ΔA_{ij}, Hulten's theorem tells us that they net out to zero (at least to first order). Instead, the sufficient statistic that is needed is simply $\Psi_{H,ij}$, the exposure on the one flow ij that is *directly* affected in the sense that its *productivity* is affected by the shock.[6] This is beneficial for simplicity (fewer effects to be quantified), for robustness (fewer assumptions required to specify and estimate the underlying primitives of the economy), and for removing the risk of mistakenly double-counting benefits (a concern highlighted by Mohring, 1993, for example).

The result in equation (1) allows us to write the first-order approximation to the welfare change due to an infrastructure improvement as

$$\Delta \log W \cong \Psi_{H,ij} \Delta \log A_{ij},$$

for sufficiently small improvements $\Delta \log A_{ij}$. One useful property of this formula is that it is linear in $\Delta \log A_{ij}$. Suppose we wish to evaluate an infrastructure project that raised the productivity of not only one particular flow ij but all such flows within a set of "infrastructure improvements" denoted $\mathcal{II}$. And suppose, for simplicity, that all such proportional technology improvements within this set take the same value, denoted $\Delta \log A_{II}$. Then we have

$$\Delta \log W \cong \Psi_{H,II} \Delta \log A_{II}, \tag{2}$$

where $\Psi_{H,II} \equiv \sum_{ij \in \mathcal{II}} \Psi_{H,ij}$ is the sum of exposures on all flows whose infrastructure is undergoing the improvement of interest. Notably, this implies that (to first-order) there is no sense in which distinct segments of improvements in an infrastructure network are complementary with, or substitutable for, one another. While commenting on Fogel (1964), David (1969) referred to this feature—since it implies that nothing would be lost by randomly choosing transport segments for

[6] Studies reviewed below have investigated the scope for infrastructure improvements to affect local firms' productivity at producing as a whole (rather than their ability to ship from ij as discussed here). Such effects do not happen in undistorted economies since all firms are always using inputs (including inputs designed to find more efficient technologies such as R&D spending) in the most efficient way of producing output possible, rendering further productivity improvements impossible. We therefore return to this point in Section 9 below when discussing distorted environments.

improvement, at least conditional on their Domar weights—by pointing out that "adherents to the theory of strategic bombing may be especially upset...".

The linearity in equation (2) is extremely useful for evaluating infrastructure projects for two reasons. The first is simplicity: even though a given road network expansion will improve the productivity of transport among many pairs of affected locations, the micro-level details of such an expansion are irrelevant. All that we need to keep track of is the sum of such exposure, $\Psi_{H,II}$. The second concerns data limitations. As discussed further in Section 6, analysts may be unable to observe all of the individual micro-level flows ij within the affected set $\mathcal{II}$, but linearity implies that no aggregation bias occurs when using macro-level data on $\Psi_{H,II}$.

As described in more detail below, the result in (2) has played (and continues to play) an extremely important role in the evaluation of transportation infrastructure projects in ostensibly undistorted economies. If we know the value of $\Delta \log A_{II}$ then all that is needed to conduct an ex-ante evaluation—a forecast of an intervention before it has happened, or one that is being contemplated but may never happen—is to know the current value of the exposure to affected flows, $\Psi_{H,II}$. For example, if we think that a given subway line will make a particular set of commuters' journey time fall by 10% then equation (2) implies that (for an undistorted economy and as a first-order approximation) the change in welfare due to this subway line is given by 0.1 (i.e. $\Delta \log A_{II} \cong 0.1$) times the total share of the value of commuters' time in the total economy-wide income (i.e. $\Psi_{H,II}$).

The discussion so far has emphasized the considerable attractions of evaluation approaches grounded in Hulten's theorem. However, three limitations deserve emphasis. First, this result applies only to the welfare of this economy's single agent. Questions about other features of this economy, such as the number of manufacturing jobs, or the spatial concentration of firms, cannot in general by answered by any sort of equivalently sufficient statistic result. Relatedly, if we are interested in distributional questions concerning the economic welfare of two different agents (such as the skill premium or spatial inequality) then Hulten's theorem is similarly of no use. Second, as stressed, we have so far only considered undistorted settings. The important limitations of Hulten's theorem in distorted environments will be treated in subsequent discussions below. Finally, a third limitation of Hulten's theorem concerns the fact that it is only a first-order approximation. We turn to this point now.

3.2 Second-order effects

Consider now the second-order approximation to the change ΔW caused by the infrastructure improvement ΔA_{ij}, namely

$$\Delta \log W \cong \Psi_{H,ij} \Delta \log A_{ij} + \frac{1}{2} \frac{\mathrm{d}^2 \log W}{\mathrm{d} \log A_{ij}^2} (\Delta \log A_{ij})^2. \tag{3}$$

However, following Baqaee and Farhi (2019), since equation (1) says that $\frac{\mathrm{dlog}W}{\mathrm{dlog}A_{ij}} = \Psi_{H,ij}$, it is clear that the derivative in the second-order term of (3) is given simply by

$$\frac{\mathrm{d}^2 \log W}{\mathrm{d} \log A_{ij}^2} = \frac{\mathrm{d}\Psi_{H,ij}}{\mathrm{d} \log A_{ij}}$$

and hence the second-order approximation sought is given by

$$\Delta \log W \cong \Psi_{H,ij} \Delta \log A_{ij} + \frac{1}{2} \frac{\mathrm{d}\Psi_{H,ij}}{\mathrm{d} \log A_{ij}} (\Delta \log A_{ij})^2. \tag{4}$$

As a final step we can, as above, imagine that all infrastructure segments ij in the set $\mathcal{II}$ are being upgraded. This would cause a change in welfare (up to a second-order approximation) that is equal to

$$\Delta \log W \cong \sum_{ij \in \mathcal{II}} \Psi_{H,ij} \Delta \log A_{ij} + \frac{1}{2} \sum_{ij \in \mathcal{II}} \frac{\mathrm{d}\Psi_{H,ij}}{\mathrm{d} \log A_{ij}} (\Delta \log A_{ij})^2 + \frac{1}{2} \sum_{ij,kl \in \mathcal{II}, ij \neq kl} \frac{\mathrm{d}\Psi_{H,ij}}{\mathrm{d} \log A_{kl}} \Delta \log A_{ij} \Delta \log A_{kl}. \tag{5}$$

Here we see that, unlike in (2) above, there is now the potential for interactions—both positive and negative—across two different improved infrastructure segments "ij" and "kl" inside the set $\mathcal{II}$. The second-order terms (the second and third terms in 5) capture the intuition that productivity improvements generate relatively large welfare impacts whenever they induce relatively large increases in the flows $\Psi_{H,ij}$, aggregated across all of the segments in $\mathcal{II}$ that are being improved.[7]

In many ways this is just as important an implication of the welfare-maximization that happens in an undistorted economy as is the envelope theorem in (2). Again, whether a shock ΔA_{ij} is small or large it will cause a cascade of endogenous responses to potentially every allocation in

[7] A polar case in which second-order (and higher-order) effects are zero is when all elasticities of substitution in the economy are unitary, such that all producers' and consumer expenditure shares are constant and hence $\mathrm{d}\Psi_{H,ij} = 0$.

the economy. We know from (2) that the first-order effect of all such adjustments is zero. And equation (5) says that the second-order effect of all adjustments in the economy is given simply by how much the directly-affected flows adjust. This means that, for example, a study of the impact of improving one highway corridor (a set of segments in $\mathcal{II}$) should pay close attention to changes in transport flows $\mathrm{d}\Psi_{H,ij}$ along that corridor, but it can ignore all other endogenous responses in the economy. Even though improvements in one corridor will surely cause a change in flows along other corridors, those effects are irrelevant to second-order. And the same is true for all other types of changes in "flows" in this economy, such as changes in output or employment in factories nearby the highway segment in question. Equally, the only interactions that do occur in (5) are across segments within $\mathcal{II}$.

The fact that second-order effects are summarized by simple sufficient statistics (the change in directly affected flows, $d\Psi_{H,ij}$) is just as useful as the fact that first-order effects are (the level of directly affected flows, $\Psi_{H,ij}$). However, these two sets of sufficient statistics require different types of knowledge on the part of the analyst. Measuring $\Psi_{H,ij}$ is often a simple act of observation, and as discussed above such an observation is valid for both ex-ante and ex-post evaluations. The change $d\Psi_{H,ij}$ is a different story.

For example, when doing ex-ante evaluation, the response $d\Psi_{H,ij}$ is unobservable (since the intervention hasn't occurred yet) so it must instead be predicted by the analyst. Doing so is usually challenging, and not aided in any substantial way by the assumptions (of an undistorted, representative agent economy) that have been so helpful thus far.[8]

Turning to ex-post evaluation, this may seem more straightforward, since at least the analyst may have data on $\Psi_{H,ij}$ at two points in time, before and after the intervention, enabling an observation we could label $\Delta\Psi_{H,ij}$. But it is crucial to recall that the thought experiment used to derive (5) is one where the only change in the economy's exogenous characteristics is the set of infrastructure improvements. In the real world we can expect the data on the observed change $\Delta\Psi_{H,ij}$ to be driven by many other changes as well. So it is a non-trivial task to parse the observed change $\Delta\Psi_{H,ij}$ into that caused by the shocks of interest (such as ΔA_{ij}) and that caused by everything else. Empirical tools of causal inference are natural for this task, as discussed in Section 7.

[8]As discussed above, the envelope theorem is of great use under these assumptions but only when the estimand is the aggregate welfare change. When we seek to know other changes, such as $d\Psi_{H,ij}$, that theorem is of no use.

3.3 Summary and the Fogel template

Equation (4) suggests a stylized template for ex-ante and ex-post infrastructure evaluations—at least, those in settings where the goal is to measure the effect on the welfare of a representative agent, where the economy is thought to be undistorted, and where the changes of interest ΔA_{ij} are thought to be small enough that a second-order approximation is sufficiently accurate. This template works as follows, and its organization drives much of the remaining sections of this chapter:

1. Identify the set $\mathcal{II}$ of all segments ij over which the infrastructure improvement will have a direct productivity effect, $\Delta A_{ij} \neq 0$.

2. Measure the size of each of the productivity increases ΔA_{ij} caused by the infrastructure improvement. Approaches to this step are discussed in Section 5.

3. Measure the flow exposure values $\Psi_{H,ij}$ for each segment ij in $\mathcal{II}$. This is discussed in Section 6.

4. Simulate (if ex ante) or measure (if ex post, and if possible) $d\Psi_{H,ij}$, the causal impacts on $\Psi_{H,ij}$ of the set of infrastructure improvements in $\mathcal{II}$, and repeat for each segment ij in $\mathcal{II}$. Section 7 describes examples of this in existing work.

5. Use the outputs of Steps 1-4 to calculate $\Delta \log W$ in equation (5).

This procedure is similar to the procedure followed by Fogel (1964) in his seminal "social savings" evaluation of the US railroad system's late nineteenth century expansion (and whose same steps have been applied countless times in studies since). But innovations in data access, theory, computation, estimation, and causal inference mean that, in practice, the Fogel template can be substantially improved upon in modern work. In particular, Fogel's work used a procedure in which each of the above steps took the following form:

1. Fogel took the segments of interest $\mathcal{II}$ to relate to the flow of merchandise only. This omitted any impacts on the flow of services, people, and many forms of capital.

2. Fogel measured the effect of railroads on the productivity of exchanging commodities (i.e. ΔA_{ij}) by calculating the difference between the (quality-adjusted) freight rate for rail and pre-rail transport options. This assumed that there was perfect pass-through of transport productivity changes into transport prices. It also assumed that all of the observed change was due to the causal impact of the railroad system.

3. Fogel measured the share of merchandise trade in total income (i.e. $\Psi_{H,II}$) in the pre-rail era from the best data sources available to him.

4. Fogel measured the change in merchandise trade due to the railroad expansion (i.e. $d\Psi_{H,II}$) by simply calculating the difference between the best available estimate of $\Psi_{H,II}$ in the rail and pre-rail eras. This assumed that all such changes were entirely due to the expansion of the railroad network, rather than also partly the result of other changes in the US economy over the same time period.

5. Fogel combined the above steps to conclude that the impact of the railroad on US total income was estimated to be no more than a few percentage points of.[9]

4 Impacts of Infrastructure and Policy in Distorted Economies

Whether because of taxation, imperfect competition, or externalities, the presence of distortions is innate to the study of spatial economies (Starrett, 1978). They were also core to many of the critiques of Fogel's estimates (see his summary and responses in Fogel, 1979). It is therefore important to consider how the above template needs to be augmented in settings with the arbitrary distortions (i.e. $\mu_{ij} \neq 1$) that featured in Section 2.

4.1 First-order effect

We continue to investigate the effect of transportation infrastructure improvements (such as ΔA_{ij}) on aggregate welfare, and continue to allow all allocations and prices to adjust in the face of such improvements. But in addition we allow for changes in the distortions $\Delta\mu_{ij}$ themselves. This

[9] In practice, rather than use the second-order expansion formula in equation (5), Fogel argued that his estimates were an upper-bound on (and hence a conservative approach to) the exact welfare effect.

incorporates the study of exogenous policy changes of interest (such as a change in congestion taxes). It also allows us to incorporate effects in which a shock to infrastructure or policy induces changes in distortions elsewhere in the economy (such as markups).

Consider, then, a shock to both ΔA_{ij} and $\Delta\mu_{ij}$, starting at a position of arbitrary distortions throughout the economy. In such a case, following Baqaee and Farhi (2020), the resulting change in welfare can be approximated (to first-order) by

$$\Delta \log W \cong \widetilde{\Psi}_{H,ij} \mathrm{d} \log A_{ij} + \mathrm{d}\Pi^{\mathrm{fixed}\mu} \tag{6}$$

where

$$\mathrm{d}\Pi^{\mathrm{fixed}\mu} \equiv \sum_{ij} (\mu_{ij} - 1) \mathrm{d}(\lambda_j \Omega_{ij}). \tag{7}$$

If we now go further and allow for a whole set $\mathcal{II}$ of infrastructure improvements $\Delta \log A_{ij}$, as well as a potential change to the distortion on every exchange in the economy, equation (6) implies that the resulting change in welfare is

$$\Delta \log W \cong \sum_{ij \in \mathcal{II}} \widetilde{\Psi}_{H,ij} \Delta \log A_{ij} + \mathrm{d}\Pi^{\mathrm{fixed}\mu}. \tag{8}$$

Comparing this expression with that of the undistorted case in (2) highlights two alterations. First, the measures of exposure that matter now are the cost-based versions $\widetilde{\Psi}_{H,ij}$ rather than the revenue-based ones, $\Psi_{H,ij}$.[10] This reflects the fact that the increase in output enabled by a productivity improvement would be given by goods valued at their marginal cost of production rather than their distorted sales prices. And second, of course, the profit term $\mathrm{d}\Pi^{\mathrm{fixed}\mu}$ is now potentially non-zero whenever some $\mu_{ij} \neq 1$, as equation (7) makes clear.

The term $\mathrm{d}\Pi^{\mathrm{fixed}\mu}$ captures the change in economy-wide "profits" (revenue from distortions), expressed as a share of total income, that would occur purely as a result of the allocational changes $\mathrm{d}(\lambda_j \Omega_{ij})$. While actual profits are also affected by any change in the distortions themselves, the direct effect of distortion changes on profits is merely a transfer from consumers to producers (or other collectors of distortion revenue, such as the government), so this effect cancels out of

[10] In the undistorted economy case these two exposure terms align since $\widetilde{\Psi}_{H,ij} = \Psi_{H,ij}$ for any exchange ij.

equation (6) and all that remains is the effect due to $d\Pi^{\text{fixed}\mu}$.

The expression for $d\Pi^{\text{fixed}\mu}$ in equation (7) can be thought of as a correlation between initial wedges $(\mu_{ij} - 1)$ and changes in allocations $d(\lambda_j \Omega_{ij})$. In particular, if the exogenous shock (to infrastructure or policy or both) happens to result in relatively large flow changes along those exchanges that are relatively most distorted then the amount of misallocation in the economy will fall and this is good for welfare. Distorted activities are those that are initially too small from a welfare-maximizing standpoint.

The $d\Pi^{\text{fixed}\mu}$ effect arises because the distorted economy is no longer maximizing welfare, so no sort of aggregate envelope theorem can be applied. This means that all of the endogenous allocation changes that we would expect to occur as a result of an infrastructure improvement can now matter, even to first order, as they appear inside equation (7). This first-order role for reallocations is in stark contrast to what happens in any initially undistorted economy.

Three obvious but important points about the $d\Pi^{\text{fixed}\mu}$ effect are worth noting. First, it is entirely possible that $d\Pi^{\text{fixed}\mu} = 0$. This could happen if the flows along distorted exchanges are unaffected by the shock. It could also occur if the flow changes that do occur are orthogonal to the distortions. Hence, the case of equation (2) in the undistorted environment of Section 3 is slightly more general than it might at first seem, since it can also be true (modulo the use of $\widetilde{\Psi}_{H,ij}$ to value direct effects instead of $\Psi_{H,ij}$) in distorted economies.

Second, there can be no simple presumptions about the sign of the $d\Pi^{\text{fixed}\mu}$ effect. It might be tempting to think that when transport improvements occur in distorted environments they will have a virtuous "multiplier" effect—causing both positive direct effects and beneficial indirect effects (i.e. $d\Pi^{\text{fixed}\mu} > 0$). Alternatively, it might be tempting to think that market failures dampen the effects that might occur in an otherwise undistorted economy (i.e. $d\Pi^{\text{fixed}\mu} < 0$). These are certainly possibilities, but the general formulation in (7) reminds us that neither of these lines of thought can be generally true. What matters is the interaction between initial wedges and flow changes, and the mere fact that the economy is distorted says nothing about what that correlation might be.

Finally, there is nothing in the distorted economy case that rules out $\Delta W < 0$ even when the underlying shock involves pure technological progress (i.e. improvements in infrastructure). This is simply the case whenever the direct effect of the seemingly beneficial technological improvement happens to worsen the economy's misallocation (by moving quantities away from relatively distorted goods), and so much so that it overwhelms the direct effect of better technology. Another seemingly paradoxical possibility is that a strict reduction in distortions can actually reduce welfare, but this follows a similar logic—a standard application of the theory of the second-best (Lipsey and Lancaster, 1956).

As with the discussion in Section 3, one can view sufficient statistic results—like the role played by $d\Pi^{\text{fixed}\mu}$ in (8)—in two different ways. From the perspective of an ex-post evaluation, this result tells us what needs to be measured: we need to know the causal effect of the infrastructure and/or policy changes on (fixed-wedge) profits. Similarly, from an ex-ante evaluation perspective it tells us that the entire purpose of a counterfactual simulation (at least to first-order) is to predict what this change in (fixed-wedge) profits will be. Regardless of the vantage point, the logic of equations (8) and (7) is the same. We need to understand (a) which goods have the largest distortions on them and (b) whether those goods will expand or shrink as a result of the infrastructure or policy shock in question. Finally, equation (7) highlights how the constituent parts of $d\Pi^{\text{fixed}\mu}$ contain their own dimensionality-reduction: the endogenous flow changes on initially undistorted exchanges do not need to be measured or predicted.

4.2 Summary and an expanded Fogel template

The case of the distorted economy considered here requires some modifications to the template offered in Section 3.3. In particular, even the first-order approximation is altered by the presence of $\mu_{ij} \neq 1$, and the role of policy and/or distortion changes $\Delta\mu_{ij}$ now becomes relevant alongside those of pure productivity improvements in infrastructure provision (i.e. ΔA_{ij}). An augmented template, to incorporate initial distortions and changes in distortions, works as follows:

1. Identify the set $\mathcal{II}$ of all bilateral segments ij for which the infrastructure improvement will have a direct productivity effect, $\Delta A_{ij} \neq 0$.

2. Measure the size of each of the productivity increases ΔA_{ij} caused by the infrastructure improvements on segments ij in the set $\mathcal{II}$. Approaches to this step are discussed in Section 5.

3. Measure the cost-based flow exposure values $\widetilde{\Psi}_{H,ij}$, for each segment ij in $\mathcal{II}$. This is discussed in Section 6.

4. Simulate (if ex ante) or measure (if ex post, and if possible) $\mathrm{d}\Pi^{\text{fixed}\mu} = \sum_{ij}(\mu_{ij} - 1)\mathrm{d}(\lambda_j \Omega_{ij})$. This captures the causal impact of the productivity and policy shocks of interest on allocative efficiency, which is revealed by studying how pre-existing distortions ($\mu_{ij} \neq 1$) interact with changes in the allocation $\mathrm{d}(\lambda_j \Omega_{ij})$ caused by the shocks of interest. These changes in distorted allocations can occur both within the transport sector (as reviewed in Section 8) and in the wider economy (Section 9).

5. Use these ingredients to calculate $\Delta \log W$ in equation (8).

The remaining sections of this chapter then build on this analysis by asking two follow-up questions. First, in Section 10, what sorts of infrastructure improvements would yield the greatest $\Delta \log W$ (given some budget or construction cost constraints), and hence be the optimal way to improve a transportation network? And second, in Section 11, what are the challenges in actually constructing, delivering, and maintaining a given infrastructure investment ΔA_{ij}?

5 Measuring Impacts of Infrastructure Improvements on Transport Costs

The primary input into the analysis above—and into many other questions besides—is the change in the productivity of transport caused by an infrastructure improvement, denoted ΔA_{ij}, on any segment ij. As discussed above, such changes can equivalently be thought of as reductions in the (user) cost of transport.[11] Yet this presents the analyst with an immediate challenge. How can such transport costs be measured at all, given that they are unlikely to correspond to any simple price measure that is reported in conventional datasets? Going further, how can analysts attempt to learn the causal effect of a transportation infrastructure investment on these costs of transport? This section reviews answers to these two questions that appear in recent work. This endeavor comprises Steps 1 and 2 of either the Fogel template of Section 3.3 or the expanded Fogel template of Section 4.2.

[11] Recall, if markups exist then this corresponds to the case of a distorted economy, to which we return below.

5.1 Measuring transport costs

The discussion above focused on transport flows of three types: goods, people, and capital. Here, we cover only the first two of these distinct forms, starting with the cost of transporting goods; the study of the cost of spatial capital mobility is comparatively thin.

5.1.1 Costs of transporting goods

Methods for measuring transport costs (for goods and other flows) that have appeared in the literature take three forms. I refer to the first as "direct measurement"—cases in which direct records of the cost of transport (such as ticket prices or freight rates per ton-mile) are available. The second draws on data about the spatial differences in prices of the item that is flowing in order to make inferences about the cost of moving the item. And the third draws on data about the quantities flowing in order to make inferences about the cost of such flows. We discuss these methods in turn. A dedicated review of (international) trade cost measurement can be found in Anderson and Van Wincoop (2004).

Measuring transport costs directly. In some settings it may be reasonable for researchers to argue that direct proxies for transport costs are available. For example, Fogel (1964) argued that an important component of the user cost of transporting goods is simply the freight charges (per ton-mile, for example, and potentially with ton-mile rates varying across regions, transport companies, or types of goods) applicable to each mode of transport. These freight charges leave a direct paper trail—since they are the price at which shipping services are exchanged—and are therefore straightforward to measure, in principle at least.

Modern research has leveraged advances in GIS computing to implement this idea in granular detail. One approach is typified in Donaldson and Hornbeck (2016). Those authors use digitized maps of the US rail network at various snapshots in time Atack (2018) along with the estimated freight rates per ton-mile prepared by Fogel (1964), to calculate an estimate of what a given shipment along the rail network—from any origin node to any destination node—would have cost (per ton) in each year. However, since shipments could (and often needed to, especially early in the authors' sample period) use non-rail transport modes, such as water-based shipping, they overlaid additional mode-specific networks with mode-specific shipping costs per unit distance. Finally, they used Dijkstra's shortest-path algorithm to calculate efficiently the least-cost route (LCR) for every pair of locations and every time period.[12] The cost of this LCR then provides a spatially granular estimate of the transport costs of goods, per ton shipped, that is built up from direct measurements. Other papers applying this approach include Fajgelbaum and Redding (2022) in Argentina, and Jedwab and Storeygard (2022) in sub-Saharan Africa.

A powerful alternative to the network-based approach described above is the fast-marching method (FMM) whose application in spatial economics was pioneered by Allen and Arkolakis (2014). This method begins with a "raster" approximation to an underlying network model. Whereas a network consists of a list of nodes (points in space) and the set of links that offer direct connections (at stated costs) between them, a raster divides space up into a complete set of small grid cells and then specifies the cost of traversing over each cell. As long as the density of grid cells is high, this can well approximate many underlying networks of interest.[13] Armed with a raster representation, the FMM can efficiently calculate the entire surface of optimized routes from any grid cell to any other, which can, in turn, be used to quantify the cost of the LCR between any pair of locations. This is particularly attractive for researchers who begin with scans of historical transport maps, since it is far easier to convert such scans into appropriate raster inputs (for FMM) than network files (for a Dijkstra-like approach). This has strong synergies with advances in map digitization that have taken place simultaneously.

One challenge that both of these approaches face arises when the full cost of transport is a combination of cost components. For example, Fogel (1964) stressed how the user cost of transport involves not only the freight rate of a shipment, but also the opportunity cost of travel time, the cost of uncertainty about arrival times, and the risk of damage to goods in transit. As another example, Glaeser et al. (2023) document the roughness of roads in the United States (which not only slows travel speeds, but also results in vehicle damage) at high resolution by using smartphone vibrations recorded by Uber drivers.[14] All approaches to direct measurement of transport costs

[12] In practice, depending on the context, all-pairs shortest path algorithms (such as the Floyd-Marshall algorithm) can be more efficient than looping Dijkstra over many desired origin-destination pairs.

[13] One challenge is posed by settings in which the underlying network is not spatially smooth (for example in the case of air routes, which can only be accessed at origin and destination airports rather than at any point along the journey). However, in these cases one could use a hybrid of network- and raster-based approaches.

[14] Gertler et al. (2024) provide a detailed study of various economic impacts of road quality improvements and maintenance in Indonesia.

require researchers to monetize and combine these additional components, and do so separately by mode prior to feeding mode-specific costs into the network- and FMM-based tools described above.

Finally, it is important to note that the methods described here assume that all transport takes place via route segments (both along a given mode and across modes) that are viewed by users as perfect substitutes for one another. This stands in contrast to a long tradition in travel demand modeling (e.g. Oppenheim (1995)) in which modes of transportation are imperfectly substitutable (although segments of trips along a given mode still remain perfect substitutes) at the aggregate demand level. Quantifying imperfect substitutability is possible with access to data on transport flows, so I discuss this further below.

Measuring transport costs using spatial price gaps. A distinct approach to the measurement of transport costs draws on data documenting the prices of the goods being transported (rather than freight rates, the price of transportation services, discussed above). A standard argument implies that the difference in the price of a homogenous good at two points in space must be—as long as arbitrage is thought to be cheap and unimpeded by policy—no larger than the cost of transporting that good between those two points. And further, to the extent that trade is actually occurring between those two points in space, this lower bound on the transport cost is actually binding.

A long tradition uses this logic to quantify transport costs for plausibly homogeneous goods (to which the above logic applies). Fackler and Goodwin (2001) provide a methodological survey as well as discussions of numerous agricultural applications. However, a common obstacle that arises (especially in intra-national contexts) is that it the researcher does not know which pairs of locations are actually trading (and hence on which pairs the arbitrage *equality* can be used).[15]

[15] When this information is unknown, it is not clear how to map nonparametrically from price gaps to transport costs. Modern econometric tools designed for partially identified models, such as this one, may offer a solution, but not one that has been applied, to my knowledge. More parametric solutions are covered in Fackler and Goodwin (2001). A different type of solution is pursued in Eaton and Kortum (2002), who pointed out that if all tradable goods have the same transport costs and (more easily available) aggregate trade data is available (and indicates that at least one good is being traded within a location pair) then the maximum within-pair price gap (with the maximum taken across the different tradable goods for which price data is available) should equal the transport cost. Simonovska and Waugh (2014) further refined this approach by incorporating a model-based correction for the fact that, in any finite sample of product types, there is a positive probability that the maximum price gap across observed goods is not the maximum price gap across all goods.

Donaldson (2018) proposed to solve this problem by focusing on a narrow set of goods that were identified by their origin locations (in the same way that Bordeaux wines should only be made in Bordeaux). A similar idea appears in Atkin and Donaldson (2016), who observed that many modern sources of price data identify goods at a sufficiently precise level that a researcher can learn their potentially unique factory (or port of import) locations.

An obvious concern with the price gap approach to measuring transport costs is that the price of an identical good may vary across locations because of differences in markups as well as differences in costs of transport from its origin. In other words, free spatial arbitrage may fail. Atkin and Donaldson (2016) propose a solution to this problem that begins by noting that one source of spatial markup variation arises exactly because of transport costs: in the presence of imperfect competition, sellers may not pass through all of a cost (such as the transport consumer) to their consumer. Flipping this logic, knowledge of a seller's pass-through rate could be used to undo this transport cost bias. The authors go on to estimate such product- and destination-specific pass-through rates by estimating the extent to which origin price shocks (which reflect the marginal costs of production or importing) appear to cause shifts in destination prices. A typical finding is that pass-through is incomplete (i.e. below one) and hence that the perfect competition (free arbitrage) assumption may lead researchers to underestimate transport costs.

Measuring transport costs using transport flows. A third option—and one that has seen particularly widespread use—for measuring transport costs is to draw on transport flow data. The key assumption here is that the researcher has already estimated a demand system for transported goods—perhaps of the form discussed in Section 7—and hence knows how the demand for such goods depends on their prices. Inverting this system at any observed vector of quantities hence reveals the vector of (relative) prices that the demander faces in any given origin. By repeating this argument at the origin location the researcher can arrive at an estimate of the spatial price gap for any given traded good, and the above procedures on inferring transport costs from price gaps can be applied. Classic references include McCallum (1995) and Anderson and Van Wincoop (2003). A particularly rich example is provided by Coşar et al. (2015), who estimate a model of oligopolistic competition in Denmark and Germany's wind turbine industry. This simultaneously provides demand system estimates and estimates of markup differences that can be used to infer transport costs.

5.1.2 Costs of transporting people

Many of the same ideas relating to measuring the cost of transporting goods have been applied to the analogous setting of moving people.

Measuring transport costs directly. An explosion of recent computational advances has dramatically improved researchers' ability to build up direct measurements of important components of the costs of transporting people. A predominant version is to assume that the cost of transport is governed by the same known function of the time that it takes to travel, and then use a value of time estimate to convert travel time measurements into monetary units.[16]

The direct measurement of travel times can draw on numerous sources. For example, many modes of travel (such as rail and air) typically follow posted schedules, and these have been used by researchers such as Leunig (2006), Berry and Jia (2010), Heblich et al. (2020), and Fajgelbaum et al. (2023). Other creative approaches to measuring travel times include the use of travel surveys (as in, for example, Couture et al., 2018), and speeds inferred from services such as Google Maps (e.g. Akbar et al., 2023b).

These methods typically result in estimates of travel times that are specific to particular modes of transport. Just like the case of goods transport above, researchers often aggregate mode-specific costs into an overall cost index by assuming that travelers take the least-cost route (LCR). As before, the LCR between any pair of locations can be computed through either a network-based approach (as in Ahlfeldt et al., 2015 or Akbar et al., 2023b) or a raster-based approach (e.g. Allen and Donaldson, 2020).

Travel costs naturally include more than the time incurred during travel. For example, within each transport mode, one may wish to aggregate both the time cost of a mode-specific journey (given a value of time) and the financial cost of the journey (such as the ticket price).

Measuring transport costs using spatial price gaps. I am unaware of work that aims to infer the cost of transporting people by the explicit use of an arbitrage-based argument (analogous to those

[16] Buchholz et al. (2020) offer a particularly creative take on the issue of how to value time by studying the offers that ride-hail users accept and reject as these offers differ in terms of wait time and price.

employed above when estimating costs of moving goods). To be sure, spatial equilibrium models typically rest on the assumption of individual rationality – that travelers do what is best for them and thereby exploit the arbitrage opportunities around them. However, empirical attempts to invert this logic and infer travel costs from spatial price gaps run into the obvious challenge of measuring the relevant "price gaps" (which would clearly involve many components such as the prices of goods, earning opportunities, and amenities).

Measuring transport costs using transport flows. Similarly to the case of goods transport, a final tool for measuring transport costs of people involves the use of data on travel flows and an estimated travel demand system. At a high level, one can distinguish between two strands of this endeavor (though of course they overlap). Single-mode travel demand systems place emphasis on the decision of where a traveler will chose to travel, whereas multi-mode systems instead focus more on the decision of which mode of travel will be used to execute the journey. Within a given origin-destination pair these systems are differentiated only by the fact that the multi-mode approach will first use a cross-mode demand system to aggregate a vector of mode-specific costs into one scalar cost index for the given origin-destination pair.[17]

A simple form of single-mode travel demand system includes work such as Ahlfeldt et al. (2015) and Heblich et al. (2020), where data on commuting flows are used (in conjunction with an estimated location demand system) to estimate an unknown parameter that converts travel time to travel costs. In more complicated cases, the travel demand involves agents making dynamic decisions under uncertainty, for example as in Artuc et al. (2010), Kennan and Walker, 2011, Caliendo et al. (2019), and Porcher et al. (2024).

Multi-mode travel demand systems have a long tradition in the study of spatial economics (e.g. Oppenheim, 1995). In a recent example, Tsivanidis (2022) uses travel survey data to estimate the demand for travel (between any pair of locations) by mode within Bogota. Here, the within-mode travel cost for each location pair is measured using the "direct" approach (and Dijkstra's algorithm). And then the total cost index for each pair reflects the appropriate amount of imperfect substitutability (according to the estimated demand system) between the two modes. Other recent

[17]The case in which modes are assumed to be perfect substitutes (e.g. when the Dijkstra or FMM methods are applied across nodes) is therefore actually a particularly simple version of the multi-mode approach.

examples of this approach include Almagro et al. (2024) with four modes of travel among location pairs in Chicago, and Barwick et al. (2019) who use a mixed-logit travel demand system that models each mode as a bundle of attributes (price, travel time, and reliability).

An additional cost component that has received attention in recent work is the cost of users searching for available transport options. For example, this occurs in both the taxi markets of Frechette et al., 2019 and Buchholz, 2022, and the shipping ports of Brancaccio et al., 2024. A key contribution of such studies is to measure the amount of time (and other resources, such as brokerage fees) that users spend searching.

5.2 Measuring impacts of infrastructure on transport costs

The work reviewed in Section 5.1 concerned the level of transport costs for goods and people within any pair of locations and at any point in time. However, Step 2 of the Fogel (or expanded Fogel) template described above calls for a different sort of knowledge: the causal impact of an infrastructure improvement on such transport costs, which we have denoted ΔA_{ij}.

How can researchers measure ΔA_{ij}? In many cases they do so naturally by following the assumptions already invoked in measuring transport cost levels. For example, the direct measurement methods described in Section 5.1 typically measure mode- and link-specific costs, so the changes in total origin-destination transport costs caused by any change in the transport infrastructure network can be simulated by turning the relevant modes and links on or off. In a sense, the researcher has a complete model of how transport costs arise, so it is straightforward to evaluate that model at counterfactual points in the provision of infrastructure in order to learn about the causal impact of infrastructure investments.

When it comes to the flow of goods, an aggregate version of this method was used by Fogel (1964), and considerably more disaggregated versions appear in, for example, Allen and Arkolakis (2014), Donaldson and Hornbeck (2016), Fajgelbaum and Redding (2022), and Jedwab and Storeygard (2022). Similarly, Heblich et al. (2020), Tsivanidis (2022), and Fajgelbaum et al. (2023) use disaggregated network models for the cost of transporting people in order to model the causal effect of changing infrastructure provision on those costs.

A distinct method for estimating causal effects of infrastructure is to use program evaluation tools to estimate an (average) statistical relationship between measures of transport costs and measures of exposure to the infrastructure in question. Under standard exogeneity conditions this can provide more flexible estimates than the direct approach described above. A simple version of this appears in Atkin and Donaldson (2016). Having estimated transport costs for moving goods using a price gap-based method, those authors estimate the association between such transport costs and road provision and quality. A more involved version using price gaps appears in Donaldson (2018), who worked with a causal model of transport costs that included unknown parameters such as the extent to which (an equivalent unit of) rail transport was cheaper than road transport, which stands in contrast to the direct approach in which such transport model parameters are known. In this case, an assumption about exogeneity of infrastructure expansion identifies the unknown parameters, at which point the causal effect of infrastructure on transport costs can be simulated in a straightforward manner.

Analogously, Akbar et al. (2023b) and Akbar et al. (2023a) arrive at measures of the cost of moving people along highway segments directly from observations about the speed of highway travel (that is, under the assumption that travel times reflect passenger travel costs). They then go on to estimate correlates of city-level average speeds such as city-level major road density and network orientation. Crucially for our purposes here, this is done using travel times observed at off-peak hours (such as at 3am) so the resulting speed estimates may more closely reflect the raw productivity of average road links (rather than sum of such productivity effects and endogenous congestion effects, which is what would likely be observed at peak travel hours), to which we return in Section 8. A similar goal is achieved in Couture et al. (2018) by the authors' use of travel surveys (containing information on travel times) and an estimated model of city-level travel supply curves; the estimated intercept of such curves provides a measure of city-level travel productivity, and the authors go on to estimate how this measure correlates with city-level road characteristics.

6 Measuring Baseline Flows

As discussed in Section 2, three crucial forms of movement in spatial economics settings are intranational movements of (a) goods and services, (b) people, and (c) capital. The baseline—that is, pre-intervention—values of these flows comprise aspects of the Ω matrix introduced above, and they play a central role in any analysis of improvements in infrastructure or changes in spatial policies. In particular, the measurement of such flows contributes to Step 3 of the Fogel template (or its expanded form) discussed above.

In particular, equation (8) highlights how in the distorted environment of Section 4 the baseline level of $\widetilde{\Psi}_{H,ij}$ is critical for evaluating how a set of productivity changes $\Delta \log A_{ij}$ will affect welfare. Naturally, a special case of this is the undistorted environment of Section 3, where the absence of distortions implies that, for any flow, the cost-based exposure $\widetilde{\Psi}_{H,ij}$ collapses to the revenue-based exposure $\Psi_{H,ij}$. Of these two measures, the revenue-based version is considerably simpler to measure. Not only can it be computed without detailed knowledge of distortions (since, by assumption, those are known to be zero in this case). But in addition, as discussed above the revenue-based measure is simply the share of the value of the ij flow in total income (i.e. the Domar weight for that flow), implying that it can be computed with limited knowledge of the rest of the economy's entire flow matrix Ω. By contrast, computing the cost-based measure requires the ability to see flows (and the distortions on them) all the way up the supply chain from the ij flow to primary factors. As such, it requires considerably greater (though not necessarily complete) knowledge about Ω.[18]

Regardless of the precise nature of what is required, the logic behind equation (8), either in the undistorted case or in the distorted one, implies that even first-order effects of infrastructure improvements can't be computed without knowledge of the flows that determine $\Psi_{H,ij}$ and $\widetilde{\Psi}_{H,ij}$. Any path to answering this question leads through the researcher's ability to obtain data on such flows or otherwise make the assumptions necessary to learn them via some form of interpolation, extrapolation, or calibration.

The goal of this section is therefore to describe a range of strategies that researchers have deployed to measure the baseline values of these flows. (Section 7 below then builds on this discussion to cover the case of measuring the changes in such flows that were caused by the policy intervention of interest.) Of course it is beyond the scope of this chapter to catalog all measures of flows that are available to researchers. But an array of examples can convey a sense of the op-

[18] The bulk of the discussion in this section refers to spatial flows, as relates to the Domar weight. In the distorted economy case, however, the required inputs go beyond purely spatial flows, since they require knowledge of all wedges upstream of the segment of interest.

portunities and challenges that exist. In addition, elsewhere in this Handbook, Abramitzky et al. (2025) catalog the revolution in data availability (concerning both historical and contemporaneous economic activity) that has transformed the study of regional and urban economics in recent decades. Many of the novel data sources covered by that chapter are, of course, highly relevant to the discussion here as well.

In parallel with the treatment of transport costs in Section 5, we divide our coverage of data on flows into that concerning goods (including services) and people in the sections that follow.

6.1 Flows of goods

Flows of merchandise: The traditional way that researchers have measured the spatial flows of goods has placed particular emphasis on the flow of physical merchandise, rather than services (discussed below). This is largely due to data limitations. For example, for decades the US Census has conducted a periodic random survey (the Commodity Flow Survey, whose microdata were first analyzed by Hillberry and Hummels (2008)) on the movement of merchandise across space within the US, and the flows of international exports and imports of merchandise can be tracked to the state of origin/destination (as used, for example, by Adão et al., 2023). But no equivalent records of services flows exist.

Similar surveys and censuses of intranational merchandise flows appear to be rarely available around the world and earlier in time. Exceptions include, for example, the aggregated rail flow statistics assembled by Fogel (1964) for the late 19th Century US context, the disaggregated flow statistics for merchandise trade in an out of Chicago in at the turn of the 20th Century assembled by Allen and Donaldson, 2020, the historical rail statistics assembled by Crafts and Klein, 2014, the historical statistics on movement of merchandise (by rail, river and sea) throughout British India assembled by Donaldson (2018), the Canadian interprovincial merchandise trade statistics first used by McCallum (1995), and the unit-level micro data on rail shipments in the Ukraine used by Korovkin et al. (2024).

A recent wave of research has creatively deployed data on the movement (obtained from GPS trackers) of individual trucks that are known to be engaged in the movement of merchandise across space. This includes Harris and Nguyen (2025), Alder et al. (2025), Lall et al. (2022), Allen et al. (2024), and Barnwal et al. (2024). While a limitation of such data is that the contents of such trucks (i.e. what is being shipped on any journey, let alone the prices of such shipments) are often unknown, sources such as these nevertheless offer unparalleled detail.

Flows of services: As alluded to above, it has traditionally been a challenge to measure intra-national flows of services. One impressive attempt to fill this gap for the case of the United States is offered by Eckert (2019), which itself built on the pioneering ideas of Gervais and Jensen (2019). This approach uses regional data on sales and employment by sector, coupled with the assumptions of a gravity-based model for trade in services and merchandise, to infer the flows of goods by sector (for both services and merchandise sectors) for any pair of regions. The result is a coherent set of spatial flow measures that match (by design) available origin- and destination-specific aggregates.

Flows of both merchandise and services: In certain contexts, the challenges of obtaining data on flows of merchandise and services have recently been relaxed by three new sources of data that have become available to researchers.

The first such source is administrative (tax) microdata that is sometimes collected as part of countries' value-added tax (or, occasionally, goods and services tax) systems. Examples of countries that have shared such data with researchers include Belgium (Dhyne et al., 2021; Bernard et al., 2022; Amiti et al., 2024), Chile (Huneeus, 2018; Atkin et al., 2024; Burstein et al., 2024), China (Egger et al., 2024), Costa Rica (Alfaro-Urena et al., 2022), Dominican Republic (Cardoza et al., 2020), Ecuador (Adao et al., 2022), Hungary (Diem et al., 2024), certain states of India (Panigrahi, 2021; Khanna et al., 2022), Kenya (Chacha et al., 2022), Rwanda (Spray and Wolf, 2018), Turkey (Coşar et al., 2022; Atalay et al., 2023; Demir et al., 2024), Uganda (Spray, 2021), and Uruguay (Amodio et al., 2025). These databases typically record unique firm identifiers, which can often be merged with other records on firm location to provide information on spatial flows across space. They may also include other granular identifiers such as precise time-stamps, unique transaction identifiers, and even details of the goods (and their prices) being exchanged. However, one limitation that arises in many (but not all) of these countries is that intra-firm flows across establishments are not taxed and hence are not recorded. This may limit the accuracy of such records, particularly in sectors such as retail where large firms have many spatially-dispersed establishments.

A second novel source of data on goods trade comes from privately owned databases. For example, Atalay et al. (2011), Bernard et al. (2019), and Andersen et al. (2022) use databases maintained by credit bureaus that describe features of firm-to-firm trade. Similarly, Fox (2018) draws on an industry database that tracks firm-to-firm trade in US auto parts. These resources are typically subject to the same caveats about the distinction between firm-to-firm and establishment-to-establishment trade discussed above.

The data sources described above track trades that take place between producers, but they typically fail to capture the final step from firm to consumer. In some cases this can be overcome with access to separate data on each firm's total sales, provided that the sales to firms is complete (and includes sales to the government and to export markets), since consumer sales could be backed out as the residual and apportioned spatially to the best of the researcher's ability. One promising exception on the tax data side is currently available in a handful of countries (e.g. Gerard and Naritomi, 2021 and Atkin et al., 2024) in which firm-to-consumer tracking is attempted. However, some relatively new sources of private data offer a different window on firm-to-consumer sales. For example, databases on payment (i.e. credit or debit) card transactions (e.g. Agarwal et al., 2017, Allen et al., 2020, Barwick et al., 2023, and Dolfen et al., 2023) document matched transactions between individual cards and merchants, often with a flag that indicates whether the transaction was in-person or not. In some cases (e.g. Relihan, 2022, Allen et al., 2020, Diamond and Moretti, 2021, and Andersen et al., 2022) researchers have also been able to match payment card records to their owners' wider bank records, allowing an even wider view of purchases (as well as additional details such as earnings, assets and liabilities). Finally, in a small number of cases, researchers such as Townsend (2016), Burke et al. (2019), and Walker et al. (2024) have carried out their own ambitious surveys to measure flows of goods throughout local economies in some low-income country settings.

From origin-destination flows to link-specific flows: The majority of data sources discussed so far describe end-to-end trade flows from the origin of a shipment to its destination. However, in the natural case in which only particular transport segments are being improved, what is needed by the analyst is segment-level flow data, not origin-destination-level flow data. In an important paper, Allen and Arkolakis (2022) offer an elegant solution to the problem of how to convert origin-destination flows into link-level flows (or vice versa) without solving for the complicated origin-destination least-cost path and allocating the flow to that path. This technique applies equally well to the other flows, such as that of people considered next.

6.2 Flows of people

A range of sources have been used by researchers aiming to track the flow of people across locations within countries.

Travel diaries: A core source of information about human mobility has traditionally been found in a type of survey known as a travel diary. The idea here is to survey individuals (and/or businesses) about the trips they have taken recently. A great example of the use to which such surveys can be put is Couture et al. (2018), and a recent initiative that has homogenized such surveys across a wide range of countries (including low-income ones) is due to Deffebach et al. (2025). One limitation of such surveys is that they are inevitably sparse records of the universe of trips. This is not a concern for statistical analysis that is built on a relatively small number of moments of the survey data, but it can preclude (without further adjustment) the type of high-dimensional analysis that has become common in calibrations of many-location models.

Commuting records: A second type of information on travel has been derived from information about commuting behavior. Here, the typical information available to the researcher documents origins and destinations of commutes. For example, the analysis in Tyndall (2021) and Severen (2023) draws on administrative records (the LODES database) about workplace and residence locations for American workers. Similarly, Monte et al. (2018) draw on questions in the American Community Survey that ask workers about their commuting patterns. Coupled with an assumption about commuting behavior—that is, about the number of commuting days per year, for example—one can arrive at an impression of the total volume of commuting flows between each pair of locations. The same idea underpins historical statistics about commuting behavior as used in Heblich et al. (2020) (primarily about early nineteenth century London, but further supplemented by historical data on Berlin, Paris, Boston, Chicago, New York, and Philadelphia).

Migration records: Just as with the case of commuting, data on migration can be used to infer flows of people over space. This is the strategy used in, for example, studies of the United States from Kennan and Walker (2011) (via the NLSY survey), Yagan (2019) (via IRS administrative data), Abramitzky et al. (2014) (via historical full-count Census data), and Allen et al. (2018) (via consular records of Mexican immigrants by their location of residence in the US). Beyond such government-based sources, researchers have recently turned to rich micro-data obtained from private data providers. For example, Diamond et al. (2019) use data (from the company Infutor) about the near-complete address histories of US residents going back to the 1980s, and Ding et al. (2016) and Baum-Snow et al. (2019) draw on the location histories tracked by Equifax, a credit bureau.

All this being said, despite the attention paid to the study of migration decisions, and the importance of this process for many questions in the field, in the framework of Section 3 the notion of human flows that matters is the share of total income that is "spent" on travel. Presumably such expenditure is relatively low for the case of the relatively infrequent flows that facilitate migration, at least when compared to the case of the far more frequent journeys that people make for commuting and leisure purposes.

Mobility Finally, a set of new and highly promising data sources on broad notions of human mobility have become increasingly available to researchers over the past decade. At the most ambitious end of the spectrum, records from smartphone users can be used to provide immensely detailed traces of their phone's locations. Such databases have been used, for example, by Atkin et al. (2022) to quantify face-to-face interactions in the San Francisco Bay area, by Athey et al. (2021) and Couture et al. (2025) to measure the similarity of locations visited by different types of demographic groups, and by Miyauchi et al. (2021) to quantify multi-stop trip behavior. In a similar vein, cellular phone call data records (the so-called "pings" that such phones make to a nearby tower when used for calls and text messages) can provide a lower spatial and temporal resolution version of such data. These have been used by, for example, Blumenstock (2012), Kirchberger (2021), and Kreindler and Miyauchi (2023).

Beyond mobile phone records, a range of other techniques have been used by researchers to provide high-resolution spatial impressions of human mobility. One promising vein has derived from payment card records, which document not only the flow of goods and services (as discussed above) but also the locations of their human card owners (when the purchase is one made in person). For example, Allen et al. (2020) use such records to great effect to understand the movements of both tourists and locals in Barcelona, and Barwick et al. (2024) use them to measure the inter-city movements (under the assumption that a card gets used at least once by their owner on any given inter-city trip) of all payment card owners throughout China.

Alongside smartphone records, the data obtained from users of particular smartphone apps has proven useful for understanding numerous aspects of human mobility. For example, Goldszmidt et al. (2020) and Gonzalez-Navarro et al. (2021) use internal company data from ride-hailing apps. These efforts, in turn, build on prior work about flows inferred from taxi usage by, for example, Mangrum and Molnar (2017) and Frechette et al. (2019). Likewise, the user-level trip records from the navigation app company Waze have been used by Hausman et al. (2023). In the case of one pioneering study Kreindler (2024) a researcher even designed his own tracking app to be used by participants of a congestion-pricing randomized controlled trial in India.

Finally, internal records of some cities' public transport systems have been used by researchers to quantify travelers' behavior. This offers the greatest potential when the system uses payment cards that are integrated across modes and also require users to register both the start and end of their journeys, as is partially the case in London (Larcom et al., 2017) and Jakarta (Kreindler et al., 2023).

6.3 Flows of Capital

To the best of my knowledge, there do not exist records of capital or financial mobility that are as rich as those that researchers have used to study the movement of goods and people. One partial window on such flows has become available in Brazil, for example, where researchers (for example, Bustos et al., 2020) see banking records that link branch location-level savings to branch location-level loans within the same banking company. In reference to a different notion of capital, numerous countries link the individual owners of companies (which can often be matched to records of owners' residences) to individual companies (which can then be matched to records of firms locations). These so-called ownership matrices have been used by **?** in Ecuador, Egger et al. (2024) in China, Atkin et al. (2024) in Chile, and Walker et al. (2024) in Kenya.

7 Measuring Transport Flow Responses to Infrastructure Improvements

In this section we turn to the question of how transport flows (in goods, people, and capital) appear to respond to a productivity improvement in the transport sector (such as that due to infrastructure improvements). The primary motivation for this exercise comes from Step 4 of the Fogel template from Section 3: the second-order approximation to the welfare change of an infrastructure improvement in undistorted settings hinges on the extent to which transport flows on directly-affected links grow. Additional motivation follows in the case of distorted economies. As discussed in Section 4, if transport activities are themselves distorted then any growth in transport flows in such sectors caused by infrastructure improvements, or policy changes, has additional welfare effects beyond what one would see in an undistorted economy. We take up the analysis of such cases in Section 8 below.

7.1 Goods Flow Responses

Section 6 above discussed efforts to measure the pre-existing levels of all flows (goods, people, and capital) in an economy at baseline, prior to the productivity shock(s) of interest. Our interest, now, is in the causal effect of such shocks on these flows, beginning in this subsection with the category of goods flows. Recall, equation 5 suggests that researchers who seek to go beyond the first-order impacts of productivity shocks (as dictated by Hulten's theorem) focus on the term:

$$\frac{1}{2}\sum_{ij,kl\in\mathcal{I}\mathcal{I}}\frac{\mathrm{d}\Psi_{H,ij}}{\mathrm{d}\log A_{kl}}\Delta\log A_{ij}\Delta\log A_{kl}, \tag{9}$$

which describes the correction needed (relative to Hulten's theorem) to reflect the second-order aggregate welfare consequences of the set of improvements in $\mathcal{I}\mathcal{I}$. This expression makes clear that the set of causal responses that need to be known are those confined to $\mathcal{I}\mathcal{I}$.

The literature has offered a range of approaches to the challenge of knowing this set of causal effects. At a high level, a distinction first needs to be drawn between those settings where the researcher seeks to understand impacts of infrastructure investments ex-ante (that is, prior to those investments having been made, including when they may be mere proposals or a range of options) and those where the researcher is conducting an ex-post evaluation (that is, one conducted after the investments have been implemented). We begin with a discussion of ex-post approaches.

Ex-post estimation A classic example of ex-post estimation in this context is due to Fogel (1964). Writing about the late nineteenth century US railroad expansion almost seventy years later, Fogel sought to know the causal responses in equation (9). These responses answer the question: in a year of interest (such as 1890) what would be the change in trade flows along segments such as ij that would result if the 1890 railroad system were removed?[19] Fogel's answer to this question involved two steps. First, he assumed that (within any affected set of flow segments) the productivity change due to railroads was the same (i.e. $\Delta A_{ij} = \Delta A_{kl}$ for all ij and kl in the set $\mathcal{II}$). This collapsed the object of interest from the high-dimensional set in equation (9) to the simpler matter of knowing the change in total flows across all transport segments affected by the rail expansion. Then, second, he assumed that the actual change in the flow of goods from the pre-rail era to 1890 was the causal effect of the construction of the rail system (and not the result of any other change in the US economy over the time period).

It is particularly easy to critique the second of these assumptions from the perspective of modern approaches to casual inference. In particular, it seems unlikely that all other shocks in the US economy happen to just cancel out such that the pure time-series change in trade flows reveals the desired ceteris paribus impact. Were this change being compared to a "control group" (one thought to receive similar non-rail shocks as the US economy did) it would be more compelling, since the assumption that shocks do not occur at all is surely weaker than the assumption that they occur in a similar manner in two locations.

With this view of causal inference in mind, one approach to ex-post analysis in modern studies seeks to identify treatment-control comparisons that offer a plausible impression of the desired causal effect. A leading example of the causal inference approach to learning about trade flow responses to infrastructure improvements can be found in the work of Duranton et al. (2014). While the authors are interested in such effects for their own sake, rather than as an explicit input to equation (9), their approach is highly relevant to the discussion here.

The Duranton et al. (2014) study aims to estimate the average effect of highways on merchan-

[19] In practice, Fogel (1964) was primarily interested in a modified version of this counterfactual question in which the railroad system was not just removed but also replaced by a non-rail system (such as one involving counterfactual canal expansions) of the same cost.

dise trade within the US. The core idea is to compare US cities with relatively favorable highway penetration (in terms of both intra-city density and inter-city connectivity to other nearby cities of large economic size) to those within the US that have less favorable highway infrastructure. Such comparisons offer the promise of learning the average treatment effect of relatively strong infrastructure under two assumptions. First, that cities with strong and weak highway density would have identical trade flows (on average) in a hypothetical scenario that they did actually have the same highway density. This unconfoundedness or independence assumption regarding treatment (i.e. highway) assignment may sound a priori implausible, since infrastructure is unlikely to have been as good as randomly assigned to US cities. However, Duranton et al. (2014) marshal a battery of evidence (and instrumental variables approaches) that do much to assuage such concerns.

A second strong assumption is that there are no treatment spillovers from other cities' highway access onto the trade of the relatively low-access cities. Such violations of the so-called Stable Unit Treatment Value Assumption (SUTVA—see, e.g., Imbens and Rubin, 2015) are required by any approach that draws inferences about the average (and hence, up tot the number of observations, also the total) impact of a treatment from the relative impacts that can be seen by comparing treated units to untreated ones. In many ways, the assumption that there are no SUTVA violations in a cross-sectional, within-country approach is the direct opposite of Fogel's assumption that the entire national time-series change is due to the treatments of interest. The former assumes no spillovers but is robust to the presence of other shocks that hit all regions equally, whereas the latter assumes there are no other shocks but is robust to the presence of all forms of (domestic) spillovers.

An active area of recent research in spatial (and other areas of) economics concerns the development of methods for striking a balance between these two extremes. At some level, all such approaches involve the use of economic theory to structure the nature of the spillovers that can occur. In the context of spatial spillovers that work via trade—a case that is not only one of natural interest in the field, but in the context of this subsection the object of direct enquiry—an extremely common approach is to leverage a parametric model of the derived demand for trade in goods across regions.

For example, Donaldson (2018) aimed to understand the impact of colonial Indian railroads on the trade of merchandise in that region and used a gravity model (in particular, one that was based on a multi-sector Eaton and Kortum (2002) model of Ricardian trade). Rather than estimating an overall (and appropriately weighted) average treatment effect of (some measure of) a region's rail access on the region's (inbound and outbound) trade flows, as equation (9) would suggest, this study estimated sector-specific values of the gravity model's so-called "trade elasticity." This structural parameter governs the stable (and common, across all origins and destinations) elasticity with which any region's bilateral imports (for the sector in question) from any two origins respond to the relative cost of importing from those two origins (again, for the sector). In the gravity model, as long as all regions are non-autarkic, any trade cost reduction (such as that due to new railroad connections) would induce changes in all regions' trade flows, so there would be no scope for a model-free way of avoiding the SUTVA problem. On the other hand, it is natural to be skeptical of any claim that the gravity model is the uniquely valid way, despite its numerous strong assumptions, to correct SUTVA violations.

This discussion has argued that there is unlikely to be any purely theory-free approach to ex-post estimation (at least, in the common scenario in which there are no "pure-control", or purely autarkic, spatial units). However, this does not rule out the use of testing or validation procedures that can at least probe the internal consistency of the theory-based approaches that do get used. Adão et al. (2025) develop this idea. They propose a test statistic for the validation of structural models that is focused on the specific counterfactual question that the model is being used to answer. For example, in the case of a researcher using a gravity model (with no distortions) to conduct an ex-post analysis of the welfare effects of transport infrastructure, the purpose of the model is (at least to a second-order approximation) effectively to populate the weighted average set of causal responses in equation (5). This suggests that the model's predictions about such a (weighted) average of all model-implied treatment effects should be compared to the (weighted) average of relative treatment effects what we see in the data. Put simply, if the role of the model is to populate the "missing intercept" of a regression line in a cross-sectional study, there is no way to check that the model got the intercept right since it is not identified. However, the researcher can nevertheless at least confirm that the model-implied slope across regions (i.e. the line that the model effectively uses to populate the unobserved intercept) aligns with the slope of relative effects that *is* observable in the data.

One interesting question about the use of structural models even in ex-post settings (which are used purely, as argued, because of the need to overcome the SUTVA problem) is how a researcher should explicitly tailor their modeling assumptions to the question at hand. While it is difficult to imagine an all-purpose answer to this question, equation (5) nevertheless invites a researcher to recognize that their structural model's misspsecifications will only matter—at least, to the extent that they maintain the high-level assumption that the economic environment is undistorted, and that a second-order approximation is sufficiently accurate for the purposes at hand—if they mistakenly populate the weighted average treatment effect in equation (5). That is, misspecification in other aspects of the model that do not enter equation (5) need not concern this researcher. One intriguing point here is that even though the question of interest concerns how infrastructure affects trade, equation (5) points out that the researcher only needs to correctly model the way that trade flows on directly affected units change. This observation could potentially be used by the researcher to focus modeling and parameter estimation attention on the trade flows that actually matter, rather than all such flows. I am unaware of explicit discussions of this point in the literature.

Ex-ante estimation In many ways, the ex-ante estimation setting is actually no different from that of many ex-post settings. As described above, if SUTVA violations are thought to be problematic, there is no available unbiased approach to counterfactual inference even in the ex-post case where the infrastructure improvement of interest has already taken place. The result is the need to use modeling assumptions to fill in counterfactual scenarios even in the ex-post case, and this is not inherently any different from the challenges faced by the ex-ante researcher who seeks to predict the effects of an infrastructure program that hasn't even been built yet.

It is therefore no surprise to observe that the solutions chosen by researchers when faced with ex-ante predication problems are similar to those described above for ex-post problems. Ex-ante predictions require a quantitative model that is calibrated in a manner thought to be suitable for the context at hand. Given such a model, the analyst can simulate any given exogenous shock of interest while holding all other exogenous features constant and thereby speak to the causal impacts that are desired. This may appear no different from the ex-post approach that also uses a quantitative model to evaluate a program that has already occurred.

One potential distinction between ex-post and ex-ante structural model evaluations concerns the way that model parameters are estimated. In principle, ex-post settings allow the analyst to estimate their model's parameters (such as the trade elasticity) from the infrastructure variation that has already taken place. This seems likely to enhance the internal validity of the ex-post evaluation relative to an approach to parameter estimation or calibration based on other settings. By contrast, in the ex-ante case there is no choice but to learn the model's parameters from other settings and assume that they are valid for the context of interest.

7.2 Person Flow Responses

Just as with the study of goods flows, a wide array of work has sought to quantify the impact on human mobility of various sorts of infrastructure improvements.

Ex-post estimation: We can again begin with the context of ex-post estimation: how the flow of people across space appears to change as a result of some set of transport infrastructure investments of interest that have already taken place. Another great example comes from the work of Duranton and Turner (2011), who estimate how highway availability within US cities leads to the promotion of travel within such cities. They do so using city-year regressions, in which SUTVA requires that one city's highway access does not affect the amount of travel taking place within any other city. An analogous regression appears in the remarkable analysis of Gonzalez-Navarro and Turner (2018), which estimates the effect of subway infrastructure expansion on public transport ridership in a sample of cities that spans the vast majority of subway systems globally. Turning to inter-city flows of people, Barwick et al. (2023), for example, estimate the effect of Chinese high-speed rail system access on the total propensity for a city's residents to be found on any given day outside of their "home" city (i.e. their most frequent location in recent months).

What is distinctive about the above examples is that they explicitly focus on people flows that are directly measured. From the perspective of equation (9) this is an important focus, since the beyond-Hulten gains from infrastructure improvement come exactly from creating more flow on the segments being improved. But as discussed in Section 6, a complementary approach to the direct study of people flows is to examine changes in the sorts of behavior—such as commuting and migration—that, by their very nature, correspond to spatial flow. In this vein, an important study is due to Baum-Snow (2010), which estimates how the amount and nature of commuting changed within US cities as a result of the interstate highway system. This is complemented by the analysis of, for example, Baum-Snow (2007) and Baum-Snow et al. (2017) concerning the impact of transport infrastructure on the centralization of employment and residential density in US and Chinese cities—since changes in city structure have clear implications for commuting patterns, and hence for human travel.

On the migration side, one can get a sense for the range of remarkable cases that have been considered in the literature via examples such as Brooks and Donovan (2020) (who estimate the effect of randomized bridge access on short-term migration in Nicaragua), Morten and Oliveira (2024) (who estimate the effect of inter-city highway access on permanent migration in Brazil), and Milsom (2023) (who estimates the effect of road access on migration in several African countries, with a particular emphasis on movement towards locations that offer higher intergenerational mobility).

Analogously to the discussion of goods flows in Section 7.1, all of the analysis mentioned so far relies on the assumption that there are no SUTVA-violating spatial spillovers across locations in the analysis. Even though this may be a plausible assumption in many contexts, in parallel with the goods flow case a complementary stream of ex-post evaluation analysis has sought to use quantitative models to predict the causal effects of a ceteris paribus infrastructure shock (even one that has already happened) on spatial flows of people throughout an economic system, and hence arrive at estimates that are not subject to SUTVA concerns. A wonderful example of this approach is due to Heblich et al. (2020), which developed a rich model of spatial labor demand (i.e. firm location) and spatial labor supply (i.e. worker location and commuting choices) and estimated it in the context of London's dramatic expansion of intra-city passenger-rail infrastructure in the early twentieth century. Another pioneering example can be found in the Tsivanidis (2022) study of Bogota's rapid bus transit system.

An interesting point of discussion in this sort of analysis concerns the way that the model handles other forms of economic adjustment. For example, Heblich et al. (2020) explicitly focus on cases of their model with and without endogenous housing supply. To the extent that any changes in housing supply are resourced by factors (labor, land and capital) within a closed spatial economy such as the city under study, these adjustments do not enter equation (5) directly in an undistorted economy. Of course, the fact that housing supply can adjust will change the spatial flows of passengers in important ways—to take an extreme example, building commuter railways to unbuilt suburbs is unlikely to result in many commuters—but the resulting passenger flow changes remain the sufficient statistic for the overall welfare impact of an infrastructure improvement, regardless of *why* those flows are small or large.

As discussed above, this focus on commuting behavior is a natural and clearly important means of quantifying spatial flows of people, but it may omit impacts on other motives for flow such as leisure-oriented travel. For this reason, some studies of infrastructure impacts on human mobility have sought to build models of demand for all trips. An elegant example is due to Couture et al. (2018), where the analysis estimates a model of trip demand (based on the type of travel diary survey described in Section 6) and then asks how cities differ in the abilities of their road networks at supplying high-speed trips (for all purposes).

Ex-ante estimation: Finally, as above we can consider the case of purely ex-ante attempts to evaluate infrastructure projects that have yet to come into existence. An example of such analysis that has recently advanced the frontier in many ways is due to Fajgelbaum et al. (2023). These authors seek to understand the political economy of high-speed rail provision in California, a theme to which we return in Section 11 below. However, as part of their analysis of voters' preferences over alternative system proposals, the authors build a model that predicts the causal impact of any given proposal. The model emphasizes person flows—as is natural, given that high-speed rail is primarily used for passenger travel—and contains motives for both short-run and long-run trips, for both leisure and business purposes. The travelers being modeled can use multiple modes of transport (e.g car and air, in addition to rail) and the high-speed rail networks being evaluated include not only actual construction proposals but a range of alternatives.

8 Impacts in the presence of distorted transport

We turn now to examples of studies that consider cases where the transport sector itself is distorted. As discussed in Section 4, in these cases there is an additional (beyond Hulten) first-order benefit of transport infrastructure improvements on economic welfare if (and only if) the improvement causes more transport in these places where it is distorted. All of the examples discussed here highlight a distortion in the transport sector, though they differ in the extent to which they explore the impact of infrastructure shocks on the allocative inefficiency caused by the distortion.

We begin by restating equations (8) and (7) as:

$$\Delta \log W \cong \sum_{ij \in \mathcal{I}\mathcal{I}} \widetilde{\Psi}_{H,ij} \Delta \log A_{ij} + \sum_{ij} (\mu_{ij} - 1) \mathrm{d}(\lambda_j \Omega_{ij}) \tag{10}$$

which describes the total welfare consequences of a shock that changes productivities in the transport sector (due to an infrastructure improvement, for example) as well as any changes in the policy environment (or other distortions). Our interest in this section is the second term in (10), which captures the additional gains or losses due to the presence of distortions relative to the undistorted environment considered earlier. Further, our interest (for now) is on the case in which the only distortions in the economy are on transport activities themselves. This means that all of the quantity changes in equation (10) are irrelevant (because their wedge-determined weights $\mu_{ij} - 1$ are zero) apart from those quantity or flow adjustments that occur along segments of the transportation system on which distortions are present.

A first step in analysis motivated by equation (10) is to construct estimates of the distortion wedges that prevail across different segments of the transport system. We organize our discussion below by the type of such distortion that underpins these wedges. This is natural, given the way that the literature tends to approach its analysis one wedge at a time. However, a caveat is in order. Equation (10) suggests that it is the joint presence of all wedges in the transportation sector that will matter. For example, one could imagine an event that causes transportation flows to move towards segments that are heavily distorted from the perspective of one type of distortion (e.g. congestion), put in parallel also cause a movement away from activities that are distorted from the perspective of a different source of distortion (e.g. environmental pollution). The adjustment would look beneficial from the perspective of the first distortion and harmful from the perspective of the second one, yet in practice it may actually be neutral overall.

Armed with estimates of distortion wedges, equation (10) highlights that the object of interest should be on the causal response indicated by $\sum_{ij} (\mu_{ij} - 1) \mathrm{d}(\lambda_j \Omega_{ij})$. That is, the effect of the shock(s) of interest on the travel flows that are weighted by the size of their pre-shock distortion wedges. Crucially, it is neither the size of the wedges nor the size of the change in flows that matters, but the extent to which the shock causes a correlation between the initial wedges and the change in flows. While explicitly documenting that correlation has been an uncommon strategy in existing work, I believe that doing so would do much to improve the transparency of studies (in distorted environments) that seek to measure the welfare effects of infrastructure and policy shocks. As equation (10) makes clear, this correlation is precisely at the heart of the matter.

Transport congestion: We begin with the case of a core source of distortion in transportation economics: congestion in the transport system. As is well known, whenever the use of transportation infrastructure is un-priced, users effectively pay the average cost of production rather than the marginal cost of production as would be the case at the efficient volume (see, for example, the textbook presentation in Small et al., 2024). Put another way, the purchasers of a good in a competitive industry are being induced to internalize the effects of their demand on costs because the price they pay reflects the marginal cost of production; but on a (no-toll) public road, the "price" of a user's journey (which, say, reflects their travel time) is driven by the total number of users on the road (via the speed with which traffic can move on the road) rather than the marginal reduction in speed caused by the marginal user.

Naturally, it is challenging to quantify the wedge that exists between the effective price and marginal cost of using each transport segment in a transport system, let alone separately at various points in time (or various points along the cost curve). Nevertheless, substantial recent progress has been made towards this goal by the work of Couture et al. (2018), Akbar et al. (2023b), Akbar et al. (2023a). In particular, in the latter case these authors use data from Google Maps to extract estimates of travel speeds and volumes for many road segments within many cities around the world, and to do so at points in time when flows are both relatively low (such as in the middle of the night) and relatively high (such as at rush-hour). This allows estimation of a cost curve for any city's roads (i.e. the relationship between user travel times and total number of travelers) and hence a comparison of the wedge between marginal cost and price (i.e. average cost) for any city at any point in time. While in principle this sort of analysis could be done separately by road segment, the natural limits of data availability and statistical power necessitate a focus on city averages.

Related work has used natural experiments to estimate the impact of travel demand on travel speeds, which can be thought of as using a demand-side instrumental variable to estimate a cost function. Examples of such a strategy include Anderson (2014), Anderson and Davis (2020), Mangrum and Molnar (2017), and Kreindler (2024).

Such estimates can then be used as inputs into equation (10), towards the goal of estimating the causal impact of a transportation infrastructure improvement as discussed above. They can also be used to evaluate the effect of policy changes on aggregate efficiency in the presence of congestion wedges—by assessing whether, the covariance between initial wedges and flow changes (caused by the policy change) is positive or negative. Indeed, often the policy changes of interest are exactly those that aim to address the consequences of allocative inefficiency due to congestion. For example, recent work such as Hanna et al. (2017), Hall (2018), Barwick et al. (2024), Hall (2021), Cook and Li (2024), Hahn et al. (2023), Herzog (2024), Kreindler (2024), and Cook et al. (2025) can be thought of in this light. Relatedly, Kashner and Ross (2025) quantify the total impact of local traffic density on house values—an effect that includes consequences of traffic congestion, but also many of the other externalities discussed in this section.

Transport policies: The case of transport policies also fits into the logic of equation (10), albeit with some care in terminology. First, suppose the researcher's interest is in evaluating the aggregate welfare effects of a transportation infrastructure shock, but this researcher knows full well that the environment in question is one in which some transport flows are subject to transport policies—for example, that some roads may incur tolls or lane restrictions, or that access to entire zones of roads may incur congestion charges. This case is no different from that discussed above in the context of distortions caused by congestion externalities: there may be pre-existing wedges on certain flows and the infrastructure evaluation needs to use equation (10) to take this fact into account. However, this example does highlight that it is the total (or net) wedge on each flow segment that matters in equation (10). For example, in the ideal case in which a toll has been deliberately put in place so as to counteract the distortion caused by congestion, there is in fact no net distortion at all, so the infrastructure evaluation need not pay attention (at least to first order) to changes in flows along such zero-distortion links.

A second sense in which we care about transport policies is when their changes are under evaluation too. Consider the case where the researcher aims to evaluate the effects of a change in the toll on a given road. Then equation (10) states that the first-order welfare effect is given by the causal effect of that toll change on the flows along all roads that have tolls (or other distortions) on them. Naturally, the optimal level of the policy under evaluation (holding other policies fixed) is reached when this change being observed has no first-order welfare effect at all.

Beginning with the case of infrastructure evaluation in the presence of pre-existing transport policies, the key question is how large is the transport policy-induced distortion along each transport flow. In the case of pecuniary tax- and subsidy-like distortions this may be relatively straightforward to measure, as a matter of "mere" legislation. For example, in the context of public transit a distortion (in terms of aggregate efficiency) arises when user prices (inclusive of fare evasion) fall below the marginal cost of operating the system. So estimates concerning the impact of infrastructure- and policy-based events on transport flows on public transport systems—such as Gendron-Carrier et al. (2022) as discussed above, Hall et al. (2018), and Kreindler et al. (2023)—can be used to evaluate components of changes in overall efficiency. To take another broad example, in the context of private vehicle transport, vehicle ownership and fuel purchase taxes can be incorporated.

Turning to optimal policy, the treatment here will be brief because this theme is the focus of the a separate contribution to this volume (Fajgelbaum and Gaubert, 2025). But two impressive examples of frontier-expanding work can be found in Almagro et al. (2024) and Hierons (2024). These works make new theoretical contributions on characterizing features of optimal urban transport policies (such as road pricing and public transit fares in the former, congestion pricing in the latter) for the correction of distortions such as congestion and environmental externalities, as well as the lack of first-best methods for raising city government revenue. This builds on important prior work such as that by Small et al. (2005).

Pollution caused by transport: Another major form of externality associated with transport is environmental pollution (including of course carbon emissions). This theme is again the purview of a separate chapter in this volume (Balboni and Shapiro, 2025), as well as the review in (Parry et al., 2007), so our coverage here will again be limited. As before, the challenge is to measure the distortion wedge associated with pollution. In the unregulated settings that often prevail, especially in the transport sector, the marginal cost of polluting to the polluter may be so low as to be well approximated as zero. By contrast, the marginal cost of polluting to the rest of society (that is, the negative "price" of the pollution that is effectively paid by atmospheric "consumers" of pollution when it gets produced) is often found to be substantial.

Naturally, in the absence of markets for pollution these social costs need to be estimated by direct measurement of all damages to the non-polluting party. As challenging as such work can be, a robust body of estimates of pollution damages caused by transportation is now available, including studies such as Barth and Boriboonsomsin (2008), Currie and Walker (2011), Knittel et al. (2016), Anderson (2020), and Simeonova et al. (2021). Moving beyond the estimation of damages (i.e. the wedges in equation 10) we require estimates of the extent to which infrastructure investments cause transport use to change, and in particular to change in a pro- or anti-pollution direction. For example, Gendron-Carrier et al. (2022) take up this cause in their study of how global subway system expansions cause changes in urban air pollution (i.e. particulate matter), and Davis (2008) documents the limited extent to which driving restrictions in Mexico City actually lead to reductions in vehicle pollution.

Safety: Another clear externality associated with transport is the risk of damage to bystanders (other travelers, pedestrians, etc.). To the extent that those who cause such accidents are not held fully responsible for compensating victims, this risk introduces another wedge between the private and social cost of transport. Examples of recent empirical work that can be thought of as estimating this wedge (to varying degrees) includes Cohen and Einav (2003), Ashenfelter and Greenstone (2004), Jacobsen (2013), Abouk and Adams (2013), DeAngelo and Hansen (2014), Anderson and Auffhammer (2014), and Anderson and Davis (2023).

Imperfect competition: Beyond externalities due to missing markets, the presence of market power is another important source of distortion in the transportation sector. Naturally, a great deal has been written about this sector in the industrial organization literature, but relatively little recent attention has been devoted to the matter of interest here: how transportation infrastructure and policy changes will affect allocative efficiency in the presence of pre-existing distortions (in this case due to market power) in the transport sector.

One classic example of work on this theme can be found in Holmes and Schmitz Jr (2001).

These authors point out that the pre-rail water shipping industry in the United States is likely to have exercised substantial market power. This means that an additional benefit of the adoption of railroads, beyond the pure productivity effects (considered by Fogel, as above), is a reduction in the misallocation (due to markups of shipping prices over shipping marginal costs) in the transport industry. Again, the first-order benefits here can be calculated as simply the initial water shipping markup times the change in the amount of shipping produced in that sector (or any other, such as rail shipping, that is thought to be a perfect substitute for it). At its simplest, the point here is that infrastructure improvements may act in some settings as entry of a mode of transport into the transport industry, so in settings where entry is thought to have pro-competitive effects there is an allocative efficiency improvement. Qin et al. (2024) offer a rich empirical exploration of this idea in the context of China's high-speed rail system (which offered a new competitor for the air travel industry).

Search: Recent work has also drawn attention to another distortion in the transport sector: that many segments of it appear to operate via decentralized search, in which there is no market. As a result, any congestion effects in the search process—which are natural, since one party's search is likely to be successful partly at the expense of some other searching party's failure—are externalities. A classic example comes from New York's taxi sector, where it is common for drivers and riders to search for one another on the city's streets. Recent work by Frechette et al. (2019) and Buchholz (2022) has sought to estimate the search process in this setting in a highly nonparametric manner, resulting in better estimates of the distortion wedge due to (this type of) search than have been previously available. Relatedly, Brancaccio et al. (2023) have developed a novel characterization of the optimal policy response to this sort of externality.

Corruption: A final example of market failure that may be intertwined with the transport sector, in some settings, is corruption. Olken and Barron (2009) make this case elegantly in a unique audit study of the (substantial) bribes that truckers in Indonesia pay (alongside other delays, threats and hassles that they endure) as they complete long-distance shipments throughout that country. While a benign view of corruption is that it entails pure transfers across agents but no loss in efficiency, the rent-seeking that presumably accompanies such activities is likely to entail directly unproductive activities.

9 Impacts in the presence of wider distortions

We turn now to cases in which there is good reason to expect that the "wider economy" (i.e. beyond the transport sector itself) contains allocations that are distorted, and that infrastructure improvements seem likely to affect that allocation (and hence lead to first-order welfare effects). This is nothing more than a continuation of the theme discussed in Section 8 in the context of distortions in the transport sector, but now applied to the rest of the economy.

One qualitative distinction here, however, is that it may be considerably harder for researchers to evaluate whether infrastructure improvements or policy changes cause changes in the allocation of output throughout the wider economy. For example, it seems clear that infrastructure improvements will have a positive effect on the amount of transport, so the only question remaining for evaluating equation (10) in the context of Section 8 was about the size (and sign) of any distortion wedges of interest in the transport sector. In the current section, however, even when the sizes of certain wedges may be clear, the effect of infrastructure improvements on the production of the goods that are relatively distorted may be difficult to know a priori. Nevertheless, the clear importance of this endeavor has generated a substantial body of work to discuss.

Urban agglomeration and congestion externalities: A natural distortion with which to start, given the urban context of much work on transportation, involves the very essence of urban life: agents experience positive and negative externalities, in both production and consumption, due to the density of interactions with other city residents. In many ways, this is the bread and butter of structural model-based approaches to the evaluation (whether ex-ante or ex-post) of infrastructure and policy changes. Commonly used versions of these models have agglomeration externalities in them by design, so (at least to first-order, and beyond the Hulten effect) the entire point of such models is to answer the question asked by equation (10): will the shock of interest move density to low-agglomeration wedge locations or high-wedge ones? For example, the ex-post evaluation of the US interstate highway system performed by Allen and Arkolakis (2014), the ex-post analysis of the US railroad system by Nagy (2023), the ex-ante evaluation of potential widening of segments of that system performed by Allen and Arkolakis (2022), and the ex-post transit investments evaluated in Tsivanidis (2022), will all hinge on the answer to this question. A similar point applies to models with homophillic externalities across types of agents (as in, for example, Weiwu, 2024).

The same broad idea of production externalities (or other production-related market failures) that differ by locations applies across sectors whose footprints may differ across locations. For example, some traditional theories of development emphasize a lower marginal product of labor in traditional sectors (such as agriculture) than in modern sectors (such as manufacturing). For this reason, the impact of transport infrastructure on the sectoral composition of production—as evaluated in Berger (2019), Lindgren et al. (2021), Frye (2024), Herzog et al. (2024), Kaboski et al. (2024), and Adamopoulos (2025), for example—may have first-order welfare consequences.

Innovation and firm-level productivity: Closely related to the production-side agglomeration externalities discussed above concerns the innovation sector itself. One clear externality is that of knowledge flows. The mere existence of these effects suggests that innovation activities are likely to be under-provided by the market (to the extent that innovation subsidies have not fully corrected this market failure). In addition, many models of innovation have the feature that it is under-provided by the market even without knowledge spillovers because the monopolist inventor of even a patented innovation may not appropriate all of the surplus of their creations. For this reason, studies such as Sokoloff (1988), Perlman (2016), Agarwal et al. (2017), Andersson et al. (2023), and Ravalli (2024), which estimate the effects of infrastructure investments on innovative activity (such as patenting and R&D expenses), are important steps towards the goal suggested by equation (10).

In addition to affecting incentives for innovation, transport may also affect the actual flow knowledge about existing innovations. To the extent these are uncompensated, such knowledge flows would amount to externalities. Agarwal et al. (2017) and Pauly and Stipanicic (2023) offer careful investigations of such effects in the context of road and air travel, respectively.

Recent work has also documented spatial variation in the marginal productivity of individual inventors across space—and in particular that such heterogeneity is correlated with local inventor density. See Moretti (2021). This implies that effects, such as those documented by Berger and

Prawitz (2024), on how transport infrastructure can change inventors' location choices may have important (positive or negative) allocative efficiency effects, as suggested by equation (10).

Another take on these themes is provided by studies that examine individual firms' productivity levels. For example, the emphasis of Abeberese and Chen (2022) is to measure the effect of road improvements on productivity measured at the firm level. Similarly, Chen et al. (2022) and Americo (2023) examine changes in technology adoption. In many ways, it would be surprising to find that such effects contribute to allocative efficiency (beyond any resulting spillovers onto other firms, which we have already considered above). One might expect that profit-maximizing firms are perfectly capable of equating the marginal revenue gains of productivity-enhancing investments to the marginal cost of such investments, in which case there would be no distortion wedge, as a result of such investments, in equation (10). On the other hand, the essence of the "X-inefficiency" view of productivity enhancements is that firms may not be able to produce in profit-maximizing ways—for example, due to internal agency problems—and hence changes in a firm's environment could actually raise its profits (beyond mechanical effects achieved even when allocations within the firm are constant).

Environmental externalities: As discussed in Section 8, another natural type of market failure arises in the context of environmental damages. While the focus above was on the direct creation of such damages by transport activities themselves, this may only be a small piece of the picture. There are many reasons to expect that the dominant impact of transport on environmental externalities arises through the impact that it has on the location and nature of wider economic activities.

Existing work has offered a range of examples through which this can happen (though again, for a full treatment see the Balboni and Shapiro, 2025 contribution to this volume). One natural case of interest occurs in the context of deforestation, particularly in settings such as rain forest ecosystems where the marginal social value of the forest may vastly exceed the actual private value earned by deforesters from deforested land (plus timber). Work such as Gollin and Wolfersberger (2024) and Araujo et al. (2023) document such effects, for example.

A different type of environmental mechanism occurs with the environmentally-biased location of some infrastructure investments themselves. For example, Balboni (2025) focuses on the case of Vietnam's recent road infrastructure plans, which, she argues, appear to be naively targeted towards coastal regions in which global warming-induced coastal flooding is expected to occur. At face value, this is not a traditional externality *per se*, since it is a government decision. But there are many reasons to suspect that private investments will be made in the wake of this public investment, and those private investments may insufficiently internalize the potential damages from environmental harm (especially in settings where the government cannot commit to refraining from providing public aid in the wake of any damages, a source of ex-ante moral hazard).

Finally, Barwick et al. (2023) propose what appears to be a new form of environment-transport interaction in the literature. These authors' focus is on the way that China's high-speed rail system may potentially allow urban residents to use travel as a means for avoiding exposure to air pollution. They use location data revealed from payment card usage to measure the realized pollution exposure of each card owner on any given date, given the intersection of daily pollution data and daily locations at which cards are used in person. Given the external costs of air pollution, a positive impact of infrastructure investments on pollution exposure amounts to a positive allocative efficiency term in equation (10).

Imperfect competition: Just as discussed above in the transport context, a large body of work documents the presence of market power—in both output and input markets—in many sectors, not just the transport sector. This has obvious implications for the evaluation of infrastructure projects and transport policies. On the theory side, these phenomena have seen intense focuse, particularly in the monopolistic competition framework; see, for example, Baldwin et al. (2011) and Laird and Venables (2017).

On the empirical side, existing work on this topic has focused on two types of phenomena. The first can be thought of as an investigation of whether transport improvements cause economic production to be reallocated towards relatively high markup activities. This is directly connected to the discussion of equation (10) above. A fantastic empirical case study is explored by Hornbeck and Rotemberg (2024) in their study of US railroad expansion from 1870-1890. While these authors estimate a large shift in relative manufacturing production across space, the allocative efficiency effects of such reallocations appear to be relatively small given the limited correlation between the shifts involved and the initial markup levels across activities and locations. However, when the shock of interest causes an expansion in a country's total factor "endowments" (by encouraging net immigration, for example) the per-capita efficiency benefits hinge on the overall average markup even if there were no markup dispersion at all. As Hornbeck and Rotemberg (2024) document, this effect contributes to very large benefits to the United States from the expansion of population caused by the railroad system. That said, if immigrants' origin countries had average markups that were similar to those of the US then global allocative efficiency would not change as a result of this channel.

A related take on this mechanism can be found in Atkin and Donaldson (2016). This study provides evidence from retail price pass-through estimates (obtained separately by product and location) that road-remote locations in Ethiopia and Nigeria appear to suffer from greater deadweight loss than less remote ones. This suggests that if a road system improvement were to be targeted towards reducing the costs of accessing remote locations it would result in an allocative efficiency improvement.

A second direction of enquiry has been the impact of transport infrastructure investments on the extent of competition itself. Aghion and Schankerman (1999) provide a classic treatment, and this was recently investigated to great effect in the context of India's "golden quadrilateral" road investment scheme by Asturias et al. (2018). Relatedly, Chen et al. (2023) examine how road infrastructure in China promoted beneficial firm selection on productivity. Finally, Brooks et al. (2021) estimate the joint effects of Indian road expansions on firms' market power in both output and input markets. While neither markup nor markdown changes caused by infrastructure improvements appear directly in equation (10), we can expect in many settings that a pro-competitive shock that shrinks market power among the firms that have it will result in a greater price reduction among such firms and hence a beneficial reallocation of output towards the firms with initially large markups—that is, a positive correlation in (10).

Taxation and regulation: As Glaeser (2020) emphasizes, the interactions between infrastructure expansions and pre-existing policies in the wider economy are numerous and important. One key example is his observation that infrastructure expansions in areas with binding land use regulations may see very little expansion in either transport flows or the types of distorted activity (land use) that hinder efficiency. Another case in point from recent work is in the developing country context, where a source of enormous heterogeneity in effective tax rates across businesses arises because many businesses operate informally, or outside the purview of the formal tax (and regulation) system. While this heterogeneity need not have obvious spatial consequences, Zárate (2022) documents the striking extent to which business formalization appears to grow in Mexico City in the vicinity of new subway stations. Since formal activity has (by definition) a larger tax on it than informal activity does, this is a clear example of a positive correlation that is good for allocative efficiency in equation (10).

Other types of distortions: Recent research has documented empirical connections between transport infrastructure and policy and a wide array of additional sources of economic distortion beyond those discussed above. For example, Allen and Atkin (2022) study imperfect credit and insurance markets. Bergquist et al. (2024) and Vitali (2023) provide detailed studies of buyer and seller search behavior in the product market, which may feature similar search congestion externalities as discussed in Section 8. Dell (2015), Acemoglu et al. (2015), Couttenier et al. (2024), Cao and Chen (2022), and Khanna et al. (2023) focus on local conflict and crime. Pilegaard and Fosgerau (2008), Abebe et al. (2021), and Mulalic et al. (2019) consider the case of inefficiencies in spatial labor search and matching, respectively. Leduc and Wilson (2013), Ramey (2021), and Chang and Lee (2024) examine the possibility that business cycle downturns may entail inefficiently low employment, implying that infrastructure construction expenditures themselves may raise aggregate efficiency in such time periods. And finally, Ahmed et al. (2024) examine child labor (and school attendance), which may potentially reflect inefficient decisions about human capital acquisition.

10 Optimal transportation infrastructure design

Having seen examples and mechanisms through which transport infrastructure projects can raise economic welfare by both relatively small and large amounts, natural questions arise. What is the best way to start investing in infrastructure expansions from any given point? What would the network look like at the optimum at which such improvements would stop? What synergies might we expect between different types of investments? A recent literature has revived interest in these questions. This discussion will be relatively brief, given the additional coverage of these themes in Fajgelbaum and Gaubert (2025) chapter in this Handbook.

Optimal investments: Recent years have seen rapid progress in the understanding of optimal improvements in complex transport networks. A key contribution to this literature is provided by Fajgelbaum and Schaal (2020). These authors set up a many-region, many-sector economy in which consumers and firms exchange goods with arbitrary preferences and technologies under competitive conditions (and any market failures have been corrected with the use of optimal Pigouvian policies). Trade takes place over a rich transport network in which all nodes can be connected by links, and links have production functions that transport goods across space but with diminishing returns to scale (i.e. there is physical congestion on each link). Finally, the infrastructure planner can make a continuous amount of investment in improving the efficiency of any link, but with diminishing returns to investment (for any given level of transport flow) on any link. Against this backdrop, the planner's problem is how to best allocate a fixed budget of total investment resources in making improvements to the network towards the goal of raising national welfare.

This is a dauntingly complex problem. Even a very small subset of it—such as the question of how a planner might want to send a given unit of a given good from a given origin where there is a given total amount of production to a given destination at which there is a given total amount of demand—is the province of a large operations research literature. Yet the Fajgelbaum and Schaal (2020) problem optimizes over all such routings, as well as over the meta-question of how many units of each good should be produced and consumed in each location, and finally over the meta-meta-question of where best to improve the network subject to the investment budget constraint. The key result is to derive conditions under which the planner's problem is a convex optimization problem, which then opens the door to the use of numerical algorithms that are well suited to find guaranteed optima to such problems. One key additional step is to observe that the equilibrium prices (of both goods and infrastructure investment capital) at each point in space also characterize the allocation (of both goods and infrastructure investment capital) and the optimization problem therefore has a lower-dimensional dual form in the space (of prices) rather than the higher-dimensional primal form (of quantities, particularly including of goods flows on each link). This contribution has been applied in a range of settings, including Santamaria (2020), Graff (2024), and Gorton and Ianchovichina (2025).

A simplification used in the Fajgelbaum and Schaal (2020) problem is that all infrastructure investments considered are able to be made in continuous amounts. This may be a reasonable approach in many settings where, at least at the level of aggregation already imposed on the problem by the available data, the question about infrastructure boils down to an abstract notion of widening link capacity rather than building entirely new links. However, recent work has also made progress on the genuinely discrete network case by exploring the behavior of heuristic optimization procedures in this context. A leading example is Alder (2016) (see also Balboni (2025)), which starts from a complete network of links and iteratively deletes those links that offer the lowest marginal losses when they are deleted. Even though this algorithm is not guaranteed to arrive at the optimum (to an as-of-yet infeasible problem), it of course nevertheless arrives at a conservative estimate of the gains that infrastructure improvements can offer in any given setting.

Finally, a different type of infrastructure problem that has been assessed in recent work is that of laying out an optimal public transport network. Kreindler et al. (2023) tackle this case in the context of Jakarta's rapid bus system. This problem involves features such as the total number of available buses, the frequency with which they are run on various routes, and the placement (and hence intersection) of the routes themselves. Inevitably, for any reasonably sized city it is completely hopeless to simple enumerate every option. Kreindler et al. (2023) therefore instead propose to draw uniformly—using a range of techniques drawn from the Monte Carlo Markov Chain literature—from the set of all possible networks in order to characterize the probability distribution of networks (and their properties). This is not only useful for understanding likely features of the optimum, but it also makes it relatively straightforward to characterize how features of the optimal network are likely to change as parameters inside the problem (such as riders' desire for bus frequency relative to cost) change.

Complementarities with other spatial investments and policies: Whether via the second-order approximation (equation 5) in undistorted settings, or the first-order approximation (8) in distorted ones, infrastructure investments may look quite different in the presence of simultaneous interventions in other productivity improvements or policies. One leading example of such interactions that is commonly voiced in developing country settings is that a “big push” of coordinated public investments may yield greater social returns than equivalent investments would on their own. In this vein, Moneke (2020), Abbasi et al. (2023), and Vanden Eynde and Wren-Lewis (2024) each document evidence for positive complementarities between road and electricity infrastructure in Ethiopia, throughout much of Africa, and in India, respectively.[20]

Here, the theoretical case for complementarities is easy to make—at least to second-order. That is, we know from equation (2) that to first order there are no interactions (in undistorted economies) to be expected across different types of productivity enhancements, but as per equation (5) there are indeed such interactions in the second-order approximation. As discussed in Section 7, these interactions are simple: if the electrification intervention means that the economy can experience a greater trade flow response to the road intervention, then the gains from the road intervention will be larger to second order.

11 Infrastructure provision

The discussion of transport infrastructure has so far imagined a completely sanitized world in which the infrastructure investment of interest achieves a productivity gain of ΔA_{ij}, and does so without cost. Further, any policy change of interest is simply given by the value $\Delta \mu_{ij}$. This view may be fine for ex-post evaluations that seek to measure what happened (i.e. ΔA_{ij} and $\Delta \mu_{ij}$) and then quantify its impacts. But for the case of ex-ante analysis, especially ex-ante optimal analysis, real-world challenges intervene. What political processes affect the policies and investments that actually happen? And how costly can we expect various interventions to be to enact? This section discusses the burgeoning work that begins to take on these important questions.

The political economy of infrastructure provision: Infrastructure investments are inherently place-based decisions. It is therefore natural that politicians may use the allocation of these investments to curry favor with their often place-based constituents. Many such interactions are explored theoretically by Glaeser and Ponzetto (2018). And recent empirical work has documented the strength of this relationship in a wide range of contexts, including post-independence Kenya

[20] A similar complementarity is found by Gebresilasse (2023), who studies interactions between roads and agricultural extension services (e.g. free advice for farmers) in Ethiopia.

(Burgess et al., 2015), Nazi Germany (Voigtländer and Voth, 2014), nineteenth century Italy (Bonfatti et al., 2021), and colonial Africa (Bonfatti et al., 2019).

A new approach to the study of the political economy of infrastructure placement comes from Fajgelbaum et al. (2023)'s pathbreaking study of California's high-speed rail system. This paper achieves two important goals. The first is to assess whether voters' preferences for a project like this, as inferred from their votes on the referendum that approved it, appear to be aligned with a quantitative model of the projected impacts. Reassuringly, there does appear to be substantial alignment. A second goal is then to assess what politicians were effectively maximizing when the developed the specific network proposal that went to referendum. Such an "inverse optimum" approach has parallels in fields such as public finance (Jacobs et al., 2017) and international trade (Goldberg and Maggi, 1999; **?**), where the policy being assessed is typically a set of continuous variables (such as the income tax schedule, or tariff policy). Here, by contrast, the set of policies is perhaps better thought of as the discrete set of feasible networks. Conducting an inverse optimum exercise in such a setting requires new tools based on moment inequality estimates. Essentially, such a procedure asks what parameter values (i.e. weights on different objectives in the political objective function) would rationalize the fact that the actual proposed network is better (at those weights) than all alternatives. Moments that encode such an inequality can then be constructed by simulating the impacts of a range of alternative proposals, particularly those that explore regions of the network space that are particularly powerful in the sense that they would only be optimal under quite different weight values.

A distinct notion of political economy that has featured in recent work concerns interactions across spatial units (typically within the same country). Transport infrastructure provision in one region will clearly affect nearby regions—both through product and factor market interactions and through the simple fact that infrastructure is typically a non-excludable good, so there is nothing a region can do to stop the residents from other regions from using its own infrastructure (for example, on a journey between two other regions). Felbermayr and Tarasov (2022) explores this theme within the context of infrastructure that promotes (intra- and international trade) and Loumeau (2023) and Bordeu (2023) do so for the context of commuting-based human mobility across regions of the same labor market. These inter-jurisdictional interactions raise natural concerns about a resulting incentive for regions to under-invest in infrastructure from the perspective of the economy as a whole.

The costs of infrastructure provision: The cost side of infrastructure provision has received less attention than the benefits side, but thanks to important recent work this imbalance is starting to change. One key step has involved careful measurement of actual quality-adjusted infrastructure provision and its trends over time. For example, building on Small and Winston (1988) and Brooks and Liscow (2023), Mehrotra et al. (2024) advance a method for estimating the cost function of the US interstate highway system—that is, the discounted present value of public expenditures required to deliver a given quantity of flow (e.g. vehicle-miles) in every time period—from publicly available data. This notion of costs appropriately includes not only operating expenses (such as maintenance) and depreciation, but also the opportunity cost of the capital and land invested in the system.[21] One puzzle that emerges from this analysis is the dramatic increase (at least a doubling from 1994 to 2008) in the price of building new lane miles in the US, which does not appear to be the result of factors related to the location of new miles in increasingly challenging locations (such as cities or environmentally sensitive domains). An intriguing possibility is that the political economy forces described above place cost pressure on the routes that do get built, not just the redistribution of who gets the benefits from where they materialize. Another fascinating study of construction and maintenance costs, but across 99 low- and middle-income countries, is provided by Collier et al. (2016). These authors document the striking cross-project variance (both within and between countries) in such costs, as well as that costs tend to be higher in high-conflict and high-corruption settings.

Moving beyond measurement, an area of natural interest is the actual process by which infrastructure projects get built. The details of highway construction and procurement have attracted significant recent attention from economists with interests in auction design (Somaini, 2020), market power estimation (Lamadon et al., 2022), and corruption enforcement (Olken, 2007). But to the best of my knowledge the lessons from these approaches are not typically integrated into the study of the costs and benefits of infrastructure projects.

[21] Likewise, recent work by Guerra et al. (2025) has quantified the total amount of land value that is occupied by urban roads in the United States.

12 Directions for Future Work

The primary goal of this chapter has been to survey existing work in transport infrastructure and policy evaluation, but to do so through the lens of a particular framework. Inevitably this means that a natural direction for future work is to deepen the connections that have been highlighted here. For example, even in the basic case of the first-order (Hulten) effect in undistorted environments, the discussion here has highlighted that the goal of ex-post evaluation is essentially to arrive at an unbiased estimate of the weighted average treatment effect of the program on transport productivity (averaged across all improved segments), with weights given by each segment's Domar weight in the economy. I am unaware of work that is presented as an attempt to arrive at such an estimand via the application of modern program evaluation techniques.

Moving beyond the framework offered here, one clear and important omission is dynamics (though see the chapter in this Handbook by Desmet and Parro, 2025 for more on this theme). This seems especially relevant in the context of evidence for persistent (or even path-dependent) impacts of many features of spatial economies, including the impact of transport infrastructure (Bleakley and Lin, 2012). There is nothing inherently static about the general economy modeled in Section 2, since the space of goods and factors can be broadly interpreted as differentiated by date. However, virtually all of the applications referred to are essentially static. And the application of sufficient statistic approaches (such as those of Hulten, 1978, based on Domar weights) in dynamic environments will require the user to take a stand on any relevant Domar weights that are still in the future at the time of data collection.[22] Equally, many inherently dynamic settings invoke new sources of potential distortion such as imperfect intellectual property rights, credit constraints, and dynastic discounting.

Another major omission from this chapter has been the study of inequality. Whether this is across types of people within locations, across people with heterogenous ties across locations, or even to the intergenerational notion of equality of opportunity, the unequal consequences of public policies and investments are of obvious concern. As discussed above, the sufficient statistics for undistorted economies that we have exploited in this chapter are not applicable to other outcomes of interest, such as inequality. On the other hand, the notion of a distortion used here has

[22]Important recent progress on this issue has been made by Coşar et al. (2024).

been broad enough to embrace situations in which (for lack of other policy instruments) the economy arrives at a point where relative marginal utilities of income across people are not equal to some desired notion of social marginal utilities. The resulting gap between any such measures of marginal utility can be treated as a "wedge" as above—as in, for example, Adão et al., 2023.

Finally, and in reflection of the literature, the work covered in this chapter has been primarily focused on the spatial economics of relatively developed economies. Lower income environments bring in a range of new phenomena, not only their distinct transport sectors, but also in the wider distortions that are likely to be important (Atkin and Donaldson, 2022). The chapter in this Handbook by Bryan et al. (2025) offers numerous connections, as do Bryan et al. (2020) and Gonzalez-Navarro et al. (2023), for those seeking to push forward our understanding of the economics of transportation around the world.

13 Concluding Remarks

The ability to accurately forecast the impact of transport infrastructure investments and changes in transport policies remains a core objective of spatial economics. The goal of this chapter has been to offer a survey of the tools that spatial economists bring to the table when approaching this challenge. A key message is that modern data and causal inference tools are poised to dramatically expand the applicability of the theoretical insights that Fogel (1964) and others brought to their seminal studies of major infrastructure programs.

Fogel's template was elegant and profound. In essence, it amounted to three steps. First, quantify the productivity improvements (caused by the infrastructure investment of interest) occurring on each segment of the transportation network. Second, quantify the share of total income in the economy spent (prior to the improvement) on transport flows along each improved segment and multiply these by the productivity gains from Step 1. And third, quantify the changes in all flows along improved segments that were the result of the improvements made in the system as a whole. The literature surveyed here has advanced economists' abilities to execute all three of these steps in undistorted (i.e. allocatively efficient) economies. Going further, researchers and practitioners are increasingly aware of the wide range of contexts in which one can expect transport improvements and pre-existing economic distortions—both within the transport sector and beyond—to interact in important ways. The augmented Fogel template described in this chapter, and its partial implementation in the literature so far, highlights how in this way, too, Fogel's insights can be updated to the modern era.

References

Abbasi, M., M. S. M. Lebrand, A. Bluette Mongoue, R. Pongou, and F. Zhang (2023): "Roads, Electricity, and Jobs: Evidence of Infrastructure Complementarity in Sub-Saharan," Available at SSRN 4516253.

Abebe, G., A. S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2021): "Anonymity or distance? Job search and labour market exclusion in a growing African city," *The Review of Economic Studies*, 88, 1279–1310.

Abeberese, A. B. and M. Chen (2022): "Intranational trade costs, product scope and productivity: Evidence from India's Golden Quadrilateral project," *Journal of Development Economics*, 156, 102791.

Abouk, R. and S. Adams (2013): "Texting bans and fatal accidents on roadways: do they work? Or do drivers just react to announcements of bans?" *American Economic Journal: Applied Economics*, 5, 179–199.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014): "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 122, 467–506.

Abramitzky, R., L. P. Boustan, and A. Storeygard (2025): "New Data and Insights in Regional and Urban Economics," National Bureau of Economic Research Working Paper Number 33561.

Acemoglu, D., C. García-Jimeno, and J. A. Robinson (2015): "State capacity and economic development: A network approach," *American Economic Review*, 105, 2364–2409.

Adamopoulos, T. (2025): "Spatial integration and agricultural productivity: Quantifying the impact of new roads," *American Economic Journal: Macroeconomics*, 17, 343–378.

Adao, R., P. Carillo, A. Costinot, D. Donaldson, and D. Pomeranz (2022): "Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence," *The Quarterly Journal of Economics*, 137, 1553–1614.

Adão, R., A. Costinot, and D. Donaldson (2025): "Putting Quantitative Models to the Test: An Application to the US-China Trade War," *The Quarterly Journal of Economics*, 140, 1471–1524.

Adão, R., A. Costinot, D. Donaldson, and J. A. Sturm (2023): "Why is trade not free? A revealed preference approach," National Bureau of Economic Research Working Paper Number 31798.

Agarwal, S., J. B. Jensen, and F. Monte (2017): "The geography of consumption," Available at SSRN 3002231.

Aghion, P. and M. Schankerman (1999): "Competition, entry and the social returns to infrastructure in transition economies," *Economics of Transition*, 7, 79–101.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

AHMED, M. H., W. M. TESFAYE, S. DIETRICH, AND F. GASSMANN (2024): "Rural Roads, Child Labor, and Schooling in Rural Ethiopia," *The Journal of Development Studies*, 60, 1675–1690.

AKBAR, P. A., V. COUTURE, G. DURANTON, AND A. STOREYGARD (2023a): "The fast, the slow, and the congested: Urban transportation in rich and poor countries," National Bureau of Economic Research Working Paper Number 31642.

——— (2023b): "Mobility and congestion in urban India," *American Economic Review*, 113, 1083–1111.

ALDER, S. (2016): "Chinese roads in India: The effect of transport infrastructure on economic development," Available at SSRN 2856050.

ALDER, S., Z. SONG, AND Z. ZHU (2025): "On (Un)Congested Roads: A Quantitative Analysis of Infrastructure Investment Efficiency using Truck GPS Data," in *ASSA Annual Meeting Paper Session: Industrial Policy, Infrastructure, and Productivity Growth in China*.

ALFARO-URENA, A., I. MANELICI, AND J. P. VASQUEZ (2022): "The effects of joining multinational supply chains: New evidence from firm-to-firm linkages," *The Quarterly Journal of Economics*, 137, 1495–1552.

ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129, 1085–1140.

——— (2022): "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 89, 2911–2957.

ALLEN, T. AND D. ATKIN (2022): "Volatility and the Gains from Trade," *Econometrica*, 90, 2053–2092.

ALLEN, T., D. ATKIN, S. C. CLEVES, AND C. E. HERNANDEZ (2024): "The Traveling Trucker Problem," in *AEA Papers and Proceedings*, vol. 114, 334–339.

ALLEN, T., C. DE CASTRO DOBBIN, AND M. MORTEN (2018): "Border walls," National Bureau of Economic Research Working Paper Number 25267.

ALLEN, T. AND D. DONALDSON (2020): "Persistence and path dependence in the spatial economy," National Bureau of Economic Research Working Paper Number 28059.

ALLEN, T., S. FUCHS, S. GANAPATI, A. GRAZIANO, R. MADERA, AND J. MONTORIOL-GARRIGA (2020): "Is tourism good for locals? Evidence from Barcelona," Dartmouth College Working Paper.

ALMAGRO, M., F. BARBIERI, J. C. CASTILLO, N. G. HICKOK, AND T. SALZ (2024): "Optimal Urban Transportation Policy: Evidence from Chicago," National Bureau of Economic Research Working Paper Number 32185.

AMERICO, P. (2023): "The Industrialization Path: Railroads, Technology Adoption, and Structural Transformation in Brazil," Available at SSRN 4942062.

AMITI, M., C. DUPREZ, J. KONINGS, AND J. VAN REENEN (2024): "FDI and superstar spillovers: Evidence from firm-to-firm transactions," *Journal of International Economics*, 152, 103972.

AMODIO, F., G. CHIOVELLI, AND S. FRACHE (2025): "Beefing Up the Service Sector: Commodity Exports to China and Production Network Spillovers," Yale University Working Paper.

ANDERSEN, A. L., K. HUBER, N. JOHANNESEN, L. STRAUB, AND E. T. VESTERGAARD (2022): "Disaggregated economic accounts," National Bureau of Economic Research Working Paper Number 30630.

ANDERSON, J. E. AND E. VAN WINCOOP (2003): "Gravity with gravitas: A solution to the border puzzle," *American Economic Review*, 93, 170–192.

——— (2004): "Trade costs," *Journal of Economic Literature*, 42, 691–751.

ANDERSON, M. L. (2014): "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion," *American Economic Review*, 104, 2763–2796.

——— (2020): "As the wind blows: The effects of long-term exposure to air pollution on mortality," *Journal of the European Economic Association*, 18, 1886–1927.

ANDERSON, M. L. AND M. AUFFHAMMER (2014): "Pounds that kill: The external costs of vehicle weight," *The Review of Economic Studies*, 81, 535–571.

ANDERSON, M. L. AND L. W. DAVIS (2020): "An empirical test of hypercongestion in highway bottlenecks," *Journal of Public Economics*, 187, 104197.

——— (2023): "Uber and traffic fatalities," *Review of Economics and Statistics*, 1–30.

ANDERSSON, D., T. BERGER, AND E. PRAWITZ (2023): "Making a market: Infrastructure, integration, and the rise of innovation," *Review of Economics and Statistics*, 105, 258–274.

ARAUJO, R., J. ASSUNÇÃO, AND A. A. BRAGANÇA (2023): "The effects of transportation infrastructure on deforestation in the amazon: A general equilibrium approach," World Bank Working Paper Number 10415.

ARTUC, E., S. CHAUDHURI, AND J. MCLAREN (2010): "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 100, 1008–1045.

ASHENFELTER, O. AND M. GREENSTONE (2004): "Using mandated speed limits to measure the value of a statistical life," *Journal of Political Economy*, 112, S226–S267.

ASTURIAS, J., M. GARCIA-SANTANA, AND R. RAMOS (2018): "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 17, 1881–1940.

ATACK, J. (2018): "Creating historical transportation shapefiles of navigable rivers, canals, and railroads for the United States before World War I," in *The Routledge Companion to Spatial History*, Routledge, 169–184.

——— (2024): "Railroads," in *Handbook of Cliometrics*, Springer, 2329–2357.

ATALAY, E., A. HORTACSU, J. ROBERTS, AND C. SYVERSON (2011): "Network structure of production," *Proceedings of the National Academy of Sciences*, 108, 5199–5202.

ATALAY, E., A. HORTAÇSU, M. RUNYUN, C. SYVERSON, AND M. F. ULU (2023): "Micro- and macroeconomic impacts of a place-based industrial policy," National Bureau of Economic Research Working Paper Number 31293.

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt (2021): "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 118, e2026160118.

Atkin, D., B. Bernadac, D. Donaldson, T. Garg, and F. Huneeus (2024): "The Incidence of Distortions," Central Bank of Chile Working Paper Number 1026.

Atkin, D., M. K. Chen, and A. Popov (2022): "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," National Bureau of Economic Research Working Paper Number 30147.

Atkin, D. and D. Donaldson (2016): "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," National Bureau of Economic Research Working Paper Number 21439.

——— (2022): "The role of trade in economic development," in *Handbook of International Economics*, Elsevier, vol. 5, 1–59.

Balboni, C. (2025): "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities," *American Economic Review*, 115, 77–116.

Balboni, C. A. and J. S. Shapiro (2025): "Spatial Environmental Economics," National Bureau of Economic Research Working Paper Number 33377.

Baldwin, R., R. Forslid, F. Robert-Nicoud, G. Ottaviano, and P. Martin (2011): *Economic geography and public policy*, Princeton University Press.

Baqaee, D. R. and E. Farhi (2019): "The macroeconomic impact of microeconomic shocks: Beyond Hulten's theorem," *Econometrica*, 87, 1155–1203.

——— (2020): "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

Barnwal, P., J. I. Dingel, D. Iurchenko, P. Krishna, and E. Van Leemput (2024): "Internal Trade Barriers in India," Columbia University Working Paper.

Barth, M. and K. Boriboonsomsin (2008): "Real-world carbon dioxide impacts of traffic congestion," *Transportation Research Record*, 2058, 163–171.

Barwick, P. J., D. Donaldson, S. Li, and Y. Lin (2019): "The Welfare Effects of Passenger Transportation Infrastructure: Evidence from China's HSR Network," Working Paper.

Barwick, P. J., S. Li, D. Donaldson, D. Rao, and Y. Lin (2023): "Transportation Networks, Short-term Mobility, and Pollution Exposure: Evidence from High-Speed Rail in China," National Bureau of Economic Research Working Paper Number 30462.

Barwick, P. J., S. Li, A. R. Waxman, J. Wu, and T. Xia (2024): "Efficiency and equity impacts of urban transportation policies with equilibrium sorting," *American Economic Review*, 104, 3161–3205.

Baum-Snow, N. (2007): "Did highways cause suburbanization?" *The Quarterly Journal of Economics*, 122, 775–805.

——— (2010): "Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960–2000," *American Economic Review: Papers and Proceedings*, 100, 378–382.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2017): "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 99, 435–448.

Baum-Snow, N., D. Hartley, and K. O. Lee (2019): "The long-run effects of neighborhood change on incumbent families," Federal Reserve Bank of Chicago Working Paper Number 2019-02.

Berger, T. (2019): "Railroads and rural industrialization: Evidence from a historical policy experiment," *Explorations in Economic History*, 74, 101277.

Berger, T. and E. Prawitz (2024): "Collaboration and connectivity: Historical evidence from patent records," *Journal of Urban Economics*, 139, 103629.

Bergquist, L. F., C. McIntosh, and M. Startz (2024): "Search costs, intermediation, and trade: Experimental evidence from Ugandan agricultural markets," National Bureau of Economic Research Working Paper Number 33221.

Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2022): "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, 130, 1765–1804.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2019): "Production networks, geography, and firm performance," *Journal of Political Economy*, 127, 639–688.

Berry, S. and P. Jia (2010): "Tracing the woes: An empirical analysis of the airline industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Bleakley, H. and J. Lin (2012): "Portage and path dependence," *The Quarterly Journal of Economics*, 127, 587–644.

Blumenstock, J. E. (2012): "Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda," *Information Technology for Development*, 18, 107–125.

Bonfatti, R., G. Facchini, A. Tarasov, G. L. Tedeschi, and C. Testa (2021): "Pork, infrastructure and growth: Evidence from the Italian railway expansion," Centre for Economic Policy Research Discussion Paper Number DP16462.

Bonfatti, R., Y. Gu, and S. Poelhekke (2019): "Priority roads: The political economy of Africa's interior-to-coast roads," Tinbergen Institute Discussion Paper Number TI 2019-006/VIII.

Bordeu, O. (2023): "Commuting infrastructure in fragmented cities," Working Paper.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2024): "Investment in infrastructure and trade: The case of ports," National Bureau of Economic Research Working Paper Number 32503.

Brancaccio, G., M. Kalouptsidi, T. Papageorgiou, and N. Rosaia (2023): "Search frictions and efficiency in decentralized transport markets," *The Quarterly Journal of Economics*, 138, 2451–2503.

BRINKMAN, J. AND J. LIN (2024): "Freeway revolts! The quality of life effects of highways," *Review of Economics and Statistics*, 106, 1268–1284.

BROOKS, L. AND Z. LISCOW (2023): "Infrastructure costs," *American Economic Journal: Applied Economics*, 15, 1–30.

BROOKS, W. AND K. DONOVAN (2020): "Eliminating uncertainty in market access: The impact of new bridges in rural Nicaragua," *Econometrica*, 88, 1965–1997.

BROOKS, W., J. P. KABOSKI, I. O. KONDO, Y. A. LI, AND W. QIAN (2021): "Infrastructure investment and labor monopsony power," National Bureau of Economic Research Working Paper Number 28977.

BRYAN, G., E. GLAESER, AND N. TSIVANIDIS (2020): "Cities in the developing world," *Annual Review of Economics*, 12, 273–297.

BRYAN, G. T., K. FRYE, AND M. MORTEN (2025): "Spatial economics for low-and middle-income countries," National Bureau of Economic Research Working Paper Number 33453.

BUCHHOLZ, N. (2022): "Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry," *The Review of Economic Studies*, 89, 556–591.

BUCHHOLZ, N., L. DOVAL, J. KASTL, F. MATĚJKA, AND T. SALZ (2020): "Personalized Pricing and the Value of Time: Evidence from Auctioned Cab Rides," National Bureau of Economic Research Working Paper Number 27087.

BURGESS, R., R. JEDWAB, E. MIGUEL, A. MORJARIA, AND G. PADRÓ I MIQUEL (2015): "The value of democracy: evidence from road building in Kenya," *American Economic Review*, 105, 1817–1851.

BURKE, M., L. F. BERGQUIST, AND T. MIGUEL (2019): "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets," *The Quarterly Journal of Economics*, 134, 785–842.

BURSTEIN, A., J. CRAVINO, AND M. ROJAS (2024): "Input price dispersion across buyers and misallocation," National Bureau of Economic Research Working Paper Number 33128.

BUSTOS, P., G. GARBER, AND J. PONTICELLI (2020): "Capital accumulation and structural transformation," *The Quarterly Journal of Economics*, 135, 1037–1094.

CALIENDO, L., M. DVORKIN, AND F. PARRO (2019): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87, 741–835.

CAO, Y. AND S. CHEN (2022): "Rebel on the canal: Disrupted trade access and social conflict in China, 1650–1911," *American Economic Review*, 112, 1555–1590.

CARDOZA, M., F. GRIGOLI, N. PIERRI, AND C. RUANE (2020): "Worker mobility and domestic production networks," International Monetary Fund Working Paper Number 20/205.

CHACHA, P., B. KIRUI, AND V. WIEDEMANN (2022): "Mapping kenya's production network transaction by transaction," Available at SSRN 4315810.

CHANDRA, A. AND E. THOMPSON (2000): "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, 30, 457–490.

CHANG, M. AND H. LEE (2024): "Bridging Micro and Macro Production Functions: The Fiscal Multiplier of Infrastructure Investment," Working Paper.

CHEN, J., Y. NIU, AND M. SHUM (2023): "Reassessing the Firm Selection Hypothesis: New Evidence from Chinese Highways," Available at SSRN 3670452.

CHEN, T., J. WANG, AND H. S. QI (2022): "Railways, Telegraph and Technology Adoption: The Introduction of American Cotton in Early 20th Century China," Available at SSRN 4868857.

COHEN, A. AND L. EINAV (2003): "The effects of mandatory seat belt laws on driving behavior and traffic fatalities," *Review of Economics and Statistics*, 85, 828–843.

COLLIER, P., M. KIRCHBERGER, AND M. SÖDERBOM (2016): "The cost of road infrastructure in low-and middle-income countries," *The World Bank Economic Review*, 30, 522–548.

COOK, C., A. KREIDIEH, S. VASSERMAN, H. ALLCOTT, N. ARORA, F. VAN SAMBEEK, A. TOMKINS, AND E. TURKEL (2025): "The Short-Run Effects of Congestion Pricing in New York City," National Bureau of Economic Research Working Paper Number 33584.

COOK, C. AND P. Z. LI (2024): "Value pricing or lexus lanes? The distributional effects of dynamic tolling," Yale University Working Paper.

COŞAR, A. K., B. DEMIR, D. GHOSE, AND N. YOUNG (2022): "Road capacity, domestic trade and regional outcomes," *Journal of Economic Geography*, 22, 901–929.

COŞAR, A. K., P. L. GRIECO, AND F. TINTELNOT (2015): "Borders, geography, and oligopoly: evidence from the wind turbine industry," *Review of Economics and Statistics*, 97, 623–637.

COŞAR, A. K., S. OSOTIMEHIN, AND L. POPOV (2024): "The Long-run Effects of Transportation Productivity on the US Economy," National Bureau of Economic Research Working Paper Number 3248.

COUTTENIER, M., J. MARCOUX, T. MAYER, AND M. THOENIG (2024): "The Gravity of Violence," Sciences Po Discussion Paper Number 2024-10.

COUTURE, V., J. I. DINGEL, A. E. GREEN, AND J. HANDBURY (2025): "Demographic preferences and income segregation," National Bureau of Economic Research Working Paper Number 33386.

COUTURE, V., G. DURANTON, AND M. A. TURNER (2018): "Speed," *Review of Economics and Statistics*, 100, 725–739.

CRAFTS, N. (2004): "Social savings as a measure of the contribution of a new technology to economic growth," London School of Economics Working Paper Number 06/04.

CRAFTS, N. AND A. KLEIN (2014): "Geography and intra-national home bias: US domestic trade in 1949 and 2007," *Journal of Economic Geography*, 15, 477–497.

CURRIE, J. AND R. WALKER (2011): "Traffic congestion and infant health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 3, 65–90.

DAVID, P. A. (1969): "Transport innovation and economic growth: Professor Fogel on and off the rails," *The Economic History Review*, 22, 506–525.

DAVIS, L. W. (2008): "The effect of driving restrictions on air quality in Mexico City," *Journal of Political Economy*, 116, 38–81.

DEANGELO, G. AND B. HANSEN (2014): "Life and death in the fast lane: Police enforcement and traffic fatalities," *American Economic Journal: Economic Policy*, 6, 231–257.

DEFFEBACH, P., D. LAGAKOS, Y. MIYAUCHI, AND E. YAMADA (2025): "The Spatial Distribution of Income in Cities: Cross-Country Evidence and Theory," Boston University Working Paper.

DELL, M. (2015): "Trafficking networks and the Mexican drug war," *American Economic Review*, 105, 1738–1779.

DEMIR, B., A. C. FIELER, D. Y. XU, AND K. K. YANG (2024): "O-ring production networks," *Journal of Political Economy*, 132, 200–247.

DESMET, K. AND F. PARRO (2025): "Spatial Dynamics," National Bureau of Economic Research Working Paper Number 33443.

DHYNE, E., A. K. KIKKAWA, M. MOGSTAD, AND F. TINTELNOT (2021): "Trade and domestic production networks," *The Review of Economic Studies*, 88, 643–668.

DIAMOND, R., T. MCQUADE, AND F. QIAN (2019): "The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco," *American Economic Review*, 109, 3365–3394.

DIAMOND, R. AND E. MORETTI (2021): "Where is standard of living the highest? Local prices and the geography of consumption," National Bureau of Economic Research Working Paper Number 29533.

DIEM, C., A. BORSOS, T. REISCH, J. KERTÉSZ, AND S. THURNER (2024): "Estimating the loss of economic predictability from aggregating firm-level production networks," *PNAS Nexus*, 3, pgae064.

DING, L., J. HWANG, AND E. DIVRINGI (2016): "Gentrification and residential mobility in Philadelphia," *Regional Science and Urban Economics*, 61, 38–51.

DIXIT, A. K. AND V. NORMAN (1980): *Theory of International Trade*, Cambridge University Press.

DOLFEN, P., L. EINAV, P. J. KLENOW, B. KLOPACK, J. D. LEVIN, L. LEVIN, AND W. BEST (2023): "Assessing the gains from e-commerce," *American Economic Journal: Macroeconomics*, 15, 342–370.

DONALD, E., M. FUKUI, AND Y. MIYAUCHI (2024): "Unpacking aggregate welfare in a spatial economy," Boston University Working Paper.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 131, 799–858.

DURANTON, G., P. M. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence from the US," *The Review of Economic Studies*, 81, 681–724.

DURANTON, G. AND M. A. TURNER (2011): "The fundamental law of road congestion: Evidence from US cities," *American Economic Review*, 101, 2616–2652.

——— (2012): "Urban growth and transportation," *The Review of Economic Studies*, 79, 1407–1440.

DURANTON, G. AND A. J. VENABLES (2019): "Place-Based Policies for Development," National Bureau of Economic Research Working Paper Number 24562.

EATON, J. AND S. KORTUM (2002): "Technology, Geography and Trade," *Econometrica*, 70, 1741–1779.

ECKERT, F. (2019): "Growing apart: Tradable services and the fragmentation of the U.S. economy," Yale University Working Paper.

EGGER, D., B. FABER, M. LI, AND W. LIN (2024): "Rural-Urban Migration and Market Integration," University of California, Berkeley Working Paper.

ELIASSON, J. AND M. FOSGERAU (2019): "Cost-benefit analysis of transport improvements in the presence of spillovers, matching and an income tax," *Economics of Transportation*, 18, 1–9.

FABER, B. (2014): "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 81, 1046–1070.

FACKLER, P. L. AND B. K. GOODWIN (2001): "Spatial price analysis," in *Handbook of Agricultural Economics*, Elsevier, vol. 1, 971–1024.

FAJGELBAUM, P. AND S. J. REDDING (2022): "Trade, structural transformation, and development: Evidence from Argentina 1869–1914," *Journal of Political Economy*, 130, 1249–1318.

FAJGELBAUM, P. D. AND C. GAUBERT (2025): "Optimal spatial policies," National Bureau of Economic Research Working Paper Number 33493.

FAJGELBAUM, P. D., C. GAUBERT, N. GORTON, E. MORALES, AND E. SCHAAL (2023): "Political Preferences and Transport Infrastructure: Evidence from California's High-Speed Rail," National Bureau of Economic Research Working Paper Number 31438.

FAJGELBAUM, P. D. AND E. SCHAAL (2020): "Optimal transport networks in spatial equilibrium," *Econometrica*, 88, 1411–1452.

FELBERMAYR, G. J. AND A. TARASOV (2022): "Trade and the spatial distribution of transport infrastructure," *Journal of Urban Economics*, 130, 103473.

FOGEL, R. W. (1964): *Railroads and American Economic Growth*, Johns Hopkins University Press.

——— (1979): "Notes on the social saving controversy," *The Journal of Economic History*, 39, 1–54.

FOX, J. T. (2018): "Estimating matching games with transfers," *Quantitative Economics*, 9, 1–38.

FRECHETTE, G. R., A. LIZZERI, AND T. SALZ (2019): "Frictions in a competitive, regulated market: Evidence from taxis," *American Economic Review*, 109, 2954–2992.

FRYE, D. (2024): "Transportation networks and the geographic concentration of employment," *Review of Economics and Statistics*, 1–34.

GEBRESILASSE, M. (2023): "Rural roads, agricultural extension, and productivity," *Journal of Development Economics*, 162, 103048.

GENDRON-CARRIER, N., M. GONZALEZ-NAVARRO, S. POLLONI, AND M. A. TURNER (2022): "Subways and urban air pollution," *American Economic Journal: Applied Economics*, 14, 164–196.

GERARD, F. AND J. NARITOMI (2021): "Job displacement insurance and (the lack of) consumption-smoothing," *American Economic Review*, 111, 899–942.

GERTLER, P. J., M. GONZALEZ-NAVARRO, T. GRAČNER, AND A. D. ROTHENBERG (2024): "Road maintenance and local economic development: Evidence from Indonesia's highways," *Journal of Urban Economics*, 143, 103687.

GERVAIS, A. AND J. B. JENSEN (2019): "The tradability of services: Geographic concentration and trade costs," *Journal of International Economics*, 118, 331–350.

GIBBONS, S., S. HEBLICH, AND E. W. PINCHBECK (2024): "The spatial impacts of a massive rail disinvestment program: The Beeching Axe," *Journal of Urban Economics*, 143, 103691.

GLAESER, E. L. (2020): "Infrastructure and urban form," National Bureau of Economic Research Working Paper Number 28287.

GLAESER, E. L., L. CURRIER, AND G. E. KREINDLER (2023): "Infrastructure inequality: Who pays the cost of road roughness?" National Bureau of Economic Research Working Paper Number 31981.

GLAESER, E. L. AND G. A. PONZETTO (2018): "The political economy of transportation investment," *Economics of Transportation*, 13, 4–26.

GLAESER, E. L. AND J. M. POTERBA (2020): "Economic analysis and infrastructure investment," National Bureau of Economic Research Working Paper Number 28215.

GOLDBERG, P. K. AND G. MAGGI (1999): "Protection for Sale: An Empirical Investigation," *American Economic Review*, 89, 1135–1155.

GOLDSZMIDT, A., J. A. LIST, R. D. METCALFE, I. MUIR, V. K. SMITH, AND J. WANG (2020): "The value of time in the United States: Estimates from nationwide natural field experiments," National Bureau of Economic Research Working Paper Number 28208.

GOLLIN, D. AND J. WOLFERSBERGER (2024): "Agricultural Trade and Deforestation: the Role of New Roads," AgroParisTech Working Paper Number hal-04171124.

GONZALEZ-NAVARRO, M., J. D. HALL, R. WILLIAMS, H. WHEELER, AND R. WILLIAMS (2021): "Uber versus Trains? Worldwide Evidence from Transit Expansions," Networks, Electronic Commerce and Telecommunications Institute Working Paper Number 21–11.

GONZALEZ-NAVARRO, M. AND M. A. TURNER (2018): "Subways and urban growth: Evidence from earth," *Journal of Urban Economics*, 108, 85–106.

GONZALEZ-NAVARRO, M., R. D. ZARATE, R. JEDWAB, AND N. TSIVANIDIS (2023): "Land transport infrastructure," *VoxDevLit*, 9.

GORTON, N. AND E. IANCHOVICHINA (2025): "Trade networks in latin america: Spatial inefficiencies and optimal expansions," *The World Bank Economic Review*, lhae054.

GRAFF, T. (2024): "Spatial inefficiencies in Africa's trade network," *Journal of Development Economics*, 171, 103319.

GUERRA, E., G. DURANTON, AND X. MA (2025): "Urban roadway in America: The amount, extent, and value," *Journal of the American Planning Association*, 91, 102–116.

HAHN, R. W., R. D. METCALFE, AND E. TAM (2023): "Welfare estimates of shifting peak travel," National Bureau of Economic Research Working Paper Number 31629.

HALL, J. D. (2018): "Pareto improvements from Lexus Lanes: The effects of pricing a portion of the lanes on congested highways," *Journal of Public Economics*, 158, 113–125.

——— (2021): "Can tolling help everyone? Estimating the aggregate and distributional consequences of congestion pricing," *Journal of the European Economic Association*, 19, 441–474.

HALL, J. D., C. PALSSON, AND J. PRICE (2018): "Is Uber a substitute or complement for public transit?" *Journal of Urban Economics*, 108, 36–50.

HANNA, R., G. KREINDLER, AND B. A. OLKEN (2017): "Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta," *Science*, 357, 89–93.

HARRIS, A. AND T. M. A. NGUYEN (2025): "Long-term relationships and the spot market: Evidence from U.S. trucking," Working Paper.

HAUSMAN, N., P. R. SAMUELS, M. C. COHEN, AND R. SASSON (2023): "Urban pull: The roles of amenities and employment," Available at SSRN 4798884.

HEBLICH, S., S. J. REDDING, AND D. M. STURM (2020): "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 135, 2059–2133.

HERZOG, I. (2024): "The city-wide effects of tolling downtown drivers: Evidence from London's congestion charge," *Journal of Urban Economics*, 144, 103714.

HERZOG, I., S. LIU, AND Y. YU (2024): "National Road Upgrading and Structural Transformation: Evidence from Ugandan Households," Huron University Working Paper.

HIERONS, T. (2024): "Spreading the jam: Optimal congestion pricing in general equilibrium," University of Chicago Working Paper.

HILLBERRY, R. AND D. HUMMELS (2008): "Trade responses to geographic frictions: A decomposition using micro-data," *European Economic Review*, 52, 527–550.

HOLMES, T. J. AND J. A. SCHMITZ JR (2001): "Competition at work: Railroads vs. monopoly in the US shipping industry," *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 3–29.

HORNBECK, R. AND M. ROTEMBERG (2024): "Growth off the rails: Aggregate productivity growth in distorted economies," *Journal of Political Economy*, 132, 3547–3602.

HULTEN, C. R. (1978): "Growth Accounting with Intermediate Inputs," *The Review of Economic Studies*, 45, 511–518.

HUNEEUS, F. (2018): "Production Network Dynamics and the Propagation of Shocks," Central Bank of Chile Working Paper.

IMBENS, G. W. AND D. B. RUBIN (2015): *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

JACOBS, B., E. L. JONGEN, AND F. T. ZOUTMAN (2017): "Revealed social preferences of Dutch political parties," *Journal of Public Economics*, 156, 81–100.

JACOBSEN, M. R. (2013): "Fuel economy and safety: The influences of vehicle class and driver behavior," *American Economic Journal: Applied Economics*, 5, 1–26.

JEDWAB, R. AND A. STOREYGARD (2022): "The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010," *Journal of the European Economic Association*, 20, 1–38.

KABOSKI, J. P., J. LU, W. QIAN, AND L. REN (2024): "Melitz Meets Lewis: The Impacts of Roads on Structural Transformation and Businesses," National Bureau of Economic Research Working Paper Number 32448.

KASHNER, Z. AND B. ROSS (2025): "Measuring and Mitigating Traffic Externalities," Stanford Graduate School of Business Working Paper.

KENNAN, J. AND J. R. WALKER (2011): "The effect of expected income on individual migration decisions," *Econometrica*, 79, 211–251.

KHANNA, G., N. MORALES, AND N. PANDALAI-NAYAR (2022): "Supply Chain Resilience: Evidence from Indian Firms," National Bureau of Economic Research Working Paper Number 30689.

KHANNA, G., A. NYSHADHAM, D. RAMOS-MENCHELLI, J. A. TAMAYO, AND A. TIEW (2023): "Spatial mobility, economic opportunity, and crime," Harvard Business School Working Paper Number 24-016.

KIRCHBERGER, M. (2021): "Measuring internal migration," *Regional Science and Urban Economics*, 91, 103714.

KNITTEL, C. R., D. L. MILLER, AND N. J. SANDERS (2016): "Caution, drivers! Children present: Traffic, pollution, and infant health," *Review of Economics and Statistics*, 98, 350–366.

KOCKELMAN, K., T. D. CHEN, AND B. NICHOLS (2013): *The Economics of Transportation Systems: A Reference for Practitioners*, Texas Department of Transportation.

KOROVKIN, V., A. MAKARIN, AND Y. MIYAUCHI (2024): "Supply Chain Disruption and Reorganization: Theory and Evidence from Ukraine's War," Massachusetts Institute of Technology Sloan Research Paper Number 7068-24.

KREINDLER, G. (2024): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, 92, 1233–1268.

KREINDLER, G., A. GADUH, T. GRAFF, R. HANNA, AND B. A. OLKEN (2023): "Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta," National Bureau of Economic Research Working Paper Number 31369.

KREINDLER, G. E. AND Y. MIYAUCHI (2023): "Measuring commuting and economic activity inside cities with cell phone records," *Review of Economics and Statistics*, 105, 899–909.

LAIRD, J. AND E. TVETER (2023): "Agglomeration and transport appraisal: New developments and research directions," Report to the Department for Transport. Inverness: Peak Economics.

LAIRD, J. J. AND A. J. VENABLES (2017): "Transport investment and economic performance: A framework for project appraisal," *Transport Policy*, 56, 1–11.

LALL, S., S. SINHA-ROY, AND F. SHILPI (2022): "Trucking Costs and the Margins of Internal Trade," World Bank Policy Research Working Paper Number 10059.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2022): "Imperfect competition, compensating differentials, and rent sharing in the US labor market," *American Economic Review*, 112, 169–212.

LARCOM, S., F. RAUCH, AND T. WILLEMS (2017): "The benefits of forced experimentation: Striking evidence from the London underground network," *The Quarterly Journal of Economics*, 132, 2019–2055.

LEDUC, S. AND D. WILSON (2013): "Roads to prosperity or bridges to nowhere? Theory and evidence on the impact of public infrastructure investment," *NBER Macroeconomics Annual*, 27, 89–142.

LEUNIG, T. (2006): "Time is money: A re-assessment of the passenger social savings from Victorian British railways," *The Journal of Economic History*, 66, 635–673.

——— (2010): "Social savings," *Journal of Economic Surveys*, 24, 775–800.

LINDGREN, E., P. PETTERSSON-LIDBOM, AND B. TYREFORS (2021): "The causal effect of transport infrastructure: Evidence from a new historical database," Institute of Industrial Economics Working Paper Number 1407.

LIPSEY, R. AND K. LANCASTER (1956): "The General Theory of Second Best," *The Review of Economic Studies*, 24, 11–32.

LOUMEAU, G. (2023): "Regional Borders, Commuting and Transport Network Integration," *Review of Economics and Statistics*, 1–45.

MANGRUM, D. AND A. MOLNAR (2017): "The marginal congestion of a taxi in New York City," Vanderbilt University Working Paper.

MCCALLUM, J. (1995): "National borders matter: Canada-US regional trade patterns," *American Economic Review*, 85, 615–623.

MEHROTRA, N., M. A. TURNER, AND J. P. URIBE (2024): "Does the US have an infrastructure cost problem? Evidence from the interstate highway system," *Journal of Urban Economics*, 143, 103681.

MILSOM, L. H. (2023): "Moving opportunity: local connectivity and spatial inequality," Center for Economic Research and Applications Working Paper Number 2303.

MIYAUCHI, Y., K. NAKAJIMA, AND S. J. REDDING (2021): "The economics of spatial mobility: Theory and evidence using smartphone data," National Bureau of Economic Research Working Paper Number 28497.

MOHRING, H. (1993): "Maximizing, measuring, and not double counting transportation-improvement benefits: A primer on closed-and open-economy cost-benefit analysis," *Transportation Research Part B: Methodological*, 27, 413–424.

MONEKE, N. (2020): "Can big push infrastructure unlock development? Evidence from Ethiopia," Structural Transformation and Economic Growth Working Paper.

MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): "Commuting, migration, and local employment elasticities," *American Economic Review*, 108, 3855–3890.

MORETTI, E. (2021): "The effect of high-tech clusters on the productivity of top inventors," *American Economic Review*, 111, 3328–3375.

MORTEN, M. AND J. OLIVEIRA (2024): "The Effects of Roads on Trade and Migration : Evidence from a Planned Capital City," *American Economic Journal: Applied Economics*, 16, 389–421.

MULALIC, I., B. D. BORGER, AND J. ROUWENDAL (2019): "Productivity and Wage Effects of an Exogenous Improvement in Transport Infrastructure: Accessibility and the Great Belt Bridge," Available at SSRN 5015669.

NAGY, D. K. (2023): "Hinterlands, city formation and growth: Evidence from the US westward expansion," *The Review of Economic Studies*, 90, 3238–3281.

OLKEN, B. A. (2007): "Monitoring corruption: evidence from a field experiment in Indonesia," *Journal of Political Economy*, 115, 200–249.

OLKEN, B. A. AND P. BARRON (2009): "The simple economics of extortion: evidence from trucking in Aceh," *Journal of Political Economy*, 117, 417–452.

OPPENHEIM, N. (1995): *Urban travel demand modeling: from individual choices to general equilibrium.*, John Wiley and Sons.

PANIGRAHI, P. (2021): "Endogenous spatial production networks: Quantitative implications for trade and productivity," Center for Economic Studies and ifo Institute Working Paper Number 9466.

PARRY, I. W. H., M. WALLS, AND W. HARRINGTON (2007): "Automobile externalities and policies," *Journal of Economic Literature*, 45, 373–399.

PAULY, S. AND F. STIPANICIC (2023): "The creation and diffusion of knowledge: Evidence from the Jet Age," Sciences Po Working Paper Number hal-04067326.

PERLMAN, E. R. (2016): "Dense enough to be brilliant: patents, urbanization, and transportation in nineteenth century America," Boston University Working Paper.

PILEGAARD, N. AND M. FOSGERAU (2008): "Cost benefit analysis of a transport improvement in the case of search unemployment," *Journal of Transport Economics and Policy (JTEP)*, 42, 23–42.

PORCHER, C., E. MORALES, AND T. FUJIWARA (2024): "Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach," National Bureau of Economic Research Working Paper 32413.

QIN, M., M. A. VITORINO, AND G. JOHN (2024): "Planes, trains, and co-opetition: Evidence from China," Available at SSRN 4194019.

RAMEY, V. A. (2021): "The Macroeconomic Consequences of Infrastructure Investment," in *Economic Analysis and Infrastructure Investment*, University of Chicago Press.

RAVALLI, G. (2024): "The Effect of Transportation Infrastructure on Innovation: The Role of Market Access in the English Railway Boom," London School of Economics Working Paper.

REDDING, S. J. (2025): "Evaluating Transport Improvements in Spatial Equilibrium," National Bureau of Economic Research Working Paper Number 33532.

REDDING, S. J. AND M. A. TURNER (2015): "Transportation costs and the spatial organization of economic activity," in *Handbook of Regional and Urban Economics*, Elsevier, vol. 5, 1339–1398.

RELIHAN, L. (2022): "Is Online Retail Killing Coffee Shops? Estimating the Winners and Losers of Online Retail Using Customer Transaction Microdata," London School of Economics and Political Science Working Paper Number 117805.

SANTAMARIA, M. (2020): "Reshaping Infrastructure: Evidence from the division of Germany," University of Warwick Working Paper Number 1244.

SEVEREN, C. (2023): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *Review of Economics and Statistics*, 105, 1073–1091.

SIMEONOVA, E., J. CURRIE, P. NILSSON, AND R. WALKER (2021): "Congestion pricing, air pollution, and children's health," *Journal of Human Resources*, 56, 971–996.

SIMONOVSKA, I. AND M. WAUGH (2014): "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 92, 34–50.

SMALL, K. A., E. T. VERHOEF, AND R. LINDSEY (2024): *The Economics of Urban Transportation*, Routledge.

SMALL, K. A. AND C. WINSTON (1988): "Optimal highway durability," *American Economic Review*, 78, 560–569.

SMALL, K. A., C. WINSTON, AND J. YAN (2005): "Uncovering the distribution of motorists' preferences for travel time and reliability," *Econometrica*, 73, 1367–1382.

SOKOLOFF, K. L. (1988): "Inventive activity in early industrial America: evidence from patent records, 1790–1846," *The Journal of Economic History*, 48, 813–850.

SOMAINI, P. (2020): "Identification in auction models with interdependent costs," *Journal of Political Economy*, 128, 3820–3871.

SPRAY, J. (2021): "Search externalities in firm-to-firm trade," International Monetary Fund Working Paper Number 2021/091.

SPRAY, J. AND S. WOLF (2018): "Industries without smokestacks in Uganda and Rwanda," in *Industries without smokestacks: Industrialization in Africa reconsidered*, Oxford University Press, 341–363.

STARRETT, D. (1978): "Market allocations of location choice in a model with free mobility," *Journal of Economic Theory*, 17, 21–37.

STARRETT, D. A. (1988): *Foundations in Public Economics*, Cambridge University Press.

TOWNSEND, R. M. (2016): "Village and larger economies: The theory and measurement of the Townsend Thai project," *Journal of Economic Perspectives*, 30, 199–220.

TSIVANIDIS, N. (2022): "Evaluating the impact of urban transit infrastructure: Evidence from Bogota's TransMilenio," University of California, Berkeley Working Paper.

TYNDALL, J. (2021): "The local labour market effects of light rail transit," *Journal of Urban Economics*, 124, 103350.

VANDEN EYNDE, O. AND L. WREN-LEWIS (2024): "Complementarities in Infrastructure: Evidence from Indian Agriculture," Center for Open Science Working Paper Number ejb8x.

VITALI, A. (2023): "Consumer search and firm location: Theory and evidence from the garment sector in Uganda," Dartmouth College Working Paper.

VOIGTLÄNDER, N. AND H.-J. VOTH (2014): "Highway to Hitler," National Bureau of Economic Research Working Paper Number 20150.

WALKER, M. W., N. SHAH, E. MIGUEL, D. EGGER, F. S. SOLIMAN, AND T. GRAFF (2024): "Slack and economic development," National Bureau of Economic Research Working Paper Number 33055.

WEIWU, L. (2024): "Unequal access: Racial segregation and the distributional impacts of interstate highways in cities," Stanford University Working Paper.

YAGAN, D. (2019): "Employment hysteresis from the great recession," *Journal of Political Economy*, 127, 2505–2558.

ZÁRATE, R. D. (2022): "Spatial misallocation, informality, and transit improvements," World Bank Policy Research Working Paper Number 9990.