Putting Quantitative Models to the Test: An Application to Trump's Trade War*

Rodrigo Adão
Chicago Booth

Arnaud Costinot
MIT

Dave Donaldson
MIT

Abstract

The primary motivation behind quantitative modeling in international trade and many other fields is to shed light on the economic consequences of policy changes. To help assess and potentially strengthen the credibility of such quantitative predictions we introduce an IV-based goodness-of-fit measure that provides the basis for testing causal predictions in arbitrary general-equilibrium environments as well as for estimating the average misspecification in these predictions. As an illustration of how to use our IV-based goodness-of-fit measure in practice, we revisit the welfare consequences of Trump's trade war predicted by Fajgelbaum et al. (2020).

*This version: May 2023. Author contacts: rodrigo.adao@chicagobooth.edu, costinot@mit.edu, and ddonald@mit.edu. We are grateful to Nimisha Gupta, Thomas Hierons, Robin Li, John Sturm and Akash Thakkar for outstanding research assistance, and to Isaiah Andrews, Lorenzo Caliendo, Ben Faber, Pablo Fajgelbaum, Cecile Gaubert, Sam Kortum, Eduardo Morales, Ralph Ossa, Esteban Rossi-Hansberg, Bas Sanders, Felix Tintelnot, Dan Trefler, Christian Wolf, and numerous seminar audiences for helpful comments and discussions.

1 Introduction

One of the *raisons d'être* of quantitative modeling in economics is to provide guidance about policy choices by producing counterfactual simulations of how economic conditions may change if a given policy were to be implemented. There is little doubt that the numbers provided by these simulations fill a demand for concrete inputs into important policy discussions, from the economic consequences of Brexit to those of global carbon taxation. There is, however, much more debate about the empirical credibility of these simulations, with standard concerns ranging from unrealistic assumptions to a general lack of transparency, as discussed in Dawkins et al. (2001).

The goal of this paper is to help assess and potentially strengthen the empirical credibility of the predictions derived from quantitative models in international trade and other related fields. To do so, we introduce an instrumental variable (IV)-based goodness-of-fit measure that provides the basis for testing causal predictions in arbitrary general-equilibrium environments as well as for estimating the average misspecification in these predictions. Following the aphorism that "all models are wrong, but some are useful," this measure is not designed to evaluate whether a quantitative model is "right" or "wrong" but whether it is "useful" in the sense of accurately answering some counterfactual question of interest. As an illustration of how to use our goodness-of-fit measure in practice, we revisit the welfare consequences of Trump's trade war predicted by Fajgelbaum, Goldberg, Kennedy and Khandelwal (2020) (henceforth FGKK).[1]

The starting point of our IV-based test is the same as that of pioneering tests of quantitative trade models due to Kehoe et al. (1995), Kehoe (2005), and Kehoe et al. (2017). After a policy change has been implemented, such as the enactment of import tariffs by the Trump administration, one may wish to test a model's quantitative predictions by directly comparing predicted and observed changes for some outcome variables. The key feature of our IV-based test is to recognize that observed changes reflect two distinct forces: (i) the causal impact of the policy change of interest; and (ii) the causal impact of all other shocks that may have occurred contemporaneously. The latter is a nuisance, whereas the former is the answer to the counterfactual question of interest, such as how different would the US economy have been absent the Trump administration's tariffs?

Many existing model fit and validation procedures ignore the previous distinction and instead simply ask whether the researcher's model can forecast future outcome variables or backcast past ones. In so doing such procedures may conclude that a model performs

[1]Like welfare predictions from other influential quantitative models, FGKK's results have been discussed broadly outside academia, e.g. Hiltzik (2019), The Economist (2019), and Council of Economic Advisers (2020).

poorly not because the model misspecifies the causal impact of the policy change under study, but because it misspecifies the causal impact of other shocks or the distribution of these shocks, which counterfactual predictions of interest are agnostic about. In fact, recognizing that much of the variation in the data may derive from the causal impact of other shocks, one may feel compelled to abandon testing altogether and instead to pursue "theory with numbers" by saturating quantitative models with a large enough number of free parameters that they exactly match the data at any point in time. This is the view articulated early on by Shoven and Whalley (1984). It is now the prevalent approach in the fields of international trade and spatial economics more broadly, see e.g. Costinot and Rodríguez-Clare (2014) and Redding and Rossi-Hansberg (2017).

Consistent with standard approaches to counterfactual analysis in the aforementioned fields, the testing procedure that we develop in this paper acknowledges that there may be other shocks beside the policy change of interest; that the exact distribution of these other shocks may be left unspecified by the researcher's model; and that there may be specific realizations of these other shocks such that the researcher's model exactly matches the data. The basic idea is to leverage, in a general equilibrium context, the same type of exclusion restrictions that empirical researchers have previously used to estimate partial equilibrium elasticities, namely that other shocks are independent of either policy changes or observable shifters of such changes. More specifically, our procedure uses the observation that if the causal impact of policy changes in the researcher's model is correct, then the difference between observed and predicted changes should be equal to the causal impact of other shocks. Accordingly, such a difference should be uncorrelated with any instrumental variable (IV) constructed from exogenous policy shifters alone. These are the moment restrictions that we will build our test upon.

Section 2 formalizes our IV-based test. It is similar in spirit to what Cameron and Trivedi (2005) refer to as an M-test, but adapted to our general equilibrium environment.[2] We focus on a situation where the causal impact of interest—in our application, the welfare impact of a change in the tariffs covering a variety of products—can itself be expressed as a linear combination of causal impacts over a subset of outcome variables—in our application, the impact of this set of tariff changes on import prices, export prices, and tariff revenues. We define the goodness of fit (or lack thereof) of the researcher's predictions, when testing based on any given IV, as the covariance between that IV and the difference between observed and predicted changes for these outcome variables of interest. For any IV constructed as a linear combination of exogenous policy shifters, we

[2]M-tests also encompass many common tests of overidentifying restrictions such as J-tests. We discuss the relationship between our IV-based test and J-tests in detail in Section 2.5.

show that the expectation of our goodness-of-fit measure is equal to a weighted sum of the misspecifications in the policy impact on each of these variables. If there is no such misspecification, the expectation of our goodness-of-fit measure must be zero.

How can one test such a moment condition? Intuitively, the higher our goodness-of-fit measure is, the less likely it is that it was generated by a mean-zero distribution, and so the more likely one should be to reject the null that the researcher's causal predictions are correct. To operationalize that idea, however, one must deal with general equilibrium considerations that create linkages between various economic variables of interest and are at the core of the quantitative models whose predictions we wish to test. A given aggregate shock may have heterogeneous effects on the prices and quantities of a large number of goods. Likewise, a single tax change may not only affect the price and quantity of the good directly subject to that tax, but also the prices and quantities of all other goods. These considerations rule out a purely reduced-form approach to the estimation of the overall causal effect of tax changes such as tariffs, as discussed in Goldberg and Pavcnik (2016), as well as create systematic dependence across variables of interest that makes off-the-shelf test statistics unavailable. Starting from the structural decomposition between the causal impact of the policy of interest and other shocks, our second analytical result shows how to compute the asymptotic distribution of our IV-based test statistic as the number of policy shifters is taken to infinity by adapting results from Adao et al. (2019) and Borusyak et al. (2022) about shift-share designs. To make our testing procedure fully compatible with standard empirical practices, we show how to compute the previous distribution both when exclusion restrictions only hold conditional on a set of controls and when prior estimation leads to uncertainty in the structural parameters of the researcher's model.

How should policy shifters be combined into an IV? While our IV-based test can be applied to any valid IV, not all potential IV-based tests have the same economic interpretation or the same statistical power. Our final analytical results provide guidance on how to choose the shares that enter our shift-share IVs. For the purposes of providing economic interpretation, we derive sufficient conditions on the form of misspecification in the researcher's model under which our goodness-of-fit measure, when calculated using appropriate shares, is an unbiased estimator of the average misspecification in the researcher's causal impact of interest. For the purposes of increasing power, we suggest choosing shares that leverage the full general-equilibrium structure of the researcher's model; and whenever estimation occurs prior to testing, we propose to alleviate concerns of mechanical fit, and hence low power, by choosing shares such that estimation moments are less informative, in the sense of Andrews et al. (2020), about our testing moments.

Section 3 explores the properties of IV-based tests through a series of Monte Carlo simulations in which we control the true data generating process. We take the researcher's model to be the quantitative trade model that FGKK developed in order to quantify the impact of Trump's trade war on the US economy. For a sequence of model economies, we randomly draw US import tariffs, foreign import tariffs, as well other structural shocks, both when FGKK's model is correctly specified and when it is not. We then study the causal impact of tariff changes on US welfare, expressed as a linear combination of their impact on export prices, import prices, and tariff revenues.

We divide our simulations into two parts. First, we use our simulations to compare the performance of correlation- and IV-based tests. Unlike the former, we show that the rejection rates of IV-based tests are not affected by the relative importance of non-tariff shocks when the model is correctly specified; and precisely because correlation-based tests are sensitive to the relative importance of non-tariff shocks, we show that the correlation between data and prediction may actually go up when the average welfare misspecification in the researcher's model increases. Second, we compare the performance of alternative IV-based tests. In line with our analytical results, we show that our preferred IV-based test statistic has both a valid economic interpretation—in that its average value is very close to the average welfare bias across policy realizations—and higher statistical power than other "naive" IV-based tests, especially when estimation takes place before testing and leverages the same exogenous policy shifters, as will be the case in our empirical application and, we expect, many others.

Section 4 turns to the consequences of Trump's trade war. We again focus on the predictions of FGKK's model for US welfare and show how IV-based tests can be used as an add-on to their analysis. Instead of generating data for hypothetical shocks, as we did in our simulations, we now feed into our testing procedure the actual changes in the three welfare-relevant outcomes—export prices, import prices, and tariff revenues—as well as the actual changes in US and foreign tariffs over the period 2016-19. In line with FGKK's estimation procedure, we assume that actual tariff changes are independent of other non-tariff shocks and use these changes as the policy shifters that enter our IVs.

Our preferred IV-based test yields a goodness-of-fit value of -0.09 and a p-value of 0.63. The second of these two numbers implies that, under the null that the impact of Trump's trade war on all welfare-relevant variables was correct, one cannot reject FGKK's prediction that "the aggregate real income loss was \$7.2 billion, or 0.04% of GDP" at standard significance levels. The first of these two numbers further implies that, for the sources of model misspecification that our preferred IV accommodates, the welfare loss may be lower on average by 0.09% of GDP, an amount that seems modest in absolute terms. We therefore view FGKK's quantitative model as useful for answering the counterfactual question being posed of it, despite the fact that, as we also document, it can be rejected for a subset of outcomes.

Related Literature

There is a large empirical literature estimating the effects of trade policy. In their review, Goldberg and Pavcnik (2016) contrast structural work based on quantitative trade models whose "estimated effects [...] depend on the assumption of the underlying structural model and the consistency of the estimated behavioral parameters of demand, supply, and implied trade elasticities" and reduced-form work exploiting a quasi-experimental research design, which "depends less on specific functional form assumptions about the underlying demand, production, and market structure" and can be used "to estimate the direct causal effect of actual trade policy on the outcomes of interest" but "is not suited to evaluate welfare implications of actual trade policy changes or the overall effects of trade policy change, both of which require fully specified structural or quantitative models." Examples of reduced-form work estimating the direct causal effect of actual trade policy includes Attanasio et al. (2004) for Colombia; Topalova (2010) for India; McCaig (2011) for Vietnam; and Kovak (2013) for Brazil, among many others. We view IV-based tests as a useful add-on to the existing literature for researchers interested in combining quantitative structural work and reduced-form empirical work. After estimating "direct causal effects" using quasi-experimental variation and simulating using a quantitative model, we advocate testing the "overall [causal] effects" that this model predicts by leveraging the same quasi-experimental variation.[3]

It is common in many areas of economics to test or "validate" models before undertaking counterfactual and welfare analysis.[4] When fully specified, economic models generate distributions over economic variables. To "validate" a model, one may thus compare the distribution that it predicts to the one that is observed in practice. A special case of this general approach consists in selecting and comparing a subset of moments using both model-generated and true data. The RBC literature offers a famous example (e.g.

[3]In FGKK, for instance, the authors estimate the direct causal effect of Trump tariffs on US import prices by comparing, within narrowly defined product categories, the prices of goods from China relative to those from other countries. The "overall [causal] effect" of the Trump tariffs further includes the indirect effect of *any* single US tariff on *all* products from both China and the rest of the world. These indirect effects are too high-dimensional to be estimated directly, hence the need for a general-equilibrium model that puts structure on them, and for testing these general-equilibrium restrictions after estimation.

[4]As mentioned earlier, that practice is much less frequent in the fields of international trade and spatial economics, presumably because of the widespread strategy to saturate quantitative models with enough parameters to exactly match available data.

Heckman and Hansen, 1996). The tests in Kehoe et al. (1995), Kehoe (2005), and Kehoe et al. (2017) fit in this tradition. They fully specify all shocks—for example that the only shocks occurring are the changes in the policy of interest—and then compare observed changes to predicted changes. By restricting shocks in this way, one bundles together two conceptually different questions: (*i*) Is the causal impact of the policy change predicted by the researcher's model accurate? and (*ii*) Is the researcher's model able to forecast the changes in economic variables between the pre- and post-policy period?

As discussed earlier, our IV-based test gives center stage to that distinction. We are interested in the first question, not the second. Answers to the first question depend on the structure of the researcher's model, but not the distribution of other shocks. Answers to the second question depend on both. Our test is designed to test the causal impact of the policy change of interest, while remaining agnostic about other shocks. In contrast, model validation procedures that focus on the R-squared of a regression of observed changes on predicted changes, its root mean square error, or the correlation between these two variables mix up causal analysis and forecasting. That is, their success or failure may derive as much from misspecification in the causal mechanism of interest as from the (un)importance of the policy change of interest in driving the variation in the data (see e.g. Lai and Trefler, 2002, Desmet et al., 2018, Dingel and Tintelnot, 2021).

Among existing validation exercises, our IV-based test is most closely related to papers that first estimate the direct causal impact of a shock—e.g. a monetary shock in Christiano et al. (2005), government spending in Nakamura and Steinsson 2014, the Berlin wall in Ahlfeldt et al. 2015, or foreign shocks in Adao et al. (2020) and Adao et al. (2022)—and then check whether the model can reproduce the same causal impact.[5] This is what Christiano et al. (1999) refer to as the Lucas (1980) program, who argues that economists "need to test them (models) as useful imitations of reality by subjecting them to shocks for which we are fairly certain how actual economies or parts of economies would react. The more dimensions on which the model mimics the answers actual economies give to simple questions, the more we trust its answers to harder questions."[6]

We view our paper as part of the same broad program outlined by Lucas (1980). The

[5]Our IV-based test also relates to tests of assumptions about market conduct in IO (e.g. Bresnahan 1982, Berry and Haile 2014, or more recently, Backus et al. 2021). The logic of such tests is that, given estimates of demand and assumptions about conduct, one can infer firms' marginal costs from observed prices by subtracting the markups that are implied by those demand estimates and conduct assumptions. Provided that there are demand-side IVs assumed to be orthogonal to marginal cost shocks, one can then test whether inferred marginal costs are indeed orthogonal to demand-side IVs. Our test follows a similar logic in a general-equilibrium environment, with the causal impact of the tariff predicted by the researcher's model playing the role of the markup and the causal impact of other shocks playing the role of the marginal cost.

[6]Early expressions of this idea can also be found in urban economics (Wise, 1985) and development economics (Todd and Wolpin, 2006) when RCTs have been used to test structural models.

distinctive feature of our analysis is to point out that starting from a causal question of interest—in our application, what is the welfare impact of Trump's trade war?—not all "dimensions on which the model mimics actual economies" are made equal. By designing specific IVs, one may strengthen the credibility of the overall causal effect predicted by the researcher's model. Although our application focuses on a trade question, we expect our approach to model validation to be useful in any general-equilibrium environment where "answers to harder [causal] questions" cannot be directly estimated from the data.

2 Putting Quantitative Models to the Test

2.1 A Bird's-Eye View of Quantitative Models

A quantitative model imposes restrictions on the behavior of endogenous variables, typically prices and quantities, as a function of exogenous variables, typically preference and productivity shocks as well as various taxes. For ease of exposition, suppose that this quantitative model is static, as is often the case in the trade literature.[7] Then in any given period t, we can describe it compactly as a mapping f such that

$$y_t = f(\tau_t, \epsilon_t), \tag{1}$$

where $y_t \equiv \{y_{n,t}\}$ denotes the vector of all endogenous variables, either quantities or prices; $\tau_t \equiv \{\tau_{k,t}\}$ denotes the vector of policies of interest that are imposed at date t, which in our applications will be import tariffs; and ϵ_t denotes the vector of all other time-varying shocks. Different assumptions about preferences, technology, and market structure lead to different mappings or "reduced-form" f that summarize the general equilibrium effects of policies and other shocks, τ_t and ϵ_t, according to the researcher's model.[8]

To state the obvious, the set of potential mappings f is *very* large. Even if one is only interested in the impact of tax changes, general equilibrium linkages imply that a tax imposed on any given good may affect the prices and quantities of all other goods. Thus,

[7]The general points that we make about testing do not depend on this assumption. Focusing on a static model, however, simplifies notation, and is consistent with our FGKK application. For expositional purposes, and in line with the rest of our analysis, we also ignore issues related to multiplicity of equilibria in which the predictions of a quantitative model may be sets rather than points.

[8]The mapping f is the "reduced-form" of the model in a Cowles Commission sense: it solves for all the endogenous variables as a function of the exogenous variables τ_t and ϵ_t, the same way one can explicitly solve for price and quantity in partial equilibrium as a function of supply and demand shifters—the counterparts of τ_t and ϵ_t—rather than describe them as the implicit solution of supply and demand equations.

direct estimation of $f(\cdot, \epsilon_t)$ requires time series variation and there is little hope of ever getting a sufficiently long series to trace it out non-parametrically.[9] The typical approach to obtain knowledge of f is therefore to start from a considerably lower-dimensional, micro-founded model where consumers maximize utility, typically of the nested CES form, firms maximize their profits, with production functions also typically of the nested CES form, and markets clear.[10]

Since knowledge of f acquired in this way relies at least in part on many a priori assumptions, it is not clear why a given quantitative model would actually be a good approximation to the true data-generating process,

$$y_t = f^*(\tau_t, \epsilon_t^*). \tag{2}$$

The question that we are interested in is whether, despite the fact that f abstracts from many features of reality and invokes strong functional form assumptions, its predictions about the causal impact of policy changes $\Delta x \equiv f(\tau_{t+1}, \epsilon_{t+1}) - f(\tau_t, \epsilon_{t+1})$ on some statistic of interest $W(\Delta x)$ are "close" to the true causal impact of policy changes $\Delta x^* \equiv f^*(\tau_{t+1}, \epsilon_{t+1}^*) - f^*(\tau_t, \epsilon_{t+1}^*)$ for that same statistic $W(\Delta x^*)$.[11] Throughout our analysis, we will focus on the case where the statistic of interest W is a linear combination of the causal impact of policy changes on a subset of outcome variables $n \in \mathcal{N}_W$,

$$W(\Delta x) \equiv \sum_{n \in \mathcal{N}_W} \omega_n \Delta x_n,$$

where the weights $\omega \equiv \{\omega_n\}$ may be arbitrary functions of data in period t. Our application will consider changes in export prices, import prices, and import tariff revenues and set the associated weights equal to expenditure and revenue shares so that $W(\Delta x)$ is equal to the first-order approximation of aggregate welfare in FGKK's model as well as in any other model in which tariffs are the only distortion.

[9]The issue goes beyond estimating flexible non-linear functions. Even if f were assumed to be linear in τ_t, e.g., $f_n(\tau_t, \epsilon_t) \equiv \sum_k \beta_{nk}^\tau \tau_{k,t} + g_n(\epsilon_t)$, general equilibrium linkages imply that one would still need to estimate as many parameters β_{nk}^τ as there are combinations of outcome variables n and policies of interest k, a type of curse of dimensionality.

[10]Having specified the model in this way, the researcher would then obtain knowledge of f by estimating utility and production functions. For expositional purposes, we first present our test while abstracting from estimation and come back to the potential interaction between testing and estimation in Section 2.3.

[11]Instead of defining the causal impact of policy starting from date $t+1$'s equilibrium as we do here, i.e. $\Delta x^* \equiv f^*(\tau_{t+1}, \epsilon_{t+1}^*) - f^*(\tau_t, \epsilon_{t+1}^*)$, one could define it starting from date t's equilibrium, i.e. $\Delta \tilde{x}^* \equiv f^*(\tau_{t+1}, \epsilon_t^*) - f^*(\tau_t, \epsilon_t^*)$. If the model is additively separable between policies and other shocks, as in the case of our FGKK application, the two definitions are equivalent. More generally, they are not and the choice of one definition over the other affects the specific exclusion restrictions that must be invoked for testing, a point we come back to below.

2.2 An IV-Based Measure of Goodness of Fit

As discussed earlier, there are many potential ways to "validate" a model. Our approach to model validation is motivated by the fact that we are ultimately *not* interested in assessing whether f is "close" to f^*—in which case one may use any shock and any variable—but instead we are interested in assessing whether the causal impact of the policy change predicted by the model on a statistic of interest $W(\Delta x)$ is "close" to $W(\Delta x^*)$.

Our goodness-of-fit measure uses as its main input the differences between a model's predicted causal impact, Δx_n, and the change observed in the data, $\Delta y_n \equiv y_{n,t+1} - y_{n,t}$ for all relevant variables $n \in \mathcal{N}_W$. The practical problem with using the raw differences, $\{\Delta y_n - \Delta x_n\}$, as the basis of a goodness-of-fit measure is that even around well-known episodes of policy change, there may be other non-policy shocks occurring whose magnitude may be large and whose distribution may be unknown. Formally, one can always structurally decompose changes observed in the data as

$$\Delta y = [f^*(\tau_{t+1}, \epsilon^*_{t+1}) - f^*(\tau_t, \epsilon^*_{t+1})] + [f^*(\tau_t, \epsilon^*_{t+1}) - f^*(\tau_t, \epsilon^*_t)] = \Delta x^* + \Delta\eta^*, \quad (3)$$

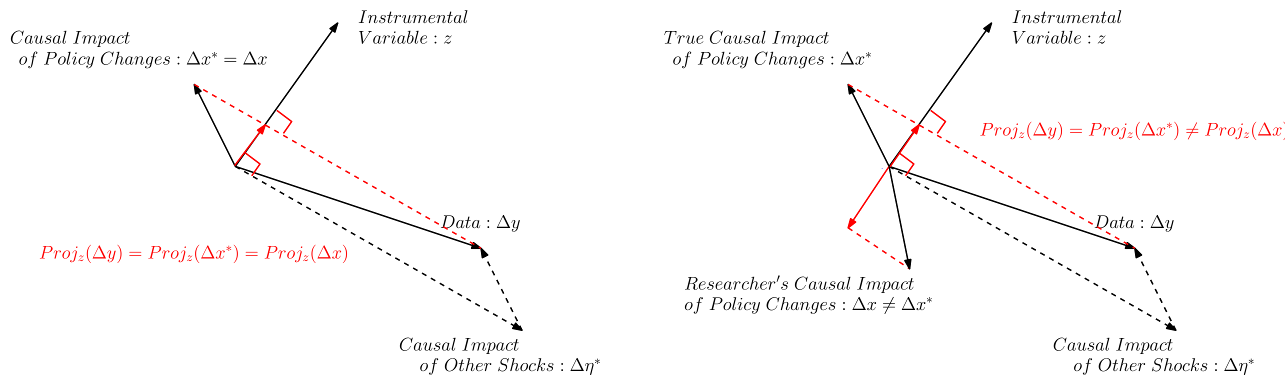
where $\Delta\eta^* \equiv f^*(\tau_t, \epsilon^*_{t+1}) - f^*(\tau_t, \epsilon^*_t)$ denotes the causal impact of all other non-policy shocks. Hence, the difference $\Delta y - \Delta x = (\Delta x^* - \Delta x) + \Delta\eta^*$ may merely reveal the magnitude of non-policy shocks, $\Delta\eta^* \neq 0$, rather than misspecification in the causal impact of the policy predicted by the researcher's model, $\Delta x^* \neq \Delta x$. This is the perspective at the core of the "exact hat" approach commonly adopted in the quantitative trade and spatial literature where observed changes Δy are implicitly used to identify other shocks from ϵ^*_t to ϵ^*_{t+1} under the assumption that the researcher's model is the true model, $f = f^*$.

To deal with the challenges posed by the presence of unobserved non-policy shocks, we propose to use the fact that, even though we may not have strong priors about the magnitude and distribution of such shocks, we may be confident, depending on the particular setting, that they are orthogonal to some exogenous shifters of policy changes, $\Delta\tau_{IV}$. This is the perspective adopted in the empirical literature estimating the "direct causal effects" of tariffs (Goldberg and Pavcnik, 2016). Depending on the context, the shifter may be the change in the tariff itself, $\Delta\tau_{IV} = \Delta\tau$, as in FGKK. Hence, while we cannot directly compare Δx to Δx^*—since Δx^* is unobserved and differs from Δy—we can compare the projections of Δx and Δx^* on an instrumental variable (IV) that is built from exogenous policy shifters $z \equiv \tilde{z}(\Delta\tau_{IV})$—since the latter coincides with the projection of Δy on the IV z under the assumption that z and $\Delta\eta^*$ remain orthogonal. If the researcher's and the true model's predictions about relevant variables coincide (i.e., $\Delta x_n = \Delta x^*_n$ for all $n \in \mathcal{N}_W$), these two projections will be identical, irrespective of how large $\Delta\eta^*$ is. This

Figure 1: An IV-Based Measure of Goodness of fit

(a) No misspecification in causal impact

(b) Misspecification in causal impact



Notes: The figure shows how one can compare the true causal impact of policy changes Δx^* to the causal impact of policy changes in the researcher's model Δx by comparing the projection of Δx on the IV z to the projection of Δy on the same IV z. Panel (a) focuses on the situation where the two causal impacts coincide, $\Delta x^* = \Delta x$, so that $Proj_z(\Delta y) = Proj_z(\Delta x^*) = Proj_z(\Delta x)$. Panel (b) focuses on the situation where the two causal impacts differ, $\Delta x^* \neq \Delta x$, and $Proj_z(\Delta y) = Proj_z(\Delta x^*) \neq Proj_z(\Delta x)$.

scenario is depicted in Panel (a) of Figure 1. If instead the two projections differ, then the researcher's and the true model's predictions must be distinct (i.e., $\Delta x_n \neq \Delta x_n^*$ for some $n \in \mathcal{N}_W$). This is the scenario depicted in Panel (b) of Figure 1.

These observations motivate the following goodness-of-fit measure.

Definition 1. *For any causal prediction Δx and statistic of interest W, we define the goodness of fit of the researcher's prediction along a candidate IV z as*

$$\hat{\beta}_z \equiv \frac{1}{N_W} \sum\nolimits_{n \in \mathcal{N}_W} z_n (\Delta y_n - \Delta x_n), \tag{4}$$

where $N_W \equiv |\mathcal{N}_W|$ denotes the number of observations entering the statistic of interest W.

Throughout the rest of our analysis, we assume that the candidate IV z is a linear function of a vector of $k = 1, ..., K$ exogenous policy shifters, $\Delta \tau_{IV} \equiv \{\Delta \tau_{IV,k}\}$, each with mean zero.[12]

A1. *Conditional on the realization of period t's policy τ_t and other shocks ϵ_t^*, policy shifters are mean zero and independent of other shocks in period $t+1$: $\Delta \tau_{IV} \perp\!\!\!\perp \epsilon_{t+1}^* | \epsilon_t^*, \tau_t$.*

[12] By the shifters being mean zero, we formally mean that $E_t[\Delta \tau_{IV,k}] = 0$ for all k, where $E_t[\cdot]$ denotes expectations conditional on (ϵ_t^*, τ_t). In the case where $\Delta \tau_{IV,k}$ are i.i.d. across k, as we assume below, one can always guarantee that shifters are mean zero by starting from $\Delta \tau_{IV,k}$ and subtracting the sample mean, $\frac{1}{K} \sum_k \Delta \tau_{IV,k}$.

A2. *For any $n \in \mathcal{N}_W$, the instrumental variable takes the form $z_n = \sum_k s_{nk} \Delta\tau_{IV,k}$, where the vector of "shares" $\{s_{nk}\}$ may be a function of, and only of, the realization of period t's policy τ_t and other shocks ϵ_t^*.*

Though some of our results hold more generally, as we point out below, we prefer to focus on shift-share IVs that satisfy A1 and A2 for expositional purposes. These are the IVs that will be the focus of our applications in Sections 3 and 4.[13]

2.3 An IV-Based Test

We propose to use our IV-based measure of goodness of fit to test the null that the causal impact of policy changes predicted by the researcher's model is correct.

A3. *For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \Delta x_n$.*

If A3 does not hold, then there is a subset of relevant variables $n \in \mathcal{N}_W$ and a subset of shock realizations such that $f(\tau_{t+1}, \epsilon_{t+1}) - f(\tau_t, \epsilon_{t+1}) \neq f^*(\tau_{t+1}, \epsilon_{t+1}^*) - f^*(\tau_t, \epsilon_{t+1}^*)$. In such situations, $W(\Delta x) \neq W(\Delta x^*)$, except for the measure-zero set of welfare weights $\{\omega_n\}$ defined by $\sum_{n \in \mathcal{N}_W} \omega_n [f(\tau_{t+1}, \epsilon_{t+1}) - f(\tau_t, \epsilon_{t+1}) - (f^*(\tau_{t+1}, \epsilon_{t+1}^*) - f^*(\tau_t, \epsilon_{t+1}^*))] = 0$. Hence A3 is not just sufficient for the researcher's counterfactual answer to be correct for all shock realizations, but also almost always necessary.

Moment Restriction. Our first proposition focuses on the expected value of our goodness-of-fit measure. The formal proof as well as all subsequent proofs can be found in Appendix A.

Proposition 1. *Take any IV z that satisfies A1 and A2. If A3 holds, then $E_t[\hat{\beta}_z] = 0$.*

A1 and A2 imply that the IV z is orthogonal to other shocks, $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta\eta_n^*] = 0$. Starting from Definition 1 and substituting for Δy_n using (3), we can therefore express the expectation of our goodness-of-fit measure as

$$E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\textstyle\sum_{n \in \mathcal{N}_W} z_n (\Delta x_n^* - \Delta x_n)], \tag{5}$$

[13] As mentioned in footnote 11, we could have considered an alternative structural decomposition of the changes observed in the data, $\Delta y = \Delta\tilde{x}^* + \Delta\tilde{\eta}^*$, where $\Delta\tilde{x}^* \equiv f^*(\tau_{t+1}, \epsilon_t^*) - f^*(\tau_t, \epsilon_t^*)$ and $\Delta\tilde{\eta}^* \equiv f^*(\tau_{t+1}, \epsilon_{t+1}^*) - f^*(\tau_{t+1}, \epsilon_t^*)$. For this decomposition, our results go through if A1 and A2 are adjusted accordingly so that they instead relate to the distribution of $\Delta\tau_{IV}$ conditional on $(\epsilon_t^*, \tau_{t+1})$ rather than (ϵ_t^*, τ_t). Which of the two decompositions to prefer depends on the context in which our test is deployed. In the case of the Trump tariffs studied by FGKK, tariffs were uniformly low in the pre-Trump period. So we find restrictions on $\Delta\tau_{IV} = \Delta\tau$ more plausible conditional on (ϵ_t^*, τ_t) than $(\epsilon_t^*, \tau_{t+1})$. But in the case of India's trade liberalization studied by Topalova (2010), for instance, tariffs were uniformly low post-liberalization. So one may instead invoke restrictions on $\Delta\tau_{IV} = -\tau_t$ conditional on $(\epsilon_t^*, \tau_{t+1})$ rather than (ϵ_t^*, τ_t).

which is a weighted sum of misspecification, $\Delta x_n^* - \Delta x_n$, along all relevant variables. Under A3, $\Delta x_n^* - \Delta x_n = 0$, so that $E_t[\hat{\beta}_z] = 0$.[14] Hence, the expected value of our goodness-of-fit measure $\hat{\beta}_z$ coincides with the degree of misspecification in the counterfactual answer of interest under A3, i.e. $W(\Delta x) - W(\Delta x^*) = 0$. If instead $E_t[\hat{\beta}_z] \neq 0$ for an IV z that satisfies A1 and A2, then A3 must be violated for some $n \in \mathcal{N}_W$ and some shock realizations, implying that the model misspecifies causal responses of interest and, in turn, that $W(\Delta x) \neq W(\Delta x^*)$ for almost all weights ω.[15]

Asymptotic Distribution. For the purposes of testing the null that the causal impact of policy changes predicted by the researcher's model is correct, we move beyond the previous observation and characterize the distribution of $\hat{\beta}_z$ under that hypothesis. We propose to do so asymptotically, as we let the number of policy shifters K go to infinity. In the applications of Sections 3 and 4, K will be the total number of products subject to tariffs, both in the US and the rest of the world, and as K goes to infinity the total number of relevant outcomes associated with the price and quantities of these products, N_W, will go to infinity as well. Because of general equilibrium linkages, neither the changes in the variables of interest Δy_n nor the researcher's prediction Δx_n are i.i.d. across outcomes n. Rather, they are n-specific functions of the same underlying change in the vector of taxes and other shocks. This raises non-trivial questions about whether a law of large numbers can be invoked for the consistency of $\hat{\beta}_z$ and whether a central limit theorem can be used for computing the null distribution, a critical step for testing. Given the shift-share structure of our IV, however, we can extend to our setting the results of Adao et al. (2019) and Borusyak et al. (2022).[16] This leads to the following proposition.

Proposition 2. *Take any IV z that satisfies A1 and A2. If A3 holds and (i) $\Delta\tau_{IV,k}$ are i.i.d. across $k = 1, ..., K$, (ii) $\frac{1}{N_W^2}\sum_k (S_k)^2 \to 0$ with $S_k \equiv \sum_n |s_{nk}|$, and (iii) $Var_t[\Delta\tau_{IV,k}]$ and $\Delta\eta_n^*$ are uniformly bounded, then $\hat{\beta}_z \to_p 0$. Furthermore, if (iv) $\frac{\max_k (S_k)^2}{\sum_k S_k^2} \to 0$, (v) $E_t[(\Delta\tau_{IV,k})^4]$ is*

[14] It should be clear that this first result does not rely on the linearity of the IV. More generally, if we assume non-linear IVs of the form $z_n = g_n(\Delta\tau_{IV}) - E_t[g_n(\Delta\tau_{IV})]$, as in Borusyak and Hull (2022), then we still have $E_t[z_n \Delta\eta_n^*] = 0$ under A1 and A2, leading to $E_t[\hat{\beta}_z] = 0$ under A3, as stated in Proposition 1.

[15] It should be clear, though, that $\Delta x \neq \Delta x^*$ does not necessarily imply $E_t[\hat{\beta}_z] \neq 0$. If $\Delta x - \Delta x^*$ is orthogonal to z, then $E_t[\hat{\beta}_z] = 0$ and our test has no statistical power to detect misspecification in the counterfactual answer of interest. This is true in the special case where the researcher's model and the true model happen to differ exactly by the same constant across all outcome variables, i.e., $\Delta x_n = \Delta x_n^* + \text{constant}$ for all $n \in \mathcal{N}_W$, since z is mean zero under A1 and A2.

[16] Our setting differs from that considered in Borusyak et al. (2022) and Adao et al. (2019) in two ways. First, the estimator of interest $\hat{\beta}_z$ is the covariance between z_n and $(\Delta y_n - \Delta x_n)$, not the OLS coefficient of a regression of $(\Delta y_n - \Delta x_n)$ on z_n. Second, we do not restrict the "shares" s_{nk} that enter $z_n = \sum_k s_{nk}\Delta\tau_{IV,k}$ to be between 0 and 1. Both departures require minor adjustments to the formal arguments in Adao et al. (2019) and Borusyak et al. (2022).

uniformly bounded, and (vi) $\frac{1}{\sum_k S_k^2} Var_t[\sum_{n \in \mathcal{N}_W} z_n \Delta\eta_n^* | \epsilon_{t+1}^*] \to_p \bar{V}_\beta > 0$ *and non-random, then* $r_\beta \hat{\beta}_z \to_d \mathcal{N}(0, \bar{V}_\beta)$ *with* $r_\beta \equiv N_W / \sqrt{\sum_k S_k^2}$.

The first part of Proposition 2 implies that $\hat{\beta}_z$ converges towards its expectation, which is zero under A1-A3 by Proposition 1. Condition (i) implies that while the IV entries z_n are not i.i.d. across n, they can be expressed as (n-specific) linear combinations of i.i.d. variables $\Delta\tau_{IV,k}$, which by A1 and A2 are independent of the structural residual $\Delta\eta_n^*$. Provided that these linear combinations do not all tend to "load on" the same policy shifters, as condition (ii) guarantees, and that standard regularity conditions hold, as stated in condition (iii), the previous assumptions provide the random variation required to establish a law of large numbers theorem, so that $\frac{1}{N_W} \sum_{n \in \mathcal{N}_W} z_n \Delta\eta_n^* \to_p 0$ and, in turn by A3, $\hat{\beta}_z \to_p E_t[\hat{\beta}_z] = 0$.

The second part of Proposition 2 states that $\hat{\beta}_z$, when appropriately scaled, is asymptotically normally distributed. The extra conditions (iv) and (v) are required to invoke the so-called Lyapunov condition. Following Adao et al. (2019), we construct a consistent estimator of the asymptotic variance of $\hat{\beta}_z$ by substituting for the structural residual $\Delta\eta_n^* = \Delta y_n - \Delta x_n^*$ with $\Delta y_n - \Delta x_n$, which must be equal to $\Delta\eta_n^*$ under A3, to obtain

$$\hat{V}[\hat{\beta}_z] = \sum_k (\Delta\tau_{IV,k})^2 [\sum_{n \in \mathcal{N}_W} s_{nk} (\Delta y_n - \Delta x_n)/N_W]^2. \tag{6}$$

Knowledge of the asymptotic distribution of $\hat{\beta}_z$ under the assumptions of Proposition 2 opens up the possibility of testing. Intuitively, the higher the value of $|\hat{\beta}_z| / \sqrt{\hat{V}[\hat{\beta}_z]}$ one observes, the lower is the probability that it was generated by the mean-zero distribution implied by A1-A3. For a given significance level, this may then lead to a rejection of the null that these assumptions hold. We will undertake this test in Sections 3 and 4.

Generalizations. For expositional purposes, we have assumed in A1 that the policy shifter $\Delta\tau_{IV}$ was independent of ϵ_{t+1}^*. As we formally establish in Appendix A.4, one can weaken A1 so that the independence of the policy shifters $\Delta\tau_{IV}$ only applies after controlling for linear determinants of the non-policy shocks, $\Delta\eta_n^*$. Propositions 1 and 2 then continue to hold, once applied to a shift-share IV $(z)_{\text{res}}$ whose shares $\{(s_{nk})_{\text{res}}\}$ have been residualized with respect to these controls. For the same reasons as in Adao et al. (2019) and Borusyak et al. (2022), one can also weaken A1 by assuming the mean-independence of $\Delta\tau_{IV}$ vis-a-vis ϵ_{t+1}^* (again conditional on τ_t and ϵ_t^*). Finally, note that we have assumed in Proposition 2 that $\Delta\tau_{IV,k}$ are i.i.d. across $k = 1, ..., K$. One can instead allow for clustering and only require $\Delta\tau_{IV,k}$ to be independent between groups of observations, but not within, as in Adao et al. (2019). We will do so in our empirical application in Section 4.

Estimation Uncertainty. We have thus far assumed that f was known. So, conditional on the realization of period t's policy and non-policy shocks, τ_t and ϵ_t^*, randomness in the goodness-of-fit measure $\hat{\beta}_z$ only reflects randomness in the policy shifters $\Delta\tau_{IV}$ and the period $t+1$ shocks, τ_{t+1} and ϵ_{t+1}^*. In practice, knowledge of f may itself derive from a prior estimation stage leading to another source of uncertainty. Formally, suppose that the reduced-form of the quantitative model of interest can be decomposed into

$$f(\tau_t, \epsilon_t) \equiv g(\tau_t, \epsilon_t|\theta) \neq g(\tau_t, \epsilon_t|\hat{\theta}),$$

where $\hat{\theta}$ is the researcher's estimator of the true structural parameter θ.[17] The estimation of θ introduces a distinction between $\hat{\beta}_z \equiv \frac{1}{N_W}\sum_{n\in\mathcal{N}_W} z_n(\Delta y_n - \Delta x_n)$, whose asymptotic behavior we have characterized in Proposition 2, and $\hat{\beta}_z(\hat{\theta}) \equiv \frac{1}{N_W}\sum_{n\in\mathcal{N}_W} z_n(\Delta y_n - \Delta x_n(\hat{\theta}))$, which is the goodness-of-fit measure associated with the causal impact of tariffs in the researcher's model when evaluated at the estimated parameter $\hat{\theta}$, with $\Delta x_n(\hat{\theta}) \equiv g(\tau_{t+1}, \epsilon_{t+1}|\hat{\theta}) - g(\tau_t, \epsilon_{t+1}|\hat{\theta})$. In order to implement our test when the researcher's model has been estimated, Appendix A.5 provides extensions of Proposition 2 that characterize the asymptotic distribution of $\hat{\beta}_z(\hat{\theta})$, both when $\hat{\theta}$ is independent of $\hat{\beta}_z(\theta)$ (e.g. when estimation has been conducted on a different sample) and when $\hat{\theta}$ is an IV estimator that might be based on the same policy shifters (which will be the case in our application). This extra step leverages the randomness of policy shifters to extend standard arguments used for the derivation of M-test statistics (e.g. Cameron and Trivedi, 2005 pp. 263-264) to an environment with general equilibrium linkages between outcome variables. We come back to these results when discussing issues of "mechanical fit" in the next subsection.

2.4 Choice of the IV

Testing can be conducted using any goodness-of-fit statistic $\hat{\beta}_z$ that is based on an IV z satisfying the assumptions of Proposition 2. Even for a fixed set of exogenous policy shifters $\Delta\tau_{IV}$, many valid IVs may be available in practice via the design of alternative shares $\{s_{nk}\}$. This section discusses two considerations for choosing those. The first relates to the economic interpretation of $\hat{\beta}_z$ and the second to its statistical power.

[17] The mapping g, in turn, reflects all other assumptions in the model that do not derive from estimation, for instance that some groups of factors and goods are perfect substitutes, that input-output linkages are Cobb-Douglas, or that markets are perfectly or monopolistically competitive.

Economic interpretation. In practice, researchers may not only be interested in whether they can or cannot reject the null that the causal impact of policy changes predicted by their model, $W(\Delta x)$, is correct. They may care, more broadly, about "how far" $W(\Delta x)$ is from the true causal impact, $W(\Delta x^*)$. Motivated by this observation, we now discuss how, starting from a candidate IV $\tilde{z}$, one may construct an alternative IV z such that $\hat{\beta}_z$ is an unbiased estimator of $E_t[W(\Delta x^*) - W(\Delta x)]$.

To do so, we need to relax A3 and take a stand on how the researcher's model may be misspecified. As a starting point, we propose the following generalization of A3.

A3′. *For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \alpha_n \Delta x_n$, with the misspecification parameter α_n a function of, and only of, period t's shocks, (ϵ_t^*, τ_t).*

According to A3′, the researcher's model and the true model may differ in unrestricted ways across various outcome variables, via α_n, but they both agree on the relative importance of various policy changes. Whether a policy τ_k or another policy $\tau_{k'}$ changes from period t to period $t+1$, the ratio $\Delta x_n^*/\Delta x_n$ is unchanged. Broadly speaking, A3′ requires that the researcher has correctly specified the general-equilibrium implications of all policy changes of interest, up to their ultimate incidence on each outcome variable. In this sense, misspecification remains local to the researcher's model. In the simulations of Section 3, we offer a specific example of misspecification that satisfies A3′ based on the magnitude of the pass-though from statutory to applied tariffs.[18]

For an arbitrary IV $\tilde{z}$ that satisfies A1 and A2, we know from the proof of Proposition 1 that the expected value of our goodness-of-fit measure $\hat{\beta}_{\tilde{z}}$ is equal to $\frac{1}{N_W} E_t[\sum_{n \in \mathcal{N}_W} \tilde{z}_n (\Delta x_n^* - \Delta x_n)]$, as displayed in equation (5). The problem is that when $\Delta x_n^* - \Delta x_n \neq 0$, this value is in general distinct from the expected welfare difference, $E_t[W(\Delta x^*) - W(\Delta x)] = E_t[\sum_{n \in \mathcal{N}_W} \omega_n (\Delta x_n^* - \Delta x_n)]$, which is the difference of interest from an economic standpoint.[19] To align our goodness-of-fit measure with the expected welfare difference, we need to find IVs that continue to satisfy A1 and A2, but are similar, on average, to the vector of welfare weights ω. Our next proposition demonstrates how to do so.

Proposition 3. *Take any IV $\tilde{z}$ that satisfies A1 and A2. If A3′ holds, then one can construct an adjusted IV $z \equiv \{\zeta_n \tilde{z}_n\}$, with the adjustment $\zeta_n \equiv N_W \omega_n E_t[\Delta x_n]/E_t[\tilde{z}_n \Delta x_n]$ for all $n \in \mathcal{N}_W$,*

[18] One may be concerned more generally that the researcher's model does *not* get the general-equilibrium implications right because it abstracts from specific features that one believes are important. If so, one could start from an alternative model that does incorporate these features and then impose A3′ around that alternative model. That is, A3′ is a complement to, not a substitute for, standard sensitivity analyses that consider alternative primitive assumptions on preferences, technology, or market structure.

[19] The gap between $E_t[\sum_{n \in \mathcal{N}_W} \tilde{z}_n (\Delta x_n^* - \Delta x_n)]$ and $E_t[\sum_{n \in \mathcal{N}_W} \omega_n (\Delta x_n^* - \Delta x_n)]$ arises for the same reasons that IV estimators may uncover what Imbens and Angrist (1994) refer to as a local average treatment effect rather than the average treatment effect. In our analysis, the misspecification error $\Delta x_n^* - \Delta x_n = (\alpha_n - 1)\Delta x_n$ plays the role of the heterogeneous treatment effect of interest.

such that z satisfies A1 and A2 and $E_t[\hat{\beta}_z] = E_t[W(\Delta x^*) - W(\Delta x)]$.

Intuitively, A1 and A2 give us the flexibility to rotate the original IV $\tilde{z}$, while still satisfying the orthogonality condition between the new IV z and the other shocks. To be more precise, if $\tilde{z}$ satisfies A1 and A2, then any alternative shift-share instrument z such that $z_n = \zeta_n \tilde{z}_n$ for some fixed vector $\zeta \equiv \{\zeta_n\}$ must also satisfy A1 and A2 and, in turn, yield $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta\eta_n^*] = 0$. While the dependence of z on the policy shifters $\Delta\tau_{IV}$ implies that z cannot be equal to ω for all realizations—an unavoidable feature if z is to be orthogonal to other shocks $\Delta\eta^*$—the freedom of choosing ζ allows the projections of $\Delta y - \Delta x$ on z and ω to be equal on average by setting $\zeta_n = N_W \omega_n E_t[\Delta x_n]/E_t[\tilde{z}_n \Delta x_n]$.[20]

Note that the previous adjustment procedure requires that the original IV satisfies $E_t[\tilde{z}_n \Delta x_n] \neq 0$ for all n. We discuss below how to satisfy this requirement. Note also that the previous adjustment procedure can be generalized to accommodate situations where the independence of the policy shifters only holds after controlling for linear determinants of the non-tariff shocks, as established in Appendix A.6. Finally, note that different forms of misspecification, i.e. distinct from A3′, may call for different manipulations of the original IV in order to provide a goodness-of-fit statistic whose magnitude has an economic interpretation. For the interested reader, Appendix A.7 considers a strictly weaker version of A3′ in which we only restrict the worst-case bias between Δx_n^* and Δx_n. In this case, we show that, while one can no longer construct z such that $\hat{\beta}_z + W(\Delta x)$ is an unbiased estimator of $E_t[W(\Delta x^*)]$, as we do in Proposition 3, one can still construct z such that the worst-case bias of $\hat{\beta}_z + W(\Delta x)$ is strictly lower than the worst-case bias of $W(\Delta x)$, as we formally show in Proposition 4.

Statistical Power. For IV-based tests to be useful, they should have high power, that is, a high probability of rejecting A3 when the causal impact of policies predicted by researcher's model is indeed incorrect. While general statements about a test's power require restrictions on the alternative to the null that is being tested—a challenging requirement in the case of model misspecification—we offer three guidelines for improving the power of IV-based tests.

First, one simple reason that an IV-based test may have low power is because, without further restrictions on the shares $\{s_{nk}\}$, the candidate IV z may be pure noise. If so,

[20] Characterizing the asymptotic distribution of our preferred goodness-of-fit measure $\hat{\beta}_z$ is a more challenging problem under A3′ than A3. Compared to the case presented in Section 2.3, the asymptotic distribution of $\hat{\beta}_z$ under A3′ depends not only on the behavior of $\sum_{n \in \mathcal{N}_W} z_n \Delta\eta_n^*$—a linear combination of the i.i.d. policy shifters $\Delta\tau_{IV}$ that enter the IV z—but also on $\sum_{n \in \mathcal{N}_W} z_n(\alpha_n - 1)\Delta x_n$—a potentially arbitrary function of such shifters. Provided that the researcher's causal impact Δx_n is itself a linear function of the policy vector, however, inference can be conducted in the same way as in the case of heterogeneous treatment effects in Adao et al. (2019). Conveniently, this linearity restriction is satisfied by our application below.

$E_t[z_n \Delta x_n] = E_t[z_n \Delta y_n] = 0$, regardless of whether or not $\Delta x_n = \Delta x_n^*$, in which case there is nothing to be learned from comparing the projections of Δx and Δy on the IV. To deal with this potential lack of power, we propose to anchor our testing procedure around a candidate IV for which we expect to have a "first-stage," i.e. $E_t[z_n \Delta x_n] \neq 0$. Specifically, our preferred IV in Sections 3 and 4 will start from shares $\{s_{nk}\}$ that are equal to the Jacobian of the researcher's model with respect to tariffs, $s_{nk} = \partial f_n / \partial \tau_k$, evaluated at (τ_t, ϵ_t). This implies that the associated IV $z_n = \sum_k (\partial f_n / \partial \tau_k) \Delta \tau_{IV,k}$ is equal to the causal impact of the policy shifter on variable n according to the researcher's model, up to a first-order approximation. By design, such a candidate IV combines existing exclusion restrictions from the empirical literature, as reflected in the choice of $\Delta \tau_{IV}$, with the general equilibrium structure of the researcher's model, as summarized by the Jacobian, $D_\tau(f) \equiv \{\partial f_n / \partial \tau_k\}$.

Second, the power of our test may be low because the confidence interval of $\hat{\beta}_z$, computed under the null using equation (6), may be large. To deal with precision, our preferred IV will also residualize the previous shares with respect to a vector of controls. This is equivalent to computing $\hat{\beta}_z$ after residualizing $\Delta y - \Delta x$, which mechanically lowers the variance of our IV-based test statistic, as described in Appendix A.4.

Finally, the power of our test may be low, *ceteris paribus*, when there has been estimation prior to testing, and the testing moment described in Definition 1 is mechanically related to moments that have already been set to zero for the purposes of estimation, as already discussed in Appendix A.5. In the extreme, if testing and estimating moments were exactly the same, then our test would always pass, and hence would have zero power. More generally, the more estimation moments are informative about $\hat{\beta}_z$ in the sense of Andrews et al. (2020)—i.e., the higher the R-squared of a hypothetical regression of $\hat{\beta}_z$ on the estimation moments—the lower we expect the power of our test to be, as illustrated in Appendix A.8. We will explore the potential importance of favoring IV-based tests for which estimation moments are relatively less informative through our simulations in Section 3.

2.5 Discussion

One can view our IV-based test as the second step of a general procedure that first uses a subset of moments for estimation and another subset of untargeted moments to build confidence in the predictions of the researcher's model. Many testing or validation procedures that have been deployed in the literature fit this general description. Here we describe the alternatives and discuss what we view as the main advantages of our test.

A moment is a moment? We have proposed the test statistic $\hat{\beta}_z$ as a way to test causal predictions. It offers a formal answer to the specific question: "can we reject the null that $W(\Delta x) = W(\Delta x^*)$?" rather than the broader: "can we reject the null that $f = f^*$?" A number of popular approaches to model validation do not draw this distinction. They implicitly embrace the researcher's model in its entirety. As a result, such approaches may reject a model even though its causal answer of interest is correct—because other shocks in the model are assumed to be absent or misspecified. Or they may fail to reject a model even though its causal answer is incorrect—because they focus on moments with little relevance to the causal mechanism of interest.

The preferred testing procedures in Kehoe et al. (1995), Kehoe (2005), and Kehoe et al. (2017) are subject to the first of these two concerns. They focus on the correlation between the change in outcomes Δy_n, such as trade flows in each industry n, that occurred around an episode of trade policy change, such as NAFTA, and the change in outcomes Δx_n predicted by the researcher's model in the absence of any other shock. This correlation, however, depends as much on misspecification in the causal answer of interest, $\Delta x_n^* \neq \Delta x_n$, as on the existence of other shocks at the time of the policy change, $\Delta \eta_n^* \neq 0$. More specifically, even when $\Delta x_n^* = \Delta x_n$ and policy changes are uncorrelated with other shocks,

$$|\text{corr}(\Delta y_n, \Delta x_n)| = \left(1 + \frac{\text{var}(\Delta \eta_n^*)}{\text{var}(\Delta x_n)}\right)^{-1/2}.$$

Despite the causal impact of trade policy predicted by the researcher's model being correct, $\text{corr}(\Delta y_n, \Delta x_n)$ therefore differs from one if $\text{var}(\Delta \eta_n^*) \neq 0$ and actually converges to zero as $\text{var}(\Delta \eta_n^*)/\text{var}(\Delta x_n)$ goes to infinity. The same concern applies to validation procedures that focus on the R-squared of the OLS regression of Δy_n on Δx_n, since $R^2 = |\text{corr}(\Delta y_n, \Delta x_n)|^2$, or the mean squared error, since $MSE = \text{var}(\Delta \eta_n^*)$, under the previous assumptions. By contrast, our proposed test based on $\hat{\beta}_z$ is valid for testing $W(\Delta x) = W(\Delta x^*)$ irrespective of the magnitude of $\text{var}(\Delta \eta_n^*)$.

Another commonly used strategy for strengthening the credibility of a quantitative model involves demonstrating that, after targeting a subset of moments of the initial cross-section for parameter estimation, the researcher's estimated model can also match untargeted moments from the same cross-section. Examples from the trade literature include Edmond et al. (2015), Costinot et al. (2016) and Antras et al. (2017). Such procedures are subject to the second concern mentioned above: they may fail to reject when the causal answer of interest is incorrect. For example, while Costinot et al. (2016) report a "reassuring" (p. 229) fit of their model's equilibrium revenue predictions (using a cross-section of crops and countries), it is possible that misspecification in such aspects of their model may be very different from misspecification in the causal impact of the shock that motivates the need for their model (designed to predict the causal impact of climate change on agricultural markets). Appendix A.9 describes an extreme example where the ability of the researcher's model to match untargeted moments from the cross-section can be arbitrarily good, yet the degree of misspecification in the counterfactual answer of interest can be arbitrarily large.[21] Whenever possible, we suggest to focus instead on a goodness-of-fit measure $\hat{\beta}_z$ that is deliberately designed around the causal impact of interest.

An IV is an IV? A distinct approach to testing draws instead on the availability of an instrument z that is believed to be independent of other shocks. As illustrated in Figure 1, armed with such an instrument one can test whether the causal impact of the policy predicted by the researcher's model is correctly specified by comparing the projection of Δy on z with that of Δx on z. Our test statistic $\hat{\beta}_z$ summarizes exactly such a comparison.

Going back to Lucas's (1980) call for "subjecting (models) to shocks for which we are fairly certain how actual economies or parts of economies would react," this idea has several antecedents. For example, in their study of the division of Berlin, Ahlfeldt et al. (2015) present a comparison of the coefficients obtained from two regressions: first, from regressing the change in block-level floor prices (i.e. Δy_n) on distance to the city's pre-War center (i.e. z_n); and second, from regressing their model's predicted causal impact of division on floor prices (i.e. Δx_n) on z_n.[22] A number of questions about the specific implementation of such a test, however, necessarily arise: (*i*) Is the choice of outcome and IV consistent with the causal prediction of interest? (*ii*) How should the test account for general-equilibrium linkages and prior estimation? (*iii*) How should the magnitude of the test statistic be interpreted? And (*iv*) How should the IV be designed to improve power and avoid mechanical fit? In this section, we have shown how one may go about

[21] The same general concern applies to validation procedures using the causal impact of shocks that differ from the policy of interest, say rainfall when the policy of interest is tariffs. The fit of the researcher's model may be arbitrarily good at predicting causal responses to rainfall shocks, yet provide little information about the difference between Δx and Δx^* that the researcher is ultimately interested in.

[22] Another closely related test is the "slope test" (see, e.g., Davis and Weinstein, 2001, Kehoe, 2005, or Kovak, 2013), which estimates an OLS regression of Δy on Δx and rejects when the slope coefficient $\hat{\beta}^{OLS}$ differs from 1. To compare this to our IV-based procedure, we begin by noting that our test statistic $\hat{\beta}_z$ is closely related to the coefficient one would obtain from an IV regression of Δy on Δx when z is used as the instrument. The coefficient $\hat{\beta}^{IV}$ from such a regression is related to $\hat{\beta}_z$ via $\hat{\beta}_z = (\frac{1}{N_W}\sum_{n\in\mathcal{N}_W} z_n \Delta x_n)(\hat{\beta}^{IV} - 1)$, which shows that, up to the first-stage covariance, a test of $\hat{\beta}^{IV} = 1$ is equivalent to that of $\hat{\beta}_z = 0$. In turn, a test of $\hat{\beta}^{OLS} = 1$ is equivalent to that of $\hat{\beta}_{\Delta x} = 0$, i.e. the special case where the researcher's prediction is assumed to be a valid IV $z = \Delta x$. This equivalence notwithstanding, it should be clear that the previous literature does not provide conditions under which one may characterize the asymptotic distribution of $\hat{\beta}^{OLS}$ in the presence of general-equilibrium linkages, a critical input for formal testing that Proposition 2 delivers for our IV-based test.

answering all these questions.

Why a two-step procedure? Our proposed test statistic $\hat{\beta}_z$ measures the difference between the moment $\frac{1}{N}\sum_{n\in\mathcal{N}_W} z_n(\Delta y_n - \Delta x_n)$ and zero. Even if one agreed on the importance of that moment for the counterfactual question of interest, one may still question why we propose to implement our test as a second step rather than assume that the moment condition $\hat{\beta}_z = 0$ holds and "stack" it with the moment conditions already used for the estimation of the vector of structural parameters θ, at which point one could compute the optimal GMM estimator of θ and perform a J-test.

This alternative one-step procedure offers the benefit of efficient estimation, in the sense of minimizing the asymptotic variance of $\hat{\theta}$ under the null that the researcher's prediction is correct, as discussed in Hansen (2008). But in regards to testing, we see two potential advantages of our approach. The first one relates to economic interpretation. The J statistic is a weighted sum of violations of the stacked vector of moment conditions. Analogously to the discussion of untargeted moments above, it is not always clear that a model's failure to match a given moment is relevant for the model's ability to make accurate statements about the causal question of interest. In contrast, Proposition 3 offers conditions under which $\hat{\beta}_z$ is an unbiased estimator of $E_t[W(\Delta x^*) - W(\Delta x)]$.

The second issue relates to power. In quantitative trade models, the moments used for estimation are typically partial equilibrium in nature and fit the data with relatively low variance. So-called gravity equations are a famous example. By contrast, the extra moment that we propose for testing is general equilibrium in nature and may therefore be expected to display relatively high variance. In such a setting, the J-test, which weights moment violations according to the inverse of their variance, will give greater weight to partial equilibrium considerations, meaning that the test may lack power to detect general equilibrium misspecification. In the extreme where the ratio of the variance of the estimation moments to our testing moment goes to zero, a J-test would never reject, provided the parameters θ are just-identified by the estimation moments. Our test instead focuses on the general equilibrium considerations embodied in $\hat{\beta}_z$, and so would potentially be able to reject an incorrect causal answer. At the very least, by offering an estimate of $\hat{\beta}_z$ and its standard error, our two-step procedure provides direct information about the power (or lack thereof) to reject the researcher's predictions.

3 Monte Carlo Simulations

We now explore the properties of IV-based tests through a series of Monte Carlo simulations in which we control the true data generating process. The researcher's model that we propose to test is the quantitative trade model developed by Fajgelbaum et al. (2020) (FGKK) to analyze the impact of Trump's trade war on the US economy. These simulations serve the dual role of illustrating the main themes of Section 2—relative performance of correlation- and IV-based tests, economic interpretation and statistical power of alternative IV-based tests, and issues of mechanical fit—as well as exploring the finite-sample properties of the IV-based tests that we will implement empirically in Section 4.

3.1 Simulation Procedure

The researcher's model in all our simulations is FGKK's model. This is a competitive trade model centered on the US economy. Appendix B.1 describes FGKK's primitive assumptions about US preferences, US technology, foreign import and export behavior, as well as the definition of a competitive equilibrium. Like in FGKK, we use a first-order approximation of the equilibrium conditions in all our simulations.[23] The five key structural parameters of the model are: the inverse elasticity of foreign export supply, ω_F; the elasticity of foreign import demand, σ_F; and three elasticities of substitution that determine US demand, κ, η, and σ. All other structural parameters are potentially time-varying Cobb-Douglas shares, preference and productivity shifters, and labor endowments. We let $\theta \equiv \{\sigma, \omega_F, \sigma_F, \eta, \kappa\}$ denote the vector of elasticities and ϵ_t denote the vector of other time-varying structural parameters. The policies of interest are US import tariffs, $\tau_t^H \equiv \{\tau_{iv,t}^H\}$, imposed by the US on one of 10,228 products v imported from one of 71 foreign countries i, and foreign import tariffs, $\tau_t^F \equiv \{\tau_{iv,t}^F\}$, imposed by these foreign countries on any product exported from the US.

For our first series of simulations, in Sections 3.2 and 3.3, we abstract from estimation uncertainty and assume that θ and ϵ_t are both known to the researcher. As further described in Appendix B.2, we set θ equal to FGKK's estimates and set ϵ_t such that, given the elasticities θ and the initial tariffs $\tau_t \equiv \{\tau_t^H, \tau_t^F\}$, FGKK's model exactly matches US

[23] Using the notation of Section 2, the researcher's model f is therefore formally given by

$$f(\tau, \epsilon) = \tilde{f}\big|_{\tau_t, \epsilon_t} + D_\tau(\tilde{f})\big|_{\tau_t, \epsilon_t} \cdot (\tau - \tau_t) + D_\epsilon(\tilde{f})\big|_{\tau_t, \epsilon_t} \cdot (\epsilon - \epsilon_t),$$

where $\tilde{f}$ is the non-linear model implied by the primitive assumptions in Appendix B.1 and (τ_t, ϵ_t) are the values of the shocks around which we linearize. There is therefore no distinction between $\Delta x \equiv f(\tau_{t+1}, \epsilon_{t+1}) - f(\tau_t, \epsilon_{t+1})$ and $\Delta \tilde{x} \equiv f(\tau_{t+1}, \epsilon_t) - f(\tau_t, \epsilon_t)$. Both are equal to $D_\tau(\tilde{f})\big|_{\tau_t, \epsilon_t} \cdot (\tau_{t+1} - \tau_t)$.

trade and production data in 2016. Like in FGKK, we let the initial tariffs τ_t be equal to their average statutory values in 2016. We will return to the specific issues associated with the estimation of θ in Section 3.4.

For each of our simulations, we consider 2,500 counterfactual economies, indexed by the superscript $b = 1, ..., 2500$, and subject them to counterfactual tariff changes, $(\Delta\tau)^b$, as well as counterfactual changes to productivity and preference parameters, $(\Delta\epsilon)^b$.[24] All changes are independently drawn from normal distributions. Tariff changes have mean $\mu_{\Delta\tau} = 0.2$ and standard deviation $\sigma_{\Delta\tau} = 0.7$. All other shocks have mean $\mu_\epsilon = 0$ and standard deviation $\sigma_\epsilon = 0.7$ unless otherwise specified. To generate counterfactual equilibrium prices and quantities, we assume that the true model is either FGKK's model, in which case $f^*(\tau, \epsilon) = f(\tau, \epsilon)$, or FGKK's model up to some misspecification to be described below, in which case $f^*(\tau, \epsilon) \neq f(\tau, \epsilon)$. Given knowledge of the true model and the researcher's model, we can then compute the counterfactual changes in all equilibrium variables, $(\Delta y)^b$, as well as the causal impact of changes in tariffs in the true model and the researcher's model, $(\Delta x^*)^b$ and $(\Delta x)^b$, respectively.

In line with the main goal of FGKK's analysis, our statistic of interest $W(\Delta x)$ is the causal impact of tariff changes on US welfare. Expressed as a percentage of US GDP in 2016, the change in welfare computed by FGKK is equal to

$$[W(\Delta x)]^b = \sum_{i,v} [\omega_{iv}^X (\Delta x_{iv}^X)^b + \omega_{iv}^M (-\Delta x_{iv}^M)^b + \omega_{iv}^R (\Delta x_{iv}^R)^b], \tag{7}$$

where X, M, R denote outcomes related to exports, imports, and tariff revenues, respectively. Specifically, Δx_{iv}^X measures changes in the log of US export prices of product v in country i (pre-foreign tariff), with ω_{iv}^X the associated percentage of export revenues in US GDP in 2016; Δx_{iv}^M measures changes in the log of US import prices of product v from country i (post-US tariff), with ω_{iv}^M the associated percentage of import spending in US GDP in 2016; and Δx_{iv}^R measures changes in US tariff revenues on product v from country i in proportion to initial import spending, with $\omega_{iv}^R = \omega_{iv}^M$, as described in Appendix B.3.

Equation (7) is a first-order approximation that follows from a standard application of the envelope theorem.[25] It holds in any model that, like FGKK's, assumes no distortions in the US economy other than the presence of import tariffs. For example, the fact

[24] Using the notation of Appendix B.1, the counterfactual changes to $(\Delta\epsilon)^b$ that we simulate derive from changes in foreign preferences for domestic varieties, $\log a_{iv,t}^F$; domestic preferences for foreign varieties, $\log a_{iv,t}$; and foreign productivity, $\log z_{iv,t}^F$.

[25] Although we restrict the statistic of interest to be linear throughout our analysis, it should be clear that one could study higher-order approximations by simply adding terms in the Taylor expansion. For example, the second-order approximation would extend the summation $W(\Delta x) \equiv \sum_{n \in \mathcal{N}_W} \omega_n \Delta x_n$ to include entries of the vector Δx created from interactions between all price and quantity changes.

that changes in export quantities do not affect welfare in (7) follows from the high-level assumption that there are no taxes nor markups on US exporting firms, not from any specific assumption about technology or preferences. As a result, all models that feature the same domestic distortions as FGKK will agree that first-order welfare changes follow equation (7), and agree on how to calculate the weights from data in 2016, even though they may disagree on the predicted price and tariff revenue changes that follow from a given set of tariff changes. Within that broad class of models, focusing on $W(\Delta x)$ is therefore sufficient to detect welfare misspecification.

As a final remark about dimensionality, we note that, in terms of Section 2's notation, each observation n here corresponds to a triplet $n = (i, v, o)$, with $o \in \{X, M, R\}$, whereas each tariff line k corresponds to a triplet $k = (i, v, c)$, with $c \in \{H, F\}$. Thus the total number of welfare-relevant variables is $71 \times 10,228 \times 3$ and the total number of tariff lines and tariff shifters is $71 \times 10,228 \times 2$. For computational reasons, we focus in our simulations on the country-product pairs that account for 90% of the value of US exports and imports. Without risk of confusion, we simply refer to the subset of welfare-relevant variables that we are left with as $\mathcal{N}_W$ and to their number as $N_W = 35,985$. The associated number of shifters that we are left with is $K = 22,590$.

3.2 Correlation- versus IV-Based Tests

Counterfactual analysis is about causal effects, not forecasting. The former holds everything else fixed, whereas the latter takes a stand on how everything else changes. Our IV-based test is designed to test the causal impact of a policy change predicted by the researcher's model, without having to specify how "everything else changes," above and beyond the fact that such changes are independent of the IV. For our first series of simulations, we illustrate how our IV-based test compares to a correlation-based test that does not aim to unbundle the size of the causal effect of interest from that of other shocks.

To implement our IV-based test, we need exogenous tariff shifters and shares that satisfy A1 and A2. Since tariff changes are independently drawn in our simulations, we simply use a normalized version of tariff changes as our tariff shifters,

$$(\Delta\tau_{IV,k})^b = \frac{(\Delta\tau_k)^b - \mu_{\Delta\tau}}{\sigma_{\Delta\tau}}. \tag{8}$$

To build shares, we follow the discussion of Section 2.4. In order to increase statistical power, we start from the Jacobian of the researcher's model with respect to tariffs and residualize the shares with respect to controls; and in order to provide an economic interpretation, we adjust the shares to target average welfare misspecification. In matrix notation, our preferred shares can be described as

$$\{s_{nk}^{\text{pref}}\} = P_c \times \text{Diag}_{\zeta} \times D_{\tau}(f), \tag{9}$$

where $D_{\tau}(f) \equiv \{\partial f_n / \partial \tau_k\}$ is the Jacobian of the researcher's model with respect to tariffs; Diag_{ζ} denotes the diagonal matrix associated with the welfare-relevant adjustment $\zeta \equiv \{\zeta_n\}$ from Proposition 3 (with controls); and P_c denotes the residual projection matrix associated with the vector of controls, as further described in Appendix B.4.[26] We let $(z_n^{\text{pref}})^b = \sum_k s_{nk}^{\text{pref}} (\Delta\tau_{IV,k})^b$ denote our preferred IV.

To test the null that the causal impact of tariffs predicted by the researcher's model is correct, we use $(\hat{\beta}_{z^{\text{pref}}})^b = \frac{1}{N_W} \sum_{n \in \mathcal{N}_W} (z_n^{\text{pref}})^b [(\Delta y_n)^b - (\Delta x_n)^b]$. We implement our test with a 5% significance level using the asymptotic results of Proposition 2—that is, our test rejects in simulation b when $(\hat{\beta}_{z^{\text{pref}}})^b / \sqrt{(\hat{V}[\hat{\beta}_{z^{\text{pref}}}])^b}$ lies either below the 2.5th or above the 97.5th percentile of a standardized normal distribution.

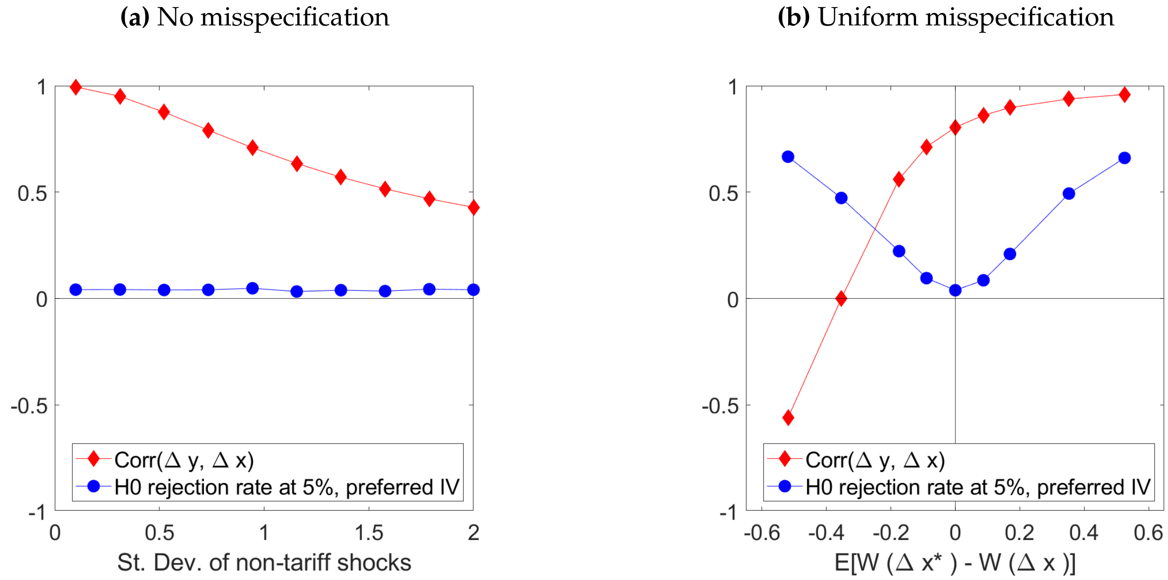
Figure 2a focuses on the case where FGKK's model is the true model, but the importance of tariff shocks relative to other shocks changes as we vary the standard deviation of all other shocks σ_{ϵ}. Since A3 holds, provided that the other conditions in Proposition 2 are satisfied, our test should therefore reject in about 5% of the 2,500 hypothetical economies that we consider, independently of the value of σ_{ϵ}. Reassuringly, this is indeed the case. In contrast, Figure 2a shows that the correlation between $(\Delta y_n)^b$ and $(\Delta x_n)^b$, averaged across hypothetical economies, is very sensitive to the value of σ_{ϵ}, despite the researcher's model being true in all these simulations. Not surprisingly, the less the variation in the data comes from tariff shocks, the lower is the correlation between observed changes and the causal impact of tariffs predicted by the researcher's model.

Figure 2b turns to the case where FGKK's model is misspecified. We focus here on the incidence of the statutory (i.e. *de jure*) tariffs. In FGKK's model, statutory tariffs are equal to the tariffs that are actually applied. In the true model, we assume that the pass-through from one to the other occurs at the rate α, so that $f^*(\tau, \epsilon) = f(\alpha\tau, \epsilon) \neq f(\tau, \epsilon)$.[27] It follows that A3' holds, with $\alpha_n = \alpha$ for all $n \in \mathcal{N}_W$, whereas A3 does not unless $\alpha = 1$. We refer

[26] Unless otherwise specified, controls include three dummy variables, for whether or not $n = (i, v, o)$ satisfies $o(n) = X$, $o(n) = M$, and $o(n) = R$, as well as each outcome n's total exposure to tariff changes, $\sum_k \partial f_n / \partial \tau_k$. For the interested reader, Figure B.1 in Appendix B.5 offers the counterpart of Figure 3 below with different sets of controls. In line with our discussion in Section 2.4, fewer controls imply a higher variance for our test statistic and less power.

[27] Besides its simplicity, the assumption that the pass-through rate of statutory tariffs into applied tariffs may not be equal to one is consistent with the empirical findings of FGKK. Specifically, Table IV in FGKK shows that the difference between the tariff pass-through rate on import prices excluding tariff duties (in column 3) and including tariff duties (in column 4) is significantly different from one.

Figure 2: Correlation- versus IV-based Tests



Notes: This figure reports the rejection rate of an IV-based test at a 5% significance level (blue circles) and the correlation between $(\Delta y_n)^b$ and $(\Delta x_n)^b$ (red diamonds) across $b = 1, ..., 2500$ simulated economies. The shifters and shares entering the IV $(z_n^{\text{pref}})^b$ are described in equations (8) and (9). Figure 2a focuses on the case where the researcher's model is the true model and varies the standard deviation of other shocks σ_ϵ used in the simulations. Figure 2b instead assumes that the researcher's model misspecifies the pass-through rate of all statutory tariffs into applied tariffs and varies the true pass-through rate used in the simulations, which results in the corresponding amount of average welfare misspecification $E_t[W(\Delta x^*) - W(\Delta x)]$ shown on the x-axis.

to this benchmark case as "uniform misspecification."

Figure 2b reports the results of simulations that vary the degree of misspecification α. For ease of interpretation, we display results for which the x-axis is in units of $E_t[W(\Delta x^*) - W(\Delta x)]$, the expectation of the difference between the causal impact of tariffs predicted by the true model and the researcher's model associated with a given value of α (expressed as a percentage of US GDP). Reassuringly, Figure 2b shows that the power of our test—that is, the propensity to reject the predictions of this misspecified model—increases with the degree of welfare misspecification. As $E_t[W(\Delta x^*) - W(\Delta x)]$ increases in absolute value, so does the probability of rejecting the null that the welfare effects predicted by the researcher's model are correct. Perhaps surprisingly, however, we see that the correlation between $(\Delta y_n)^b$ and $(\Delta x_n)^b$ instead *increases* as the degree of misspecification in the counterfactual answer of interest becomes more and more positive. The reason is that tariff changes account for a larger share of observed changes as α increases, which raises the correlation between $(\Delta y_n)^b$ and $(\Delta x_n)^b$.

3.3 Alternative IV-Based Tests

In Section 2.4, we have argued how different dimensions of an IV may help increase the statistical power of our test as well as offer an economic interpretation for our goodness-of-fit measure. We now illustrate these issues via the comparison of two different IV-based tests. The first IV z^{pref} is the same as in the previous simulations, with shifters and shares described in equations (8) and (9). The second IV z^{naive}, which we will refer to as a "naive" IV, uses the same shifters, but different shares,

$$\{s_{nk}^{\text{naive}}\} = P_c \times \text{Diag}_{\omega} \times I. \tag{10}$$

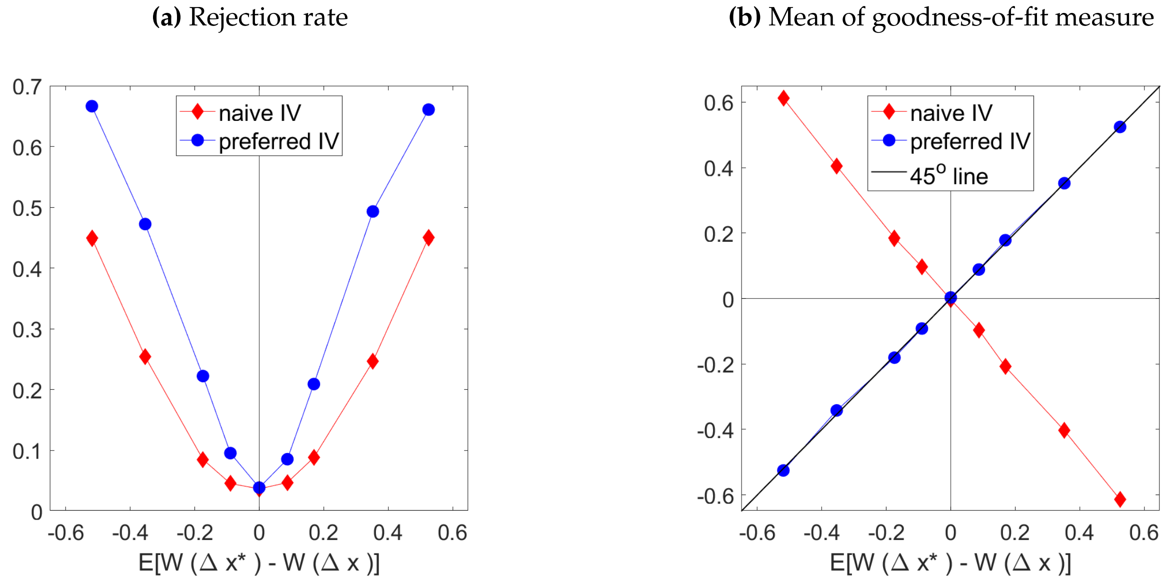
Compared to our preferred shares in equation (9), the naive shares ignore the general-equilibrium structure of the model and instead only use the policy shifter associated with the outcome of interest, either the US tariff for import and revenue outcomes or the foreign tariff for export outcomes, as reflected in the matrix I instead of the Jacobian $D_{\tau}(f)$.[28] In order to ease comparison, we residualize the naive IV using the same controls as our preferred IV, hence the same projection matrix P_c, and assume that it incorporates welfare weights, but does so in an ad-hoc manner, as reflected in Diag_{ω} rather than Diag_{ζ}.

Figure 3 returns to the same uniform misspecification as in the previous subsection. We know from Proposition 2 that since both instruments z^{pref} and z^{naive} satisfy A1 and A2, both IV-based tests that use them should have the correct size (probability of rejecting equal to stated significance level) when the model is correctly specified. Reassuringly, this is true in our simulations, as is apparent in Figure 3a, in which the y-axis displays rejection rates and the x-axis again reports the expectation of the counterfactual welfare mistake $E_t[W(\Delta x^*) - W(\Delta x)]$. However, although both of these IV-based tests reject models that are misspecified, we see that the test based on our preferred IV z^{pref} has an advantage in terms of power, since the blue circles are higher than the red diamonds (greater rejection rate) everywhere but at $E_t[W(\Delta x^*) - W(\Delta x)] = 0$.

The difference between the two goodness-of-fit measures in terms of economic interpretation is starker. Figure 3b reports how the average values of $(\hat{\beta}_{z^{\text{pref}}})^b$ and $(\hat{\beta}_{z^{\text{naive}}})^b$ vary with the degree of misspecification. In line with Proposition 3, the average value of $(\hat{\beta}_{z^{\text{pref}}})^b$ tracks $E_t[W(\Delta x^*) - W(\Delta x)]$, since the blue circles follow the 45° line. The average of $(\hat{\beta}_{z^{\text{naive}}})^b$, on the other hand, is actually negatively related to $E_t[W(\Delta x^*) - W(\Delta x)]$ even though one might hope that incorporating welfare weights in z^{naive} would go some

[28] Formally, I corresponds to the $N_W \times K$ matrix whose entry (n, k) is a dummy variable that equals one if, and only if, $i(n) = i(k)$ and $v(n) = v(k)$ (i.e., outcome n and shifter k correspond to the same country-product pair) as well as either $o(n) = X$ and $c(k) = F$ (i.e., n is an export price and k a foreign tariff shifter) or $o(n) \in \{M, R\}$ and $c(k) = H$ (i.e., n is an import price or tariff revenue and k a US tariff shifter).

Figure 3: Alternative IV-based Tests Under Uniform Misspecification



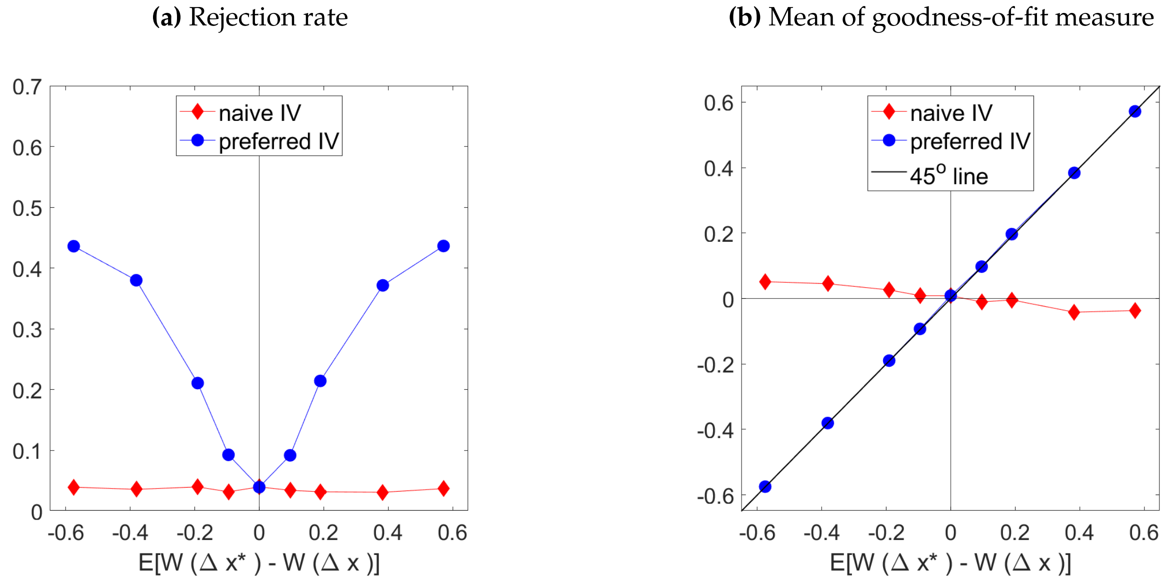
Notes: This figure reports the resuls from $b = 1, ..., 2500$ simulated economies in which the researcher's model misspecifies the pass-through rate of all statutory tariffs into applied tariffs by an amount α that is uniform across import, export, and tariff revenue outcomes. Simulations differ in terms of the size of the underlying α to generate the variation in $E_t[W(\Delta x^*) - W(\Delta x)]$ reported on the x-axes. Figure 3a plots the rejection rate on the y-axis for IV-based tests using our preferred IV (z^{pref}, blue circles) and a naive IV (z^{naive}, red diamonds). Figure 3b does the same but with the average value of our goodness-of-fit measure ($\hat{\beta}_{z^{\text{pref}}}$ and $\hat{\beta}_{z^{\text{naive}}}$, respectively) on the y-axis.

way towards the economic interpretability of $(\hat{\beta}_{z^{\text{naive}}})^b$.[29]

Figure 4 turns to misspecification that only affects a subset of variables. Specifically, we assume that the degree of misspecification for export outcomes is the same as before (i.e. $(\Delta x_{iv}^{X*})^b = \alpha(\Delta x_{iv}^{X})^b$) but we now rule out misspecification in import and revenue outcomes (i.e. $(\Delta x_{iv}^{M*})^b = (\Delta x_{iv}^{M})^b$ and $(\Delta x_{iv}^{R*})^b = (\Delta x_{iv}^{R})^b$). Such a form of misspecification continues to satisfy A3', and so it is no surprise to see in Figure 4b that the economic interpretability of $\hat{\beta}_{z^{\text{pref}}}$ continues to hold true. More interestingly, we see in Figure 4a that the naive IV-based test using z^{naive} now has much less power to reject the misspecified model. Intuitively, the naive test does not take into account the fact that in FGKK's model there is less variation across export prices than import prices and tariff revenues, which leads the naive test to underweight misspecification in export outcomes and, in turn, to under-reject. We return to this distinction between testing export- and import-based out-

[29] In this example, it is interesting to note that there is a trade-off between economic interpretability and statistical power. Using the welfare weights for the purposes of economic interpretability comes at the cost of a less precise test statistic. As can be seen from Figure B.2 in Appendix B.6, using a variant of our preferred IV that weighs all observations equally—that is, which computes Diag_ζ under the assumption that $\omega_n = 1$ for all $n \in \mathcal{N}_W$—implies an average estimate that is less sensitive to $E_t[W(\Delta x^*) - W(\Delta x)]$, but the test based on it has a slightly higher rejection rate due to a lower variance.

Figure 4: Alternative IV-based Tests Under Export Misspecification



Notes: This figure reports the results from $b = 1, \ldots, 2500$ simulated economies in which the researcher's model misspecifies the response of export outcomes to tariffs by an amount α but the responses of import and revenue outcomes are correctly specified. Simulations differ in terms of the size of the underlying α to generate the variation in $E_t[W(\Delta x^*) - W(\Delta x)]$ reported on the x-axes. Figure 4a plots the rejection rate on the y-axis for IV-based tests using our preferred IV (z^{pref}, blue circles) and a naive IV (z^{naive}, red diamonds). Figure 4b does the same but with the average value of our goodness-of-fit measure ($\hat{\beta}_{z^{\text{pref}}}$ and $\hat{\beta}_{z^{\text{naive}}}$, respectively) on the y-axis.

comes in Section 4.3.

3.4 Estimation, Informativeness, and Mechanical Fit

Our final set of simulations focuses on the issue of mechanical fit. We do so by contrasting two potential approaches to counterfactual analysis and testing. In the first one, like in the previous subsections, we assume that the researcher calibrates her model using estimates from the literature. In the second one, we assume that the researcher uses moments associated with the policy change of interest to estimate structural parameters of her model. The second approach has the benefit of recognizing that estimates obtained in one context may not be valid in another, but it raises the issue of mechanically failing to reject incorrect predictions because, by design, estimates of structural parameters have already been selected to match (some) observed responses to the policy change of interest. Intuitively, the severity of this issue depends on how closely related the moments used for estimation and testing are. Here, we follow Section 2.4 and use the notion of informativeness introduced by Andrews et al. (2020) as a way to measure the distance between estimation and testing moments and, in turn, to construct IV-based tests that are less likely to be subject to mechanical fit.

In our simulations, we contemplate the possibility of a researcher estimating the US import demand elasticity, an exercise with a long tradition in the field and one revisited by FGKK. However, unknown to this researcher, there is misspecification on the import side. Specifically, we assume that the true model is the one described by FGKK, except for the response of US import quantities to tariff changes, which is assumed to be α times larger in the true model than in FGKK's model; all other responses to tariff changes are assumed to be the same in the true model and in FGKK's model.[30] When estimation takes place in simulation b, the researcher infers the value $(\hat{\sigma})^b$ of the US import demand elasticity via a linear regression as in FGKK; she regresses changes in log import quantities on changes in log import prices using the change in US tariff as an IV, controlling for product fixed effects, and weighing all observations equally. However, because of misspecification in import quantities, the average value of $\hat{\sigma}$ that she obtains is not equal to the true underlying preference parameter σ, but rather to $\alpha \times \sigma$. In line with FGKK's empirical analysis, we set the true preference parameter $\sigma = 2.53/\alpha$ so that the average value of $\hat{\sigma}$ is equal to FGKK's point estimate. When estimation does not take place, we keep the true value of $\sigma = 2.53/\alpha$ unchanged, but assume that the researcher wrongly calibrates it to 4 based on existing estimates of import demand elasticities available in the literature. Throughout this last round of simulations, we set $\alpha = 10$.

To illustrate the potential importance of mechanical fit, our simulations compare tests based on our preferred IV z^{pref} to those based on a new naive instrument z^{mech} that is designed to be closer to the IV used in estimation. Namely, it takes a similar form as z^{naive}, but is residualized with respect to product fixed effects and unweighted. Formally, the new matrix of shares is equal to

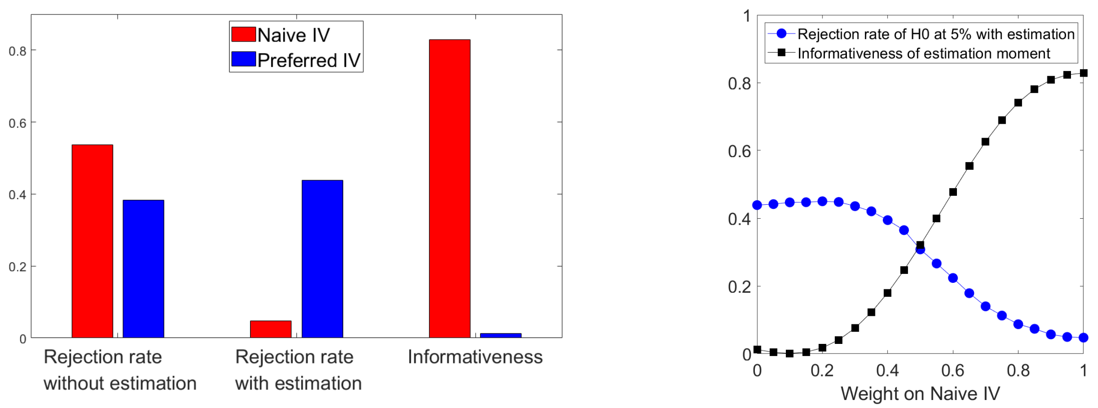
$$\{s_{nk}^{\text{mech}}\} = P_{c^{\text{mech}}} \times Id \times I, \tag{11}$$

where $P_{c^{\text{mech}}}$ denotes the residual projection matrix associated with the product fixed effects and Id is the identity matrix. When no estimation takes place, the average welfare misspecification is around 0.2% of GDP and, as Figure 5a shows, both tests reject around 50% of the time. When estimation takes place, the average welfare misspecification remains around 0.2% of GDP, but only our preferred IV-based test continues to reject at a similar rate, whereas the naive IV-based test now dramatically loses power. Accordingly, a researcher who switches from calibrating $\sigma = 4$ to estimating $\hat{\sigma} = 2.5$ and goes

[30]This particular misspecification is equivalent to assuming that there is a distinction between US import prices, as observed in the trade data, and the prices ultimately faced by US buyers, with the latter being α times larger.

Figure 5: Estimation, Informativeness, and Mechanical Fit

(a) Naive vs. preferred IV **(b)** Informativeness and power



Notes: This figure reports the resuls from $b = 1, ..., 2500$ simulated economies in which the researcher's model misspecifies the response of import quantities to tariffs by an amount $\alpha = 10$ but not the responses of other outcomes. The true US preference parameter is $\sigma = 2.5/\alpha$, but the researcher either calibrates it to 4 (in the case without estimation) or estimates it to be on average equal to 2.5 based on a linear regression of import quantities on import prices intrumented with US tariff changes, controlling for product fixed effects, and weighing all observations equally (in the case with estimation). Figure 5a plots rejection rates for the naive IV z^{mech} (in red) and the preferred IV z^{pref} (in blue), both without estimation and with estimation, as well as informativeness of the estimating moment for $\hat{\beta}_{z^{\text{mech}}}$ (in red) and $\hat{\beta}_{z^{\text{pref}}}$ (in blue). Figure 5b plots informativeness (black squares) and rejection rates (blue circles) associated with hybrid IVs z^{λ} that are linear combinations of z^{pref} and z^{mech} with the weight λ on z^{mech} reported on the x-axis.

from failing to passing the naive test may wrongly infer that incorrect calibration was the problem—instead, her test is *mechanically* successful when testing follows estimation.

The last two bars in Figure 5a explain why. They plot the informativeness of the estimation moment $\hat{\gamma}$ for our two testing moments, $\hat{\beta}_{z^{\text{pref}}}$ and $\hat{\beta}_{z^{\text{mech}}}$, which we recover from the covariance matrix of the estimation and testing moments.[31] As can be seen from Figure 5a, the estimation moment is much more informative for $\hat{\beta}_{z^{\text{naive}}}$ than $\hat{\beta}_{z^{\text{pref}}}$. Thus, when the estimating moment $\hat{\gamma}$ is set to zero for the purposes of obtaining $\hat{\sigma}$, we expect $\hat{\beta}_{z^{\text{naive}}}$ to become much less sensitive to misspecification than $\hat{\beta}_{z^{\text{pref}}}$, thereby lowering its power.

Figure 5b further illustrates the previous intuition by considering hybrid IVs $z^{\lambda} = \lambda z^{\text{mech}} + (1-\lambda) z^{\text{pref}}$ that are linear combinations of the naive and preferred IVs, with

[31] Formally, for a given counterfactual economy $b = 1, ..., 2500$, following Andrews et al. (2020) we compute the informativeness of $\hat{\gamma}$ for a testing moment $\hat{\beta} \in \{\hat{\beta}_{z^{\text{pref}}}, \hat{\beta}_{z^{\text{mech}}}\}$ as $[(\hat{\mathcal{V}}_{\beta\gamma})^b]^2/[(\hat{\mathcal{V}}_{\gamma})^b(\hat{\mathcal{V}}_{\beta})^b]$. Figure 5a then reports the average across all counterfactual economies. The estimator of the covariance matrix is the same one used to account for estimation uncertainty in Appendix A.5, as described in equations (A.25)-(A.27). In the context of our simulations, we can also compute informativeness directly by running a regression of $(\hat{\beta})^b$ on $(\hat{\gamma})^b$ across simulated economies. The R-squared values that we obtain from this procedure are 0.01 and 0.74, for $(\hat{\beta})^b = (\hat{\beta}_{z^{\text{pref}}})^b$ and $(\hat{\beta})^b = (\hat{\beta}_{z^{\text{mech}}})^b$ respectively, which are close to the average values of 0.01 and 0.83 obtained when using our estimator of the covariance matrices. This is reassuring since the previous estimator is derived ignoring misspecification in the researcher's model.

$\lambda \in [0,1]$. By construction, the informativeness of the estimation moment for the testing moment $\hat{\beta}_{z^\lambda}$ tends to increase with the weight λ put on the naive instrument, as illustrated in the black curve. And as informativeness increases, the rejection rate of the test based on the hybrid IV steadily decreases, as illustrated in the blue curve. As a practical way to strengthen the credibility of a test and alleviate concern of mechanical fit when estimation occurs prior to testing, we therefore suggest reporting the informativeness of the estimation moments for the testing moment. We will do so next in our empirical application.

4 Application to Trump's Trade War

Section 3 has illustrated the properties of our IV-based test in the context of simulated policy changes. We now turn to a recent and prominent actual policy change: the Trump Administration's 2018 actions that increased US import tariffs on many products and caused several trading partners to retaliate with tariff increases of their own. FGKK's seminal analysis of the 2018 tariff war concluded that "the aggregate real income loss was \$7.2 billion, or 0.04% of GDP." In this section we show how to use IV-based tests as a way to explore the empirical credibility of this conclusion.

4.1 From Simulated to Actual Shocks

The researcher's model is FGKK's model and the counterfactual question of interest is the same as in Section 3. To stay as close as possible to FGKK's original analysis, we again use the first-order approximation of the model presented in Appendix B.1 as well as the parameters θ and ϵ_t described in Appendix B.2. This implies that, like in Section 3, FGKK's model exactly matches trade and production data from the 2016 US economy, by construction. We use FGKK's model to predict changes in three outcome variables: (i) the log of US export prices of product v in country i (pre-foreign tariff), Δx^X_{iv}; (ii) the log of US import prices of product v from country i (post-US tariff), Δx^M_{iv}; and (iii) changes in US tariff revenues on product v from country i in proportion to initial import spending, Δx^T_{iv}. As discussed in Section 3, changes in these three variables are sufficient to evaluate changes in US welfare, up to a first-order approximation, provided that import tariffs are the only source of distortions in the US economy, as assumed by FGKK.

We only depart from the simulations in Section 3 in terms of the shocks fed into the researcher's model, which are now taken from the data rather than simulated. Changes in outcomes Δy_n that enter our testing procedure are the actual changes over the period

2016-19, with $y_{n,t}$ and $y_{n,t+1}$ an average of monthly data over all months of 2016 and 2019, respectively. Predicted changes Δx_n are computed by subjecting the model economy to actual tariffs in 2016 and 2019 for the US and its trading partners. For each 10-digit HS code product v and foreign country i, the 2016 tariffs $\tau^H_{iv,t}$ and $\tau^F_{iv,t}$ imposed by the US on country i and by country i on the US, respectively, are equal to the average of monthly tariffs during that year, whereas the 2019 tariffs $\tau^H_{iv,t+1}$ and $\tau^F_{iv,t+1}$ are equal to their values in April 2019, the final month in FGKK's estimation sample. This set of tariff changes is what we refer to as Trump's trade war.

Like in our simulations, we use a normalized version of actual tariff changes as shifters,

$$\Delta\tau_{IV,k} = \frac{\Delta\tau_k - \mu_{\Delta\tau}}{\sigma_{\Delta\tau}}, \text{ for any tariff line } k = (i, v, c), \tag{12}$$

where $\mu_{\Delta\tau}$ denotes the sample mean of tariff changes and $\sigma_{\Delta\tau}$ their standard deviation. To invoke the results of Propositions 1-3 and implement our IV-based test, we therefore require tariff changes to be independent of other shocks (conditional on controls) as well as i.i.d. across tariff lines. Both requirements are consistent with the exclusion restrictions invoked by FGKK to estimate the vector of structural parameters θ entering their model.[32] This makes our test a natural add-on to their analysis.

4.2 The Impact of Trump's Trade War on US Welfare: -0.04% of GDP?

The results of our test are presented in Table 1. To set the stage, column (1) examines a simple correlation between the model's prediction Δx_n and what actually happened in the data Δy_n. We find that the correlation between Δy_n and Δx_n for all welfare-relevant outcomes is just 0.08—or equivalently, that the R^2 from a regression of one on the other would be less than 0.01. This implies that the variance of the impact of tariff changes in this setting is orders of magnitude lower than that of other shocks ϵ_t. But as discussed above such a finding has no bearing on whether the researcher's model is successful at its stated goal of predicting the causal impact of tariff changes on US welfare.

By contrast, the test that we have developed in Section 2 is designed to be applicable even in settings like these where the variance of other shocks is relatively large. Box 1 summarizes the generic steps required to implement our test. We follow these steps in parallel for two separate IV-based tests based on the "preferred" and "naive" IVs, z^{pref}

[32] Although A1 is consistent with FGKK's exclusion restrictions, it is stronger in that it also requires shifters to be independent of shocks to labor endowments and various Cobb-Douglas shares. This is because FGKK's counterfactual predictions depend on these structural parameters but their estimating equations do not.

Table 1: Testing Predictions about the Welfare Impact of Trump's Trade War

Goodness-of-fit measure:	Correlation	IV-Based Test	
		Naive IV	Preferred IV
	$Corr\left(\Delta y_n, \Delta x_n(\hat{\theta})\right)$	$\hat{\beta}_{z^{\text{naive}}}(\hat{\theta})$	$\hat{\beta}_{z^{\text{pref}}}(\hat{\theta})$
	(1)	(2)	(3)
Point estimate	0.08	-0.01	-0.09
Inference ignoring estimation uncertainty			
Std. error		0.18	0.15
p-value of H0: $\hat{\beta} = 0$		0.96	0.56
Inference accounting for estimation uncertainty			
Std. error		0.24	0.18
p-value of H0: $\hat{\beta} = 0$		0.97	0.63

Notes: All statistics are based on the pooled sample of changes in 25,115 welfare-relevant outcomes between 2016 and 2019, and FGKK's estimates $\hat{\theta}$, as described in Section 3.1. Column (1) reports the correlation between actual changes and the predicted impact of Trump's trade war across all outcomes. Columns (2) and (3) implement our IV-based test using the naive IV z^{naive}, as defined by equations (10) and (12), and our preferred IV z^{pref}, as defined by equations (9) and (12). Inference ignoring estimation uncertainty is as described in Section 2.3. Inference accounting for estimation uncertainty is as described in Appendix C.1.

and z^{naive}, introduced in Section 3. In step 1 we obtain policy shifters for both IVs as described in equation (12). Step 2 uses the shares described in equations (9) and (10). Both of these shares adjust for controls to increase power; the preferred IV z^{pref} goes further by using the model's Jacobian and applying the adjustment in Proposition 3 (with controls) to allow for economic interpretability.

We next follow step 3 and compute our IV-based goodness-of-fit measure $\hat{\beta}_z(\hat{\theta})$. This statistic is reported in the first row of Table 1, with those based on $z = z^{\text{pref}}$ and $z = z^{\text{naive}}$ in columns (2) and (3), respectively. The corresponding standard errors $\sqrt{\hat{V}[\hat{\beta}_z(\hat{\theta})]}$, computed via step 4, and p-values as per step 5, are reported next. Like in Section 2.3, we begin with versions of these for which the asymptotic variance estimate $\hat{V}[\hat{\beta}_z(\hat{\theta})]$ being used treats $\hat{\theta}$ as a known parameter, as is common in many calibration exercises. With p-values of 0.96 and 0.56, the test of A3 does not reject when based on either IV.

The last row of Table 1 reports standard errors and p-values for the case, as is relevant here, in which the parameter vector $\hat{\theta}$ is subject to estimation uncertainty. As described in Appendix C.1, FGKK obtain $\hat{\theta}$ from an IV estimator that draws on the same tariff changes underlying our IV-based test. Accordingly, we can use the results in Appendix A.5 to compute standard errors accounting for the uncertainty in $\hat{\theta}$.[33] The adjusted standard

[33] An idiosyncratic feature of FGKK's analysis is that estimation is conducted at the monthly level, whereas counterfactual analysis is conducted at the annual level. Appendix C.1 shows how to deal with

Box 1: A blueprint for testing the causal predictions of quantitative models

After obtaining an estimate $\hat{\theta}$ of the structural parameters of her model, a researcher predicts that the causal impact of a policy change on some statistic of interest is

$$W(\Delta x(\hat{\theta})) = \sum_{n \in \mathcal{N}_W} \omega_n \Delta x_n(\hat{\theta}),$$

where $\Delta x_n(\hat{\theta})$ is the predicted change in any given outcome n and ω_n is its weight in the statistic of interest. The following steps show how to test this causal prediction.

Step 1: Select shifters, $\{\Delta\tau_{IV,k}\}$. Choose policy shifters that are independent of other (non-policy) shocks, potentially those have already been used in the estimation of θ.

Step 2: Select shares, $\{s_{nk}\}$. All else equal, prefer shares that allow for economic interpretability of the test statistic (e.g. by using Proposition 3) and raise statistical power (e.g. by using the model's Jacobian, by residualizing with respect to controls, or by using shares for which estimating moments are less informative of testing moments).

Step 3: Compute IV-based goodness-of-fit measure, $\hat{\beta}_z(\hat{\theta})$. Combine shifters and shares to build the IV $z_n = \sum_k s_{nk}\Delta\tau_{IV,k}$ and compute $\hat{\beta}_z(\hat{\theta}) \equiv \frac{1}{N_W}\sum_{n\in\mathcal{N}_W} z_n(\Delta y_n - \Delta x_n(\hat{\theta}))$.

Step 4: Compute asymptotic variance of $\hat{\beta}_z(\hat{\theta})$. Calculate $\hat{V}[\hat{\beta}_z(\hat{\theta})]$ treating $\hat{\theta}$ as known (as in Section 2) or taking into account estimation uncertainty (as in Appendix A.5).

Step 5: Test causal prediction. Compute the p-value under the null that all predicted changes $\Delta x_n(\hat{\theta})$ entering W are correct: p-value $= 2(1 - \Phi^{-1}(|t_z|))$ with $t_z \equiv \hat{\beta}_z(\hat{\theta})/\sqrt{\hat{V}[\hat{\beta}_z(\hat{\theta})]}$ and Φ the cdf of a standard normal distribution, by Proposition 2.

Step 6: When possible, discuss economic significance of $\hat{\beta}_z(\hat{\theta})$. If the IV z has been adjusted along the lines of Proposition 3, for instance, then $\hat{\beta}_z(\hat{\theta})$ is an estimate of the average bias in the causal prediction of interest.

errors for the test statistics based on z^{pref} and z^{naive} are 20% and 33% larger than those ignoring estimation uncertainty. As a result, the p-values of our test increase from the already high values obtained without accounting for estimation uncertainty. Regardless of whether we use z^{pref} and z^{naive}, the informativeness of the estimation moments for the testing moments remains very low, 1.2×10^{-4} for z^{naive} and 0.5×10^{-4} for z^{naive}, suggesting that mechanical fit is not a concern in this context.[34] Based on these results, we cannot reject that the true causal response of US real income to Trump's Trade War was $W(\Delta x) = -0.04\%$, as predicted by FGKK's model.

this issue. In line with the rest of our analysis, our test requires independence across tariff lines, but does not impose any restriction on the month-to-month variation within a given tariff line.

[34]This is a consequence of three differences between estimation and testing moments. First, FGKK's estimating moments are computed on a different subsample and using different weights. Second, they entail a larger set of controls. Third, they depend on the realization of a smaller set of structural shocks.

As a final step, we turn to the economic significance of estimates in Table 1. Doing so is not an option for the naive IV-based test statistic in column (2). In contrast, since the goodness-of-fit measure $\hat{\beta}_{z^{\text{pref}}}(\hat{\theta})$ in column (3) leverages Proposition 3, it has economic interpretability. Under the assumption that misspecification in the researcher's model satisfies A3′, $\hat{\beta}_{z^{\text{pref}}}(\hat{\theta})$ provides an estimate of the average value of $W(\Delta x^*) - W(\Delta x)$ across realizations of the policy shifters. The value of $\hat{\beta}_z(\hat{\theta}) = -0.09$ reported in column (3) therefore implies that, under this assumption, the true welfare loss $|W(\Delta x^*)|$ is on average higher by 0.09% of US GDP than what FGKK's model predicts.[35]

4.3 A Final Diagnosis

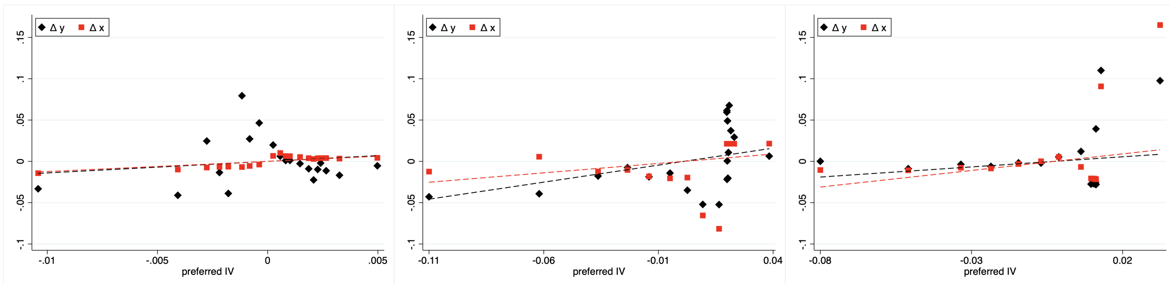
We have designed our IV-based test with one goal in mind: to assess and potentially strengthen the credibility of a quantitative model's answer to a causal question of interest. We recognize that in practice, researchers may also be interested in understanding economic mechanisms and exploring further which subsets of their predictions may be inconsistent with data. As Kehoe and Prescott (1995) note, "shortcomings in [counterfactual] predictions of a model would then provide motivation for further theoretical development and further testing." Motivated by this observation, we conclude with a more diagnostic approach that applies our IV-based test separately for each of the three outcomes that have been pooled so far: export prices, import prices, and tariff revenues.

Figure 6 presents bin-scatter plots of the observed changes Δy_n and predicted changes Δx_n against our preferred IV z_n^{pref}, and does so separately for export prices (in Figure 5a), import prices (in Figure 5b), and tariff revenues (in Figure 5c). Recall that the basic idea of our IV-based goodness-of-fit measure is to compare the projection of Δy_n on z_n^{pref} to the projection of $\Delta x_n(\hat{\theta})$ on z_n^{pref}, as illustrated in Figure 1. Under the null that the researcher's model predictions are correct, these two projections should coincide. The lines in Figure 6—which are drawn to have intercepts of zero and slopes equal to $\frac{1}{|\mathcal{N}_o|}\sum_{n\in\mathcal{N}_o} z_n^{\text{pref}} \Delta y_n$ and $\frac{1}{|\mathcal{N}_o|}\sum_{n\in\mathcal{N}_o} z_n^{\text{pref}} \Delta x_n(\hat{\theta})$, respectively, with $\mathcal{N}_o$ the set of observations associated with each outcome—suggest that that this is approximately the case for export prices, but much less so for import prices and tariff revenues. Although the instrument is correlated with both Δy_n and Δx_n for all three outcomes, there is a gap between the two projections in the case of import prices and tariff revenues, suggesting that while one is not able to reject FGKK's predictions when they are pooled across all welfare-relevant outcomes, one may be able

[35] The "on average" qualifier is important. Under the assumptions of Proposition 3, it should be clear that we cannot estimate the value of $W(\Delta x^*)$ associated with the actual tariff changes. Thus we cannot conclude that the true impact on US welfare of Trump's trade war was a reduction by $0.04 + 0.09 = 0.13\%$, only that FGKK tends to under-predict losses across all possible tariff realizations by 0.09% of US GDP.

Figure 6: Observed versus Predicted Changes

(a) Export Prices **(b)** Import Prices **(c)** Tariff Revenues



Notes: The figure plots observed changes Δy_n and predicted changes $\Delta x_n(\hat{\theta})$ against our preferred IV z_n^{pref} separately for the three types of outcomes that enter equation (7): export prices (in Figure 5a), import prices (in Figure 5b), and tariff revenues (in Figure 5c). Each figure displays a binned scatter plot in which the underlying product-country observations are grouped into 20 bins in terms of z_n^{pref}. The illustrated lines have slopes equal to $\frac{1}{|\mathcal{N}_o|}\sum_{n\in\mathcal{N}_o} z_n^{\text{pref}}\Delta y_n$ and $\frac{1}{|\mathcal{N}_o|}\sum_{n\in\mathcal{N}_o} z_n^{\text{pref}}\Delta x_n(\hat{\theta})$, respectively, and intercepts of zero, with $\mathcal{N}_o$ the set of observations associated with each outcome.

to reject them for a subset of such outcomes.[36] Table C.1 in Appendix C.2 shows that this is indeed the case.

Together with the results of Table 1, this finding therefore paints a picture of FGKK's quantitative model as "wrong, but useful." From a statistical standpoint, there are outcomes, even welfare-relevant ones, along which FGKK's predictions can be rejected. But from an economic standpoint, the average bias in the causal prediction of interest can nevertheless be viewed as modest in absolute terms, of the order of 0.09% of US GDP.

5 Concluding Remarks

Policymakers around the world will continue to face the choice to liberalize trade or not. They may choose to cut their tariffs unilaterally, like India and Brazil in the 1990s, form regional trade agreements, like the EU, NAFTA, and Mercosur, or join the World Trade Organization, like China in 2001. Or they may choose to raise their tariffs, as the Trump administration did in 2018. To provide guidance about these various policy choices, trade economists, like economists in many other areas, have developed quantitative models.

[36]Note that the relevant range of the x-axis in Figure 5a is much smaller than that in Figures 5b and 5c. Even though z_n^{pref} in each case is built from the same tariff shifters, the shares that enter the IV for the case of export prices in Panel a display much less variation because, in FGKK's model, tariff changes affect export prices in the same manner across all products within broad sectors. As such, all else equal, when examining export prices, one expects the power to detect a statistically significant gap between data-based and model-based projections to be considerably lower.

These models produce counterfactual predictions of how a country's economic conditions and ultimately welfare may change if, everything else being equal, a given policy were to be implemented.

To help assess and potentially strengthen the credibility of such quantitative predictions we have developed an IV-based goodness-of-fit measure and shown how it could be used for testing causal predictions in arbitrary general-equilibrium environments as well as for estimating the average misspecification in the prediction of interest. Our testing procedure gives center stage to the distinction between "the researcher's model being right or wrong" and "the answer to the counterfactual question of interest being right or wrong." It is purposefully designed to shed light on the latter rather than the former. Compared to a number of existing model validation procedures, our test recognizes that many other shocks, beside the changes in the policy of interest, may be driving the changes observed in the data and that such shocks may dilute the ability of the researcher's model to forecast future changes or backcast past ones. Interestingly, while we document that the correlation between FGKK's predictions and what actually happened in the data during Trump trade war's is low, we cannot reject that the answer of FGKK's model to the question being posed of it is correct.

Although we have focused here on a quantitative trade model and its implications for a well-known change in trade policy, it should be clear that the toolkit developed in this paper could be applied more broadly. The shock of interest, in particular, need not be policy-related. The focus may be instead on mergers that simultaneously affect multiple markets, as in the IO literature, or on productivity shocks that have been indirectly inferred from the researcher's model, rather than directly observed in the data, as in common in quantitative analysis of technological change in the macro and labor literatures or the analysis of infrastructure projects in the spatial literature. In all these settings, where the existence of general-equilibrium linkages means that existing quasi-experimental variation alone is insufficient to evaluate the overall causal effects of interest, we hope that our IV-based test can prove to be a useful add-on to test and ultimately improve the credibility of quantitative models' predictions.

References

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito, "General equilibrium effects in space: Theory and measurement," 2020. NBER Working paper 25544.

—, Michal Kolesar, and Eduardo Morales, "Shift-Share Designs: Theory and Inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

—, Paul Carillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz, "Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence," *Quarterly Journal of Economics*, 2022, *137* (3), 1553–1614.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nicolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, November 2015, *83* (6), 2127–2189.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro, "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 2020, *88* (6), 2231–2258.

Antras, Pol, Teresa C. Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, 2017, *107* (9), 2514–64.

Attanasio, Orazio, Pinelopi Goldberg, and Nina Pavcnik, "Trade Reforms and Wage Inequality in Colombia," *Journal of Development Economics*, 2004, *74* (2), 331–366.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson, "Common Ownership and Competition in the Ready-To-Eat Cereal Industry," 2021. NBER working paper 28350.

Berry, Steven T. and Philip A. Haile, "Identification in differentiated products markets using market level data," *Econometrica*, 2014, *82* (5), 1749–1797.

Billingsley, Patrick, *Probability and Measure*, third ed., John Wiley and Sons, 1995.

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Exogenous Shocks," 2022. NBER Working Paper 27845.

—, —, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

Bresnahan, Timothy F., "The Oligopoly Solution Concept is Identified," *Economics Letters*, 1982, *10*.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics Methods and Applications*, Cambridge University Press, 2005.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Monetary policy shocks: What have we learned and to what end?," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1, Part A Elsevier 1999, pp. 65–148.

_ , _ , and _ , "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113* (1), 1–45.

Costinot, Arnaud and Andres Rodríguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 4, New York: Elsevier, 2014.

_ , Dave Donaldson, and Cory Smith, "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from a 1.7 Million Fields Around the World," *Journal of Political Economy*, 2016, *124* (1), 205–248.

Council of Economic Advisers, *Economic Report of the President*, Council of Economic Advisers, 2020.

Davis, Donald R. and David Weinstein, "An Account of Global Factor Trade," *American Economic Review*, 2001, *91* (5), 1423–1453.

Dawkins, Christina, T.N. Srinivasan, and John Whalley, "Calibration," in J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, pp. 3653–3703.

Desmet, Klaus, David K. Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal of Political Economy*, 2018, *126* (3), 903–983.

Dingel, Jonathan and Felix Tintelnot, "Spatial Economics for Granular Settings," *NBER Working Paper 27287*, 2021.

Edmond, Chris, Virgiliu Midrigan, and Daniel Xu, "Competition, Markups and the Gains from Trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

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

Goldberg, Pinelopi K. and Nina Pavcnik, "The Effects of Trade Policy," in Kyle Bagwell and Robert W. Staiger, eds., *Handbook of Commercial Policy*, Vol. 1A, Elsevier, 2016.

Hansen, Lars Peter, "Generalized Method of Moments Estimation," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, second ed. 2008.

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

Hiltzik, Michael, "Trump's trade war is politically motivated, yet hurts consumers and GOP voters, study shows," *Los Angeles Times*, March 4 2019.

Imbens, Guido W. and Johsua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–469.

Kehoe, Timothy J., "An Evaluation of the Performance of Applied General Equilibrium Models of the Impact of NAFTA," in Timothy J. Kehoe, T.N. Srinivasan, and John Whalley, eds., *Frontiers in Applied General Equilibrium Modeling*, New York: Cambridge University Press, 2005, pp. 341–377.

_ **and Edward C. Prescott**, "Introduction to the Symposium: The discipline of applied general equilibrium," *Economic Theory*, 1995, *6*, 1–11.

_ **, Clemente Polo, and Ferran Sancho**, "An evaluation of the performance of an applied general equilibrium model of the spanish economy," *Economic Theory*, 1995, pp. 115–141.

_ **, Pau S. Pujolas, and Jack Rossbach**, "Quantitative Trade Models: Developments and Challenges," *Annual Review of Economics*, 2017, *9*, 295–325.

Kovak, Brian K., "Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization?," *American Economic Review*, 2013, *103* (5), 1960–76.

Lai, Huiwen and Daniel Trefler, "The Gains from Trade with Monopolistic Competition: Specification, Estimation, and Mis-Specification," *NBER Working Paper 9169*, 2002.

Lucas, Robert E., "Methods and problems in business cycle theory," *Journal of Money, Credit and Banking*, 1980, *12* (4), 696–715.

McCaig, Brian, "Exporting out of poverty: provincial poverty in Vietnam and U.S. market access," *Journal of International Economics*, 2011, *85* (1), 102–113.

Nakamura, Emi and Jon Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 2014, *104* (3), 753–792.

Redding, Stephen and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9*, 21–58.

Shoven, John B. and John Whalley, "Applied General-Equilibrium Models of Taxation and International Trade: An Introduction and Survey," *Journal of Economic Literature*, 1984, *22* (3), 1007–1051.

The Economist, "So far, Donald Trump's trade war has not derailed the global economy," *The Economist*, May 9 2019.

Todd, Petra E. and Kenneth I. Wolpin, "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 2006, *96* (5), 1384–1417.

Topalova, Petia, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, *2* (4), 1–41.

Wise, David A., "Behavioral Model versus Experimentation: The Effects of Housing Subsidies on Rent," in P. Brucker and R. Pauly, eds., *Methods of Operations Research*, Vol. 50, Konigstein: Verlag Anton Hain, 1985, pp. 441–489.

A Theoretical Appendix

A.1 Notation

All asymptotic results take the number of shifters K to infinity. We let $\rightarrow_d$ denote convergence in distribution, $\rightarrow_p$ denote convergence in probability, and $\rightarrow$ denote point-wise convergence. We write $X_K = o_p(1)$ if $X_K \rightarrow_p 0$. We say that X_K is uniformly bounded if there exists $M > 0$ such that $|X_K| \leq M$ for all K. We write $X_K \preceq Y_K$ if there exists $M > 0$ such that $X_K \leq MY_K$ for all K. Whenever there is no risk of confusion, we drop the subscript K from the variables we consider. We also write $\sum_n$ as a shorthand for $\sum_{n \in \mathcal{N}_W}$ unless otherwise specified.

A.2 Proof of Proposition 1

Proof. By Definition 1, $E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_n z_n(\Delta y_n - \Delta x_n)]$. Substituting for Δy_n using (3), we get

$$E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_n z_n(\Delta x_n^* - \Delta x_n)] + \frac{1}{N_W} \sum_n E_t[z_n \Delta \eta_n^*]. \tag{A.1}$$

A1 and A2 imply $E_t[z_n \Delta\eta_n^*] = \sum_k s_{nk} E_t[\Delta\tau_{IV,k}] E_t[\Delta\eta_n^*] = 0$. A3 implies $E_t[\sum_n z_n(\Delta x_n^* - \Delta x_n)] = 0$. Combining equation (A.1) with the two previous observations, we obtain $E_t[\hat{\beta}_z] = 0$. $\square$

A.3 Proof of Proposition 2

We split the proof of Proposition 2 into two lemmas. Lemma 1 focuses on the consistency of $\hat{\beta}_z$; its proof follows closely the arguments in Borusyak et al. (2022). Lemma 2 focuses on its asymptotic distribution; its proof follows closely the arguments in Adao et al. (2019).

Lemma 1 (Consistency). *Take any IV z that satisfies A1 and A2. If A3 holds and (i) $\Delta\tau_{IV,k}$ are i.i.d. across $k = 1, ..., K$, (ii) $\frac{1}{N_W^2} \sum_k (S_k)^2 \rightarrow 0$ with $S_k \equiv \sum_n |s_{nk}|$, and (iii) $Var_t[\Delta\tau_{IV,k}]$ and $\Delta\eta_n^*$ are uniformly bounded, then $\hat{\beta}_z \rightarrow_p 0$.*

Proof. Since z satisfies A1 and A2, the same argument as in the proof of Proposition 1 implies

$$E_t[\frac{1}{N_W} \sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*] = 0. \tag{A.2}$$

A2 further implies $\text{Var}_t[\frac{1}{N_W} \sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*] = \text{Var}_t[\frac{1}{N_W} \sum_k (\sum_n \Delta\eta_n^* s_{nk}) \Delta\tau_{IV,k} | \epsilon_{t+1}^*]$. Since $\Delta\tau_{IV,k}$ are i.i.d. across $k = 1, ..., K$, by condition (i), and independent of ϵ_{t+1}^* conditional on (ϵ_t^*, τ_t), by A1, we can rearrange the previous expression as

$$\text{Var}_t[\frac{1}{N_W} \sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*] = \frac{1}{N_W^2} \times \text{Var}_t[\Delta\tau_{IV,1}] \times \sum_k (\sum_n \Delta\eta_n^* s_{nk})^2 \preceq \frac{1}{N_W^2} \times \sum_k (S_k)^2, \tag{A.3}$$

where we have used $S_k \equiv \sum_n |s_{nk}|$ and the fact that $\text{Var}_t[\Delta\tau_{IV,1}]$ and $\Delta\eta_n^*$ are uniformly bounded, by condition (iii). Together with inequality (A.3), condition (ii) implies $\text{Var}_t[\frac{1}{N_W}\sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*] \to 0$. Combining this observation with equation (A.2), we get $E_t[(\frac{1}{N_W}\sum_n z_n \Delta\eta_n^*)^2 | \epsilon_{t+1}^*] \to 0$ and, in turn, $\frac{1}{N_W}\sum_n z_n \Delta\eta_n^* \to_p 0$. Since A3 implies $\hat{\beta}_z = \frac{1}{N_W}\sum_n z_n \Delta\eta_n^*$, Lemma 1 follows. □

Lemma 2 (Asymptotic Normality)**.** *Take any IV z that satisfies A1 and A2. If A3 holds and (i) $\Delta\tau_{IV,k}$ are i.i.d. across $k = 1, ..., K$, (ii) $Var_t[\Delta\tau_{IV,k}]$ and $\Delta\eta_n^*$ are uniformly bounded, (iii) $\frac{\max_k (S_k)^2}{\sum_k S_k^2} \to 0$, (iv) $E_t[(\Delta\tau_{IV,k})^4]$ is uniformly bounded, and (v) $\frac{1}{\sum_k S_k^2} Var_t[\sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*] \to_p \bar{V}_\beta > 0$ and non-random, then $r_\beta \hat{\beta}_z \to_d \mathcal{N}\left(0, \bar{V}_\beta\right)$ with $r_\beta \equiv N_W / \sqrt{\sum_k S_k^2}$.*

Proof. Start from the definitions of $\hat{\beta}_z$ and r_β, and use (3), A2, and A3 to write

$$r_\beta \hat{\beta}_z = \sum_k R_k \Delta\tau_{IV,k}, \tag{A.4}$$

$$R_k \equiv \frac{\sum_n s_{nk} \Delta\eta_n^*}{\sqrt{\sum_k S_k^2}}. \tag{A.5}$$

Define $\mathcal{Y}_k \equiv R_k \Delta\tau_{IV,k}$. Since $\Delta\tau_{IV,k}$ are independent of ϵ_{t+1}^* conditional on (τ_t, ϵ_t^*) and mean-zero, by A1, i.i.d. across k, by condition (i), $\mathcal{Y}_k$ are independent across k conditional on (τ_t, ϵ_t^*) and

$$E_t[\mathcal{Y}_k | \epsilon_{t+1}^*] = 0, \tag{A.6}$$

$$\sum_k \text{Var}_t[\mathcal{Y}_k | \epsilon_{t+1}^*] = \text{Var}_t[r_\beta \hat{\beta}_z | \epsilon_{t+1}^*] = \frac{1}{\sum_k S_k^2} \times \text{Var}_t[\sum_n z_n \Delta\eta_n^* | \epsilon_{t+1}^*]. \tag{A.7}$$

Note also that

$$\textstyle\sum_k E_t\left[\mathcal{Y}_k^4 | \epsilon_{t+1}^*\right] \quad = \quad \frac{\sum_k (\sum_n s_{nk} \Delta\eta_n^*)^4 E_t\left[(\Delta\tau_{IV,k})^4\right]}{(\sum_k S_k^2)^2} \preceq \frac{\sum_k (\sum_n s_{nk} \Delta\eta_n^*)^4}{(\sum_k S_k^2)^2}, \tag{A.8}$$

since $E_t[(\Delta\tau_{IV,k})^4]$ is uniformly bounded, by condition (iv), and $\Delta\tau_{IV,k}$ is independent of ϵ_{t+1}^* conditional on $(\tau_t, \epsilon_t^*,)$, by A1. Furthermore, we have

$$\frac{\sum_k (\sum_n s_{nk} \Delta\eta_n^*)^4}{(\sum_k S_k^2)^2} = \frac{\sum_k \left(\sum_i \sum_j s_{jk} s_{ik} \Delta\eta_i^* \Delta\eta_j^*\right)^2}{(\sum_k S_k^2)^2} \preceq \frac{\sum_k \left(\sum_i \sum_j |s_{jk}||s_{ik}|\right)^2}{(\sum_k S_k^2)^2}, \tag{A.9}$$

where the inequality $\preceq$ derives from the fact that $\Delta\eta_n^*$ is uniformly bounded, by condition (ii). By definition of S_k, $\sum_i \sum_j |s_{jk}||s_{ik}| = (S_k)^2$. Combining this observation with inequalities (A.8) and (A.9), we get

$$\sum_k E_t\left[\mathcal{Y}_k^4 | \epsilon_{t+1}^*\right] \preceq \frac{\sum_k (S_k)^4}{(\sum_k S_k^2)^2} \leq \frac{\max_k (S_k)^2}{\sum_k (S_k)^2}.$$

Together with condition (iii), this implies $\sum_k E_t\left[\mathcal{Y}_k^4 | \epsilon_{t+1}^*\right] \to 0$. Condition (v) and equation (A.7) further imply $\sum_k \text{Var}_t[\mathcal{Y}_k|\epsilon^*_{t+1}] \to_p \bar{V}_\beta > 0$. Combining the two previous limits, we obtain

$$\sum_k E_t \left[\frac{\mathcal{Y}_k^4}{(\sum_k \text{Var}_t[\mathcal{Y}_k|\epsilon^*_{t+1}])^2} |\epsilon^*_{t+1} \right] \to 0. \tag{A.10}$$

We can therefore invoke the Lyapunov Central Limit Theorem (e.g. Billingsley, 1995, Theorem 27.3, p. 362) to conclude that $Z_K \equiv \sum_k \mathcal{Y}_k / \sqrt{\sum_k \text{Var}_t[\mathcal{Y}_k|\epsilon^*_{t+1}]} \to_d N(0,1)$. By equations (A.4) and (A.7), we know that $r_\beta \hat{\beta}_z = Z_K \sqrt{\text{Var}_t[r_\beta \hat{\beta}_z|\epsilon^*_{t+1}]}$. Since $\text{Var}_t[r_\beta \hat{\beta}_z|\epsilon^*_{t+1}] = \sum_k \text{Var}_t[\mathcal{Y}_k|\epsilon^*_{t+1}] \to_p \bar{V}_\beta$ non-random, Lemma 2 follows. $\square$

A.4 Asymptotic Results with Controls

In this appendix we provide generalizations of Propositions 1 and 2 that only require the exogeneity of the policy shifters after controls have been included. We also discuss how the introduction of controls affects the asymptotic variance of our test statistic.

Assumptions. Suppose that non-tariff shocks can be decomposed into a linear combination of structural shocks, $\{\nu_j^*\}$, and a residual, $(\Delta\eta_n^*)_{\text{res}}$,

$$\Delta\eta_n^* = \sum_j c_{nj}\nu_j^* + (\Delta\eta_n^*)_{\text{res}}, \tag{A.11}$$

where the controls $\{c_{nj}\}$ are observable and only depend on (ϵ_t^*, τ_t), and the structural shocks $\{\nu_j^*\}$ are unobserved and functions of ϵ^*_{t+1}. In this environment, one can relax A1 and instead impose the following exogeneity assumption.

A1$_{\text{controls}}$. *Conditional on the realization of period t's tariffs and other shocks, policy shifters are mean zero and independent of residualized shocks:* $\Delta\tau_{IV} \perp\!\!\!\perp (\Delta\eta^*)_{res}|\epsilon_t^*, \tau_t$.

Note that A1$_{\text{controls}}$ allows for the policy shifters $\Delta\tau_{IV}$ to be correlated with the structural shocks $\{\nu_j^*\}$ that enter (A.11) and, in turn, for $E_t[\Delta\tau_{IV,k}\Delta\eta_n^*] \neq 0$. However, A1$_{\text{controls}}$ requires that after residualizing non-tariff shocks with respect to the vector of controls $\{c_{nj}\}$, we have

$$E_t[\Delta\tau_{IV,k}(\Delta\eta_n^*)_{\text{res}}] = E_t[\Delta\tau_{IV,k}]E_t[(\Delta\eta_n^*)_{\text{res}}] = 0.$$

In what follows we let $C \equiv \{c_{nj}\}$ denote the matrix of controls and $P_c \equiv Id - C(C^TC)^{-1}C^T$ the associated residual projection matrix, with the superscript T denoting a transpose matrix. Starting from a given candidate IV $z \equiv M_s\Delta\tau_{IV}$, where $M_s \equiv \{s_{nk}\}$ denotes the matrix of shares, we can define the residualized IV $(z)_{\text{res}} \equiv P_c z = M_{(s)_{\text{res}}}\Delta\tau_{IV}$, where $M_{(s)_{\text{res}}} \equiv \{(s_{nk})_{\text{res}}\} = P_c M_s$ is the matrix of residualized shares.

Proof of Proposition 1 with Controls.

Proposition 1 (with controls). *Suppose that non-tariff shocks satisfy equation* (A.11) *and that* $A1_{controls}$ *holds. Starting from a candidate IV* $z \equiv M_s \Delta\tau_{IV}$, *take the residualized IV* $(z)_{res} = M_{(s)_{res}} \Delta\tau_{IV}$, *with* $M_{(s)_{res}} = P_c M_s$ *the matrix of residualized shares. If A3 holds, then* $E_t[\hat{\beta}_{(z)_{res}}] = 0$.

Proof. By construction, we have

$$N_W \hat{\beta}_{(z)_{\text{res}}} = (\Delta y - \Delta x) \cdot (z)_{\text{res}}$$
$$= (\Delta x^* - \Delta x) \cdot (z)_{\text{res}} + (C\nu^*) \cdot (z)_{\text{res}} + (\Delta\eta_n^*)_{\text{res}} \cdot (z)_{\text{res}}$$
$$= (\Delta x^* - \Delta x) \cdot (z)_{\text{res}} + (\Delta\eta_n^*)_{\text{res}} \cdot (z)_{\text{res}},$$

where the second equality follows from (A.11) and the third from $(C\nu^*) \cdot (z)_{\text{res}} = (\nu^*)^T C^T P_c z = 0$. The previous equation can be rearranged as

$$\hat{\beta}_{(z)_{\text{res}}} = \frac{1}{N_W} \sum_n (z_n)_{\text{res}} (\Delta x_n^* - \Delta x_n) + \frac{1}{N_W} \sum_n (z_n)_{\text{res}} (\Delta\eta_n^*)_{\text{res}}. \tag{A.12}$$

Under $\text{A1}_{\text{controls}}$, we have $E_t[(z_n)_{\text{res}}(\Delta\eta_n^*)_{\text{res}}] = \sum_k (s_{nk})_{\text{res}} E_t[\Delta\tau_{IV,k}(\Delta\eta_n^*)_{\text{res}}] = 0$. Taking expectation of (A.12) we therefore get

$$E_t[\hat{\beta}_{(z)_{\text{res}}}] = \frac{1}{N_W} \sum_n E_t[(\Delta x_n^* - \Delta x_n)(z_n)_{\text{res}}].$$

If A3 holds, we therefore obtain $E_t[\hat{\beta}_{(z)_{\text{res}}}] = 0$. □

Proof of Proposition 2 with Controls.

Proposition 2 (with controls). *Suppose that non-tariff shocks satisfy equation* (A.11) *and that* $A1_{controls}$ *holds. Starting from a candidate IV* $z \equiv M_s \Delta\tau_{IV}$, *take the residualized IV* $(z)_{res} = M_{(s)_{res}} \Delta\tau_{IV}$, *with* $M_{(s)_{res}} = P_c M_s$ *the matrix of residualized shares. If A3 holds and* (i) $\Delta\tau_{IV,k}$ *are i.i.d. across* $k = 1, ..., K$, (ii) $\frac{1}{N_W^2} \sum_k (S_k)^2_{res} \to 0$ *with* $(S_k)_{res} \equiv \sum_n |(s_{nk})_{res}|$, *and* (iii) $Var_t[\Delta\tau_{IV,k}]$ *and* $(\Delta\eta_n^*)_{res}$ *are uniformly bounded, then* $\hat{\beta}_{(z)_{res}} \to_p 0$. *Furthermore, if* (iv) $\frac{\max_k (S_k)^2_{res}}{\sum_k (S_k)^2_{res}} \to 0$, (v) $E_t[(\Delta\tau_{IV,k})^4]$ *is uniformly bounded, and* (vi) $\frac{1}{\sum_k (S_k)^2_{res}} Var_t[\sum_n (z_n)_{res} (\Delta\eta_n^*)_{res} | \epsilon_{t+1}^*] \to_p \bar{V}_{\beta_{res}} > 0$ *and non-random, then* $r_{\beta_{res}} \hat{\beta}_{(z)_{res}} \to_d \mathcal{N}(0, \bar{V}_{\beta_{res}})$ *with* $r_{\beta_{res}} \equiv N_W / \sqrt{\sum_k (S_k)^2_{res}}$.

Proof. The same argument as in the proof of Proposition 1 implies

$$E_t[\frac{1}{N_W} \sum_n (z_n)_{\text{res}} (\Delta\eta_n^*)_{\text{res}} | \epsilon_{t+1}^*] = 0.$$

Given the regularity conditions (i)-(iii), $\hat{\beta}_{(z)_{\text{res}}} \to_p 0$ follows from the same argument as in the proof of Lemma 1. Next, starting from equation (A.12) and the definition of $r_{\beta_{\text{res}}}$, we can use (3), (A.11), $(z)_{\text{res}} = M_{(s)_{\text{res}}} \Delta\tau_{IV}$, and A3 to write

$$r_{\beta_{\text{res}}} \hat{\beta}_{(z)_{\text{res}}} = \sum_k (R_k)_{\text{res}} \Delta\tau_{IV,k},$$
$$(R_k)_{\text{res}} \equiv \frac{\sum_n (s_{nk})_{\text{res}} (\Delta\eta_n^*)_{\text{res}}}{\sqrt{\sum_k (S_k)^2_{\text{res}}}}.$$

Given the regularity conditions (i)-(vi), the result that $r_{\beta_{\text{res}}} \hat{\beta}_{(z)_{\text{res}}} \to_d \mathcal{N}(0, \bar{V}_{\beta_{\text{res}}})$ then follows from the same argument as in the proof of Lemma 2. □

Asymptotic Variance with Controls. Under the previous assumptions, the estimator $\hat{V}[\hat{\beta}_{(z)_{\text{res}}}]$ of the variance of the test using the residualized IV $(z)_{\text{res}}$ is, in turn, equal to

$$\hat{V}[\hat{\beta}_{(z)_{\text{res}}}] = \sum_k (\Delta\tau_{IV,k})^2 [\sum_n (s_{nk})_{\text{res}} (\Delta y_n - \Delta x_n)/N_W]^2. \tag{A.13}$$

Note that

$$\begin{aligned}
\hat{V}[\hat{\beta}_{(z)_{\text{res}}}] &= \sum_k (\Delta\tau_{IV,k})^2 [\sum_n (s_{nk})_{\text{res}} \Delta\eta_n^* / N_W]^2 \\
&= \sum_k (\Delta\tau_{IV,k})^2 [\sum_n s_{nk} (\Delta\hat{\eta}_n^*)_{\text{res}} / N_W]^2 \\
&\leq \sum_k (\Delta\tau_{IV,k})^2 [\sum_n s_{nk} (\sum_j c_{nj} \hat{v}_j^* + (\Delta\hat{\eta}_n^*)_{\text{res}}) / N_W]^2 = \hat{V}[\hat{\beta}_z],
\end{aligned}$$

where $\hat{v}_j^*$ and $(\Delta\hat{\eta}_n^*)_{\text{res}}$ denote the OLS coefficients and fitted residuals, respectively, of a regression of $\Delta\eta_n^* = \Delta y_n - \Delta x_n$ on the vector of controls $\{c_{nj}\}$. Thus the asymptotic variance of the residualized IV-based test is lower than that of the non-residualized version $\hat{V}[\hat{\beta}_z]$ displayed in equation (6).

A.5 Estimation Uncertainty

Suppose that the researcher's model f is only known up to the estimation of a vector of structural parameters, $f(\tau_t, \epsilon_t) \equiv g(\tau_t, \epsilon_t|\theta) \neq g(\tau_t, \epsilon_t|\hat{\theta})$, where θ denotes the true value of the structural parameters and $\hat{\theta}$ denotes its estimator. We let $\Delta x(\hat{\theta}) \equiv g(\tau_{t+1}, \epsilon_{t+1}|\hat{\theta}) - g(\tau_t, \epsilon_{t+1}|\hat{\theta})$ denote the causal impact of the policy change predicted by the researcher's model when evaluated at $\hat{\theta}$. In turn, the counterpart of the goodness-of-fit measure introduced in Definition 1 is

$$\hat{\beta}_z(\hat{\theta}) \equiv \frac{1}{N_W} \sum_{n \in \mathcal{N}_W} z_n (\Delta y_n - \Delta x_n(\hat{\theta})).$$

In Proposition 2, we have characterized the asymptotic distribution of $\hat{\beta}_z = \hat{\beta}_z(\theta)$. In this appendix, we characterize the asymptotic distribution of $\hat{\beta}_z(\hat{\theta})$ when $\hat{\theta}$ is itself a random variable.

Case 1: $\hat{\theta}$ is independent of $\hat{\beta}_z(\theta)$. This may occur, for instance, if $\hat{\theta}$ is obtained from a different sample than the one used for testing. Formally, we impose the following assumptions.

A4. *The estimator $\hat{\theta}$ (i) is independent from $\hat{\beta}_z(\theta)$, (ii) satisfies $r_\theta(\hat{\theta}-\theta) \to_d \mathcal{N}(0, \bar{V}_\theta)$, with (iii) the rate of convergence r_θ such that $r_\beta / r_\theta \to_p c$ non-random.*

A5. *The gradient $D(\hat{\beta}_z(\theta)) \equiv \{\partial \hat{\beta}_z(\theta)/\partial \theta_j\}$ satisfies $D(\hat{\beta}_z(\theta)) \to_p \bar{D}_\beta$.*

When $\hat{\theta}$ and $\hat{\beta}_z(\theta)$ are independent, Proposition 2 extends as follows.

Proposition 2 (with estimation uncertainty, case 1)**.** *Suppose that, in addition to the assumptions of Proposition 2, A4 and A5 hold. Then $r_\beta \hat{\beta}_z(\hat{\theta}) \to_d \mathcal{N}\left(0, \bar{V}_\beta + c^2 \bar{D}_\beta \bar{V}_\theta \bar{D}_\beta^T\right)$.*

Proof. A first-order Taylor expansion implies

$$\begin{aligned} r_\beta \hat{\beta}_z(\hat{\theta}) &= r_\beta \hat{\beta}_z(\theta) + r_\beta \left(\hat{\beta}_z(\hat{\theta}) - \hat{\beta}_z(\theta)\right) \\ &= r_\beta \hat{\beta}_z(\theta) + r_\beta D(\hat{\beta}_z(\theta)) \cdot \left(\hat{\theta} - \theta\right) + o_p(1). \end{aligned}$$

Combining this expression with $D(\hat{\beta}_z(\theta)) \to_p \bar{D}_\beta$, by A5, and using $r_\beta / r_\theta \to_p c$, by A4 (iii), we get

$$r_\beta \hat{\beta}_z(\hat{\theta}) = r_\beta \hat{\beta}_z(\theta) + c\bar{D}_\beta \cdot [r_\theta(\hat{\theta} - \theta)] + o_p(1). \tag{A.14}$$

Since $r_\beta \hat{\beta}_z(\theta)$ is asymptotically normally distributed under the assumptions of Proposition 2, $r_\theta(\hat{\theta}-\theta)$ is asymptotically normally distributed under A4 (ii), and $\hat{\theta}$ and $\hat{\beta}_z(\theta)$ are independent, under A4 (i), we have

$$\begin{bmatrix} r_\theta(\hat{\theta} - \theta) \\ r_\beta \hat{\beta}_z(\theta) \end{bmatrix} \to_d \mathcal{N}\left(0, \begin{bmatrix} \bar{V}_\theta & 0 \\ 0 & \bar{V}_\beta \end{bmatrix}\right). \tag{A.15}$$

Given (A.14) and (A.15), we can invoke the Multivariate Convergence in Distribution Theorem (e.g. Rao 1973, p. 128) to conclude that $r_\beta \hat{\beta}_z(\hat{\theta}) \to_d \mathcal{N}\left(0, \bar{V}_\beta + c^2 \bar{D}_\beta \bar{V}_\theta \bar{D}_\beta^T\right)$. □

Under the previous assumptions, we can then estimate the asymptotic variance of $\hat{\beta}_z(\hat{\theta})$ as

$$\hat{V}[\hat{\beta}_z(\hat{\theta})] = \hat{\mathcal{V}}_\beta + \hat{D}_\beta \hat{\mathcal{V}}_\theta \hat{D}_\beta^T, \tag{A.16}$$

$$\hat{\mathcal{V}}_\beta = \sum_k (\Delta \tau_{IV,k})^2 [\sum_{n \in \mathcal{N}_W} s_{nk}(\Delta y_n - \Delta x_n)/N_W]^2, \tag{A.17}$$

$$\hat{D}_\beta = D(\hat{\beta}_z(\hat{\theta})), \tag{A.18}$$

where $\hat{\mathcal{V}}_\theta$ is a consistent estimator of the asymptotic variance of $\hat{\theta}$.

Case 2: $\hat{\theta}$ is an IV estimator. For expositional purposes, we assume that θ is a scalar and that there is a single estimating moment $\hat{\gamma}(\theta)$ such that

$$\hat{\gamma}(\theta) = \frac{1}{N_\theta} \sum_{n \in \mathcal{N}_\theta} z_n^\theta e_n(w, \theta),$$

where $N_\theta \equiv |\mathcal{N}_\theta|$ denotes the number of observations used for estimation; $w \equiv (y_t, y_{t+1}, \tau_t, \tau_{t+1}, \Delta\tau_{IV})$ denotes the vector of all observables; $e_n(w, \theta)$ is a structural residual that only depends on the realizations of $(\epsilon_t^*, \epsilon_{t+1}^*, \tau_t)$; and $z^\theta \equiv \{z_n^\theta\}$ is the IV used in the estimation of θ. Like in Section 2, we express the IV in a shift-share form and assume that it satisfies the two following conditions.[37]

A1′. *Conditional on the realization of period t's policy τ_t and other shocks ϵ_t^*, policy shifters used in estimation are mean zero and independent of other shocks in period $t+1$: $\Delta\tau_{IV}^\theta \perp\!\!\!\perp \epsilon_{t+1}^* | \epsilon_t^*, \tau_t$.*

A2′. *For any $n \in \mathcal{N}_\theta$, the instrumental variable used in estimation takes the form $z_n^\theta = \sum_k s_{nk}^\theta \Delta\tau_{IV,k}^\theta$, where the vector of "shares" $\{s_{nk}^\theta\}$ may be a function of, and only of, the realization of period t's tariffs and other shocks, (ϵ_t^*, τ_t).*

A1′ and A2′ imply $E[\hat{\gamma}(\theta)] = 0$, for the same reasons as in Proposition 1. We assume that $\hat{\theta}$ is given by the unique solution to its sample analog, $\hat{\gamma}(\hat{\theta}) = 0$. In line with our previous analysis, we further impose the following conditions.

A4′. *The estimating moment $\hat{\gamma}(\theta)$ is such that: (i) $\Delta\tau_{IV,k}^\theta$ are i.i.d. across $k = 1, ..., K$; (ii) $\frac{1}{N_\gamma^2} \sum_k (S_k^\theta)^2 \to 0$ and $\frac{\max_k (S_k^\theta)^2}{\sum_k (S_k^\theta)^2} \to 0$, with $S_k^\theta \equiv \sum_n |s_{nk}^\theta|$; (iii) e_n and $E_t[(\Delta\tau_{IV,k}^\theta)^4]$ are uniformly bounded; (iv) the rate of convergence $r_\gamma \equiv N_\theta / \sqrt{\sum_k (S_k^\theta)^2}$ satisfies $r_\beta / r_\gamma \to_p c$ non-random; and (v) $cov_t[(r_\gamma \hat{\gamma}(\theta), r_\beta \hat{\beta}_z(\theta))' | \epsilon_{t+1}^*] \to_p \begin{pmatrix} \bar{V}_\gamma & \bar{V}_{\beta\gamma} \\ \bar{V}_{\beta\gamma} & \bar{V}_\beta \end{pmatrix} \equiv \bar{V} > 0$ and non-random.*

A5′. *The derivatives $D(\hat{\beta}_z(\theta)) \equiv d\hat{\beta}_z(\theta)/d\theta$ and $D(\hat{\gamma}(\theta)) \equiv d\hat{\gamma}(\theta)/d\theta$ satisfy $D(\hat{\beta}_z(\theta)) \to_p \bar{D}_\beta$ and $D(\hat{\gamma}(\theta)) \to_p \bar{D}_\gamma \neq 0$.*

Under these assumptions, Proposition 2 extends as follows.

Proposition 2 (with estimation uncertainty, case 2)**.** *Suppose that, in addition to the assumptions of Proposition 2, A1′, A2′, A4′ and A5′ hold. Then $r_\beta \hat{\beta}_z(\hat{\theta}) \to_d \mathcal{N}\left(0, \bar{V}_\beta + c^2 \bar{D}^2 \bar{V}_\gamma + 2c\bar{D}\bar{V}_{\beta\gamma}\right)$ with $\bar{D} \equiv -\bar{D}_\beta / \bar{D}_\gamma$.*

Proof. First-order Taylor expansions of $\hat{\beta}(\hat{\theta})$ and $\hat{\gamma}(\hat{\theta})$ imply

$$\begin{aligned} r_\beta \hat{\beta}_z(\hat{\theta}) &= r_\beta \hat{\beta}_z(\theta) + r_\beta \left(\hat{\beta}_z(\hat{\theta}) - \hat{\beta}_z(\theta)\right) \\ &= r_\beta \hat{\beta}_z(\theta) + r_\beta [D(\hat{\beta}_z(\theta))] \left(\hat{\theta} - \theta\right) + o_p(1) \\ &= r_\beta \hat{\beta}_z(\theta) - \frac{r_\beta}{r_\gamma} [D(\hat{\beta}_z(\theta)) / D(\hat{\gamma}(\theta))] (r_\gamma \hat{\gamma}(\theta)) + o_p(1). \end{aligned}$$

Combining this expression with $-D(\hat{\beta}_z(\theta))/D(\hat{\gamma}(\theta)) \to_p \bar{D}$, by A5′, and $r_\beta / r_\gamma \to_p c$, by A4′ (iv), we get

$$r_\beta \hat{\beta}_z(\hat{\theta}) = r_\beta \hat{\beta}_z(\theta) + c\bar{D}(r_\gamma \hat{\gamma}(\theta)) + o_p(1). \tag{A.19}$$

[37] It should be clear that any given IV can always be expressed as the special case of a shift-share IV with shifters equal to the IV itself $\{z_n^\theta\}$ and shares equal to dummy variables $\{1[n = k]\}$. This will be the case in FGKK's estimation procedure and in our application in Sections 3 and 4.

Let us now show that $r_\beta \hat{\beta}_z(\theta) + c\bar{D}(r_\gamma \hat{\gamma}(\theta))$ is normally distributed. The steps are the same as in the proof of Proposition 2. First, rearrange the previous expression as a function of the shifters,

$$r_\beta \hat{\beta}_z + c\bar{D}(r_\gamma \hat{\gamma}(\theta)) = \sum_k \mathcal{Y}_k, \quad \mathcal{Y}_k \equiv R_{\beta,k}\Delta\tau_{IV,k} + \lambda_\gamma R_{\gamma,k}\Delta\tau^\theta_{IV,k}, \tag{A.20}$$

with $R_{\beta,k} \equiv \sum_n s_{nk}\Delta\eta^*_n / \sqrt{\sum_k S_k^2}$, $R_{\gamma,k} \equiv \sum_n s^\theta_{nk} e_n / \sqrt{\sum_k (S^\theta_k)^2}$, and $\lambda_\gamma \equiv c\bar{D}$. Since $(\Delta\tau_{IV,k}, \Delta\tau^\theta_{IV,k})$ are mean-zero and independent of ϵ^*_{t+1} conditional on (τ_t, ϵ^*_t), by A1 and A1′, i.i.d. across k, by Proposition 2's condition (i) and A4′ (i), $\mathcal{Y}_k$ must be independent across k conditional on (τ_t, ϵ^*_t) and satisfy

$$E_t[\mathcal{Y}_k | \epsilon^*_{t+1}] = 0, \tag{A.21}$$

$$\sum_k \mathrm{Var}_t[\mathcal{Y}_k | \epsilon^*_{t+1}] = \mathrm{Var}_t[r_\beta \hat{\beta}_z + \lambda_\gamma (r_\gamma \hat{\gamma}(\theta)) | \epsilon^*_{t+1}]. \tag{A.22}$$

Note also that

$$\begin{aligned}
\sum_k E_t[\mathcal{Y}_k^4 | \epsilon^*_{t+1}] = \sum_k (& R^4_{\beta,k} E_t[(\Delta\tau_{IV,k})^4] + \lambda^4_\gamma R^4_{\gamma,k} E_t[(\Delta\tau^\theta_{IV,k})^4] + 6\lambda^2_\gamma R^2_{\beta,k} R^2_{\gamma,k} E_t[(\Delta\tau_{IV,k})^2 (\Delta\tau^\theta_{IV,k})^2] \\
& + 4\lambda_\gamma R^3_{\beta,k} R_{\gamma,k} E_t[(\Delta\tau_{IV,k})^3 (\Delta\tau^\theta_{IV,k})] + 4\lambda^3_\gamma R_{\beta,k} R^3_{\gamma,k} E_t[(\Delta\tau_{IV,k})(\Delta\tau^\theta_{IV,k})^3]) \\
& \preceq \sum_k (R^4_{\beta,k} + \lambda^4_\gamma R^4_{\gamma,k} + 6\lambda^2_\gamma R^2_{\beta,k} R^2_{\gamma,k} + 4\lambda_\gamma R^3_{\beta,k} R_{\gamma,k} + 4\lambda^3_\gamma R_{\beta,k} R^3_{\gamma,k}),
\end{aligned} \tag{A.23}$$

since $E_t[(\Delta\tau_{IV,k})^4]$ is uniformly bounded, by Proposition 2's condition (v); $E_t[(\Delta\tau^\theta_{IV,k})^4]$ is uniformly bounded by A4′; $E_t[(\Delta\tau_{IV,k})^2(\Delta\tau^\theta_{IV,k})^2]$, $E_t[(\Delta\tau_{IV,k})^3(\Delta\tau^\theta_{IV,k})]$), and $E_t[(\Delta\tau_{IV,k})^3(\Delta\tau_{IV,k})]$ are uniformly bounded by Hölder's inequality; and $(\Delta\tau_{IV,k}, \Delta\tau^\theta_{IV,k})$ are independent of ϵ^*_{t+1} conditional on $(\tau_t, \epsilon^*_t,)$, by A1 and A1′. Using the same strategy as in Proposition 2, we can then bound each of the terms on the right-hand side of (A.23),

$$\sum_k R^4_{\beta,k} = \frac{\sum_k (\sum_n s_{nk}\Delta\eta^*_n)^4}{(\sum_k S_k^2)^2} \preceq \frac{\sum_k \left(\sum_i \sum_j |s_{jk}||s_{ik}|\right)^2}{(\sum_k S_k^2)^2} \leq \frac{\max_k (S_k)^2}{\sum_k (S_k)^2},$$

$$\sum_k R^4_{\gamma,k} = \frac{\sum_k (\sum_n s^\theta_{nk} e_n)^4}{(\sum_k (S^\theta_k)^2)^2} \preceq \frac{\sum_k \left(\sum_i \sum_j |s^\theta_{jk}||s^\theta_{ik}|\right)^2}{(\sum_k (S^\theta_k)^2)^2} \leq \frac{\max_k (S^\theta_k)^2}{\sum_k (S^\theta_k)^2},$$

$$\sum_k R^2_{\beta,k} R^2_{\gamma,k} = \sum_k \frac{(\sum_n s_{nk}\Delta\eta^*_n)^2}{\sum_k S_k^2} \frac{(\sum_n s^\theta_{nk} e_n)^2}{\sum_k (S^\theta_k)^2} \preceq \sum_k \frac{(\sum_n |s_{nk}|)^2}{\sum_k S_k^2} \frac{(\sum_n |s^\theta_{nk}|)^2}{\sum_k (S^\theta_k)^2} \leq \frac{\max_k (S^\theta_k)^2}{\sum_k (S^\theta_k)^2},$$

$$\sum_k R^3_{\beta,k} R_{\gamma,k} = \sum_k \frac{(\sum_n s_{nk}\Delta\eta^*_n)^3}{(\sum_k S_k^2)^{3/2}} \frac{(\sum_n s^\theta_{nk} e_n)}{(\sum_k (S^\theta_k)^2)^{1/2}} \preceq \sum_k \frac{(\sum_n |s_{nk}|)^3}{(\sum_k S_k^2)^{3/2}} \frac{(\sum_n |s^\theta_{nk}|)}{(\sum_k (S^\theta_k)^2)^{1/2}} \leq \sqrt{\frac{\max_k S_k^2 \max_k (S^\theta_k)^2}{\sum_k S_k^2 \sum_k (S^\theta_k)^2}},$$

$$\sum_k R_{\beta,k} R^3_{\gamma,k} = \sum_k \frac{(\sum_n s_{nk}\Delta\eta^*_n)}{(\sum_k S_k^2)^{1/2}} \frac{(\sum_n s^\theta_{nk} e_n)^3}{(\sum_k (S^\theta_k)^2)^{3/2}} \preceq \sum_k \frac{(\sum_n |s_{nk}|)}{(\sum_k S_k^2)^{1/2}} \frac{(\sum_n |s^\theta_{nk}|)^3}{(\sum_k (S^\theta_k)^2)^{3/2}} \leq \sqrt{\frac{\max_k S_k^2 \max_k (S^\theta_k)^2}{\sum_k S_k^2 \sum_k (S^\theta_k)^2}}.$$

The first inequality for each of these terms derives from the fact that $\Delta\eta_n^*$ and e_n are uniformly bounded, by Proposition 2's condition (iii) and A4' (iii). Together with Proposition 2's condition (iv) and A4' (iv), the previous series of inequalities implies $\sum_k E_t\left[\mathcal{Y}_k^4|\epsilon_{t+1}^*\right] \to 0$. Together with A4' (v), equation (A.22) further implies $\sum_k \mathrm{Var}_t[\mathcal{Y}_k|\epsilon_{t+1}^*] \to_p \bar{V}_\beta + (\lambda_\gamma)^2\bar{V}_\gamma + 2\lambda_\gamma\bar{V}_{\beta\gamma} > 0$. This allows us to invoke the Lyapunov Central Limit Theorem (e.g. Billingsley, 1995, Theorem 27.3, p. 362) to conclude that $Z_K \equiv \sum_k \mathcal{Y}_k / \sqrt{\sum_k \mathrm{Var}_t[\mathcal{Y}_k|\epsilon_{t+1}^*]} \to_d \mathcal{N}(0,1)$. Since $r_\beta\hat{\beta}_z + c\bar{D}(r_\gamma\hat{\gamma}(\theta)) = Z_K\sqrt{\sum_k \mathrm{Var}_t[\mathcal{Y}_k|\epsilon_{t+1}^*]}$, by equation (A.20), and $\sum_k \mathrm{Var}_t[\mathcal{Y}_k|\epsilon_{t+1}^*] \to_p \bar{V}_\beta + (\lambda_\gamma)^2\bar{V}_\gamma + 2\lambda_\gamma\bar{V}_{\beta\gamma} > 0$, Proposition 2 follows. □

Under the previous assumptions, we can estimate the asymptotic variance $\hat{V}[\hat{\beta}_z(\hat{\theta})]$ of the goodness-of-fit measure $\hat{\beta}_z(\hat{\theta})$ as

$$\hat{V}[\hat{\beta}_z(\hat{\theta})] = \hat{\mathcal{V}}_\beta + \hat{D}^2\hat{\mathcal{V}}_\gamma + 2\hat{D}\hat{\mathcal{V}}_{\beta\gamma}, \tag{A.24}$$

$$\hat{\mathcal{V}}_\beta = \sum_k [(\Delta\tau_{IV,k}) \sum_{n\in\mathcal{N}_W} s_{nk}(\Delta y_n - \Delta x_n(\hat{\theta}))/N_W]^2, \tag{A.25}$$

$$\hat{\mathcal{V}}_\gamma = \sum_k [(\Delta\tau_{IV,k}^\theta) \sum_{n\in\mathcal{N}_\theta} s_{nk}^\theta e_n(w;\hat{\theta})/N_\theta]^2, \tag{A.26}$$

$$\hat{\mathcal{V}}_{\beta\gamma} = \sum_k [(\Delta\tau_{IV,k}) \sum_{n\in\mathcal{N}_W} s_{nk}(\Delta y_n - \Delta x_n(\hat{\theta}))/N_W][(\Delta\tau_{IV,k}^\theta) \sum_{n\in\mathcal{N}_\theta} s_{nk}^\theta e_n(w;\hat{\theta})/N_\theta], \tag{A.27}$$

$$\hat{D} = -[D(\hat{\beta}_z(\hat{\theta}))][D(\hat{\gamma}(\hat{\theta}))]^{-1}. \tag{A.28}$$

A.6 Proof of Proposition 3

Baseline without controls. We first focus on the case without controls, as described in the main text of Section 2.4.

Proof. First, note that if $\tilde{z}$ satisfies A1 and A2, then z satisfies A1 and A2 as well, since $z_n = \sum_k s_{nk}\Delta\tau_k^{IV}$ with $s_{nk} \equiv N_W\omega_n\tilde{s}_{nk}E_t[\Delta x_n]/E_t[\tilde{z}_n\Delta x_n]$. Next, note that the same argument as in the proof of Proposition 1 implies

$$E_t[\beta_z] = \frac{1}{N_W}E_t[\sum\nolimits_n z_n(\Delta x_n^* - \Delta x_n)]. \tag{A.29}$$

Under A3', we know that $\Delta x_n^* - \Delta x_n = (\alpha_n - 1)\Delta x_n$. Combining this observation with (A.29) and $z_n = N_W\omega_n\tilde{z}_n E_t[\Delta x_n]/E_t[\tilde{z}_n\Delta x_n]$, we get

$$E_t[\beta_z] = \sum\nolimits_n \omega_n(\alpha_n - 1)E_t[\frac{E_t[\Delta x_n]}{E_t[\tilde{z}_n\Delta x_n]}\tilde{z}_n\Delta x_n] = E_t[\sum\nolimits_n \omega_n(\alpha_n - 1)\Delta x_n] = E_t[W(\Delta x^*) - W(\Delta x)].$$

□

Extension with controls. We now show how the proof of Proposition 3 can be adapted to deal with the case of linear controls. Like in Appendix A.4, we assume that non-tariff shocks satisfy equation (A.11) and that policy shifters satisfy $A1_{\text{controls}}$.

In order to state our next result, it is useful to introduce the following notation. For any vector $v \equiv \{v_n\}$, we let Diag_v denote the diagonal matrix with diagonal element v_n in row n. And like in Appendix A.4, we let $P_c \equiv Id - C(C^T C)^{-1} C^T$ denote the residual projection matrix associated with the controls $C \equiv \{c_{nj}\}$.

Proposition 3 (with controls)**.** *Suppose that non-tariff shocks satisfy equation* (A.11) *and that $A1_{controls}$ and A3′ hold. Then starting from a candidate IV $\tilde{z} \equiv M_{\tilde{s}} \Delta \tau_{IV}$, one can construct an adjusted IV $z \equiv \{\zeta_n \tilde{z}_n\}$, with the adjustment vector $\zeta \equiv N_W (E_t[Diag_{\Delta x} P_c Diag_{\tilde{z}}])^{-1} (Diag_\omega E_t[\Delta x])$, and a residualized IV $(z)_{res} \equiv P_c z$ such that $E_t[\hat{\beta}_{(z)_{res}}] = E_t[W(\Delta x^*) - W(\Delta x)]$.*

Proof. The same argument as in the proof of Proposition 1 with controls in Appendix A.4 implies

$$E_t[\beta_{(z)_{\text{res}}}] = \frac{1}{N_W} E_t[\textstyle\sum_n (z_n)_{\text{res}} (\Delta x_n^* - \Delta x_n)].$$

Using A3′ and the definition of $(z)_{\text{res}}$ and z, this can be rearranged as

$$E_t[\beta_{(z)_{\text{res}}}] = \frac{1}{N_W} (\alpha - 1) \cdot E_t[\text{Diag}_\delta \Delta x],$$

with $\alpha - 1 \equiv \{\alpha_n - 1\}$ and $\delta \equiv P_c \text{Diag}_\zeta \tilde{z}$. Recall that for any two vectors u and v, $\text{Diag}_u v = \text{Diag}_v u$. The definition of ζ therefore implies

$$E_t[\text{Diag}_\delta \Delta x] = E_t[\text{Diag}_{\Delta x} P_c \text{Diag}_{\tilde{z}} \zeta] = N_W E_t[\text{Diag}_\omega \Delta x].$$

Combining the two previous expressions implies

$$E_t[\beta_{(z)_{\text{res}}}] = (\alpha - 1) \cdot E_t[\text{Diag}_\omega \Delta x] = E_t[\textstyle\sum_n \omega_n (\alpha_n - 1) \Delta x_n] = E_t[W(\Delta x^*) - W(\Delta x)].$$

□

A.7 Alternative Forms of Misspecification

In Proposition 3 we have shown how to estimate and correct for the expected bias—taken across all potential realizations of the policy—of the researcher's causal answer provided that A3′ holds. In this appendix, we propose to relax A3′ and assume instead that the true causal impact of policy belongs to some set $\mathcal{X}$ such that the following condition holds.

A3″. *For any $n \in \mathcal{N}_W$, $sup_{\Delta x^* \in \mathcal{X}} |\Delta x_n^* - \Delta x_n| = b_n |\Delta x_n|$ where the maximum misspecification parameter $b_n > 0$ may be a function of, and only of, the realization of period t's shocks, (ϵ_t^*, τ_t).*

If A3′ holds, then A3″ holds as well with $b_n = |\alpha_n - 1|$. The converse, however, is not true since A3″ only applies to the worst case scenario for the researcher's model.

Let $W_z(\Delta x) \equiv W(\Delta x) + \beta_z$ denote the researcher's welfare prediction adjusted by our goodness-of-fit measure. Under A3′, we have shown that one can construct z such that $W_z(\Delta x)$ is an unbiased estimator of $E_t[W(\Delta x^*)]$, i.e. $E_t[W_z(\Delta x)] = E_t[W(\Delta x^*)]$. Under A3″, the next proposition shows that one can construct z such that the worst-case expected bias of $W_z(\Delta x)$, now potentially non-zero, is less than the worst-case expected bias of $W(\Delta x)$.

Proposition 4. *Take any IV $\tilde{z}$ that satisfies A1 and A2. If A3″ holds, then one can construct z, with $z_n \equiv N_W \tilde{z}_n \omega_n \zeta_n$ and $\zeta_n \equiv argmin_\zeta E_t[|1 - \zeta \tilde{z}_n||\Delta x_n|]$ for all $n \in \mathcal{N}_W$, such that z satisfies A1 and A2 and $sup_{\Delta x^* \in \mathcal{X}} |E_t[W(\Delta x^*) - W_z(\Delta x)]| \leq sup_{\Delta x^* \in \mathcal{X}} |E_t[W(\Delta x^*) - W(\Delta x)]|$.*

Proof. Start from

$$E_t[W(\Delta x^*) - W(\Delta x)] = \sum_n \omega_n E_t[\Delta x_n^* - \Delta x_n].$$

A3″ implies

$$\begin{aligned}
&\sup_{\Delta x^* \in \mathcal{X}} |E_t[W(\Delta x^*) - W(\Delta x)]| \\
&\qquad = \sum_n |\omega_n| \int \sup_{\Delta x^* \in \mathcal{X}} |\Delta x_n^* - \Delta x_n| dF_t(\tau_{t+1}) \\
&\qquad = \sum_n |\omega_n| \int b_n |\Delta x_n| dF_t(\tau_{t+1}) = \sum_n b_n E_t[|\omega_n||\Delta x_n|] \equiv \bar{\beta},
\end{aligned}$$

where $F_t(\tau_{t+1})$ denotes the cdf of the policy in period $t+1$. Next, consider

$$\begin{aligned}
E_t[W(\Delta x^*) - W_z(\Delta x)] &= \sum_n E_t[(\omega_n - z_n/N_W)(\Delta x_n^* - \Delta x_n)] \\
&= \sum_n \int (\omega_n - z_n/N_W) \int (\Delta x_n^* - \Delta x_n) dF_t(\tau_{t+1}|z_n) dG_{t,n}(z_n),
\end{aligned}$$

where $G_{t,n}(z_n)$ denotes the cdf of z_n and $F_t(\tau_{t+1}|z_n)$ denotes the cdf of the policy in period $t+1$ conditional on z_n. A3″ further implies

$$\begin{aligned}
\sup_{\Delta x^* \in \mathcal{X}} |E_t[W(\Delta x^*) - W_z(\Delta x)]| &= \sum_n \int |\omega_n - z_n/N_W| \int \sup_{\Delta x^* \in \mathcal{X}} |\Delta x_n^* - \Delta x_n| dF_t(\tau_{t+1}|z_n) dG_{t,n}(z_n) \\
&= \sum_n \int |\omega_n - z_n/N_W| b_n \int |\Delta x_n| dF_t(\tau_{t+1}|z_n) dG_{t,n}(z_n) \\
&= \sum_n b_n E_t[|\omega_n - z_n/N_W||\Delta x_n|] \equiv \bar{\beta}_z.
\end{aligned}$$

Next for each $n \in \mathcal{N}_W$, set $z_n = N_W \tilde{z}_n \omega_n \zeta_n$ with

$$\zeta_n \equiv \text{argmin}_\zeta E_t[|1 - \zeta \tilde{z}_n||\Delta x_n|].$$

For the same reason as in the proof of Proposition 3, z satisfies A1 and A2. Furthermore, the definitions of z_n and ζ_n imply

$$\begin{aligned}\bar{\beta}_z &= \sum_n b_n E_t[|\omega_n - \tilde{z}_n \omega_n \zeta_n||\Delta x_n|] \\ &= \sum_n b_n |\omega_n| \min_\zeta E_t[|1 - \zeta \tilde{z}_n||\Delta x_n|] \leq \sum_n b_n E_t[|\omega_n||\Delta x_n|] = \bar{\beta}.\end{aligned}$$

□

To apply the results of Proposition 4, one needs to compute $\zeta_n \equiv \operatorname{argmin}_\zeta E_t[|1 - \zeta \tilde{z}_n||\Delta x_n|]$ for each $n \in \mathcal{N}_W$. This can be done by first rearranging the previous minimization problem as

$$\begin{aligned}&\min_\zeta E_t[|1 - \zeta \tilde{z}_n||\Delta x_n|] \\ &= \min_\zeta \int_{-\infty}^{+\infty} |\frac{1}{z} - \zeta| E_t[|\tilde{z}_n \Delta x_n||\tilde{z}_n = z] dG_{t,n}(z) = \min_\zeta \int_{-\infty}^{+\infty} |u - \zeta| dH_{t,n}(u),\end{aligned}$$

where the cdf $H_{t,n}(u)$ is given by

$$H_{t,n}(u) = \frac{\int_{-\infty}^{u} E_t[|\Delta x_n||\tilde{z}_n = 1/v] g_{t,n}(1/v)|v|^{-3} dv}{\int_{-\infty}^{+\infty} E_t[|\Delta x_n||\tilde{z}_n = 1/v] g_{t,n}(1/v)|v|^{-3} dv},$$

and then noting that for a given random variable u with cdf $H_{t,n}$, the solution to

$$\min_\zeta \int_{-\infty}^{+\infty} |u - \zeta| dH_{t,n}(u)$$

is equal to its median value $H_{t,n}^{-1}(0.5)$.

A.8 Power and Informativeness

We go back to the environment considered in Case 2 of Appendix A.5. There are two moments, $\hat{\gamma}(\alpha, \theta)$ and $\hat{\beta}_z(\alpha, \theta)$, with the first one used for estimation and the second one used for testing. The joint distribution of these two moments depends on the true values of two structural parameters, α and θ. The researcher uses the first moment to estimate θ by setting

$$\hat{\gamma}(\alpha, \hat{\theta}) = 0. \tag{A.30}$$

We let $\hat{\theta}(\alpha)$ denote the estimator that solves (A.30).

The value of the other structural parameter α is restricted, perhaps wrongly, by the researcher's model to be equal to 1. For instance, one can think of misspecification taking the form of A3' with $\alpha \equiv \{\alpha_n\}$. The results in Appendix A.5 characterize the asymptotic distribution F_z of $\hat{\beta}_z(\alpha, \hat{\theta}(\alpha))$ under the null that $\alpha = 1$. Let s denote the size of our test under the null and let $\delta_{z,s}$ denote the associated critical value, i.e. $F_z(\delta_{z,s}) = 1 - s/2$. The probability of rejecting the null that $\alpha = 1$ under different values of α is equal to

$$\pi_z(\alpha) = \Pr[|\hat{\beta}_z(\alpha, \hat{\theta}(\alpha))| \geq \delta_{z,s}].$$

By construction, $\pi_z(1) = s$ under the null. The next example illustrates the idea that, everything else being equal, power $\pi_z(\alpha)$ under the alternative $\alpha \neq 1$ decreases with the informativeness of the estimation moment $\hat{\gamma}(\alpha, \theta)$ for the testing moment $\hat{\beta}_z(\alpha, \theta)$ in the sense of Andrews et al. (2020), i.e. power decreases with the R-squared of a hypothetical population-level regression of $\hat{\beta}_z(\alpha, \theta)$ on $\hat{\gamma}(\alpha, \theta)$.

Example. Consider two testing moments $\hat{\beta}_z(\alpha, \theta)$ and $\hat{\beta}_{z'}(\alpha, \theta)$. Without loss of generality, we can decompose them into

$$\begin{aligned}
\hat{\beta}_z(\alpha, \theta) &= \mu_z(\alpha, \theta)\hat{\gamma}(\alpha, \theta) + \mu_z^{\perp}(\alpha, \theta)\hat{\gamma}_z^{\perp}(\alpha, \theta), \\
\hat{\beta}_{z'}(\alpha, \theta) &= \mu_{z'}(\alpha, \theta)\hat{\gamma}(\alpha, \theta) + \mu_{z'}^{\perp}(\alpha, \theta)\hat{\gamma}_{z'}^{\perp}(\alpha, \theta),
\end{aligned}$$

where $\hat{\gamma}_z^{\perp}(\alpha, \theta)$ and $\hat{\gamma}_{z'}^{\perp}(\alpha, \theta)$ are uncorrelated with the estimation moment. Equation (A.30) implies

$$\begin{aligned}
\hat{\beta}_z(\alpha, \hat{\theta}(\alpha)) &= \mu_z^{\perp}(\alpha, \hat{\theta}(\alpha))\hat{\gamma}_z^{\perp}(\alpha, \hat{\theta}(\alpha)), \\
\hat{\beta}_{z'}(\alpha, \hat{\theta}(\alpha)) &= \mu_{z'}^{\perp}(\alpha, \hat{\theta}(\alpha))\hat{\gamma}_{z'}^{\perp}(\alpha, \hat{\theta}(\alpha)).
\end{aligned}$$

Now suppose that the two testing moments only differ in terms of how informative the estimation moment $\hat{\gamma}(\alpha, \theta)$ is. Specifically, suppose that: (i) $\delta_{z,s} = \delta_{z',s} \equiv \delta_s$; (ii) $\hat{\gamma}_z^{\perp}(\alpha, \theta) = \hat{\gamma}_{z'}^{\perp}(\alpha, \theta) \equiv \hat{\gamma}^{\perp}(\alpha, \theta)$; and (iii) $0 \leq \mu_z(\alpha, \theta) < \mu_{z'}(\alpha, \theta) \leq 1$ and $\mu_z(\alpha, \theta) + \mu_z^{\perp}(\alpha, \theta) = \mu_{z'}(\alpha, \theta) + \mu_{z'}^{\perp}(\alpha, \theta) = 1$. By conditions (ii) and (iii), the estimation moment $\hat{\gamma}(\alpha, \theta)$ is therefore less informative for the testing moment $\hat{\beta}_z(\alpha, \theta)$ than $\hat{\beta}_{z'}(\alpha, \theta)$. And by conditions (i)-(iii), power after estimation is higher using $\hat{\beta}_z(\alpha, \theta)$ than $\hat{\beta}_{z'}(\alpha, \theta)$:

$$\pi_z(\alpha) = \Pr[|1 - \mu_z(\alpha, \hat{\theta}(\alpha))||\hat{\gamma}^{\perp}(\alpha, \hat{\theta}(\alpha))| \geq \delta_s] > \Pr[|1 - \mu_{z'}(\alpha, \hat{\theta}(\alpha))||\hat{\gamma}^{\perp}(\alpha, \hat{\theta}(\alpha))| \geq \delta_s] = \pi_{z'}(\alpha).$$

A.9 Testing Using Untargeted Cross-Sectional Moments

Suppose that the researcher's model takes the form,

$$f_n(\tau_t, \epsilon_t) = \begin{cases} g_n(\tau_t|\theta) + \epsilon_{n,t}, & \text{for } n \in \mathcal{N}_T, \\ h_n(\{f_m(\tau_t, \epsilon_t)\}_{m \in \mathcal{N}_T}), & \text{for } n \in \mathcal{N}_U, \end{cases}$$

whereas the true model is

$$f_n^*(\tau_t, \epsilon_t^*) = \begin{cases} g_n^*(\tau_t|\theta) + \epsilon_{n,t}^*, & \text{for } n \in \mathcal{N}_T, \\ h_n(\{f_m^*(\tau_t, \epsilon_t^*)\}_{m \in \mathcal{N}_T}), & \text{for } n \in \mathcal{N}_U. \end{cases}$$

Whenever $g_n \neq g_n^*$ the researcher's model misspecifies the causal relationship between the vector of endogenous variables y_t and the policy vector τ_t, but it always correctly specifies the cross-sectional relationship between the endogenous variables that have been targeted in the estimation stage (in $\mathcal{N}_T$) and those that have not (in $\mathcal{N}_U$). Thus if one were to perfectly match $\{y_n\}_{n \in \mathcal{N}_T}$ via estimation, then one would also perfectly match $\{y_{n,t}\}_{n \in \mathcal{N}_U}$. Yet this observation would provide no information about the relationship (or lack thereof) between $g_n(\tau_t|\theta)$ and $g_n^*(\tau_t|\theta)$.

B Simulations Appendix

B.1 FGKK's Model

FGGK consider a world economy comprising multiple countries, indexed by $i \in \mathcal{I}$. The country of interest is the US ($i = H$). The US comprises many regions, indexed by $r \in \mathcal{R}$, whose firms can produce in many sectors, indexed by $s \in \mathcal{S}$. Time is discrete and indexed by t. Labor is the only primary factor of production. We let $L_{rs,t}$ denote the inelastic supply of labor in region r and sector s at date t and $w_{rs,t}$ denote the associated wage rate.

Preferences. There is a US representative household with nested CES preferences over non-tradables, produced in sector $s = NT$, and tradables, produced in different sectors $s \in \mathcal{S}_T$ and countries,

$$U_t = (C_{NT,t})^{\beta_{NT,t}}(C_{T,t})^{1-\beta_{NT,t}}, \tag{B.1}$$

$$C_{T,t} = \prod_{s\in\mathcal{S}_T}(C_{Ts,t})^{\beta_{s,t}}, \tag{B.2}$$

$$C_{Ts,t} = \left[\sum_{j=H,F}(A_{js,t})^{\frac{1}{\kappa}}(C_{js,t})^{\frac{\kappa-1}{\kappa}}\right]^{\frac{\kappa}{\kappa-1}}\text{, for all } s \in \mathcal{S}_T, \tag{B.3}$$

$$C_{js,t} = \left[\sum_{v\in\mathcal{P}_s}(a_{jv,t})^{\frac{1}{\eta}}(c_{jv,t})^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}\text{, for } j = H, F \text{ and } s \in \mathcal{S}_T, \tag{B.4}$$

$$c_{Fv,t} = \left[\sum_{i\neq H}(a_{iv,t})^{\frac{1}{\sigma}}(c_{iv,t})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}\text{, for all } v \in \mathcal{P}_s \text{ and } s \in \mathcal{S}_T, \tag{B.5}$$

where $\mathcal{P}_s$ is the set of all products v from a given tradable sector s; $\beta_{s,t}$, $A_{Hs,t}$, $A_{Fs,t}$, $a_{Hv,t}$, $a_{Fv,t}$, and $\{a_{iv,t}\}_{i\neq H}$ are exogenous preference shifters; $\kappa \geq 0$ is the elasticity of substitution between domestic consumption and imports within a given sector s; $\eta \geq 0$ is the elasticity of substitution between products within each of these two nests; and $\sigma \geq 0$ is the elasticity of substitution between different foreign sources within a given product.

Technology. In each US region r, firms in the non-tradable sector ($s = NT$) produce one-to-one from labor,

$$Q_{rNT,t} = Z_{rNT,t}L_{rNT,t}\text{, for all } r \in \mathcal{R}, \tag{B.6}$$

whereas firms in any tradable sector $s \in \mathcal{S}_T$ produce according to

$$Q_{rs,t} = Z_{rs,t}(M_{rs,t})^{\alpha_{Is,t}}(L_{rs,t})^{\alpha_{Ls,t}}\text{, for all } r \in \mathcal{R} \text{ and } s \in \mathcal{S}_T, \tag{B.7}$$

$$M_{rs,t} = \prod_{k\in\mathcal{S}_T}(M_{krs,t})^{\alpha_{ks,t}}\text{, for all } r \in \mathcal{R} \text{ and } s \in \mathcal{S}_T, \tag{B.8}$$

with $\alpha_{Is,t} + \alpha_{Ls,t} \leq 1$ and $\sum_{k \in \mathcal{S}_T} \alpha_{ks,t} = 1$. Tradable intermediates, $M_{krs,t}$, from sector k demanded by firms in sector s and region r are produced using domestic and foreign products in the same nested-CES manner as final consumption from that sector, $C_{Tk,t}$, as described by equations (B.3)-(B.5). For future reference, we let $q_{ivrs,t}$ denote the quantity of product v from country i demanded by firms from region r and sector s. Finally, given total sector-level output, $\sum_{r \in \mathcal{R}} Q_{rs,t}$, the vector of destination-and-product level output, $\{q_{iv,t}\}_{v \in \mathcal{P}_s, i \in \mathcal{I}}$, satisfies the following resource constraint,

$$\sum_{v \in \mathcal{P}_s, i \in \mathcal{I}} q_{iv,t}/z_{iv,t} = \sum_{r \in \mathcal{R}} Q_{rs,t} \text{ for all } s \in \mathcal{S}_T. \tag{B.9}$$

The productivity shifters, $Z_{rs,t}$ and $z_{iv,t}$, are exogenous and potentially time-varying.

Prices, Import Tariffs, and Transfers. There are no domestic transportation costs so prices of tradables are equalized across US regions. For any variety purchased in the US, we let $p^H_{iv,t}$ denote the price of product v from country i faced by US households and firms. Likewise, for any product v sold by the US in a country i, we let $\bar{p}^H_{iv,t}$ denote the price received by US firms. Home's import tariffs $\tau^H_t \equiv \{\tau^H_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$ drive a wedge between US import prices $p^H_t \equiv \{p^H_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$ and the prices received by foreigners $\bar{p}^F_t \equiv \{\bar{p}^F_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$, whereas foreign tariffs $\tau^F_t \equiv \{\tau^F_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$ drive a wedge between US export prices $\bar{p}^H_t \equiv \{\bar{p}^H_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$ and the prices $p^F_t \equiv \{p^F_{iv,t}\}_{v \in \mathcal{P}_s, i \neq H}$ paid by foreigners,

$$p^H_{iv,t} = (1 + \tau^H_{iv,t})\bar{p}^F_{iv,t}, \text{ for all } i \neq H, v \in \mathcal{P}_s, \text{ and } s \in \mathcal{S}_T, \tag{B.10}$$

$$p^F_{iv,t} = (1 + \tau^F_{iv,t})\bar{p}^H_{iv,t}, \text{ for all } i \neq H, v \in \mathcal{P}_s, \text{ and } s \in \mathcal{S}_T. \tag{B.11}$$

There are no other taxes. The US government rebates total tariff revenues as well as any foreign transfer F_t through a lump-sum transfer T_t to the representative household. The US government's budget constraint is

$$T_t = \sum_{s \in \mathcal{S}_T, v \in \mathcal{P}_s, i \in \mathcal{I}} \tau^H_{iv,t}\bar{p}^F_{iv,t}(c_{iv,t} + \sum_{r \in \mathcal{R}, s \in \mathcal{S}_T} q_{ivrs,t}) + F_t. \tag{B.12}$$

Foreign Import Demand and Export Supply. The rest of the world is modeled as a series of import demand and export supply curves that determine the quantities $c^F_{iv,t}$ and $q^F_{iv,t}$ of any product v that a country $i \neq H$ imports from and exports to Home, respectively,

$$c^F_{iv,t} = a^F_{iv,t}(p^F_{iv,t})^{-\sigma_F}, \tag{B.13}$$

$$q^F_{iv,t} = (z^F_{iv,t})^{\frac{1}{\omega_F}}(\bar{p}^F_{iv,t})^{\frac{1}{\omega_F}}, \tag{B.14}$$

where $a^F_{iv,t}$ and $z^F_{iv,t}$ are exogenous import demand and export supply shifters; σ_F is the elasticity of foreign import demand; and ω_F is the inverse of the elasticity of foreign export supply.

Market clearing. Supply equals demand for products from all origins in all destinations,

$$q_{iv,t} = c^F_{iv,t}, \text{ for all } i \neq H, v \in \mathcal{P}_s, \text{ and } s \in \mathcal{S}_T, \tag{B.15}$$

$$q_{Hv,t} = c_{Hv,t} + \sum_{r \in \mathcal{R}, s \in \mathcal{S}_T} q_{Hvrs,t}, \text{ for all } v \in \mathcal{P}_s, \text{ and } s \in \mathcal{S}_T, \tag{B.16}$$

$$q^F_{iv,t} = c_{iv,t} + \sum_{r \in \mathcal{R}, s \in \mathcal{S}_T} q_{ivrs,t}, \text{ for all } i \neq H, v \in \mathcal{P}_s, \text{ and } s \in \mathcal{S}_T, \tag{B.17}$$

$$Q_{rNT,t} = C_{rNT,t}, \text{ for all } r \in \mathcal{R}. \tag{B.18}$$

Competitive Equilibrium. Given tariffs $\tau_t \equiv \{\tau^H_t, \tau^F_t\}$, a competitive equilibrium corresponds to prices $p_t \equiv \{p^H_t, p^F_t, \bar{p}^H_t, \bar{p}^F_t, w_{rs,t}, p_{rNT,t}\}$, quantities $q_t \equiv \{c_{iv,t}, c^F_{iv,t}, q_{ivrs,t}, q_{iv,t}, q^F_{iv,t}, Q_{rNT,t}\}$, and a lump-sum lump-sum transfer T_t such that the US representative household maximizes its utility, as described in (B.1)-(B.5), subject to its budget constraint; price-taking US firms maximize their profits subject to technological constraints, as described in (B.6)-(B.9); import and export prices satisfy the non-arbitrage conditions (B.10) and (B.11); the domestic government's budget is balanced, as described in (B.12); foreigners are on their export supply and import demand curves, as described in (B.13) and (B.14); and good markets clear, as described in (B.15)-(B.18).

B.2 Calibration of FGKK's Model

Equilibrium prices and quantities (p_t, q_t) in FGKK's model depend on the policy vector τ_t as well as a vector of time-invariant elasticities and a vector of time-varying shocks,

$$\theta \equiv \{\sigma, \omega_F, \sigma_F, \eta, \kappa\},$$
$$\epsilon_t \equiv \{\beta_{s,t}, A_{Hs,t}, A_{Fs,t}, a_{Hv,t}, a_{Fv,t}, a_{iv,t}, Z_{rs,t}, \alpha_{Is,t}, \alpha_{Ls,t}, \alpha_{ks,t}, z_{iv,t}, a^F_{iv,t}, z^F_{iv,t}, F_t, L_{rs,t}\}.$$

Here we describe how we calibrate θ and ϵ_t for the simulation in Sections 3.2 and 3.3.

Demand and Supply Elasticities (θ). We use the five elasticity estimates obtained by FGKK when using changes in US tariffs, and foreign retaliatory tariffs, in 2018. The first three such elasticities describe Home households' and firms' elasticities of substitution across domestic and foreign bundles of products within each tradable sector ($\kappa = 1.19$), across different products within the domestic or foreign bundles ($\eta = 1.53$), and across different foreign origins within each foreign product bundle ($\sigma = 2.53$). The remaining two elasticities capture foreign firms' and households' elasticities of demand for imports sourced from Home ($\sigma_F = 1.04$) and foreign firms' (inverse) elasticities of supply for exports to Home ($\omega_F = 0.00$).

Time-Varying Shocks (ϵ_t). We calibrate ϵ_t such that, given the estimated supply and demand elasticities $\theta \equiv (\sigma, \sigma_F, \omega_F, \kappa, \eta)$ and the initial period tariffs τ_t, FGKK's model exactly matches trade and production data from the US in 2016. Following FGKK, we let the initial period tariffs observed in the US and the rest of the world, $\{\tau_t^H, \tau_t^F\}$, be equal to their average statutory values in 2016. All other data inputs are also the same as in FGKK and comprise: *(i)* variety-level quantities and values for both exports and imports, for 71 foreign countries i and $10,228$ tradable products v (based on 10-digit HS codes) partitioned into 88 tradable sectors s (based on a 4-digit NAICS classification); *(ii)* sector-level revenues and expenditures on both labor and intermediates for each tradable sector s; and *(iii)* region-sector employment, with each region r a US county. We obtain all these variables from FGKK's replication package.

B.3 Counterfactual Welfare Changes

Consider a counterfactual change in US and foreign tariffs. Since FGKK's model is perfectly competitive and features a representative agent in the US, a standard envelope argument implies that the first-order impact of a change in tariffs on US welfare, expressed in dollar units, is

$$dW = \sum_{i \neq H, v} q_{iv,t} d\bar{p}_{iv,t}^H - \sum_{i \neq H, v} q_{iv,t}^F d\bar{p}_{iv,t}^F + \sum_{i \neq H, v} \tau_{iv,t}^H \bar{p}_{iv,t}^F dq_{iv,t}^F,$$

where $d\bar{p}_{iv,t}^H$, $d\bar{p}_{iv,t}^F$, and $dq_{iv,t}^F$ are the changes in US export prices (pre-foreign tariff), US import prices (pre-US tariffs), and US import quantities caused by the tariff changes. Letting Y_t denote US GDP in period t, the previous expression can be rearranged as

$$\begin{aligned} \frac{dW}{Y_t} &= \sum_{i \neq H, v} \frac{\bar{p}_{iv,t}^H q_{iv,t}}{Y_t} d \ln \bar{p}_{iv,t}^H \\ &\quad + \sum_{i \neq H, v} \frac{p_{iv,t}^H q_{iv,t}^F}{Y_t} \left[-d \ln p_{iv,t}^H + \frac{d(\tau_{iv,t}^H \bar{p}_{iv,t}^F q_{iv,t}^F)}{p_{iv,t}^H q_{iv,t}^F} \right]. \end{aligned}$$

Equation (7) follows with

$$\begin{aligned} \omega_{iv}^X &\equiv 100 \frac{\bar{p}_{iv,t}^H q_{iv,t}}{Y_t}, \\ \omega_{iv}^M &\equiv 100 \frac{p_{iv,t}^H q_{iv,t}^F}{Y_t}, \\ \omega_{iv}^R &\equiv 100 \frac{p_{iv,t}^H q_{iv,t}^F}{Y_t}, \\ \Delta x_{iv}^X &\equiv d \ln \bar{p}_{iv,t}^H, \\ \Delta x_{iv}^M &\equiv d \ln p_{iv,t}^H, \\ \Delta x_{iv}^R &\equiv \frac{d(\tau_{iv,t}^H \bar{p}_{iv,t}^F q_{iv,t}^F)}{p_{iv,t}^H q_{iv,t}^F}. \end{aligned}$$

B.4 Construction of the IVs

All IVs in our simulations use the shifters described in equation (8), but differ in their shares. We now describe each of them in detail.

Preferred IV z^{pref}. In Sections 3.2-3.4, the shares $\{s_{nk}^{\text{pref}}\}$ of our preferred IV z^{pref} are computed as

$$\{s_{nk}^{\text{pref}}\} = P_c \times \text{Diag}_{\zeta} \times D_{\tau}(f),$$

where $D_{\tau}(f)$ is the Jacobian of the researcher's model with respect to tariffs; Diag_{ζ} denotes the diagonal matrix associated with the vector of welfare adjustment $\zeta \equiv \{\zeta_n\}$ described in the extension of Proposition 3 with controls in Appendix A.6,

$$\zeta \equiv N_W(E_t[\text{Diag}_{\Delta x} P_c \text{Diag}_{D_{\tau}(f)\Delta\tau}])^{-1}(\text{Diag}_{\omega} E_t[\Delta x]);$$

and P_c denotes the residual projection matrix associated with the controls $C \equiv \{c_{nj}\}$,

$$P_c \equiv Id - C(C^T C)^{-1} C^T,$$

where Id denotes the identity matrix. The four controls $j = 1, ..., 4$ are the three dummy variables for whether or not $n = (i, v, o)$ satisfies $o(n) = X$, $o(n) = M$, and $o(n) = R$ as well as each outcome n's total exposure to tariff changes, $\sum_k \partial f_n / \partial \tau_k$.

Naive IV z^{naive}. In Section 3.3, the shares $\{s_{nk}^{\text{naive}}\}$ of our naive IVz^{naive} are computed as

$$\{s_{nk}^{\text{naive}}\} = P_c \times \text{Diag}_{\omega} \times I,$$

where I is the $N_W \times K$ matrix whose entry (n, k) is a dummy variable that equals one if and only if $i(n) = i(k)$ and $v(n) = v(k)$ (i.e., outcome n and shifter k correspond to the same country-product pair) as well as either $o(n) = X$ and $c(k) = F$ (i.e., n is an export price and k a foreign tariff shifter) or $o(n) \in \{M, R\}$ and $c(k) = H$ (i.e., n is an import price or tariff revenue and k a US tariff shifter); Diag_{ω} denotes the diagonal matrix associated with the welfare weights $\omega \equiv \{\omega_n\}$; and P_c denotes the same residual projection matrix as before.

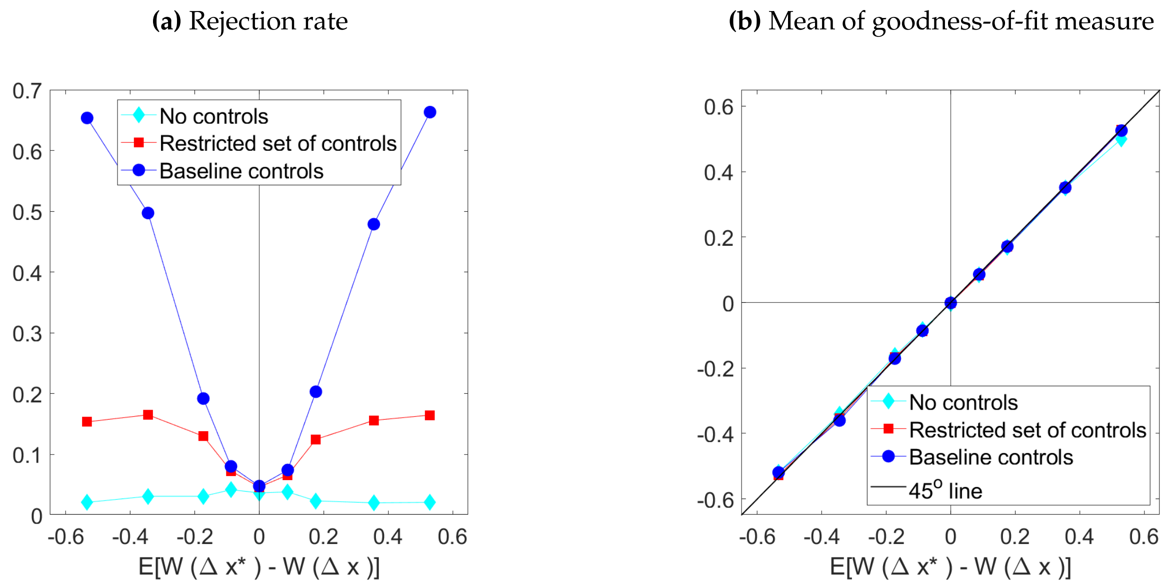
Naive IV z^{mech}. In Section 3.4, the shares $\{s_{nk}^{\text{mech}}\}$ of our naive IV z^{mech} are computed as

$$\{s_{nk}^{\text{mech}}\} = P_{c^{\text{mech}}} \times Id \times I,$$

where $P_{c^{\text{mech}}} \equiv Id - C^{\text{mech}}((C^{\text{mech}})^T C^{\text{mech}})^{-1}(C^{\text{mech}})^T$ denotes the residual projection matrix associated with the controls used in estimation, C^{mech}, i.e., dummy variables for exported and imported products.

B.5 IV-based Tests with Alternative Control Sets

Figure B.1: Alternative IV-based Tests with Fewer Controls



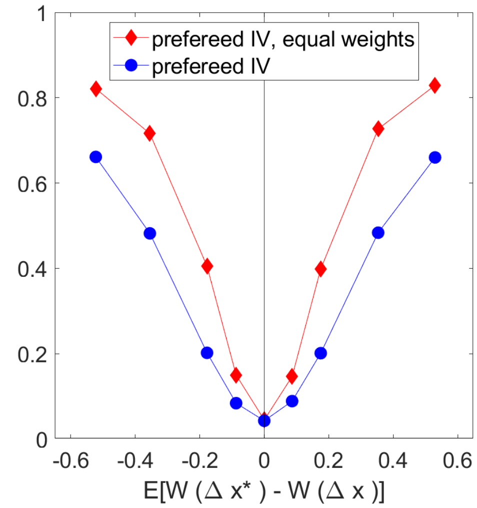
Notes: This figure reports the resuls from $b = 1, ..., 2500$ simulated economies in which the researcher's model misspecifies the pass-through rate of all statutory tariffs into applied tariffs by an amount α that is uniform across import, export, and tariff revenue outcomes. Simulations differ in terms of the size of the underlying α to generate the variation in $E_t[W(\Delta x^*) - W(\Delta x)]$ reported on the x-axes. Figure B.1a plots the rejection rate on the y-axis for IV-based tests using our preferred IV with all controls (blue circles), with a restricted set of controls that only includes dummy variables for outcome types $o \in \{X, M, R\}$ (red diamonds), and without controls (turquoise triangles). Figure B.1b does the same but with the average value of our goodness-of-fit measure $\hat{\beta}_z$ on the y-axis.

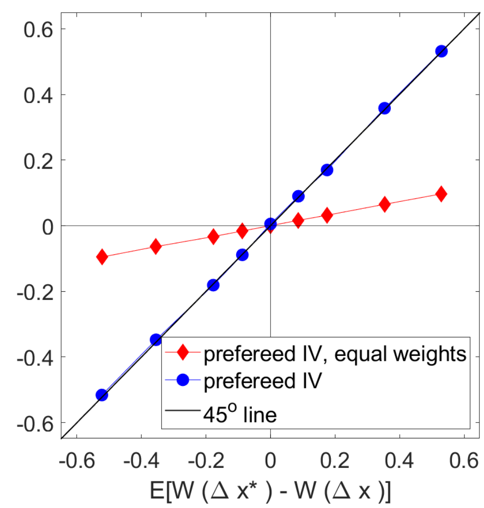
B.6 IV-based Tests without Welfare Weights

Figure B.2: Alternative IV-based Tests without Welfare Weights

(a) Rejection rate

(b) Mean of goodness-of-fit measure





Notes: This figure reports the resuls from $b = 1, ..., 2500$ simulated economies in which the researcher's model misspecifies the pass-through rate of all statutory tariffs into applied tariffs by an amount α that is uniform across import, export, and tariff revenue outcomes. Simulations differ in terms of the size of the underlying α to generate the variation in $E_t[W(\Delta x^*) - W(\Delta x)]$ reported on the x-axes. Figure B.2a plots the rejection rate on the y-axis for IV-based tests using our preferred IV (z^{pref}, blue circles) and an alternative IV (z^{alt}, red diamonds) defined as our preferred IV but with ζ computed under the assumption of equal weights (i.e., $\omega_n = 1$ for all n). Figure B.2b does the same but with the average value of our goodness-of-fit measure $\hat{\beta}_z$ on the y-axis.

C Empirical Application Appendix

C.1 Implementation

This appendix describes how to implement our test while accounting for the uncertainty in FGKK's estimates of the structural parameters $\hat{\theta} \equiv (\hat{\sigma}, \hat{\sigma}_F, \hat{\omega}_F, \hat{\kappa}, \hat{\eta})$ and allowing for correlation of policy shifters within clusters. We first follow the approach described in Appendix A.5 and characterize the estimating and testing moments required to apply Proposition 2 when $\hat{\theta}$ is an IV estimator. One idiosyncratic feature of FGKK's analysis is that estimation is conducted at the monthly level, whereas counterfactual simulations are conducted at the annual level, something that we will formally take into account by allowing outcome variables and policy shifters to vary at the monthly level as well. Finally, to account for the potential autocorrelation in tariff changes over time, we use results from Adao et al. (2019) to characterize the estimator of the asymptotic variance of our test statistic under the assumption that policy shifters are independent between groups of observations, but not within.

Estimating moments. The estimation strategy of FGKK (presented in Section III.C of their paper) consists of five linear specifications estimated with IVs built from monthly tariff changes. Using the notation from Appendix A.5, FGKK's strategy is equivalent to obtaining $\hat{\theta} \equiv (\hat{\sigma}, \hat{\omega}_F, \hat{\sigma}_F, \hat{\eta}, \hat{\kappa})$ from the moment condition $\hat{\gamma}(\hat{\theta}) = 0$ with the estimating moments such that

$$\hat{\gamma}(\theta) \equiv \begin{bmatrix} \frac{1}{|\mathcal{N}_\sigma|} \sum_{n \in \mathcal{N}_\sigma} z_n^\sigma e_n^\sigma(w, \theta) \\ \frac{1}{|\mathcal{N}_{\omega_F}|} \sum_{n \in \mathcal{N}_{\omega_F}} z_n^{\omega_F} e_n^{\omega_F}(w, \theta) \\ \frac{1}{|\mathcal{N}_{\sigma_F}|} \sum_{n \in \mathcal{N}_{\sigma_F}} z_n^{\sigma_F} e_n^{\sigma_F}(w, \theta) \\ \frac{1}{|\mathcal{N}_\eta|} \sum_{n \in \mathcal{N}_\eta} z_n^\eta e_n^\eta(w, \theta) \\ \frac{1}{|\mathcal{N}_\kappa|} \sum_{n \in \mathcal{N}_\kappa} z_n^\kappa e_n^\kappa(w, \theta) \end{bmatrix} \tag{C.1}$$

The different terms entering equation (C.1) are as follows. First, the sets of outcome variables n used in the estimation of each structural parameter are equal to

$$\begin{aligned} \mathcal{N}_\sigma &\equiv \{n = (i, v, m) \in \mathcal{I} \times \mathcal{P} \times \mathcal{M}\}, \\ \mathcal{N}_{\omega_F} &\equiv \{n = (i, v, m) \in \mathcal{I} \times \mathcal{P} \times \mathcal{M}\}, \\ \mathcal{N}_{\sigma_F} &\equiv \{n = (i, v, m) \in \mathcal{I} \times \mathcal{P} \times \mathcal{M}\}, \\ \mathcal{N}_\eta &\equiv \{n = (v, m) \in \mathcal{P} \times \mathcal{M}\}, \\ \mathcal{N}_\kappa &\equiv \{n = (s, m) \in \mathcal{S}_T \times \mathcal{M}\}, \end{aligned}$$

where $\mathcal{I}$ denotes the set of foreign countries, $\mathcal{P}$ denotes the set of products, $\mathcal{S}_T$ denotes the set of tradable sectors, as in Appendix B.1, and $\mathcal{M}$ denotes the set of months in FGKK's estimation sample, which runs from January 2017 to April 2019.

Second, the structural residuals are given by

$$e_n^{\sigma}(w,\theta) = (\Delta_{\text{monthly}} \log q_n^F)_{\text{res}} + \sigma(\Delta_{\text{monthly}} \log p_n^H)_{\text{res}}, \tag{C.2}$$

$$e_n^{\omega_F}(w,\theta) = (\Delta_{\text{monthly}} \log \bar{p}_n^F)_{\text{res}} - \omega_F(\Delta_{\text{monthly}} \log q_n^F)_{\text{res}}, \tag{C.3}$$

$$e_n^{\sigma_F}(w,\theta) = (\Delta_{\text{monthly}} \log q_n)_{\text{res}} + \sigma_F(\Delta_{\text{monthly}} \log p_n^F)_{\text{res}}, \tag{C.4}$$

$$e_n^{\eta}(w,\theta) = (\Delta_{\text{monthly}} \log B_{Fn}^H)_{\text{res}} + (\eta - 1)(\Delta_{\text{monthly}} \log P_{Fn}^H(\sigma))_{\text{res}}, \tag{C.5}$$

$$e_n^{\kappa}(w,\theta) = (\Delta_{\text{monthly}} \log[E_{Fn}^H / E_{Hn}^H])_{\text{res}} + (\kappa - 1)(\Delta_{\text{monthly}} \log[\bar{P}_{Fn}^H(\sigma,\eta)/\bar{P}_{Hn}])_{\text{res}}, \tag{C.6}$$

where $\Delta_{\text{monthly}} \log y_n \equiv \log y_{r(n)m(n)} - \log y_{r(n)m(n)-1}$ denotes the month-to-month change in any outcome $n \equiv (r,m)$, and the subscript "res" is a reminder that all variables have been residualized from the set of controls specified in each of FGKK's estimating equations.[38] In equations (C.2)-(C.4), q_n^F, p_n^H and $\bar{p}_n^F$ denote the quantity, the price faced by US households, and the price received by foreign firms, respectively, of each product imported by the US, and q_n and p_n^F denote the quantity and the price faced by foreign households, respectively, for each product exported by the US, consistent with our notation in Appendix A.5. In equation (C.5), $B_{Fn}^H \equiv \sum_{i \in \mathcal{I}} p_{in}^H q_{in}^F / E_{Fsm(n)}^H$ denotes the share of product $v(n)$ in US expenditure on imports from sector s in month $m(n)$ and $\Delta_{\text{monthly}} \log P_{Fn}^H(\sigma)$ denotes the month-to-month change in the associated CES price index. FGKK compute this as

$$\Delta_{\text{monthly}} \log P_{Fn}^H(\sigma) \equiv \frac{1}{1-\sigma} \log\left(\sum_{i \in \mathcal{C}_n} b_{in} \mathrm{e}^{(1-\sigma)\Delta \log p_{in}^H + e_{in}^{\sigma}(w,\theta)}\right) - \frac{1}{1-\sigma} \log\left(\frac{B_{m(n)}(\mathcal{C}_{\text{countries},n})}{B_{m(n)-1}(\mathcal{C}_{\text{countries},n})}\right), \tag{C.7}$$

where $b_{in} \equiv p_{in}^H q_{in}^F / \sum_{j \in \mathcal{I}} p_{jn}^H q_{jn}^F$ is the share of product $v(n)$ from country i among all countries supplying product $v(n)$ in month $m(n) - 1$ and $B_{m(n)}(\mathcal{C}_{\text{countries},n})$ is the share of spending in month $m(n)$ on the set of foreign countries $\mathcal{C}_{\text{countries},n}$ that are continuing to export product $v(n)$ between month $m(n) - 1$ and $m(n)$. Lastly, in equation (C.6), $E_{Fn}^H \equiv \sum_{v \in \mathcal{P}_{s(n)}} \sum_{i \in \mathcal{I}} p_{ivm(n)}^H q_{ivm(n)}^F$ and $E_{Hn}^H \equiv \sum_{v \in \mathcal{P}_{s(n)}} p_{Hvm(n)}^H q_{Hvm(n)}$ are US spending on imported and domestic products, respectively, in sector $s(n)$ and month $m(n)$, $\bar{P}_{Hn}$ is the US producer price index in that same sector and month, and $\Delta_{\text{monthly}} \log \bar{P}_{Fn}^H(\sigma,\eta)$ denotes the month-to-month change in the foreign counterpart of that CES price index, which FGKK compute as

$$\Delta \log P_{Fn}^H(\sigma,\eta) \equiv \frac{1}{1-\eta} \log\left(\sum_{v \in \mathcal{C}_n} b_{vn} \mathrm{e}^{(1-\eta)\Delta \log P_{Fvm(n)}^H(\sigma) + e_{vm(n)}^{\eta}(w,\theta)}\right) - \frac{1}{1-\eta} \log\left(\frac{B_{m(n)}(\mathcal{C}_{\text{products},n})}{B_{m(n)-1}(\mathcal{C}_{\text{products},n})}\right) \tag{C.8}$$

where $b_{vn} \equiv \sum_{i \in \mathcal{I}} p_{in}^H q_{in}^F / E_{Fn}^H$ is the share of imports of product v among all products in sector $s(n)$, $v \in \mathcal{P}_{s(n)}$, in month $m(n) - 1$, and $B_{m(n)}(\mathcal{C}_{\text{products},n})$ is the share of spending in month $m(n)$ on the

[38] Specifically, for the estimation of $(\sigma, \omega_F, \sigma_F)$, the controls are dummy variables for product-month, country-month, and country-sector; for the estimation of η, the controls are dummy variables for sector-month. For the estimation of κ, the controls are fixed effects for sectors and months.

set of products that are continuing to be imported in sector s between months $m(n)-1$ and $m(n)$. Finally, the instrumental variables in FGKK can be written as shift-share IVs equal to

$$z_n^{\sigma} = \sum_{k \in \mathcal{K}^{\theta}} s_{nk}^{\sigma} \Delta_{\text{monthly}} \tau_{IV,k}^{\theta}, \tag{C.9}$$

$$z_n^{\omega_F} = \sum_{k \in \mathcal{K}^{\theta}} s_{nk}^{\omega_F} \Delta_{\text{monthly}} \tau_{IV,k}^{\theta}, \tag{C.10}$$

$$z_n^{\sigma_F} = \sum_{k \in \mathcal{K}^{\theta}} s_{nk}^{\sigma_F} \Delta_{\text{monthly}} \tau_{IV,k}^{\theta}, \tag{C.11}$$

$$z_n^{\eta} = \sum_{k \in \mathcal{K}^{\theta}} s_{nk}^{\eta} \Delta_{\text{monthly}} \tau_{IV,k}^{\theta}, \tag{C.12}$$

$$z_n^{\kappa} = \sum_{k \in \mathcal{K}^{\theta}} s_{nk}^{\kappa} \Delta_{\text{monthly}} \tau_{IV,k}^{\theta}, \tag{C.13}$$

where $\mathcal{K}^{\theta} \equiv \mathcal{I} \times \mathcal{P} \times \{H, F\} \times \mathcal{M}$ denotes the set of all tariff line-month combinations $k = (i, v, c, m)$ used in the estimation of the vector of structural parameters θ; the shifters are equal to the month-to-month changes in tariffs,

$$\Delta_{\text{monthly}} \tau_{IV,k}^{\theta} \equiv \log(1 + \tau_{i(k)v(k)m(k)}^{c(k)}) - \log(1 + \tau_{i(k)v(k)m(k)-1}^{c(k)}), \tag{C.14}$$

with τ_{ivm}^{c} the tariff imposed by the US on product v from country i in month m, if $c = H$, and the tariff imposed by country i on the US, if $c = F$, as in Appendix B.1; and the shares are equal to

$$s_{nk}^{\sigma} = 1[c(k) = H] \times 1[i(n) = i(k), v(n) = v(k), m(n) = m(k)], \tag{C.15}$$

$$s_{nk}^{\omega_F} = 1[c(k) = H] \times 1[i(n) = i(k), v(n) = v(k), m(n) = m(k)], \tag{C.16}$$

$$s_{nk}^{\sigma_F} = 1[c(k) = F] \times 1[i(n) = i(k), v(n) = v(k), m(n) = m(k)], \tag{C.17}$$

$$s_{nk}^{\eta} = 1[c(k) = H] \times 1[v(n) = v(k), m(n) = m(k)] \times 1[i(k) \in \mathcal{C}_{\text{countries},n}] / |\mathcal{C}_{\text{countries},n}|, \tag{C.18}$$

$$s_{nk}^{\kappa} = 1[c(k) = H] \times 1[m(n) = m(k)] \times 1[v(k) \in \mathcal{C}_{\text{products},n}] 1[i(k) \in \mathcal{C}_{\text{countries},k}] / |\mathcal{C}_{\text{products},n}|. \tag{C.19}$$

Testing moment. By Definition 1, our testing moment is equal to

$$\hat{\beta}_z(\theta) = \frac{1}{N_W} \sum_{n \in \mathcal{N}_W} z_n (\Delta y_n - \Delta x_n(\theta)), \tag{C.20}$$

where the set of outcome variables $\mathcal{N}_W$ is the same as in our simulations and described at the end of Section 3.1. Consistent with FGKK's counterfactual question of interest, observed and predicted changes, Δy_n and $\Delta x_n(\theta)$, are computed as changes between $t = 2016$ and $t + 1 = 2019$, as described in Section 4.1. Likewise, the shift-share IV z is such that

$$z_n = \sum_{k \in \mathcal{K}} s_{nk} \Delta \tau_{IV,k} \quad \text{with} \quad \Delta \tau_{IV,k} \equiv \frac{\Delta \tau_k - \mu_{\Delta \tau}}{\sigma_{\Delta \tau}}, \tag{C.21}$$

where $\Delta\tau_k \equiv \tau_{k,t+1} - \tau_{k,t}$ is the change in the tariff line $k = (i, v, c) \in \mathcal{I} \times \mathcal{P} \times \{H, F\} \equiv \mathcal{K}$ between $t = 2016$ and $t + 1 = 2019$, as also described in Section 4.1, and $\mu_{\Delta\tau}$ and $\sigma_{\Delta\tau}$ are the average and the standard deviation, respectively, of $\Delta\tau_k$ computed across all $k \in \mathcal{K}$. To apply Proposition 2 when $\hat{\theta}$ is an IV estimator, we need the policy shifters used in estimation and testing, $\Delta_{\text{monthly}}\tau^{\theta}_{IV,k}$ and $\Delta\tau_{IV,k}$, to be defined over the same ks. At this point, however, the set of shifters used in estimation varies over $k \in \mathcal{K}^{\theta} = \mathcal{K} \times \mathcal{M}$, whereas the set of shifters used in testing vary over $k \in \mathcal{K}$. To deal with this discrepancy, we simply note that the policy shifter $\Delta\tau_{IV,k}$ used in testing can be expressed as the cumulative sum of the month-to-month changes,

$$\Delta\tau_{IV,k} = \sum_{m \in \mathcal{M}} \Delta_{\text{monthly}}\tau_{IV,km} \quad \text{with} \quad \Delta\tau_{IV,k} \equiv \frac{\Delta_{\text{monthly}}\tau^{c(k)}_{i(k)v(k)m}}{\sigma_{\Delta\tau}} - \frac{1}{|\mathcal{M}|}\frac{\mu_{\Delta\tau}}{\sigma_{\Delta\tau}}, \tag{C.22}$$

with the convention $\Delta_{\text{monthly}}\tau^{c(k)}_{i(k)v(k)\text{January2017}} \equiv \tau^{c(k)}_{i(k)v(k)\text{January2017}} - \tau_{k,2016}$. Combining equations (C.21) and (C.22), we can then rewrite the IV used for testing as

$$z_n = \sum_{k \in \mathcal{K}, m \in \mathcal{M}} s_{nk}\Delta_{\text{monthly}}\tau_{IV,km} = \sum_{k \in \mathcal{K}^{\theta}} \bar{s}_{nk}\Delta_{\text{monthly}}\tau_{IV,k}, \tag{C.23}$$

with $\bar{s}_{nk} \equiv s_{ni(k)v(k)c(k)}$ for any tariff line-month combination $k = (i, v, c, m) \in \mathcal{K}^{\theta}$.

Variance estimator. First, consider the case without clustering. Equations (C.9)–(C.19) imply that the estimating moments in (C.1) can be rearranged as

$$\hat{\gamma}(\theta) = \sum_{k \in \mathcal{K}^{\theta}} \Delta_{\text{monthly}}\tau^{\theta}_{IV,k} \times R_{\gamma,k}(\theta), \tag{C.24}$$

with

$$R_{\gamma,k}(\theta) \equiv \begin{bmatrix} \frac{1}{|\mathcal{N}_{\sigma}|}\sum_{n \in \mathcal{N}_{\sigma}} s^{\sigma}_{nk}e^{\sigma}_{n}(w, \theta) \\ \frac{1}{|\mathcal{N}_{\omega_F}|}\sum_{n \in \mathcal{N}_{\omega_F}} s^{\omega_F}_{n}e^{\omega_F}_{n}(w, \theta) \\ \frac{1}{|\mathcal{N}_{\sigma_F}|}\sum_{n \in \mathcal{N}_{\sigma_F}} s^{\sigma_F}_{n}e^{\sigma_F}_{n}(w, \theta) \\ \frac{1}{|\mathcal{N}_{\eta}|}\sum_{n \in \mathcal{N}_{\eta}} s^{\eta}_{n}e^{\eta}_{n}(w, \theta) \\ \frac{1}{|\mathcal{N}_{\kappa}|}\sum_{n \in \mathcal{N}_{\kappa}} s^{\kappa}_{n}e^{\kappa}_{n}(w, \theta) \end{bmatrix}. \tag{C.25}$$

Likewise, equation (C.23) implies that the testing moment in equation (C.20) can be rearranged as

$$\hat{\beta}_z(\theta) = \sum_{k \in \mathcal{K}^{\theta}} \Delta_{\text{monthly}}\tau_{IV,k} \times R_{\beta,k}(\theta), \tag{C.26}$$

with

$$R_{\beta,k}(\theta) \equiv \frac{1}{N_W}\sum_{n \in \mathcal{N}_W} \bar{s}_{nk}(\Delta y_n - \Delta x_n(\theta)).$$

Following the approach in Appendix A.5, now applied to the case where θ is a vector, we compute the asymptotic variance $\hat{V}[\hat{\beta}_z(\hat{\theta})]$ of the goodness-of-fit measure $\hat{\beta}_z(\hat{\theta})$ as

$$\hat{V}[\hat{\beta}_z(\hat{\theta})] = \hat{\mathcal{V}}_\beta + \hat{D}\hat{\mathcal{V}}_\gamma\hat{D}^T + 2\hat{D}\hat{\mathcal{V}}_{\beta\gamma}, \tag{C.27}$$

$$\hat{\mathcal{V}}_\beta = \sum_{k\in\mathcal{K}^\theta} [\Delta_{\text{monthly}}\tau_{IV,k}R_{\beta,k}(\theta)]^2, \tag{C.28}$$

$$\hat{\mathcal{V}}_\gamma = \sum_{k\in\mathcal{K}^\theta} [\Delta_{\text{monthly}}\tau^\theta_{IV,k}R_{\gamma,k}(\theta)][\Delta_{\text{monthly}}\tau^\theta_{IV,k}R_{\gamma,k}(\theta)]^T \tag{C.29}$$

$$\hat{\mathcal{V}}_{\beta\gamma} = \sum_{k\in\mathcal{K}^\theta} [\Delta_{\text{monthly}}\tau^\theta_{IV,k}R_{\gamma,k}(\theta)][\Delta_{\text{monthly}}\tau_{IV,k}R_{\beta,k}(\theta)] \tag{C.30}$$

$$\hat{D} = -[D(\hat{\beta}_z(\hat{\theta}))][D(\hat{\gamma}(\hat{\theta}))]^{-1}, \tag{C.31}$$

where we solve analytically the Jacobian matrix $D(\hat{\gamma}(\hat{\theta}))$ of the estimating moments in (C.24), using (C.2)-(C.6), and we evaluate numerically the gradient $D(\hat{\beta}_z(\hat{\theta}))$ of the testing moment in (C.20) by considering small changes in θ.

Next consider the case with clustering. We group the tariff line-month combinations $k = (i, v, c, m) \in \mathcal{K}^\theta$ by tariff lines $g \equiv (i, v, c) \in \mathcal{K}$. Formally, a tariff line-month combination k belongs to the group g, which we denote by $k \in \mathcal{G}_g$, if and only if $i(k) = i(g)$, $v(k) = v(g)$, and $c(k) = c(g)$. Following the approach in Adao et al. (2019), we then compute the asymptotic variance estimator allowing for arbitrary correlation of policy shifters across months within a tariff line g as

$$\hat{V}[\hat{\beta}_z(\hat{\theta})] = \hat{\mathcal{V}}_\beta + \hat{D}\hat{\mathcal{V}}_\gamma\hat{D}^T + 2\hat{D}\hat{\mathcal{V}}_{\beta\gamma}, \tag{C.32}$$

$$\hat{\mathcal{V}}_\beta = \sum_{g\in\mathcal{K}}\sum_{k,k'\in\mathcal{G}_g} [\Delta_{\text{monthly}}\tau_{IV,k}R_{\beta,k}(\theta)][\Delta_{\text{monthly}}\tau_{IV,k'}R_{\beta,k'}(\theta)], \tag{C.33}$$

$$\hat{\mathcal{V}}_\gamma = \sum_{g\in\mathcal{K}}\sum_{k,k'\in\mathcal{G}_g} [\Delta_{\text{monthly}}\tau^\theta_{IV,k}R_{\gamma,k}(\theta)][\Delta_{\text{monthly}}\tau^\theta_{IV,k'}R_{\gamma,k'}(\theta)]^T \tag{C.34}$$

$$\hat{\mathcal{V}}_{\beta\gamma} = \sum_{g\in\mathcal{K}}\sum_{k,k'\in\mathcal{G}_g} [\Delta_{\text{monthly}}\tau^\theta_{IV,k}R_{\gamma,k}(\theta)][\Delta_{\text{monthly}}\tau_{IV,k'}R_{\beta,k'}(\theta)] \tag{C.35}$$

$$\hat{D} = -[D(\hat{\beta}_z(\hat{\theta}))][D(\hat{\gamma}(\hat{\theta}))]^{-1}. \tag{C.36}$$

C.2 Additional results

Table C.1: Testing Predictions about the Impact of Trump's Trade War

Outcome:	Δy_n (1)	$\Delta x_n(\hat{\theta})$ (2)	$\Delta y_n - \Delta x_n(\hat{\theta})$ (3)
Panel A: All outcomes			
Point estimate	0.28	0.37	-0.09
Std. error	0.15	0.02	0.18
p-value of H_0: $\hat{\beta} = 0$	0.06	0.00	0.63
Panel B: Export prices			
Point estimate	1.40	1.23	0.16
Std. error	0.61	0.03	0.73
p-value of H_0: $\hat{\beta} = 0$	0.02	0.00	0.82
Panel C: Import prices			
Point estimate	0.41	0.23	0.18
Std. error	0.05	0.07	0.07
p-value of H_0: $\hat{\beta} = 0$	0.00	0.00	0.01
Panel D: Tariff revenue			
Point estimate	0.25	0.40	-0.16
Std. error	0.07	0.06	0.04
p-value of H_0: $\hat{\beta} = 0$	0.00	0.00	0.00

Notes: Sample of changes between 2016 and 2019 for 25,115 welfare-relevant outcomes in Panel A, 6,179 exported varieties with data on prices in Panel B, and 9,468 imported varieties with data on prices and duties in Panels C and D. In the sample $\mathcal{N}$ associated with each panel, we use the preferred IV z^{pref}, as defined by equations (9) and (12), to compute: $\frac{1}{|\mathcal{N}|}\sum_{n\in\mathcal{N}} z_n^{\text{pref}} \Delta y_n$ with Δy_n the actual change in outcome n, in column (1); $\frac{1}{|\mathcal{N}|}\sum_{n\in\mathcal{N}} z_n^{\text{pref}} \Delta x_n(\hat{\theta})$ with $\Delta x_n(\hat{\theta})$ the predicted change in outcome n using FGKK's estimates $\hat{\theta}$ described in Section 3.1, in column (2); the IV-based test $\frac{1}{|\mathcal{N}|}\sum_{n\in\mathcal{N}} z_n^{\text{pref}}(\Delta y_n - \Delta x_n(\hat{\theta}))$, in column (3). Inference accounting for the estimation of $\hat{\theta}$ as described in Appendix C.1.