Journal of Economic Perspectives—Volume 36, Number 3—Summer 2022—Pages 185–210

Blending Theory and Data: A Space Odyssey

Dave Donaldson

What are the effects of trade liberalization, or the recent US-China trade war? Is urban gentrification leading to spatial inequalities and an erosion of opportunities for economic mobility? Do transportation infrastructure investments justify their astronomic price tags? These are all great questions—and they comprise only a small sampling from the bread-and-butter topics of spatial economics. But readers seeking specific answers to such questions have come to the wrong place. The focus of this article is instead about how economists working in the fields of international, regional, and urban economics arrive at answers to these sorts of counterfactual—that is, inherently causal—questions. It is about the spatial journey rather than the spatial destination.

For questions like these, economic theory alone does little to narrow the range of quantitative answers. Moreover, for questions like these, nature has not granted us sufficiently rich quasi-experimental serendipity that we can draw on it as a replacement for economic theory. What is to be done? The only option on the table is to combine the lessons of economic theory with what we can glean from empirical patterns. While there are many ways to do so, my focus will be on research that pursues an explicit theory-empirics nexus. This process involves using the relationships identified by quasi-experimental variation to the full extent possible, while also recognizing that data typically will not fully answer the policy question that motivates a given research study. As such, the analyst must use additional

■ *Dave Donaldson is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is ddonald@mit.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.36.3.185.

information—modeling assumptions and the logical deductions that follow from them—to bridge the gap between what is identified and what is desired.

Just as any theoretical model is a metaphor, not an attempt to be a true representation of reality, the work I describe in this article involves researchers aiming to build an *empirical* metaphor. Economists are used to resisting the temptation to judge a model by the strength of its abstraction or its assumptions. Instead, we ask how useful a model appears to be at achieving some goal. That is, the role of a model is to provide a clear mapping from assumptions to answers to a given question, and it should be judged relative to how faithfully it delivers that answer. An empirical model is no different. It provides a clear mapping of assumptions to answers—but it does so conditional on the extra information provided by features that can be observed in the data. In this sense, a theme that appears throughout much of my discussion is one in which researchers, aiming to answer a given question, understand that theoretical assumptions will be needed to answer their question, but still do their best to minimize the need for such assumptions through the use of facts that can be extracted from the available data.

The goal of this article is to highlight, through a generic framework and a range of examples, some of the techniques deployed in spatial economics that have leaned on the complementarity between theory and data. This inevitably draws on advances made in all areas of economics, and hence relates to other recent methodological surveys that emphasize interactions between theory and data.[1] Nevertheless, the nature of spatial research often presents unique challenges due to the large number of economic interactions at work both within geographic units (among producers, consumers, and factors of production) and across them.

Models and Questions

I begin by describing a generic research problem—a question to be answered, a set of data features that are observable, a set of beliefs about sources of exogenous variation in such data, and a "model."

My discussion revolves around the following scenario. We imagine a researcher who, in some setting of interest, desires to answer the question, "What would be the change in outcome W if a change in policy X were to occur?" Notably, the goal is to quantify a causal effect: that of X on W. To continue an example from above, with knowledge of the effect of certain transportation infrastructure investments (X) on a certain government's social objective function (W), we could seek to evaluate whether those investments were money well spent.

[1] Examples include Acemoglu (2010), Andrews, Gentzkow, and Shapiro (2020), Baum-Snow and Ferreira (2015), Einav and Finkelstein (2018), Finkelstein and Hendren (2020), Hansen and Heckman (1996), Heckman (2010), Holmes and Sieg (2015), Intriligator (1983), Keane (2010), Leamer (2012), Low and Meghir (2017), Manski (2013), Matzkin (1986), Nakamura and Steinsson (2018), Nevo and Whinston (2010), Paarsch and Hong (2006), Reiss and Wolak (2007), Rust (2014), Timmins and Schlenker (2009), and Wolpin (2013).

What can a researcher observe about this setting? As a starting point, we imagine that the policy variable X is observed for each member among a set of units of observation: for example, countries, regions, firms, or households. However, in general, the object of interest W is not observed—indeed, we are often interested in concepts, such as notions of economic welfare, whose measurement from even idealized datasets can be controversial.

It is at this point that the researcher's theoretical "model" enters the picture. Both the object of interest variable W and the policy variable X are related to other variables. First, the researcher believes that the object of interest W can be viewed as a function of an additional observable, an auxiliary outcome (or, at times, several outcomes) denoted by Y. We write this function as $W = g(Y,\theta)$. In this notation $g(\cdot)$ is a model—that is, it is a function the researcher will assume because there is reason to believe it is plausible—but the model's parameters, denoted by θ, may not be known. Second, we imagine that the researcher can observe an additional variable, denoted by Z, that is connected to X and is often referred to as an "instrument." This variable will be used to study the effect of X on Y in a manner discussed below but, as is already apparent, it is of no direct relation to the researcher's model or question of interest. Its role will be important, but merely instrumental—just as a hammer is an indispensable tool for hanging a painting on a wall, but the hammer itself is not much to look at.

To give a sense of how research from this framework might operate, one can imagine the researcher striving to assemble two ingredients. The first tells us how the researcher's policy of interest X affects the auxiliary outcome Y. The second tells us how the auxiliary observable Y translates into the unobserved outcome of interest W, a mapping that depends on the researcher's model $g(\cdot)$ and the parameters θ. This two-way breakdown is central to what follows.

While the discussion so far has been deliberately abstract, a number of essential points are already apparent. First, we are starting with a question—that is, how large is the change in W caused by a change in X?—and holding the question fixed. Second, since W is unobserved, we could not answer this question without the help of our model, whose role is to tell us how the observable Y relates to the desired outcome W. Third, since both Y and X are observable, it is possible, at least in principle, to use data alone to reveal the empirical effect of X on Y. Given knowledge of such effect sizes, the parameters θ are the only unknowns that stand in the way of the researcher arriving at an answer to the question posed.

Finally, and crucially, a researcher will typically have wide latitude to choose the set of Y variables being included in the model $g(\cdot)$. This is important because the logical essence of the model changes as we condition on more auxiliary outcomes—indeed, the strength of the assumptions being invoked in $g(\cdot)$ gets weaker as more outcomes Y are included. In this sense, the parameters of any model are specific to that model. As Fernandez-Villaverde (2008) puts it, in the context of procedures that use estimates of individual-level responses in aggregate-level models: "Borrowing parameters from microeconomic models forgets that parameters do not have a life of their own as some kind of platonic entity. Instead, parameters have meaning only within the context of a particular model." While one could imagine economists building up a complete understanding of the world's economic parameters from the ground up—akin to the book full of natural constants that can be found in a chemistry lab—this isn't how most economics research actually works. We write down models that strike a balance between plausibility, parsimony, and (statistical) precision, but always relative to the question of interest and the data available. In this regard, it is no surprise to open an economics journal and find that almost any given empirical model will have a similar (and small) number of parameters to be estimated, regardless of whether they aim to reflect the Peruvian prawn industry or half of planetary production. The challenges of doing social science mean that the empirical metaphors that economists use are, unlike those in the field of chemistry, inevitably context-specific and deliberately parsimonious.

The Tyranny of Distance Between Data and Answers

We have set up the researcher's problem: a question to be answered, a set of available data, and a set of maintained assumptions that we call a "model." How can the researcher use these inputs of theory and data to answer the question that has been posed? I will build up one way of describing responses to this challenge in spatial economics, with examples along the way. These examples begin with settings in which the available data variation very closely answers the question of interest, and so the role of the researcher's model is relatively minimal. My examples then progress to settings with greater separation between data and answers, where the discussion will be organized around steps that researchers take to minimize such a distance—that is, to make the leap from data to answers under as plausible a set of theoretical assumptions as they can. These steps involve careful choices about which auxiliary outcomes Y to measure as well as an understanding of economic theory that helps inform the researcher's model.

As discussed above, the first key ingredient in all of the research described here will be the researcher's empirically grounded knowledge of how the policy of interest X affects some auxiliary outcome(s) Y. How can such knowledge be obtained? Thankfully, this problem is extremely well studied in the field of econometrics.[2] A key starting point is the researcher's belief that the instrumental variable Z satisfies an exogeneity restriction, one version of which amounts to the belief that the variation across units in this variable is as good as random. In some settings, this belief is easily justified. A good example of such a scenario would be when Z is literally a random variable, as with treatment assignment in a randomized controlled trial, where Z is a measure of whether a unit of observation in the trial received the trial's treatment or not. In other cases, the Z variable may draw on certain quasi-experimental features the researcher has isolated in a natural experiment. More

[2] See, for example, the textbook treatment in Angrist and Pischke (2009), and Matzkin (2013) and Chesher and Rosen (2020) for surveys of recent advances.

generally, the characteristics of the Z variable are such that the researcher is comfortable with an assumption of exogeneity.

When the instrument Z is exogenous, the researcher can faithfully "identify" (and hence, with a large sample, hope to arrive at an accurate estimate of) the magnitude of two causal relationships: how Z affects the policy variable X and how it affects the auxiliary outcome variable Y. Clearly, this information will be insufficient, in general, to answer the researcher's original question. There are two obvious problems: the desired outcome W is not (yet) a known function of Y, and Y is not (yet) a known function of X.

Nature's Bounty

Before continuing with the general case, we pause to discuss an idealized—though not uncommon—scenario. Suppose, first, that the parameter θ is known to the researcher. This amounts to saying that the outcome of interest W is a known transformation of the auxiliary outcome vector Y (often simply because $W = Y$, or because W is a known aggregation of individual-level values of Y). Second, the as-good-as-random instrument Z is similarly a known transformation of the policy variable X that features in the researcher's question (again, often because $X = Z$). Clearly, relative to the general case we started with, this researcher is in an extremely fortunate situation. But through painstaking effort and tons of ingenuity, some researchers have found themselves in exactly such a position, as the following example illustrates.

> *Example #1: Driven to Dhaka.* Rural laborers in low-income countries often face a choice between paying to migrate (perhaps seasonally) to a large city or working for an inferior wage on local farms. But how responsive are migration choices to changes in migration costs? Would a widespread reduction in travel costs cause sufficient migration that even local wages for those workers who stay behind might increase? Bryan, Chowdhury, and Mobarak (2014) randomly subsidized travel to a major city among a sample of rural households in Bangladesh in order to examine these questions. Akram, Chowdhury, and Mobarak (2017) followed up with a larger version that randomly varied the number of subsidized households per village. In both cases the policy of interest X (travel costs) was explicitly randomized (so in this setting, $X = Z$), and the outcomes of interest W (migration rates and village-level wages) were observed, so the effects of lower travel costs on migration and village-level wages were apparent. They were also surprising. Migration responses were enormous (and persisted even years after the one-off subsidy was gone) and wide roll-out within a village did raise local wages (despite the high density of nearby, untreated villages).

Nature's Instrumental Bounty

We now continue with a slightly less idealized setting. Continue to imagine that θ is known—so the researcher knows how to map from an observed auxiliary outcome Y to the object of interest, W, perhaps simply because Y is in fact the object of interest. But now we retreat from the happy scenario in which the instrument Z is a transformed version of the policy of interest X. Effectively, the policy of interest X is no longer as good as randomly allocated in the researcher's dataset.

While challenging, this setting is familiar for economists. As discussed above, the effects of Z on Y and of Z on X are known, thanks to the exogeneity restriction. One more assumption—the so-called exclusion restriction—allows researchers to combine these two effects into knowledge of the desired effect, which is how X affects Y. The exclusion restriction requires that all of the effect of Z on Y happens because of the fact that Z affects X, rather than a potentially distinct (but ruled out, by assumption) effect of Z on Y even as X remains unchanged. It is conceptually distinct from the process that determines Z (which may underpin the researcher's belief in the exogeneity assumption), so it needs to be assessed on its own merits. Still, this assumption is plausible in many settings, and so the exogenous and excludable variation in instrumental variables plays an essential role in all of the work discussed below. The next example provides a setting in which it continues to be the case that $W = Y$, but no longer the case that X is the same as Z.

> *Example #2: Sealing the Suez.* How much would the GDP of a typical country be harmed if it were less open to trade? During the 1967–1975 Suez Canal blockade, caused by regional conflict, some shipping routes, such as Tokyo-Amsterdam, had to be redrawn while others, such as Tokyo–Los Angeles, were unperturbed. Feyrer (2021) argued that the resulting variation in the exposure of countries to the increase in shipping distances caused by the blockade could be used (as his instrumental variable Z) in order to estimate the effect of the blockade on the policy variable of trade flows (X) and on the outcome variable, GDP (in this case the auxiliary observable Y is the same as the outcome of interest W). Putting the two together implied that, for a typical country (among those affected by the blockade), when its level of trade openness (imports and exports as a share of GDP) fell by 10 percentage points, its real GDP per capita fell by about 5 percent.

Examples such as this and the previous one offer a compelling set of answers for the research questions posed, and these answers draw less on explicit theory than the work in the remainder of this article. Before going on, however, it bears stressing that it would be wrong to imagine that economic theory plays no role at all in studies like those discussed so far. To the contrary, researchers draw on theory when designing randomized trials, when justifying their belief in the exclusion restriction of an instrument Z being used, and even when making basic decisions such as which research questions to ask (which X's and W's to investigate among the infinite set of options) in the first place. In addition, there is often a desire to extrapolate beyond the lessons from any given study and thereby hope that the estimates from one setting are "externally valid"—generalizable beyond the setting at hand—and theory provides an essential guide for doing so.

Surrogacy

Although researchers do sometimes find themselves in the fortunate position described in the previous two examples, most of the time they do not. The remainder of this article considers scenarios that feature such challenges. We continue to imagine that the researcher is using data and a valid instrument Z in order to establish the effect of the policy variable X on the auxiliary output variable(s) Y. But at this point, the researcher has gone as far as possible toward answering the basic research question without bringing in the model's theoretical assumptions. To bridge the gap between the observed auxiliary variable Y and the unobserved object of interest W, the researcher has no choice but to lean on the additional assumptions encoded in the function $W = g(Y, \theta)$. This abstract scenario exemplifies the inherent complementarity between theory and data. Without theory, the researcher would not be able to move from the auxiliary variable Y to the object of interest W. But without the quasi-experimental variation in the data, the theoretical assumptions needed would be far more ornate, time consuming, and subject to doubt if the researcher had to rely on theory, rather than data, for the empirical knowledge about the effect of X on Y.

To move from the observed Y to the unobserved object of interest W, even with the help of theory in the form of $W = g(Y, \theta)$, the researcher must know which values of the parameters in the vector θ to use in the mapping. We will now imagine that the researcher will draw on some additional data, labeled D, in order to arrive at an estimate of θ. Even though the details of this step are important, they vary across settings in ways that don't matter for the discussion here, so we shall summarize this estimation process as $D = \theta$. This implies that θ is known, thanks to the data elements embodied in D.

Summarizing the discussion so far, the researcher's model is a theoretical device for extrapolating (with the help of additional data, D) from the effect of X on Y, which is observed, to the effect of interest, of X on W, which is not observed. A simple illustration of this theory-as-extrapolation logic draws on what is referred to as a "surrogate" method in the field of statistics. Here is a classic example from the medical literature. A researcher is investigating whether a given cancer treatment drug (X in this context) improves patients' survival chances (W in this context). The researcher has enough experimental control to vary access to the drug X across members of the sample in an exogenous manner. However, measuring survival rates W is often impractical. For example, doing so may require waiting too long, or it may be the case that in-sample observations of survival rates W are just too noisy for researchers to hope to say anything conclusive about how the drug affects survival outcomes, given the sample size. Thus, the medical researcher's question cannot be answered without the help of theory and model-informed estimation.

In this case, the researcher's theory comes in the form of a model of physiology in which the mapping $W = g(Y, \theta)$, from various observable biomarkers or "surrogate outcomes" Y to the survival rate W, is already well studied, to the point where $g(\cdot)$ and θ are known. Crucially, the biomarkers Y are chosen because they are far easier to observe than the survival outcomes W. So researchers use the "surrogacy assumption"—that is, the belief that their knowledge of $g(Y, \theta)$ and θ is correct—to use the inexpensive measurements of biomarkers Y to map from the policy variable X to the object of interest W. This research strategy effectively splits the job of empirical estimation into the two parts noted earlier: estimating the effect of the drug on the biomarkers (the effect of policy variable X on auxiliary outcome variable Y) and modeling the quantitative relationship between biomarkers and survival rates. In practice, this second part may simply involve estimating a linear relationship between the object of interest W and the auxiliary variables Y (in limited but vital settings where measurement of the object of interest W is feasible) but the principle generalizes to any potential mapping $g(Y, \theta)$.

While this may sound like an idealized setting found only in clinical medical trials, many studies in spatial economics share similar elements. The following is an example.

> *Example #3: Engel's Law meets Indian Trade Liberalization.* What effect did India's 1991 tariff liberalization have on the real income of households in regions that were specialized in sectors most affected by tariff reductions relative to households in regions that were not? Real income (W) is hard to measure in the absence of high-quality price data covering all consumption, a particular problem in this context, especially given the changes in product quality and variety that are emphasized as important mechanisms underpinning the gains from trade. To overcome this challenge, Atkin et al. (2020) describe primitive assumptions under which any cross-section of utility-maximizing households will obey an Engel's Law–like relationship: that is, as real household income increases, the share of income spent on, say, meat as a share of food expenditure declines in a stable and monotonic manner that is invariant to relative prices in non-food sectors of the economy (at least among those in which price measurement is difficult). These assumptions can then be invoked as a form of surrogacy assumption. The inverse of the estimated Engel-like curve relates the hard-to-measure desired outcome (real income, W) to the easy-to-measure surrogate (meat expenditure shares within food, Y)—and this estimated relationship thereby populates the parameters θ in $W = g(Y, \theta)$. Atkin et al. (2020) go on to exploit plausibly exogenous variation in the exposure of Indian regions to tariff reductions (Z), as previously developed by Topalova (2010). This method exploits an interaction between predetermined regional specialization across sectors and the Indian government's desire to homogenize variation in tariffs (as well as reduce the overall level), which meant that sectors with initially high tariffs had farther to fall in the 1991 liberalization cuts. Armed with Topalova's instrument, one can arrive at estimates of the effect of tariff exposure (X) on food budget shares (Y) and then use

the estimated Engel-based surrogacy relationship to estimate the effects on real income. In this way, Atkin et al. (2020) estimate large negative effects of the reduced import tariffs on rural households, evenly spread throughout both rich and poor rural households. In interpreting these results, the authors are careful to stress that relative effects across regions, not the overall effect on living standards in India as a whole, are the object of interest. The exogenous variation is cross-regional, so it cannot speak to the nationwide level effect.

Surrogacy-like assumptions often provoke skepticism in both the medical and economics literatures. But they often come with the ability for testing in special settings where W (and Y and X) are observed, because the implication of the surrogacy assumption is that X should have no effect on the difference between W and $g(Y, \theta)$.[3] In addition, the primitive economic assumptions that are invoked in the model $g(Y, \theta)$ may have additional predictions that can be tested.

More Challenging Extrapolation

In the classical surrogacy case, the researcher's model function $W = g(Y, \theta)$ is linear. Example #3 stressed a more involved case, but one that rested on the intuitive economic logic of Engel curves. In wider economics applications, the model function $g(\cdot)$ is often considerably more complicated. For example, the function $g(\cdot)$ could represent the solution to a large system of nonlinear equations that describes the general equilibrium of a competitive economy or the Nash equilibrium of a game-theoretic model of interactions between firms. It could even represent the result of a search over a set of feasible economic policies, where evaluating the merits of each candidate policy involves solving for the equilibrium that would be believed to prevail in an economy as a result of enacting the policy.

Whether $g(\cdot)$ is simple or complex, there is still substantial value in drawing a distinction between the two ingredients that the researcher will learn from the data: the effect of changes in the policy variable X on some auxiliary outcome Y, and the parameters θ that enter the model's mapping $g(\cdot)$. These two ingredients do not necessarily have the same provenance. By definition, θ does not connect neatly to some estimable effect of policy X in the researcher's own study—just like in the surrogates case, where θ must be drawn from a wider body of knowledge outside of the study at hand. The following example illustrates the power of extrapolation from estimated policy effects on auxiliary outcomes to a desired goal that involves feeding those estimated effect sizes into a more complex, equilibrium model.

Example #4: Trump's terms-of-trade war. How much would aggregate US real income change from levying import tariffs? How would matters differ if

[3] Athey et al. (2019) develop tools that weaken the assumptions behind surrogate methods, as well as those that allow a researcher to calculate bounds on the potential for bias due to, and to test for, violations of the surrogacy assumption.

foreign countries were to retaliate with their own tariff hikes? Fajgelbaum et al. (2020) study the tariff changes stemming from the 2018 trade war, instigated by the Trump administration, to answer these research questions. The researchers estimate effects of plausibly exogenous variation in US and foreign tariffs (so $X = Z$ here) on certain features of four key auxiliary outcomes (Y): prices and quantities of narrowly defined products coming into the US from tariff-hit countries relative to those that were spared; and similar prices and quantities for products leaving the United States for retaliating countries relative to others. These comparisons conveyed the striking finding that, despite the relatively large size of the United States in many global markets, tariff increases were immediately passed through into import prices, with large commensurate reductions in quantities crossing borders. While these results illustrate micro-level patterns of US and foreign supply and demand, an aggregate, general equilibrium model of entire US production and consumption $g(Y, \theta)$ is needed to answer the researchers' question about aggregate real income (W). To build such a model, Fajgelbaum et al. (2020) propose that US production is competitive and that production functions and inter-sectoral preference functions take the Cobb-Douglas form. Importantly, this model features producers who benefit from protective tariffs, producers who suffer from a rise in the price of imported materials, and consumers who both suffer from a rise in consumer prices and gain from an increase in tax revenue. Together, the model's assumptions imply both how θ can be pinned down by available data (D) as well as the mapping $g(Y, \theta)$ from the auxiliary outcomes Y to real income (W). Ultimately, the researchers' empirical model implies that the average US resident lost \$22 of real income due to the tariffs (but these losses would have instead been gains, albeit very small ones of about \$1, in a hypothetical scenario without foreign retaliation).

Sufficient Statistics

The discussion so far has emphasized the unavoidable need, when many questions of interest are concerned, to use theory embodied in the $W = g(Y, \theta)$ function to extrapolate from empirical knowledge of how the policy variable X affects the auxiliary variable Y to the question of interest—namely, how that same policy variable X affects the outcome W. Our image of theory as extrapolation raises the question: how "far" are we extrapolating?

One interpretation of "distance" relates to the "narrowness" of the space of reasonable economic assumptions under which $W = g(Y, \theta)$ is the correct—or equivalently, to the "width" of the space of reasonable assumptions under which this is the incorrect—model to use to answer the question at hand. Economists will have different perceptions of the magnitudes of these distances. Recall, however, that $g(\cdot)$ is not a conventional theoretical model, but an empirical model. That is, its content changes when values of the auxiliary variable Y are observed and when the unknown parameters θ are pinned down by data. Thus, any assessment of the breadth of assumptions invoked by $g(\cdot)$ must be done while holding Y and the available data on other parameters θ fixed.

This distinction matters in practice. It is often the case that a researcher will consider using two different models that disagree on many things. However, the researcher may discover that the two models actually agree on what matters—that is, on their answers to the researcher's question of how changes in the policy variable X will affect the object of interest W—once we condition on features of the available data. Such features could derive from the estimated impact of the policy variable X on the auxiliary variable Y, and they could also derive from the values of the data that inform model parameters. Heckman (2010, p. 359) refers to this observation as "Marschak's maxim" in honor of Jacob Marschak (1953), who pioneered the understanding of situations in which low-dimensional combinations of model elements could suffice for answering a given policy question.

Another way of expressing this scenario is to say that, across the elements of some set of models, the evidence in the data acts as a "sufficient statistic" (or vector of statistics).[4] Conditioning on the available data is not just sufficient for filling in unknown elements of the model, in the usual sense of parameter estimation regarding the model's only unknown, θ. It may also be sufficient for eliminating elements of disagreement between two more plausible (but meaningfully distinct) models.

The endeavor to isolate sufficient statistics will depend on the question of interest. Asking models to agree on their answer to every question, even when we condition on available data, is a tall order. But asking models to agree when they are being used to answer a specific question is far more common and feasible. The following example illustrates the powerful logic of sufficient statistics in a spatial context.

> *Example #5: Million dollar or billion dollar plants?* When local governments offer subsidies and other incentives to attract large businesses, are their residents better off? Greenstone and Moretti (2003) describe a class of models in which workers are mobile and have identical preferences, local land is in fixed supply, other factors (such as capital) are mobile, and land markets are competitive. While the set of assumptions that defines this class is restrictive, it is far less restrictive than models that would go on to place additional restrictions on, or seek to estimate, the precise forms of firms' technologies (such as how those firms use mobile and immobile factors) and consumers' preferences (such as how consumers value the outputs made by firms and the public goods provided by local governments). Greenstone and Moretti (2003) then show that, within this class of models, paying a subsidy (X) to attract a business will impact local

[4] The application of sufficient statistics in this fashion has many parallels in other fields of economics. See Chetty (2009) and Kleven (2021) for reviews.

> residents' welfare (W) by an amount that is equal to the observed change in land values in the location (Y)—that is, within this class, the auxiliary outcome Y is a sufficient statistic for W. Notably, this finding holds true despite the researchers' ignorance about the myriad complexities arising from general equilibrium product and factor market interactions (in this location and all others), local and wider externalities in production and amenities, and the gory details of how the subsidy is financed out of local funds (which may hence change tax rates and/or public service delivery) or supra-local sources. The intuition behind this result is that when one local factor (land, here) is in fixed supply and competitively exchanged, and yet all other factors are supplied perfectly elastically to a location, then the economic incidence of all location-specific phenomena (wages, prices, productivity, taxes, transfers, and others) would accrue to the fixed factor and be measurable via the observed change in its price. Based on this argument, and a plausibly exogenous source of variation in whether US locations narrowly win bids for a "million dollar" industrial plant (their instrument Z), the authors find that a typical winning location saw an increase in property values of at least $2.7 billion (in 2021 dollars), or about $11,000 per resident, within six years.

As compelling as this example is for answering the question of interest, it also serves to highlight the question-dependent nature of the sufficient statistics deployed. For example, it is harder to imagine how the estimates could be used to study the extent to which subsidies in one location are a zero-sum (or worse) game at regional or national levels, a topic of substantial policy interest (Slattery and Zidar 2020).

Necessary Statistics?

Once we identify settings where a class of plausible models agree—after conditioning on certain potential sufficient statistics—about the question at hand, the researcher has a stark choice to make. One option is to strengthen various theoretical assumptions so as to rule out models until only one model remains, and use that model alone. The alternative is for the researcher to find data on the sufficient statistic variables and make the model discrepancy go away. Such data will not always be available to researchers. But when it is, more and more research areas are transitioning to a view that the use of such data is no longer merely sufficient, but could also be considered necessary.

One example of this logic at work occurs in settings where the outcome of interest W corresponds to the value of the objective function of some decision-making agent who is believed to be optimizing (possibly subject to a constraint). This agent could be a consumer or a firm—and the next section discusses cases in which this agent may even correspond to the hypothetical representative agent of an entire economy. As economists know well (by the so-called envelope theorem of optimization theory), when an (optimizing) agent faces an exogenous change in its environment, the first-order effect of that change on the value of its constrained optimization problem (W, here) is given by the direct effect of the change, because any indirect effects due to the agent changing its behavior are zero to first order. Crucially, this argument holds irrespective of the objective function. Thus, it can be applied even when the objective function that gives rise to W is not completely known, a natural scenario given our starting point that W is unobserved. The only knowledge required is the size of the direct effect of the change.

For the special case in which the change under consideration is to a set of prices faced by a consumer (known as Shephard's lemma), this result implies that the first-order proportional change in welfare is simply the product of any proportional price changes and the pre-change expenditure shares on the goods whose prices have changed. Thus, when the question of interest refers to a case in which the object of interest W is consumer welfare, a researcher can split up the analysis into two parts. First, the researcher could estimate the impact of the observed policy variable X on consumer prices Y. Second, the researcher could infer (to a first-order approximation) the effect of price changes on consumer welfare W with the help of data on all relevant initial expenditure shares. Formally, this approach would be invoking the assumption that (or choosing the model in which) the consumer under study is optimizing, and so as a result the effect of changes in consumer prices Y on consumer welfare W is fully revealed by the data on expenditure shares. This sufficient statistic result is useful because the space of reasonable models in which a consumer is just optimizing is extremely "wide" relative to the nested set of models in which the consumer is not just optimizing, but optimizing some particular utility function. The following example illustrates this idea at work.

> Example #6: *Pain and gain from tourists in Spain.* Who is helped and who is harmed when a location begins to export more? Allen et al. (2021) examine the recent doubling of tourist visits to Barcelona. They consider a class of models in which residents of any of the city's neighborhoods optimize a homothetic (but otherwise arbitrary) utility function over both their mix of consumption (including housing) goods from every neighborhood and their earnings from supplying labor to any neighborhood. Using data from Spain's largest consumer bank, the researchers observe data on individuals' budget and earnings shares for each of these options. These researchers therefore split up their analysis of how any individual's welfare (W) would be affected by a rise in (say) American tourists (X) into two components. First, they use plausibly exogenous variation in the timing and neighborhood concentration of certain tourists (Z) to estimate the effect of the change in tourism on prices and wages (Y) in each location. Second, they apply the insights above to argue that the effect of a given set of changes in wages and prices (Y) in any location on individual welfare (W) is a function of that individual's budget shares on each price and earnings shares on each type of income (D). These estimated effects imply that the tourism boom caused average welfare to rise for those in

peripheral city locations and to fall for those in the city center (where most tourism occurs).

Unnecessary Statistics

The discussion so far has imagined a researcher who wishes to answer an explicit counterfactual question using (because it is the only option) an explicit model. Further, the researcher has sought to leave as many of the details of that model as possible to be filled in by data features that can be conditioned upon.

One benefit of thinking this way is, as described above, the ability to minimize the extent to which the researcher's answers are driven by underlying assumptions. Another benefit is that the researcher may discover that the data requirements are actually simpler (and hence easier to collect) than may have first been apparent. Formally, this corresponds to a setting where the data requirement is a set of observable statistics that is actually a known combination of other observables. The most obvious version of this is where the data (on either Y or D) is a "macro-level" variable that is an aggregation over more micro-level statistics, as will arise when the micro-level statistics enter linearly and with uniform weight. This means that the long vector of micro data includes a set of *unnecessary statistics*, once we condition on observing the shorter vector of macro data. The following study describes an example where this logic applies.

Example #7: *Gains from trade in a gravity world*. How much does a country gain from the trading it does with the wider world? Arkolakis, Costinot, and Rodríguez-Clare (2012) consider a class of models in which consumers have constant-elasticity of substitution preferences, firms have heterogeneous but constant marginal costs of selling to any country, firms use one factor that is in fixed supply to each location, and firms compete either perfectly competitively or monopolistically competitively (with, in this latter case, fixed costs of developing a differentiated good and entering any market). In such an environment the welfare (W) cost of autarky (for example, by erecting prohibitive tariffs X) could range from zero to infinite depending on the heterogeneity in marginal and fixed costs. However, these researchers derive a surprising sufficient statistic result about a commonly used subset of models in this class known as "gravity" models—those that may differ in many underlying details but nevertheless display a constant and homogenous "trade elasticity," which is defined as the proportional change in a country's relative imports (which we could think of as an auxiliary outcome Y) from any two origins due to a proportional change in the relative tariff levied on those two origins (X). In particular, Arkolakis et al. (2012) show that the welfare cost of autarky for a given "Home" country is a function of just two statistics: the value of the trade elasticity and the current share of imports in Home's total consumption. These are both aggregate statistics, which implies that underlying micro data, such as that on the sets of firms, products, and/or countries inside Home's aggregate import share, are unnecessary statistics for the question at hand and within the class of models considered. The same is true for the response of relative imports to relative tariffs—it is the aggregate value of imports Y that matters for learning the trade elasticity. As reported in Costinot and Rodríguez-Clare (2018), under these assumptions, for a country like the United States, the welfare cost of moving to autarky in 2011 is found to be 1.5 percent. This relatively low number arises both because the United States imports relatively little and the trade elasticity is thought to be relatively high.

Sufficient Functions

The language so far has stressed cases in which the sufficient statistic is either a single statistic or a vector of statistics. But nothing in the logic rules out cases where the sufficient statistic is actually an infinite-dimensional statistic—a sufficient *function*—that a researcher could hope to estimate (nonparametrically) in order to feed into the answer of the basic research question. At a high level of abstraction, this observation is trivial, because clearly the function $g(\cdot)$ is a sufficient function for answering the researcher's question. But a more common way for an economist to visualize the model is as a collection of functions—for example, the supply and demand systems for all firms and consumers in an economy. In this respect, the promise of a useful sufficient function is one that aggregates over (or otherwise combines) some or all of the many micro-level functions inside a researcher's model to arrive at the lowest-dimensional system that is needed to answer the researcher's question. Such a scenario implies not only the usual benefits of sufficient statistics—the ability to use data to avoid theoretical debate about the appropriate model to use within some wider class—but it can also serve as a guide to researchers about the minimal set of functions that are required to be learned from that data for the purposes of the goal at hand. The following is an example of such a case.

Example #8: *Gains from trade without gravity*. Let us return to the question from the previous example: How much does a country gain from trading with the wider world? Adao, Costinot, and Donaldson (2017) consider a class of models with arbitrary preferences and arbitrary technologies used under competitive conditions. Even though countries trade goods in these models (and in the real world), for every model in this class, and for the purposes of answering questions such as the one posed here, the model is isomorphic to one in which countries instead merely trade the services of their (geographically immobile) factors. Thus, any country has a set of well-behaved but "reduced" preferences over the factor services (rather than the goods) on offer around the world. Such reduced preferences for as-if factor service exchange, if known, can therefore summarize the underlying preferences and technologies for the goods in the world and hence provide the inputs for welfare analysis. The underlying logic builds on that in Example #5: in general equilibrium, immobile factors are the objects onto which the total effects of other local phenomena accrue under competitive conditions. Putting this into practice, to the extent that there are fewer factors than goods, the summary offered by reduced preferences is dimension-reducing—and in the context of commonly used modeling environments with thousands (or even a continuum) of goods, the empirical dimension-reduction involved can be substantial. Adao, Costinot, and Donaldson (2017) use variation in transport costs (Z) to estimate reduced factor demand functions (relating factor service flows Y to trade cost shifters X). Based on such estimates, Costinot and Rodríguez-Clare (2018) calculate that the welfare (W) cost of autarky (a prohibitively high X) for the United States would be 2.3 percent (rather than the 1.5 percent mentioned in the previous example in the context of a gravity model).

Wedges, Welfare, and What-If Questions

The central theme of this article has been the interaction of economic theory and data. On the theory side, one of the most powerful ideas that economics has to offer is embodied in the first and second welfare theorems. These theorems state that, in the absence of market failures (such as externalities and market power), and with access to lump-sum transfers to address distributional concerns, along with some additional (and more technical) assumptions, the laissez-faire market allocation would be optimal.[5] The converse is equally important: in the presence of market failures, or in the absence of lump-sum taxation, market allocations are likely to be sub-optimal. This foundational theoretical result has important implications for the conduct of empirical work. Indeed, these implications resonate with many of the points raised above.

Designing Optimal Policies

Often, the researcher's counterfactual question will not just be "What would be the effect of a particular policy?" but "What is the policy that would be optimal in some well-defined sense?" How can researchers combine theory and data to answer questions such as these?

We shall begin by considering settings in which the object of interest W represents the welfare of an economy's representative agent—or equivalently, where the researcher believes it is plausible that policy could make (something close to) lump-sum transfers to agents as part of the optimal policy scheme. In such a setting, and in

[5] This statement assumes that all agents rationally pursue their best interest and so ignores policy motives deriving from a failure of agents to optimize. While such motives have featured in other branches of economics, they have seen far less focus in the areas I cover here.

the absence of market failures, the welfare theorems tell us that the optimal policy is already known: it is to step aside and let the market do its work. Obviously, in this case, there is no role for data or theoretical modeling to answer the question of interest. But the corollary is interesting: when the goal is to design optimal policies, the role that theory and data play is purely to provide a measure of the magnitude of market failures and of the consequences of real-world limits on lump-sum transfer schemes.

Consider, for example, the case of market failures. The intuition from the welfare theorems implies that optimal policy would align the prices that prevail in the actual economy with the "first-best" prices that would prevail in an economy that is equivalent—that is, an economy that features identical preferences, technologies, and endowments—but in which market failures are absent. Put differently, optimal policy would use taxes and subsidies to offset the wedges that market failures create between prices in the actual economy and those in the first-best equivalent. This framework provides the basis for proposals that call for imposing on polluters a tax equal to the wedge between the private and social cost of the pollutants they produce, or for offering subsidies to innovative producers equal to the wedge between the private and social benefits produced by their research and development efforts.

This implication of the welfare theorems is well known to economists. But it has a stark implication for the direction of empirical work on optimal policy of the sort described in this section: the goal of empirical work in such a context can focus on measuring the locus and magnitude of all relevant wedges and put other matters to the side.

How can such wedge measurement be done? We can generically think of market failures arising whenever the buyer of some "good" pays a different price for that good than the price that the seller receives. In some cases this is relatively easy to quantify, because the wedges are directly the result of taxes, subsidies, or other policies that leave a clear paper trail. For example, a 10 percent sales tax causes a clear distortion because whatever price the seller is charging for the good being exchanged, the consumer pays 10 percent more.

However, many of the wedges that concern spatial economists are not so easily observable. For example, consider the classic case of a factory that causes an externality when it expels pollution into a nearby river. Here, the "good" (technically, a "bad") changing hands is pollution, the "seller" of pollution is the factory, and the "buyer" of pollution is the nearby resident who drinks water from the river downstream of the polluting factory. Further, if the factory pays no penalty and bears none of the cost of its behavior, this pollution seller receives a price of zero when it sells this good. On the other hand, the buyer of the pollution is effectively (and involuntarily) paying a large price for the good because of the health damages caused by drinking polluted water. As before, the essence of this market failure is that the price the selling factory receives (zero) is different from the price that the buying residents are paying (large). But this wedge leaves no simple paper trail. Instead, it hinges on the (monetary equivalent of the) size of the health damages caused per unit of pollution. Nevertheless, the goal of wedge measurement in this case is clear: we need an estimate of the damage function relating health to pollution.

One way to measure wedges in these more challenging cases can be expressed as follows. Let one of the auxiliary outcome variables Y be an observed variable that measures the social benefit or cost of an agent's actions and let X denote an observed measure of the private benefit or cost, to that agent, of those same actions. As above, we imagine that the researcher has an instrument Z that allows estimation of the effect of X on Y. But such an effect is exactly a measure of the ratio of marginal social benefit to marginal social cost, which is the wedge we seek to understand. To take another example, consider the case of the markup (the ratio of price to marginal cost) that a firm with market power would charge. Here, the firm's action is the decision to produce more of its product. The marginal social value of this action, per unit produced, is simply the price it charges to consumers. The marginal private cost, to the firm, of producing is simply the cost of producing an additional unit. An estimate of the markup can be formed by estimating the treatment effect of the firm's production costs (at fixed input prices) on the firm's sales (at fixed output prices), as in Hall (1988).

Doing this for every wedge that seems relevant for the researcher's question is certainly challenging—even daunting. But a researcher can make substantial progress by replacing assumptions about wedges (including of course the assumption that they are all absent) with accurate measurement of wedges. The payoff of wedge estimation is particularly clear in the next example.

Example #9: Tennessee Valley Authority or Hudson Valley Authority? Where should place-based policies and infrastructure investments be optimally placed to maximize national output? Kline and Moretti (2014) evaluate the Depression-era investments (for example, in hydropower generation facilities) that were made in the Tennessee Valley. One clear benefit of such investments is that local firms and households had access to cheaper electricity, and perhaps the Tennessee Valley offered uniquely untapped engineering benefits as a place where new electricity generation capacity could be created relatively cheaply. But a more commonly voiced idea is that relatively underdeveloped areas, such as the TVA region, were places with untapped *economic* potential. Formally, this idea only makes sense if there are local positive externalities of production in the region—which would drive a wedge between private and social values of production and result in inefficiently low levels of output. Indeed, if such spillovers were higher in the Tennessee Valley than, say, in the Hudson Valley near Manhattan, then Tennessee would be a more efficient place to spend national investment funds than the Hudson Valley (at least on the margin). For this reason, Kline and Moretti (2014) devote significant effort to the estimation of the shape of the local spillovers (which will then govern the size of the wedge between social and private values of production at any location in the country); this amounts to estimating a non-linear relationship between local productivity (Y) and local size (X), using features of the TVA program as instruments (Z). Perhaps surprisingly, they find the local spillover function to have approximately the same elasticity in all locations. This means that both small and large locations appear to have the same extent of externalities, and hence wedges, on the margin. It follows that (apart from engineering-related considerations) the Tennessee Valley Authority investments would have generated just as much additional national output wherever in the country they were targeted. The function relating national output (W) to the sizes of locations (X) appears quite flat—so when the question of interest concerns how best to use the TVA to manipulate X so as to maximize W, the answer is that almost any allocation would be equally good.

Continuing our theme of optimal policy design, a distinct motive for market interventions (beyond the attempts to offset market failures discussed above) may arise when the distributional goals underpinning our notion of policy optimality may not be feasible because lump-sum taxes and transfers are thought to be unrealistic. One alternative focal point in the public literature concerns the more plausible scenario in which a government can levy taxes in relation to a household's earnings only—in contrast to the case of lump-sum taxation in which any desired amount could be hypothetically taken from one household and transferred to another without affecting household decisions. Income taxation incurs inefficiencies because the government cannot condition tax liabilities on notions of effort (such as hours worked) or investment (such as time spent training) that households may make in the process of earning their income. In such settings, a government may wish to tax commodities (perhaps via import tariffs or location-specific business support) to achieve redistributional objectives, even in the absence of market failures.

An obvious challenge involved in incorporating such goals into empirical models of optimal policy design is that the analyst needs to know what the government's objectives actually are. For example, what weight should the government attach to the marginal consumption of a household below the poverty line, or to the top 1 percent of income earners? Economists are naturally disinclined to even dare to answer questions such as these. An alternative is to solve for the nature of optimal policies under any given set of conceivable weights, and offer a menu to the government to choose from, but this is usually impractical. However, the following example illustrates one way around this challenge.

Example #10: International trade and the equity-efficiency trade-off. How should import tariffs be designed to achieve redistributional objectives—such as to offset the distributional consequences that Autor, Dorn, and Hanson (2013) argue have resulted from the recent surge of US manufacturing sector imports from China? Costinot and Werning (forthcoming) work with a model in which the country of interest features no market failures and is small enough that it has no reason to impose tariffs in the hopes of improving its terms-of-trade (consistent with the evidence discussed in

Example #4). As such, the only motive for a tariff is that it may provide *pre*-distribution that cannot be achieved via income taxation. Costinot and Werning (forthcoming) also assume that the government's redistributional objective is a function of incomes (rather than other taxpayer identities) and that the observed income tax schedule reflects the government's redistributional objectives. In such a setting, these authors show that the optimal tariff on Chinese imports is a function of four sufficient statistics: the marginal income tax schedule, the income distribution, elasticities of labor supply at each income level, and estimates of the impact of Chinese imports on wages at each quantile of the income distribution. Remarkably, the optimal tariff, when written this way, does not depend on the government's preferences over the distribution of income, because these are already revealed by the observed tax schedule. Nor does it depend on the underlying economic details of exactly why Chinese imports might affect earnings differently across the distribution. To apply this formula, Costinot and Werning (forthcoming) use estimates of income quantile-specific wage (Y) impacts of Chinese imports (X) from Chetverikov, Larsen, and Palmer (2016)—researchers who themselves leveraged Autor, Dorn, and Hanson's (2013) empirical strategy of (analogously to the work in Example #3) comparing regions of America that had relatively greater employment in goods with greater Chinese import growth to regions with lower such exposure (as well as a measure of plausibly exogenous Chinese import growth derived from patterns of Chinese exports to other countries, Z). While the impact of Chinese imports is thought to have differed substantially across the income distribution, the implications of this finding for the redistribution-driven optimal tax on these imports is minimal, as the implied optimal tariff rate is less than a tenth of a percent.

Impacts of Other Shocks in the Presence of Market Failures

Finally, we consider now a researcher whose object of interest W corresponds, as above, to the welfare of a representative agent. But now the research question is not about optimal policy. Instead, we return to the case in which the researcher is studying the welfare effects of a change in some other characteristic X of the economic environment. For example, this X could be a change in the economy's technology, like the installation of new infrastructure.

What does the presence or absence of market failures imply for the researcher's answer to this question? As discussed above, when there are no market failures, and when lump-sum transfers are thought to be feasible, a consequence of the welfare theorems is that the market allocation is maximizing aggregate welfare. As a result, there are no first-order benefits from changing this allocation in response to an exogenous change in the environment. This observation implies—in a result known as Hulten's (1978) theorem, an economy-wide application of the envelope theorem that we mentioned earlier—that the first-order benefits of a shock to X in an efficient economy are given by the vector of "Domar weights" (which are the value of production as a share of GDP) on all activities that are directly affected by *X* in the sense that the productivity-enhancing benefits of the shock occur in such activities. Furthermore, another implication is that even the second-order benefits are given simply by the *changes* in the Domar weights of directly-affected activities. Remarkably, the initial levels of, and endogenous changes in, prices and quantities of every other component of the economy do not need to be known to the researcher because they do not matter (up to second order).

These results may sound straightforward, but they have deep implications for empirical work in settings where researchers believe that distortions are limited. One is that an essential ingredient of any analysis will be the size of the direct productivity changes caused by the shock to *X*. This can be estimated in standard fashion: let *Y* represent the productivity of activities that *X* is plausibly directly affecting, and use the methods described earlier to estimate the effect of *X* on *Y*. Another is that, to the extent that first-order welfare changes suffice, the size of the Domar weights on those directly affected activities will be a set of sufficient statistics for the welfare impact. Finally, to additionally incorporate second-order welfare effects, it suffices to estimate the effects of *X* on an additional auxiliary outcome variable *Y*, namely the *changes* in the Domar weights. The following describes a classic example of this logic:

> *Example #11: Indispensable statistics and railroads.* How large are the economic benefits of massive investments in transportation infrastructure? Robert Fogel's (1964) landmark book, *Railroads and American Economic Growth: Essays in Econometric History*, examined the "axiom of indispensability"—that America's rapid growth in the late nineteenth century would not have happened without railroads. His analysis assumed that the economy was free of market failures and therefore focused on three goals. First, the reduction in the average user cost of transportation that the railroad network contributed relative to pre-existing transport system—this was Fogel's measure of the direct productivity benefits of railroads on the activity of transport, the directly affected activity. Second, the value of transported goods as a share of GDP before the railroad expansion—this was Fogel's measure of the Domar weight on transport. Third, the change in the value of transported goods over the time period in question—this was Fogel's measure of the change in the transport sector's Domar weight, as was necessary for quantifying the second-order welfare benefits. The methods that Fogel deployed did not apply econometric tools in the modern sense of the word. They focused on the total change in the amount of transport, and in the user cost of transport, over the period rather than an attempt to estimate the role of railroads in causing such changes. But Fogel's focus on these three indispensable statistics was clear, and it led to the provocative finding that the rail expansion increased GDP by no more than a few percentage points. As indispensable as the new technology of railroad may have looked to some observers, in an efficient economy railroads could

> not have been very beneficial unless they either drove large changes in user costs (which they probably did not), took place at a time when transport was a large sector in the economy (which it wasn't), or enacted substantial growth in the use of transport (which they probably didn't).

As powerful as Hulten's theorem can be, its logic breaks down in distorted environments. In writing about "Professor Fogel On and Off the Rails," David (1969) focused his criticism on the fact that Fogel's method was reliant on the controversial assumption that market failures were unimportant. In the presence of market failures, a first-order component of how changes in the feature X affect the object of interest W will now hinge on two additional mechanisms. The first concerns the extent to which the shock to X causes reallocations of primary factors toward those activities that have large positive wedges—that is, the activities for which social value exceeds private value. Such reallocations would generate a benefit of X that could not happen in an efficient economy. The second mechanism is the extent to which the shock actually changes the wedges themselves, which can provide additional benefits. Of course, some changes in X might have mixed or negative effects, perhaps mitigating certain market failures but exacerbating others.

This approach implies a feature of first-order welfare analysis that echoes Tolstoy's comment (at the beginning of *Anna Karenina*) that "[h]appy families are all alike; every unhappy family is unhappy in its own way." Here, all efficient economies are alike in their predicted responses of W to X (conditional on a given set of observed Domar weights). But every inefficient economy could see W respond to X in its own way (even after we condition on Domar weights). Predicting first-order welfare effects in undistorted economies hinges only on Domar weights. But doing so in distorted ones requires one to predict counterfactual reallocations, which requires strong modeling assumptions and measurement (of wedges and elasticities of agents' choice functions).

Baqaee and Farhi (2020) clarify and generalize these classical themes. Predicting the effects of counterfactual changes in X will require a full model of the economy of interest—at least any component of the economy in which reallocation could happen and in which wedges exist. For example, in a setting with only one factor of production—say, labor—and in which firms make goods that enter final output only, the extent of misallocation due to market failures in production is easy to see: it hinges on the extent to which some firms have larger value marginal products of labor than others. Further, if a shock of interest X were to have first-order reallocative effects on welfare W, it could do so if and only if it were to cause labor to move toward the firms with higher-value marginal products of labor; indeed, the first-order benefit of such a move is the size of the gap between value marginal products of labor (equal to the firms' relative wedges on output) times the change in labor reallocation that is due to the change in the policy variable X. At least in a simple setting like this one, modeling efforts would do well to focus on measuring pre-existing wedges—as per our previous discussion of optimal policy—and, just as importantly, on understanding how the shock might be expected to cause labor reallocation across activities with different wedges.

Sometimes, the researcher's question concerns the welfare effects of shocks that have already occurred. In this case, the labor allocation is, in principle, an outcome we could observe—it is a Y variable—and we could use such observations to estimate the effect of our shock variable X on this particular Y. Then, the reallocative effects that underpin how the shock to X affects welfare (at least to first order) in this context would be given by simply the product of pre-shock wedges and our estimates of how changes in X affect Y. My final example pursues such an approach.

> *Example #12: Formalizing reallocation in Vietnam.* Can an export demand shock improve allocative efficiency? McCaig and Pavcnik (2018) study the large tariff reductions on US imports from Vietnam that followed from a 2011 bilateral trade agreement. Vietnamese manufacturing industries that saw relatively large reductions in US import tariffs exported more to the United States and expanded employment, and they did so relatively more among the formal-sector firms (as opposed to informal, household enterprises) within industries that could more feasibly overcome exporting hurdles. These reallocations would have no first-order welfare consequences if value marginal products of labor were equalized across and within industries, but one source of (within-industry) non-equalization derives from the fact that formal firms face greater taxation and regulation (and so would be expected to have larger value marginal products of labor). McCaig and Pavcnik (2018) estimate that such productivity difference wedges prior to 2011 were approximately 4 percent. They then quantify the effect of the trade agreement (X) on labor reallocation (Y) and multiply this estimated effect by the labor productivity wedge. The result suggests that aggregate labor productivity (W) rose by about 6 percent as a result of the trade shock.

Concluding Remarks

The article has offered an eclectic journey through some of the ways that recent work in spatial economics has sought to blend theory and data. Combining theory and empirics in this way is hard. Unsurprisingly, attempts to do so have been controversial. Skepticism stands in the way of those who wish to extrapolate from the estimated effects provided by quasi-experimental variation to the counterfactual questions that need to be answered. Yet given the necessity of such extrapolation, it seems vital that researchers understand the data-assumptions frontier in which they invoke only the most plausible theoretical assumptions necessary to map the data they have to the questions at hand, and in which they seek to minimize reliance on modeling assumptions by drawing on data that can resolve model ambiguities to the greatest extent possible. The examples described above, drawn from a much wider body of work in the field, can be seen as pursuing that goal.

At the same time, spatial economists are witnessing a golden age of newly available sources of data. For example, troves of data tracking tax transactions, satellite imagery, smart phones, and credit card use are all being used to reveal spatial flows and linkages in previously unimaginable detail. The opportunities for blending the insights of economic theory with evidence from the spatial world around us have never been richer.

■ *I am grateful to Rodrigo Adao, Treb Allen, David Atkin, Arnaud Costinot, Gilles Duranton, and Enrico Moretti for helpful discussions about the themes of this article, and to the editors, Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams, as well as to Ben Faber, Pablo Fajgelbaum, and Amit Khandelwal, for their comments on an earlier draft.*

References

Acemoglu, Daron. 2010. "Theory, General Equilibrium, and Political Economy in Development Economics." *Journal of Economic Perspectives* 24 (3): 17–32.

Adao, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2017. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade." *American Economic Review* 107 (3): 633–89.

Akram, Agha Ali, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2017. "Effects of Emigration on Rural Labor Markets." NBER Working Paper 23929.

Allen, Treb, Simon Fuchs, Sharat Ganapati, Alberto Graziano, Rocio Madera, and Judit Montoriol-Garriga. 2021. "Urban Welfare: Tourism in Barcelona." Unpublished.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2020. "Transparency in Structural Research." *Journal of Business and Economic Statistics* 38 (4): 711–22.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102 (1): 94–130.

Athey, Susan, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. 2019. "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely." NBER Working Paper 26463.

Atkin, David, Benjamin Faber, Thibault Fally, and Marco Gonzalez-Navarro. 2020. "Measuring Welfare and Inequality with Incomplete Price Information." NBER Working Paper 26890.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

Baqaee, David Rezza, and Emmanuel Farhi. 2020. "Productivity and Misallocation in General Equilibrium." *Quarterly Journal of Economics* 135 (1): 105–63.

Baum-Snow, Nathaniel, and Fernando Ferreira. 2015. "Causal Inference in Urban and Regional Economics." *Handbook of Regional and Urban Economics,* Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 3–68. Amsterdam: Elsevier.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5): 1671–748.

Chesher, Andrew, and Adam M. Rosen. 2020. "Generalized Instrumental Variable Models, Methods, and Applications." *Handbook of Econometrics*, Vol. 7, edited by Steven Durlauf, Lars Hansen, James J. Heckman, and Rosa L. Matzkin, 1–110. Amsterdam: Elsevier.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics* 1: 451–88.

Chetverikov, Denis, Bradley Larsen, and Christopher John Palmer. 2016. "IV Quantile Regression for Group-Level Treatments, with an Application to the Distributional Effects of Trade." *Econometrica* 84 (2): 809–33.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2018. "The US Gains from Trade: Valuation Using the Demand for Foreign Factor Services." *Journal of Economic Perspectives* 32 (2): 3–24.

Costinot, Arnaud, and Ivan Werning. Forthcoming. "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation." *Review of Economic Studies.*

David, Paul A. 1969. "Transport Innovation and Economic Growth: Professor Fogel On and Off the Rails." *Economic History Review* 22 (3): 506–24.

Einav, Liran, and Amy Finkelstein. 2018. "Moral Hazard in Health Insurance: What We Know and How We Know It." *Journal of the European Economic Association* 16 (4): 957–82.

Fajgelbaum, Pablo D., Pinelopi K. Goldberg, Patrick J. Kennedy, and Amit K. Khandelwal. 2020. "The Return to Protectionism." *Quarterly Journal of Economics* 135 (1): 1–55.

Fernandez-Villaverde, Jesus. 2008. "Horizons of Understanding: A Review of Ray Fair's *Estimating How the Macroeconomy Works*." *Journal of Economic Literature* 46 (3): 685–703.

Feyrer, James. 2021. "Distance, Trade, and Income — The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment." *Journal of Development Economics* 153 (C): 1–12.

Finkelstein, Amy, and Nathan Hendren. 2020. "Welfare Analysis Meets Causal Inference." *Journal of Economic Perspectives* 34 (4): 146–67.

Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.

Greenstone, Michael, and Enrico Moretti. 2003. "Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare?" NBER Working Paper 9844.

Hall, Robert E. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96 (5): 921–47.

Hansen, Lars Peter, and James J. Heckman. 1996. "The Empirical Foundations of Calibration." *Journal of Economic Perspectives* 10 (1): 87–104.

Heckman, James J. 2010. "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature* 48 (2): 356–98.

Holmes, Thomas J., and Holger Sieg. 2015. "Structural Estimation in Urban Economics." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 69–114. Amsterdam: Elsevier.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45 (3): 511–18.

Intriligator, Michael D. 1983. "Economic and Econometric Models." In *Handbook of Econometrics*, Vol. 1, edited by Zvi Griliches and Michael D. Intriligator, 181–221. Amsterdam: Elsevier.

Keane, Michael P. 2010. "Structural vs. Atheoretic Approaches to Econometrics." *Journal of Econometrics* 156 (1): 3–20.

Kleven, Henrik J. 2021. "Sufficient Statistics Revisited." *Annual Review of Economics* 13: 515–38.

Kline, Patrick, and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *Quarterly Journal of Economics* 129 (1): 275–331.

Leamer, Edward E. 2012. *The Craft of Economics: Lessons from the Heckscher-Ohlin Framework*. Cambridge, MA: MIT Press.

Low, Hamish, and Costas Meghir. 2017. "The Use of Structural Models in Econometrics." *Journal of Economic Perspectives* 31 (2): 33–58.

Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.

Marschak, Jacob. 1953. "Economic Measurements for Policy and Prediction." In *Studies in Econometric Methods*, edited by William C. Hood and Tjalling C. Koopmans, 1–26. New York: Wiley.

Matzkin, Rosa L. 1986. “Restrictions of Economic Theory in Nonparametric Methods.” *Handbook of Econometrics*, Vol. 4, edited by Robert F. Engle and Daniel L. McFadden, 2523–558. Amsterdam: Elsevier.
Matzkin, Rosa L. 2013. “Nonparametric Identification in Structural Economic Models.” *Annual Review of Economics* (5): 457–86.
McCaig, Brian, and Nina Pavcnik. 2018. “Export Markets and Labor Allocation in a Low-Income Country.” *American Economic Review* 108 (7): 1899–941.
Nakamura, Emi, and Jón Steinsson. 2018. “Identification in Macroeconomics.” *Journal of Economic Perspectives* 32 (3): 59–86.
Nevo, Aviv, and Michael D. Whinston. 2010. “Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference.” *Journal of Economic Perspectives* 24 (2): 69–82.
Paarsch, Harry J., and Han Hong. 2006. *An Introduction to the Structural Econometrics of Auction Data.* Cambridge, MA: The MIT Press.
Reiss, Peter C., and Frank A. Wolak. 2007. “Structural Econometric Modeling: Rationales and Examples from Industrial Organization.” *Handbook of Econometrics,* Vol. 6, edited by James J. Heckman and Edward E. Leamer, 4277–415. Amsterdam: North-Holland.
Rust, John. 2014. “The Limits of Inference with Theory: A Review of Wolpin (2013).” *Journal of Economic Literature* 52 (3): 820–50.
Slattery, Cailin, and Owen Zidar. 2020. “Evaluating State and Local Business Incentives.” *Journal of Economic Perspectives* 34 (2): 90–118.
Timmins, Christopher, and Wolfram Schlenker. 2009. “Reduced-Form Versus Structural Modeling in Environmental and Resource Economics.” *Annual Review of Resource Economics* 1: 351–80.
Topalova, Petia. 2010. “Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India.” *American Economic Journal: Applied Economics* 2 (4): 1–41.
Wolpin, Kenneth I. 2013. *The Limits of Inference without Theory.* Cambridge, MA: The MIT Press.