Robustness Measures for Welfare Analysis*

Zi Yang Kang[†] Shoshana Vasserman[‡]

February 2025

Abstract

Economists routinely make functional form assumptions on demand curves to derive welfare conclusions. How sensitive are these conclusions to such assumptions? In this paper, we develop robustness measures that quantify the extent to which the true demand curve must deviate from common functional form assumptions in order to overturn a welfare conclusion. We parametrize this variability in terms of the gradient and curvature of the demand curve. By leveraging tools from information design, we show that our measures are easy to compute. Our measures are also flexible and easy to use, as we illustrate through empirical applications.

JEL classification: C14, C51, D04, D61, F14, H21, Q48

Keywords: welfare analysis, welfare bounds, robustness, functional form assumptions

* This paper is dedicated to Eddie Lazear, whose enthusiasm and support we benefited from until his untimely passing. We are grateful for comments and suggestions from Lanier Benkard, Steve Berry, Steve Cicala, Ignacio Cuesta, Rebecca Diamond, Matthias Giesecke, Bob Hahn, Nathan Hendren, Ravi Jagadeesan, Philipp Jäger, Rob Metcalfe, Neale Mahoney, Steve Redding, Andy Skrzypacz, Pietro Tebaldi, Chris Tonetti, and many generous seminar and conference participants. Finally, we thank Noah Jussila and especially Jai Singh for their excellent research assistance.

† Department of Economics, University of Toronto; zy.kang@utoronto.ca.

‡ Graduate School of Business, Stanford University and NBER; svass@stanford.edu.

1 Introduction

Many empirical evaluations of policy interventions make functional form assumptions on supply or demand curves in order to draw welfare conclusions. Often, these assumptions are connected to regression specifications. For instance, the estimated price coefficient of a linear regression can be interpreted as the gradient of a linear demand curve. These assumptions are innocuous when the magnitude of the policy intervention is small: Taylor's theorem implies that any demand curve can be locally well-approximated by a linear demand curve.[1] However, when the magnitude of the policy intervention is substantial, these assumptions are no longer innocuous: welfare conclusions might not be robust to the particular functional form assumed.

In this paper, we propose a framework to evaluate the robustness of welfare conclusions with respect to a continuum of alternatives that could have been chosen instead. Our starting point is a functional form assumption—or, equivalently, a regression specification—that is used to estimate the welfare impact of a given policy intervention and conclude whether the intervention is net beneficial. Our framework yields measures of robustness that quantify the minimum violations of the functional form assumption that is required to overturn the welfare conclusion. A conclusion is thus "more robust" if a larger violation is required to overturn it.

To fix ideas, consider the canonical example of a tax levied on a good (Harberger, 1964). There are two periods: $t = 0$ before the tax is levied, and $t = 1$ after. We assume that the demand curve does not shift between the two periods. At $t = 0$, q_0 units of the good are sold at a unit price of p_0. At $t = 1$, the posted price remains unchanged, but an *ad valorem* tax τ is introduced, yielding an effective price of $p_1 = (1 + \tau)\, p_0$ and a new quantity q_1.

A researcher who wishes to evaluate the net impact of the tax on consumer surplus faces an imputation problem. Even if she observes the points (p_0, q_0) and (p_1, q_1) perfectly, she does not observe any further information about the demand curve $D(p)$ that connects them. To resolve this, the researcher might interpolate between (p_0, q_0) and (p_1, q_1) with a straight line (as Harberger did). If q_1 is not perfectly observed, the researcher might extrapolate from an estimated average treatment effect via $\hat{q}_1 = q_0 + \hat{\beta} \cdot \tau$, for instance. In either case, the linear imputation corresponds to assuming that the relationship $D(p) = q_0 + \hat{\beta}\,(p - p_0)$ holds at every point between p_0 and p_1.

When the tax τ is small, any imputation is justified via Taylor's theorem; but no such guarantee holds when τ is large—which is often the case in practice. To assess the extent to which her results

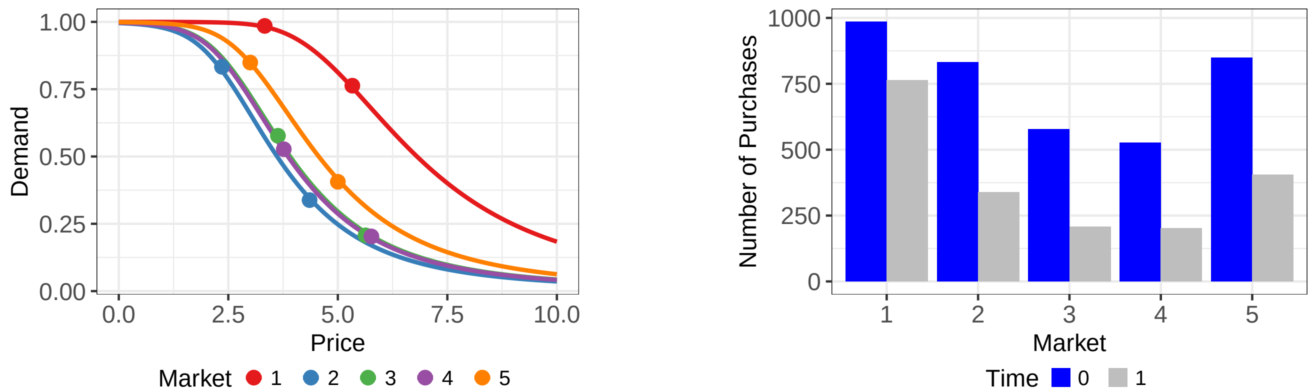
[1] Harberger (1964) employs this observation to measure the deadweight cost of taxation. A subsequent literature in sufficient statistics builds on this to analyze the welfare effects of marginal policy interventions; see Chetty (2009) and Kleven (2021) for comprehensive surveys.

would change under a different imputation—say, an exponential or isoelastic curve, instead of a linear one—the researcher might simply try alternative functional forms and compare them. But a brute-force search over possible demand curves may be challenging to compute and interpret.

Our framework provides a principled approach for quantifying the robustness of the researcher's results. Starting with a linear imputation, for example, we might ask how different a demand curve would need to be in order to overturn a conclusion based on the linearly imputed loss in consumer surplus (e.g., for the loss to be lower than some given benchmark). We parametrize difference from linearity through variability in *gradient* and *curvature*. In each case, we establish a measure of robustness, defined as the threshold index of variability at which the welfare conclusion would be reversed. This index interpolates between two extremes: It takes its maximum value when the conclusion holds for *any* demand curve that passes through (p_0, q_0) and (p_1, q_1); this means that the conclusion is maximally robust. It takes its minimum value when the conclusion holds only with a linear imputation; this means that the conclusion is minimally robust (i.e., maximally dependent on the functional form being used).

These measures of robustness are flexible and simple to use in empirical applications. On one hand, our measures extend easily to other regression-based imputations such as isoelastic demand (e.g., based on an elasticity estimate in a log-log regression). On the other hand, while we motivate these imputations through regression, our framework applies directly to cases where the treatment effect of the price change on quantities demanded is estimated in other ways. To illustrate, we apply our framework to three empirical examples drawn from published papers.

A key feature of our measures of robustness is that they account for all demand curves consistent with the observed points. In particular, our framework captures the possibility that the true (but unobserved) demand curve might be any of uncountably many demand curves that pass through the observed points. Identifying the "least variable" demand curve—relative to the imputation benchmark—in this uncountable set requires solving an *a priori* challenging infinite-dimensional optimization problem. We overcome this technical challenge by uncovering a novel connection between this problem and recent work in information design. This allows us to leverage information design tools and obtain general bounds on welfare for demands within general constraint sets. We then focus on specific constraint sets that are one-dimensional parameterized relaxations of standard functional form assumptions; this allows us to provide a recipe to compute robustness measures and graphically represent robustness sensitivity exercises around a standard framework.



(a) Unobserved underlying demand curves. *(b) Observed sales per 1000 consumers by group.*

Figure 1: Simulated demand data for the motivating example.

2 Illustration and Preview of Results

To illustrate our framework, we begin with an idealized experiment in the spirit of Angrist and Pischke (2009). Suppose that a researcher observes price and quantity data from an experiment in which prospective consumers in five different markets are randomized into two groups. Each group is assigned a different price for a good: in each market, group $t = 0$ is assigned the baseline price in that market and group $t = 1$ is assigned a surcharge of $\Delta p = \$0.50$. The researcher wishes to estimate the resulting loss in consumer surplus in order to evaluate whether to levy a tax in this amount. For simplicity, suppose that the tax revenue is used to generate a known social surplus gain of G; hence the researcher's goal is to compare the estimated loss in consumer surplus to G.

Each market might have a different baseline price and underlying consumer preferences. In this example, we suppose that consumer i in market m has preferences given by random coefficients logit demand:

$$D_{im}(p) = \frac{\exp(\xi_m - \beta_i p)}{1 + \exp(\xi_m - \beta_i p)}, \quad \beta_i \sim \mathcal{N}(-1.5, 2),$$

where ξ_m is a market-level fixed effect and β_i is an individual-level price coefficient. Figure 1(a) plots the aggregate demand curve in each market and highlights the two price-quantity pairs that are observed through the experiment in each market. For the sake of illustration, we choose parameters such that the demand curves in different markets have different amounts of curvature and the observed price-quantity pairs lie on different parts of the demand curve in each market.

Of course, in practice, the researcher observes no more than: *(i)* the price that each consumer in the experiment is assigned and *(ii)* whether or not the consumer purchased the good. Figure 1(b) summarizes what this data might look like. The blue bar in each market m represents the share of consumers in the baseline group $t = 0$ who purchased the good at the price p_m. The gray bar in each market—which is unsurprisingly lower than the corresponding blue bar—represents the share of consumers in the treated group $t = 1$ who purchased the good at the higher price $p_m + \Delta p$.

2.1 Common Approaches

Because the researcher does not observe the underlying demand curves, a common approach for estimating the loss in consumer surplus is to apply an approximation. In this idealized experiment, as prices are randomly assigned to consumers in each market, the researcher can estimate the causal impact of the surcharge on sales by comparing the quantities sold in each group for each market.

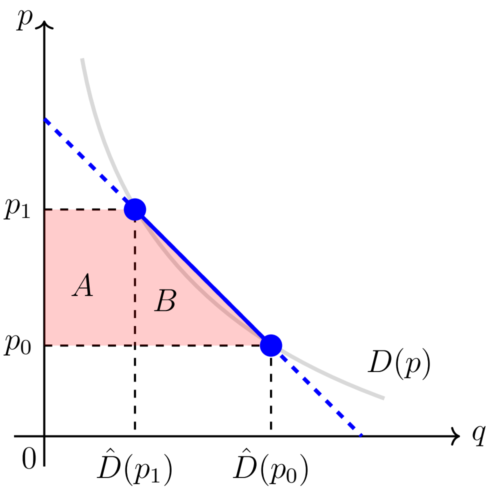
Figure 2(a) summarizes how this might look in a given market with an unknown demand curve. The demand curve $D(p)$ traces the share of consumers in the market who would purchase the good if they were offered it at each hypothetical price p. As the researcher observes only a subsample of consumers making a decision at the prices p_0 (for group $t = 0$) and p_1 (for group $t = 1$), she must infer the counterfactual behavior of each consumer if they had been assigned to the other group. As in the classical potential outcomes framework, a consistent estimator for this quantity is the sample frequency of purchases within each group, denoted by $\hat{D}(p_0)$ and $\hat{D}(p_1)$ respectively.

Having mapped the data to the hypothetical demand curve, the researcher can now estimate the loss in consumer surplus, represented as the shaded region in Figure 2(a). If the researcher knew the true demand curve $D(p)$, she would integrate it from p_0 to p_1. However, knowing only $\hat{D}(p_0)$ and $\hat{D}(p_1)$, she might instead approximate this integral by imputing a straight line between the two points, as demonstrated by the blue line in Figure 2(a).

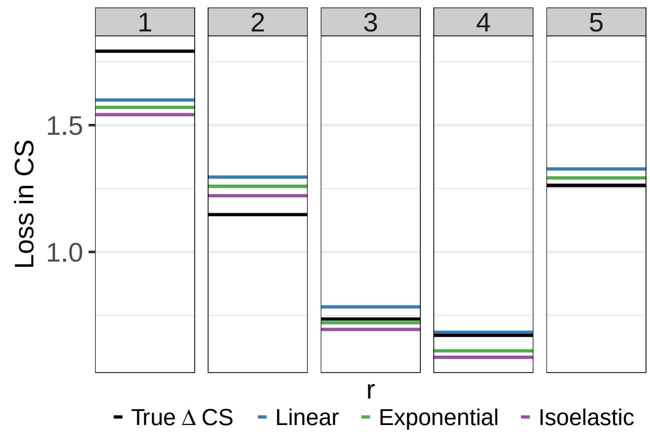
To bring our exercise closer to less idealized applications, instead of a randomized simultaneous experiment, suppose that the variation in our data came from an exogenous price shock, changing the price in each market from p_{m0} at $t = 0$ to a higher price $p_{m1} = p_{m0} + \Delta p$ at $t = 1$. In this case, the researcher might worry that unobserved factors might coincide with the treatment timing in different markets and bias the estimates of market-level treatment effects. As such, she may instead choose to estimate an average treatment effect across markets, as with the following pooled regression:

$$q_{mt} = \beta p_{mt} + \text{FE}_m + \eta_{mt}, \tag{1}$$

where η_{mt} is a mean-zero residual. The estimate of the price coefficient $\hat{\beta}$ in this regression corresponds to the average treatment effect of the price increase. In terms of Figure 2(a), $\hat{\beta}$ is the estimated gradient of the imputed linear demand curve.



(a) Imputing a linear demand curve.



(b) Estimated loss in consumer surplus by imputation.

Figure 2: Estimating loss in consumer surplus via imputations.

Once the parameters of the linear imputation have been estimated, the loss in consumer surplus can be easily approximated as the area of a trapezium:

$$\Delta\text{CS} \approx \frac{1}{2}(p_1 - p_0)\left[\hat{D}(p_0) + \hat{D}(p_1)\right] = (p_1 - p_0)\,\hat{D}(p_0) + \frac{1}{2}\hat{\beta}\,(p_1 - p_0)^2\,.$$

But what if the researcher had chosen a different imputation—say an exponential or isoelastic demand curve? Figure 2(b) plots the estimated losses in consumer surplus in each market from fitting a linear, exponential, and isoelastic demand curve (in blue, green and purple, respectively). Because this is a simulated example, we can also plot the true loss in consumer surplus in each market (in black) for comparison. While the blue, green and purple lines are often close together, they are sometimes far above or below the true loss. Indeed, as we show in Section 5, even these three commonly used imputations can be very far apart when the price shock is large.

Before we move on, it is worth noting the role of heterogeneity in this exercise. The simulated data reflects two main sources of heterogeneity: individual price coefficients β_i and market-level fixed-effects ξ_m. The regression in equation (1) aggregates individuals to the market level with a common cross-market price coefficient. This specification reflects a common constraint in empirical applications: while there may be a lot of underlying heterogeneity across units of observation, it may not be possible to identify granular responses to price shocks with precision. The consumer surplus estimates plotted in Figure 2(b) account for cross-market heterogeneity as follows: to compute them, we first predict $\hat{D}(p_0)$ and $\hat{D}(p_1)$ in each market based on its estimated fixed effect and $\hat{\beta}$; we then obtain the loss in consumer surplus between $\hat{D}(p_0)$ and $\hat{D}(p_1)$ for each functional form.[2] Because the functional forms are different, the results differ across markets even though they share a common price coefficient. Still, as the figure shows, the functional form specifications are quite restrictive. In the remainder of this section, we will build on this breakdown by market to show how our robustness measures account for heterogeneity at different levels of aggregation.

2.2 Formalizing a Measure of Robustness

In practice, the researcher cannot observe the true loss in Figure 2(b) but will still want to assess how robust her welfare results are with respect to the assumption that demand is linear. A natural approach is to fit alternative imputations, such as the exponential and isoelastic imputations in our example above. If all of these imputations produce similar estimates, then the linear assumption can be treated as innocuous and the result can be deemed robust.

To facilitate this exercise, we propose a two-step approach.

Step 1: Deriving welfare bounds. Rather than iteratively fit ad hoc imputations, we compute the range of welfare losses that are attainable by general classes of imputations. For now, we focus on the class of demand curves for which the gradient at different prices can vary within a given range, parametrized by $\underline{\beta}$ and $\overline{\beta}$:

Illustrative Assumption 1. *Given $\underline{\beta} \leq \overline{\beta} \leq 0$, the gradient of the demand curve is bounded between $\underline{\beta}$ and $\overline{\beta}$ at every price between the observed prices:*

$$D'(p) \in [\underline{\beta}, \overline{\beta}] \qquad \textit{for } p \in [p_0, p_1].$$

Depending on the values of $\underline{\beta}$ and $\overline{\beta}$, Illustrative Assumption 1 permits different imputations. In the special case where $\underline{\beta} = \overline{\beta} = \hat{\beta}$, only the linear imputation where $D'(p) = \hat{\beta}$ for $p \in [p_0, p_1]$ is allowed. But when $\underline{\beta}$ and $\overline{\beta}$ are sufficiently far apart, the exponential and isoelastic imputations are allowed as well.

It is not a priori obvious that the range of welfare losses can be computed under Illustrative Assumption 1: there are uncountably many demand curves that satisfy the assumption if $\underline{\beta} \neq \overline{\beta}$. However, we show that tight bounds for welfare losses can in fact be easily computed. By leveraging tools developed in the information design literature, we show that the largest and smallest possible

[2] We discuss directly fitting nonlinear functional forms as a benchmark in Section 2.3.

losses in consumer surplus between p_0 and p_1 under Illustrative Assumption 1, $\overline{\Delta\text{CS}}$ and $\underline{\Delta\text{CS}}$, are

$$\begin{cases} \overline{\Delta\text{CS}} & = \dfrac{(p_1 - p_0)(q_0 + q_1)}{2} + \dfrac{\left[q_0 - q_1 + \underline{\beta}(p_1 - p_0)\right]\left[q_0 - q_1 + \overline{\beta}(p_1 - p_0)\right]}{2\left(\underline{\beta} - \overline{\beta}\right)}, \\ \underline{\Delta\text{CS}} & = \dfrac{(p_1 - p_0)(q_0 + q_1)}{2} - \dfrac{\left[q_0 - q_1 + \underline{\beta}(p_1 - p_0)\right]\left[q_0 - q_1 + \overline{\beta}(p_1 - p_0)\right]}{2\left(\underline{\beta} - \overline{\beta}\right)}. \end{cases}$$

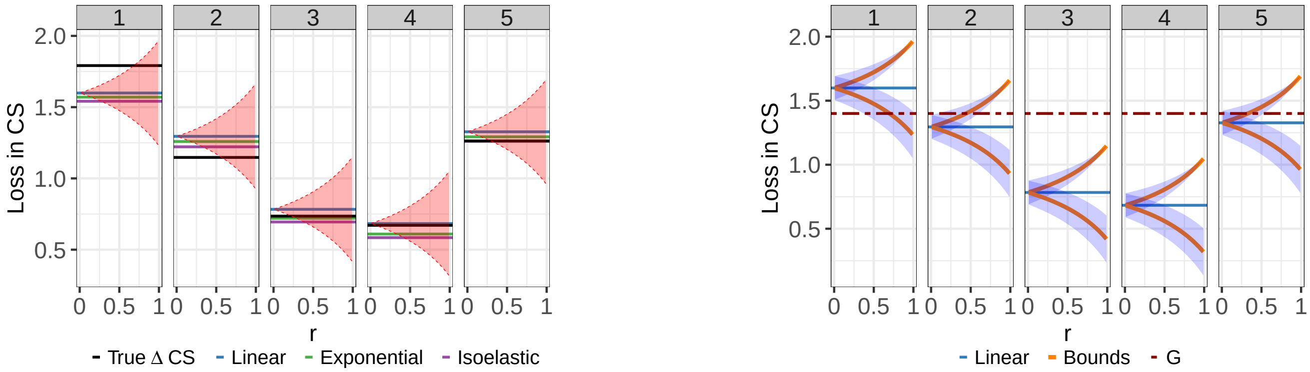
Step 2: Computing robustness measures. To make our derived welfare bounds useful for evaluating robustness, we propose one-dimensional parametrizations for classes of imputations that parsimoniously capture increasing relaxations of the linear imputation. In the case of Illustrative Assumption 1, although $\underline{\beta}$ and $\overline{\beta}$ need not be otherwise restricted, we propose the following parametrization based on the estimated gradient $\hat{\beta}$—such as from the regression in equation (1):

$$\underline{\beta} = \frac{\hat{\beta}}{1 - r} \quad \text{and} \quad \overline{\beta} = \hat{\beta}(1 - r) \qquad \text{for } r \in [0, 1].$$

Here, r represents the percentage deviation from $\hat{\beta}$: at any price between p_0 and p_1, the gradient of the demand curve is at most $r \cdot 100\%$ different in magnitude than $\hat{\beta}$. When $r = 0$, only the linear demand curve with gradient $\hat{\beta}$ is allowed by Illustrative Assumption 1. By contrast, as $r \to 1$, any downward-sloping demand curve that passes through the observed points $(p_0, \hat{D}(p_0))$ and $(p_1, \hat{D}(p_1))$ is allowed in the limit. Consequently, r measures the extent to which Illustrative Assumption 1 constrains the shape of the demand curve. This parametrization also leads to the following simple expressions for the largest and smallest losses in consumer surplus as functions of r:

$$\begin{cases} \overline{\Delta\text{CS}}(r) & = \dfrac{(p_1 - p_0)\left[q_0 + q_1(1 - r)\right]}{2 - r}, \\ \underline{\Delta\text{CS}}(r) & = \dfrac{(p_1 - p_0)\left[q_0(1 - r) + q_1\right]}{2 - r}. \end{cases}$$

In the context of our example, Figure 3(a) plots $\overline{\Delta\text{CS}}(r)$ and $\underline{\Delta\text{CS}}(r)$ in each market for each $r \in [0, 1]$. The upper and lower boundaries of the shaded area in each market correspond to $\overline{\Delta\text{CS}}(r)$ and $\underline{\Delta\text{CS}}(r)$ respectively, while the shaded area itself represents the range of admissible consumer surplus losses that are consistent with a demand curve whose gradients between p_0 and p_1 are bounded between $\frac{\hat{\beta}}{1-r}$ and $\hat{\beta}(1 - r)$. Note that the set of demand curves allowed by each r is quite permissive: the gradient of demand is not required to adhere to any specific parametric form. In this way, the bounds account for a broad range of (aggregate) demand curves that could result from heterogeneous individual preferences. Unsurprisingly, in each market, the loss in consumer



(a) Bounds against common imputations.

(b) Bounds with confidence intervals against G.

Figure 3: Welfare bounds when the demand curve has nonconstant gradient.

surplus under the true demand curve (i.e., the black line) overlaps with the shaded area. However, it may be far from the linear imputation (i.e., the blue line), which always lies fully within the shaded area; this is because the only admissible demand curve when $r = 0$ is the linear imputation itself. The intersection between the black line and the boundary of the shaded area allows us to see how much variability is required in order to rationalize the true loss in consumer surplus in each market. In market 1, for instance, this is quite high: an r above 0.75 is required, or gradients ranging up to 4 times larger than than the estimated linear price coefficient.

The researcher can use a similar exercise to assess the robustness of a welfare conclusion based on a particular imputation. To do this in our example, we must first clarify what the benchmark welfare conclusion is. Recall that the goal of our exercise is to compare the potential loss in consumer surplus from a tax increasing the price from p_0 to p_1 against a social surplus gain of G that would be generated by the tax. To illustrate robustness measures for this goal, we consider a welfare assessment in each market against G, plotted in a dashed line in Figure 3(b).[3]

We now assess the welfare conclusion in each market based on the linear imputation benchmark. In market 1, the conclusion is negative: the imputed loss in consumer surplus is above G. In markets 2–5, the conclusion is positive. To test the robustness of each conclusion, we therefore compare G to the lower bound in market 1, and to the upper bound in markets 2–5. Once we have defined the comparison, it is easy to read off the threshold necessary to reverse the conclusion in each market: we need only find the smallest value of r for which the lower or upper bound on the loss in consumer surplus crosses G. In market 2, for instance, the threshold r^* is about 0.5. This

[3] In many applications (e.g., Section 5.1), the relevant welfare benchmark is at an aggregate level across markets. Our simulated example considers market-level benchmarks in order to illustrate a wider ranges of cases.

means that the demand curve of market 2 would need to have gradients outside the range $(2\hat{\beta}, \hat{\beta}/2)$ in order to rationalize a loss in consumer surplus higher than G. As we discuss in Section 4.2, r^* can be interpreted in other intuitive ways. For instance, for the linear imputation benchmark, $r^* = 0.5$ means that a second-order Taylor approximation (i.e., a quadratic imputation) cannot overturn the benchmark conclusion; higher order terms would be necessary. In some markets—like markets 3 and 4—G is above the upper bound entirely, meaning that $r^* = 1$. In this case, no downward-sloping demand curve can overturn the welfare conclusion.

Each boundary curve, depicted in red in Figure 3(b), has a 95% confidence band around it, shaded in blue. This confidence band captures the uncertainty about what the true boundary curves—and subsequently, the true threshold r^*—are, arising from the measurement error in $\hat{\beta}$. To see where this comes from, note that our bounds provide a robust deterministic guarantee on the possible losses in consumer surplus with respect to uncertainty about the shape of the demand curve in each market—among the set of demand curves parameterized by $\hat{\beta}$ and r under Illustrative Assumption 1. However, had we measured $\hat{\beta}$ differently, we would have obtained a different set of bounds. In order to obtain a bound that is also robust to uncertainty regarding the estimate $\hat{\beta}$, we can propagate the uncertainty from the regression in equation (1) to the bounds calculation by nesting it in a bootstrap procedure. Because the bounds under Illustrative Assumption 1 turn out to be expressible in closed form, we computed the standard errors on the bounds in Figure 3(b) directly by applying the delta method with respect to $\hat{\beta}$ at each value of r.

Before concluding this subsection, it is worth noting that we could have accommodated a less idealized example for this exercise. Our regression approach in equation (1) extends easily in cases where, instead of a market fixed effect and simultaneous timing, treatment timing is staggered and heterogeneity among observations is modeled through a set of comprehensive controls. More generally, our approach does not require an exogenous price shock. It requires only a baseline quantity (at a baseline price) and a valid causal estimate of the treatment effect that the change in price would induce in the quantity demanded by the treated population. To illustrate how this might work, we extend our example to an instrumental variables (IV) setting. Instead of a certain price increase of Δp in period $t = 1$, suppose that each consumer i in market m is assigned a lottery number Z_{imt} that is correlated with the probability of being treated. In this case, a direct regression of quantities on prices would not yield a valid estimator of $\hat{\beta}$, but the following IV regression would:

$$\mathbf{1}(\text{purchase})_{imt} = \beta p_{imt} + \text{FE}_m + \eta_{imt},$$
$$p_{imt} = p_{m0} + Z_{imt}\Delta p + \nu_{imt},$$

where η_{imt} and ν_{imt} are both mean-zero residuals. While the mechanics of estimating $\hat{\beta}$ (in this case, a LATE instead of an ATE) are different than in equation (1), the interpretation in terms of bounding consumer surplus is nearly the same. Under full take up, $\hat{\beta}$ can be interpreted as the average gradient of the demand curves across the markets.[4] As such, we can apply the same analysis as in Figure 3 in much the same way.

2.3 Preview of Other Results

So far, our motivating example has focused on a measure of robustness with respect to variability in gradient given a linear imputation benchmark. In the remainder of the paper, we also develop a measure of robustness with respect to variability in curvature. In addition, we show how our measures can be applied to other commonly used imputation benchmarks.

Variability in curvature. A different relaxation of the linear demand assumption is to allow curvature in demand. Linear imputations have the special property that the second derivative of demand with respect to price is zero *everywhere* along the demand curve: $D''(p) = 0$ for $p \in [p_0, p_1]$. Instead, we consider the class of demand curves for which the second derivative at different prices can vary within a given range, parametrized by $\underline{\gamma}$ and $\overline{\gamma}$:

Illustrative Assumption 2. *Given $\underline{\gamma} \leq 0 \leq \overline{\gamma}$, the second derivative of the demand curve with respect to price is bounded between $\underline{\gamma}$ and $\overline{\gamma}$ at every price between the observed prices:*

$$D''(p) \in [\underline{\gamma}, \overline{\gamma}] \qquad \textit{for } p \in [p_0, p_1].$$

Bounding the loss in consumer surplus under Illustrative Assumption 2 is more technically challenging than under Illustrative Assumption 1. This is because the condition that $\overline{\beta} \leq 0$ in Illustrative Assumption 1 implies that the demand curve is downward-sloping. By contrast, we have to separately ensure that the demand curve is downward-sloping in addition to the restrictions

[4] If there is not full take up, then the estimated gradient does not account for a segment of the market for which the researcher has no information about revealed preferences.

on curvature in Illustrative Assumption 2. In Section 4, we show how this technical challenge can be overcome with tools developed in the information design literature.

Similar to the case of Illustrative Assumption 1, $\underline{\gamma}$ and $\overline{\gamma}$ in Illustrative Assumption 2 need not be otherwise restricted. However, we focus on the following parametrization based on the estimated gradient $\hat{\beta}$—such as from the regression in equation (1):

$$\underline{\gamma} = \kappa \cdot \frac{2\hat{\beta}}{p_1 - p_0} \quad \text{and} \quad \overline{\gamma} = -\kappa \cdot \frac{2\hat{\beta}}{p_1 - p_0} \qquad \text{for } \kappa \in \mathbb{R}_+.$$

When $\kappa = 0$, only the linear demand curve with zero curvature is allowed by Illustrative Assumption 1. As we show in Section 4, the normalization factor of $2\hat{\beta}/(p_1 - p_0)$ is chosen so that the downward-sloping constraint on the demand curve binds if and only if $\kappa \geq 1$. Moreover, as $\kappa \to \infty$, any downward-sloping demand curve that passes through the observed points $(p_0, \hat{D}(p_0))$ and $(p_1, \hat{D}(p_1))$ is allowed in the limit. Consequently, similar to r in Illustrative Assumption 1, κ measures the extent to which Illustrative Assumption 2 constrains the shape of the demand curve.

Other common imputation benchmarks. Many empirical applications (cf. Section 5) use imputation benchmarks other than linear demand. For instance, rather than a linear regression (1) that yields a gradient estimate $\hat{\beta}$, the researcher might run a log-log regression of the following form:

$$\log q_{mt} = \varepsilon \log p_{mt} + \text{FE}_m + \eta_{mt}. \tag{2}$$

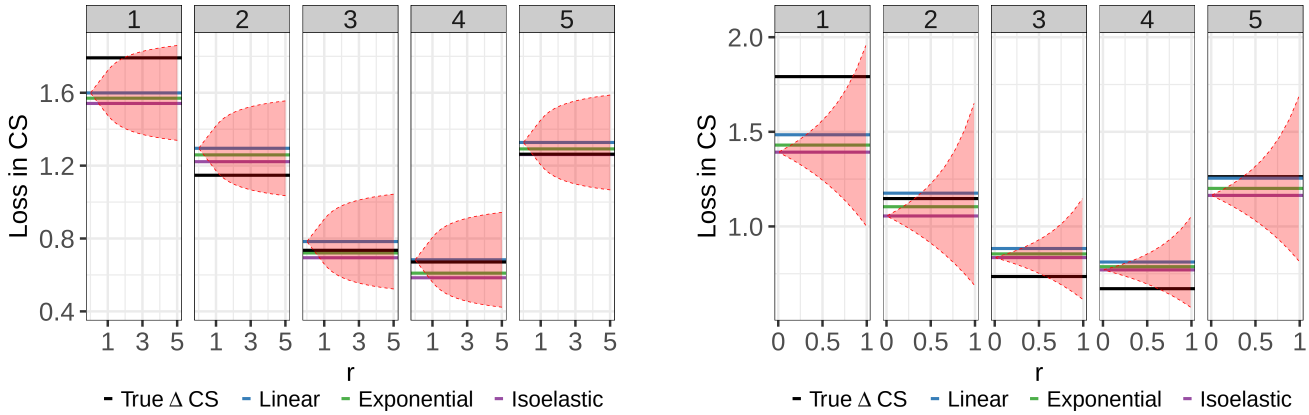
In this case, $\hat{\varepsilon}$ can be interpreted as the estimated elasticity of the isoelastic demand curve $D(p) = q_0 \cdot (p/p_0)^{\varepsilon}$. Of course, other regression specifications are possible and imply different imputation benchmarks.[5]

To allow for these applications, we consider demand curves such that $A(q)$ is affine in $B(p)$ for some predetermined increasing functions A and B. In the context of our example, this can be interpreted as running a regression of the following more general form:

$$A(q_{mt}) = \beta B(p_{mt}) + \text{FE}_m + \eta_{mt}.$$

Demand curves of this form include many commonly used imputations, such as:

[5] While common, it is not required that the benchmark imputation is derived from a regression. For instance, researchers may estimate the treatment effect on price in a separate procedure and impute a demand curve based on the estimated quantities. Only the final imputation is relevant to our proposed robustness measures.



(a) Curvature bounds.

(b) Elasticity (log-log gradient) bounds.

Figure 4: Alternative specifications of bounds in the motivating example.

- linear demand, where $A(q) = q$ and $B(p) = p$;
- isoelastic demand, where $A(q) = \log q$ and $B(p) = \log p$;
- exponential demand, where $A(q) = \log q$ and $B(p) = p$;
- logarithmic demand, where $A(q) = q$ and $B(p) = \log p$; and
- monomial demand, where $A(q) = q^n$ and $B(p) = p$ for some $n > 0$.

Our robustness exercise takes the increasing functions A and B as given: these are pinned down by the functional form assumption that the researcher makes. Using this functional form as the benchmark, we follow a similar two-step approach as in Section 2.2:

1. First, we derive welfare bounds for demand curves that satisfy analogs of Illustrative Assumptions 1 and 2 for the chosen A and B. That is, for a given range of gradients or curvatures in A–B space, we obtain tight bounds on welfare across all downward-sloping demand curves with gradients or curvatures in the allowed range.

2. Second, we compute robustness measures by proposing one-dimensional parametrizations of our analogs of Illustrative Assumptions 1 and 2. Our robustness measures thus quantify the minimum threshold deviation away from A and B that is necessary to overturn the welfare conclusion obtained under this choice of A and B.

Figure 4 demonstrates how our analysis in Section 2.2 extends to: (a) robustness of linear demand with respect to curvature; and (b) robustness of isoelastic demand with respect to gradients. As in the case of linear demand, we can obtain a full set of bounds as a function of a baseline quantity (or log-quantity) and an estimated treatment effect. We can then compare the bounds against a benchmark value to find a minimal critical value (κ in the case of curvature; r in the case of gradient) that we may interpret as small or large.

As our discussion highlights, the researcher has several degrees of freedom in determining how to apply our robustness analysis. She must first choose her benchmark functional form, and then choose the type of relaxation (in gradients or curvatures) to apply. From the perspective of our analysis, both choices are external. The choice of benchmark should reflect the baseline welfare analysis that the researcher would deem appropriate prior to considering robustness, and the choice of relaxation should reflect the notion of robustness that the researcher can best compare to empirical moments.[6] As such, we recommend that researchers apply the robustness measure that is most natural and easy to interpret in the context of their application rather than produce a comprehensive table of benchmark relaxations.

3 Framework

We study a market for a good, for which aggregate demand is downward-sloping by assumption[7] and denoted by $D(\cdot)$. Although our example in Section 2 shows that our approach can be applied more generally, we suppose for simplicity that there are two time periods, $t = 0$ and $t = 1$, and the good is exposed to an exogenous price increase (e.g., a new *ad valorem* tax) at $t = 1$. For each period t, we denote the price by p_t and the quantity demanded by $q_t = D(p_t)$.

We examine the problem faced by a researcher who wishes to evaluate the welfare impact of the price increase. Throughout most of our theoretical analysis, we focus on the loss in Marshallian consumer surplus as a measure of welfare impact; this is equal to the area below the demand curve between p_0 and p_1:

$$\Delta\text{CS} = \int_{p_0}^{p_1} D(p)\ \mathrm{d}p. \tag{3}$$

However, as we demonstrate through empirical applications in Section 5 (see also Appendix B.4), our approach extends to other measures of welfare impact, such as the change in deadweight loss,

[6] While the specific robustness measure computed for a given benchmark and relaxation will differ, there is a sense in which the robustness conclusions based on these different approaches will generally be similar: Because Illustrative Assumptions 1 and 2 and their analogs in A–B space allow the true demand curve to be any downward-sloping demand curve that passes through the observed data points (for a sufficiently large range of gradients or curvatures), the resulting welfare bounds will coincide as $r \to 1$ and $\kappa \to \infty$.

[7] This is widely maintained in empirical applications and holds if consumers have quasilinear utility in money.

compensating variation, and equivalent variation.

3.1 Benchmark

We begin by considering a benchmark where the researcher imposes a functional form relationship between quantity demanded, q, and price, p. To capture a wide variety of common approaches, we suppose that the researcher assumes that $A(q)$ is affine in $B(p)$, where $A(\cdot)$ and $B(\cdot)$ are predetermined increasing functions. As explained in Section 2, this connects naturally to regression in empirical applications and nests many familiar functional form assumptions (including linear, isoelastic, exponential, logarithmic, and monomial demand).

The functional form assumption that the researcher chooses allows her to identify a unique demand curve that passes through the two observed points, (p_0, q_0) and (p_1, q_1). For instance, under the assumption of linear demand, the implied demand curve is

$$D(p) = \frac{q_1 - q_0}{p_1 - p_0}(p - p_0) + q_0.$$

In general, for arbitrary increasing functions $A(\cdot)$ and $B(\cdot)$, the implied demand curve is

$$D(p) = A^{-1}\left(\frac{A(q_1) - A(q_0)}{B(p_1) - B(p_0)}[B(p) - B(p_0)] + A(q_0)\right).$$

Using this demand curve, the researcher compares the estimated loss in consumer surplus, ΔCS, to a known social surplus gain, $G > 0$. As in Section 2, G might represent the potential gain in social surplus generated by the tax revenue. For arbitrary increasing functions $A(\cdot)$ and $B(\cdot)$, equation (3) implies that the loss in consumer surplus can be written as

$$\Delta\text{CS} = \int_{p_0}^{p_1} A^{-1}\left(\frac{A(q_1) - A(q_0)}{B(p_1) - B(p_0)}[B(p) - B(p_0)] + A(q_0)\right) \mathrm{d}p.$$

If $\Delta\text{CS} > G$, the researcher concludes that the welfare impact of the price increase is net positive. Similarly, if $\Delta\text{CS} < G$, she concludes that the welfare impact of the price increase is net negative.

3.2 Measures of Robustness

We now develop two measures to evaluate the robustness of the researcher's welfare conclusion. To this end, we parametrize different relaxations of her functional form assumption. The measures that we develop represent the *minimum* relaxation required to reverse her welfare conclusion.

3.2.1 Variability in Gradient

One way in which the researcher's functional form assumption is special is that it requires the gradient of $A(q)$ with respect to $B(p)$ to be constant between the prices p_0 and p_1. This gradient can be expressed as

$$\hat{\beta} = \frac{A(q_1) - A(q_0)}{B(p_1) - B(p_0)}.$$

Put differently, common regression specifications often impose a constant-gradient assumption between some transformation of quantities and prices. Linear demand, for instance, requires the gradient of the demand curve—the gradient of q with respect to p—to be constant, while isoelastic demand requires the price elasticity—the gradient of $\log q$ with respect to $\log p$—to be constant.

For our first robustness measure, we consider relaxations of this assumption by allowing the gradient of $A(q)$ with respect to $B(p)$ to vary within a given range, parametrized by $\underline{\beta}$ and $\overline{\beta}$:

Assumption 1. *Given $\underline{\beta} \leq \overline{\beta} \leq 0$, the gradient of $A(D(p))$ with respect to $B(p)$ is bounded between $\underline{\beta}$ and $\overline{\beta}$ at every price between the observed prices:*

$$\frac{A'(D(p))D'(p)}{B'(p)} \in [\underline{\beta}, \overline{\beta}] \qquad \textit{for } p \in [p_0, p_1].$$

Although Assumption 1 allows $\underline{\beta}$ and $\overline{\beta}$ to vary independently of each other, it is useful to view relaxations of gradient constraints through symmetric deviations from the estimated gradient, $\hat{\beta}$. We thus focus on the following one-dimensional parametrization of $\underline{\beta}$ and $\overline{\beta}$ in Assumption 1 based on $\hat{\beta}$:

$$\underline{\beta}(r) = \frac{\hat{\beta}}{1-r} \quad \text{and} \quad \overline{\beta}(r) = \hat{\beta}\,(1-r) \qquad \text{for } r \in [0,1].$$

As noted in Section 2, this parametrization is constructed such that r represents the percentage deviation relative to $\hat{\beta}$. When $r = 0$, only the demand curve with a constant gradient $\hat{\beta}$ of $A(q)$ with respect to $B(p)$ is allowed by Assumption 1. As $r \to 1$, any downward-sloping demand curve that passes through the observed points (p_0, q_0) and (p_1, q_1) is allowed in the limit. Consequently, r measures the extent to which Assumption 1 constrains the shape of the demand curve.

Under this parametrization, we can express the largest and smallest possible losses in consumer surplus as functions of r. To this end, for any given $r \in [0,1]$, let $\mathcal{D}_1(r)$ denote the set of feasible demand curves satisfying Assumption 1 with $\underline{\beta}(r)$ and $\overline{\beta}(r)$:

$$\mathcal{D}_1(r) := \left\{ D : [p_0, p_1] \to \mathbb{R} \text{ is decreasing, } D(p_t) = q_t,\ \underline{\beta}(r) \leq \frac{A'(D(p))D'(p)}{B'(p)} \leq \overline{\beta}(r) \right\}.$$

Then the largest and smallest possible losses in consumer surplus can be written as

$$\begin{cases} \overline{\Delta \text{CS}}_1(r) & := \sup\limits_{D \in \mathcal{D}_1(r)} \displaystyle\int_{p_0}^{p_1} D(p)\ \text{d}p, \\ \underline{\Delta \text{CS}}_1(r) & := \inf\limits_{D \in \mathcal{D}_1(r)} \displaystyle\int_{p_0}^{p_1} D(p)\ \text{d}p. \end{cases}$$

This allows us to formalize a measure of robustness, r^*, that represents the minimum relaxation required to reverse the researcher's welfare conclusion. On one hand, if her benchmark conclusion (under her functional form assumption) is that the welfare impact is net positive, then r^* is the minimum relaxation such that the largest possible loss in consumer surplus exceeds the gain in social surplus:

$$r^* := \inf \left\{ r \in [0, 1] : \overline{\Delta \text{CS}}_1(r) > G \right\} \qquad \text{if } \Delta \text{CS} < G.$$

On the other hand, if her benchmark welfare conclusion is that the welfare impact is net negative, then r^* is the minimum relaxation such that the gain in social surplus exceeds the smallest possible loss in consumer surplus:

$$r^* := \inf \{ r \in [0, 1] : \underline{\Delta \text{CS}}_1(r) < G \} \qquad \text{if } \Delta \text{CS} > G.$$

Equivalently, r^* can be defined symmetrically for both cases as the minimum relaxation for which $G \in [\underline{\Delta \text{CS}}_1(r^*), \overline{\Delta \text{CS}}_1(r^*)]$.

Higher values of r^* indicate that the researcher's welfare conclusion is more robust to her functional form assumption. If $r^* = 0$, her welfare conclusion holds only when the gradient of $A(q)$ with respect to $B(p)$ is constant *everywhere* between p_0 and p_1: any violation of the constant-gradient assumption overturns the welfare conclusion. By contrast, if $r^* = 1$, then her welfare conclusion holds for any demand curve that passes through the observed points and does not rely on the constant-gradient assumption at all. We defer further discussion of how one can interpret the magnitude of r^* to Section 4.2.

3.2.2 Variability in Curvature

Another way in which the researcher's functional form assumption is special is that it requires the second derivative of $A(q)$ with respect to $B(p)$ to be zero between the prices p_0 and p_1. Put differently, the constant-gradient assumption in common regression specifications can be viewed as a zero-curvature assumption between some transformation of quantities and prices.

For our second robustness measure, we relax this assumption by considering the class of demand curves for which the second derivative of $A(q)$ with respect to $B(p)$ can vary within a given range, parametrized by $\underline{\gamma}$ and $\overline{\gamma}$:

Assumption 2. *Given $\underline{\gamma} \leq 0 \leq \overline{\gamma}$, the second derivative of $A(D(p))$ with respect to $B(p)$ is bounded between $\underline{\gamma}$ and $\overline{\gamma}$ at every price between the observed prices:*

$$\frac{1}{B'(p)} \frac{\mathrm{d}}{\mathrm{d}p} \left[\frac{A'(D(p))D'(p)}{B'(p)} \right] \in [\underline{\gamma}, \overline{\gamma}] \qquad \textit{for } p \in [p_0, p_1].$$

Like our first robustness measure, we focus on the following one-dimensional parametrization of $\underline{\gamma}$ and $\overline{\gamma}$ in Assumption 2 based on the estimated gradient $\hat{\beta}$:

$$\underline{\gamma}(\kappa) = \kappa \cdot \frac{2\hat{\beta}}{B(p_1) - B(p_0)} \quad \text{and} \quad \overline{\gamma}(\kappa) = -\kappa \cdot \frac{2\hat{\beta}}{B(p_1) - B(p_0)} \qquad \text{for } \kappa \in \mathbb{R}_+.$$

When $\kappa = 0$, only the demand curve for which $A(q)$ is affine in $B(p)$ is allowed by Assumption 2. As $\kappa \to \infty$, any downward-sloping demand curve that passes through the observed points (p_0, q_0) and (p_1, q_1) is allowed in the limit. Therefore, similar to r in Assumption 1, κ measures the extent to which Assumption 2 constrains the shape of the demand curve.

Under this parametrization, we can express the largest and smallest possible losses in consumer surplus as functions of κ. To this end, for any given $\kappa \in \mathbb{R}_+$, let $\mathcal{D}_2(\kappa)$ denote the set of feasible demand curves satisfying Assumption 2 with $\underline{\gamma}(\kappa)$ and $\overline{\gamma}(\kappa)$:

$$\mathcal{D}_2(\kappa) := \left\{ D : [p_0, p_1] \to \mathbb{R} \text{ is decreasing},\ D(p_t) = q_t,\ \underline{\gamma}(\kappa) \leq \frac{1}{B'(p)} \frac{\mathrm{d}}{\mathrm{d}p} \left[\frac{A'(D(p))D'(p)}{B'(p)} \right] \leq \overline{\gamma}(\kappa) \right\}.$$

Then the largest and smallest possible losses in consumer surplus can be written as

$$\begin{cases} \overline{\Delta\mathrm{CS}}_2(\kappa) & := \sup_{D \in \mathcal{D}_2(r)} \int_{p_0}^{p_1} D(p)\ \mathrm{d}p, \\ \underline{\Delta\mathrm{CS}}_2(\kappa) & := \inf_{D \in \mathcal{D}_2(r)} \int_{p_0}^{p_1} D(p)\ \mathrm{d}p. \end{cases}$$

As before, this allows us to formalize a measure of robustness, κ^*, that represents the minimum relaxation required to reverse the researcher's welfare conclusion. On one hand, if her benchmark welfare conclusion is that the welfare impact is net positive, then κ^* is the minimum relaxation such that the largest possible loss in consumer surplus exceeds the gain in social surplus:

$$\kappa^* := \inf\left\{\kappa \in \mathbb{R}_+ : \overline{\Delta\text{CS}}_2(\kappa) > G\right\} \qquad \text{if } \Delta\text{CS} < G.$$

On the other hand, if her benchmark welfare conclusion is that the welfare impact is net negative, then κ^* is the minimum relaxation such that the gain in social surplus exceeds the smallest possible loss in consumer surplus:

$$\kappa^* := \inf\left\{\kappa \in \mathbb{R}_+ : \underline{\Delta\text{CS}}_2(\kappa) < G\right\} \qquad \text{if } \Delta\text{CS} > G.$$

Similar to r^*, higher values of κ^* indicate that the researcher's welfare conclusion is more robust to her functional form assumption. If $\kappa^* \to 0$, her welfare conclusion holds only if the second derivative of $A(q)$ with respect to $B(p)$ is zero *everywhere* between p_0 and p_1: any violation of the zero-curvature assumption overturns the welfare conclusion. By contrast, in the limit as $\kappa^* \to \infty$, then her welfare conclusion holds for any demand curve that passes through the observed points and does not rely on the zero-curvature assumption at all. We defer further discussion of how one can interpret the magnitude of κ^* to Section 4.2.

4 Theoretical Analysis

In this section, we characterize and interpret our robustness measures, r^* and κ^*. Recall from Section 3 that r^* and κ^* are defined by

$$\begin{cases} r^* = \inf\left\{r \in [0,1] : \overline{\Delta\text{CS}}_1(r) > G\right\} \quad \text{and} \quad \kappa^* = \inf\left\{\kappa \in \mathbb{R}_+ : \overline{\Delta\text{CS}}_2(\kappa) > G\right\} & \text{if } \Delta\text{CS} < G, \\ r^* = \inf\left\{r \in [0,1] : \underline{\Delta\text{CS}}_1(r) < G\right\} \quad \text{and} \quad \kappa^* = \inf\left\{\kappa \in \mathbb{R}_+ : \underline{\Delta\text{CS}}_2(\kappa) < G\right\} & \text{if } \Delta\text{CS} > G. \end{cases}$$

Consequently, to compute r^* and κ^*, it suffices to characterize the extremal losses in consumer surplus—namely, $\overline{\Delta\text{CS}}_1$, $\underline{\Delta\text{CS}}_1$, $\overline{\Delta\text{CS}}_2$, and $\underline{\Delta\text{CS}}_2$—as functions of r and κ respectively.

4.1 Characterization of Extremal Losses in Consumer Surplus

We begin by stating and explaining our characterizations of $\overline{\Delta\text{CS}}_1$, $\underline{\Delta\text{CS}}_1$, $\overline{\Delta\text{CS}}_2$, and $\underline{\Delta\text{CS}}_2$.

4.1.1 Extremal Losses in Consumer Surplus for Variability in Gradient

Our first main result characterizes $\overline{\Delta\text{CS}}_1$ and $\underline{\Delta\text{CS}}_1$ as functions of r. To this end, we first derive the largest and smallest losses in consumer surplus under Assumption 1 for general $\underline{\beta}$ and $\overline{\beta}$, and then specialize to the one-dimensional parametrization $\underline{\beta}(r) = \frac{\hat{\beta}}{1-r}$ and $\overline{\beta}(r) = \hat{\beta}\,(1-r)$.

Theorem 1. *Define the auxiliary functions* $\beta^*, \beta_* : [B(p_0), B(p_1)] \to \mathbb{R}$ *as follows:*

$$\beta^*(s) \coloneqq \begin{cases} \underline{\beta} & \text{if } s > \frac{\overline{\beta}B(p_0) - \underline{\beta}B(p_1) - A(q_0) + A(q_1)}{\overline{\beta} - \underline{\beta}}, \\ \overline{\beta} & \text{if } s \leq \frac{\overline{\beta}B(p_0) - \underline{\beta}B(p_1) - A(q_0) + A(q_1)}{\overline{\beta} - \underline{\beta}}; \end{cases}$$

$$\beta_*(s) \coloneqq \begin{cases} \overline{\beta} & \text{if } s > \frac{\overline{\beta}B(p_1) - \underline{\beta}B(p_0) + A(q_0) - A(q_1)}{\overline{\beta} - \underline{\beta}}, \\ \underline{\beta} & \text{if } s \leq \frac{\overline{\beta}B(p_1) - \underline{\beta}B(p_0) + A(q_0) - A(q_1)}{\overline{\beta} - \underline{\beta}}. \end{cases}$$

Under Assumption 1, the largest and smallest possible losses in consumer surplus between p_0 and p_1, $\overline{\Delta\text{CS}}$ and $\underline{\Delta\text{CS}}$, are respectively:

$$\begin{cases} \overline{\Delta\text{CS}} = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta^*(s)\ \mathrm{d}s\right)\ \mathrm{d}p, \\ \underline{\Delta\text{CS}} = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta_*(s)\ \mathrm{d}s\right)\ \mathrm{d}p. \end{cases}$$

Theorem 1 derives the largest and smallest possible losses in consumer surplus by characterizing the extremal demand curves implied by Assumption 1. These extremal demand curves are given by

$$\begin{cases} D^*(p) \coloneqq A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta^*(s)\ \mathrm{d}s\right), \\ D_*(p) \coloneqq A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta_*(s)\ \mathrm{d}s\right). \end{cases}$$

Notice that D^* and D_* satisfy Assumption 1: since β^* and β_* are non-positive, D^* and D_* are downward-sloping; and it is straightforward to verify that $D^*(p_t) = D_*(p_t) = q_t$ for $t \in \{0, 1\}$. As we show in Section 4.4 below, the largest and smallest possible losses in consumer surplus under Assumption 1 are attained at D^* and D_* respectively.

Theorem 1 can be used to derive $\overline{\Delta\text{CS}}_1$ and $\underline{\Delta\text{CS}}_1$ as functions of r. Indeed, the one-dimensional parametrization, $\underline{\beta}(r) = \frac{\hat{\beta}}{1-r}$ and $\overline{\beta}(r) = \hat{\beta}\,(1-r)$, implies that the auxiliary functions β^* and β_*

in Theorem 1 can be written as

$$\beta^*(s;r) = \begin{cases} \frac{\hat{\beta}}{1-r} & \text{if } s > B(p_0) + \frac{B(p_1)-B(p_0)}{2-r}, \\ \hat{\beta}\,(1-r) & \text{if } s \le B(p_0) + \frac{B(p_1)-B(p_0)}{2-r}; \end{cases}$$

$$\beta_*(s;r) = \begin{cases} \hat{\beta}\,(1-r) & \text{if } s > B(p_1) - \frac{B(p_1)-B(p_0)}{2-r}, \\ \frac{\hat{\beta}}{1-r} & \text{if } s \le B(p_1) - \frac{B(p_1)-B(p_0)}{2-r}. \end{cases}$$

In turn, Theorem 1 implies that $\overline{\Delta\mathrm{CS}}_1$ and $\underline{\Delta\mathrm{CS}}_1$ can be expressed as

$$\begin{cases} \overline{\Delta\mathrm{CS}}_1(r) = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta^*(s;r)\ \mathrm{d}s\right)\ \mathrm{d}p, \\ \underline{\Delta\mathrm{CS}}_1(r) = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta_*(s;r)\ \mathrm{d}s\right)\ \mathrm{d}p. \end{cases} \tag{4}$$

While $\overline{\Delta\mathrm{CS}}_1$ and $\underline{\Delta\mathrm{CS}}_1$ can be computed straightforwardly (e.g., by numerical integration) in general for empirical applications, many common functional form assumptions allow $\overline{\Delta\mathrm{CS}}_1$ and $\underline{\Delta\mathrm{CS}}_1$ to be expressed in closed form. Table 1 presents a list of examples.

4.1.2 Extremal Losses in Consumer Surplus for Variability in Curvature

Our second main result characterizes $\overline{\Delta\mathrm{CS}}_2$ and $\underline{\Delta\mathrm{CS}}_2$ as functions of κ. Again, we first derive the largest and smallest losses in consumer surplus under Assumption 2 for general $\underline{\gamma}$ and $\overline{\gamma}$, and then specialize to the one-dimensional parametrization $\underline{\gamma}(\kappa) = \frac{2\hat{\beta}\kappa}{B(p_1)-B(p_0)}$ and $\overline{\gamma}(\kappa) = -\frac{2\hat{\beta}\kappa}{B(p_1)-B(p_0)}$.

Functional form	Upper bound, $\overline{\Delta \text{CS}}_1(r)$	Lower bound, $\underline{\Delta \text{CS}}_1(r)$
linear $A(q) = q,\ B(p) = p$ $\Delta\text{CS} = \frac{(p_1 - p_0)(q_0 + q_1)}{2}$	$\frac{(p_1 - p_0)\left[q_0 + q_1(1-r)\right]}{2-r}$	$\frac{(p_1 - p_0)\left[q_0(1-r) + q_1\right]}{2-r}$
isoelastic $A(q) = \log q,\ B(p) = \log p$ $\Delta\text{CS} = \frac{(p_1 q_1 - p_0 q_0)\log(p_1/p_0)}{\log(q_1/q_0) + \log(p_1/p_0)}$	$\frac{p_0 q_0}{\frac{\log(q_1/q_0)}{\log(p_1/p_0)}(1-r)+1}\left[\left(\frac{q_1}{q_0}\right)^{\frac{1-r}{2-r}}\left(\frac{p_1}{p_0}\right)^{\frac{1}{2-r}} - 1\right] + \frac{(1-r)\,p_1 q_1}{\frac{\log(q_1/q_0)}{\log(p_1/p_0)} + 1 - r}\left[1 - \left(\frac{q_0}{q_1}\right)^{\frac{1}{2-r}}\left(\frac{p_0}{p_1}\right)^{\frac{1-r}{2-r}}\right]$	$\frac{p_1 q_1}{\frac{\log(q_1/q_0)}{\log(p_1/p_0)}(1-r)+1}\left[1 - \left(\frac{q_0}{q_1}\right)^{\frac{1-r}{2-r}}\left(\frac{p_0}{p_1}\right)^{\frac{1}{2-r}}\right] + \frac{(1-r)\,p_0 q_0}{\frac{\log(q_1/q_0)}{\log(p_1/p_0)} + 1 - r}\left[\left(\frac{q_1}{q_0}\right)^{\frac{1}{2-r}}\left(\frac{p_1}{p_0}\right)^{\frac{1-r}{2-r}} - 1\right]$
exponential $A(q) = \log q,\ B(p) = p$ $\Delta\text{CS} = \frac{(p_1 - p_0)(q_0 - q_1)}{\log(q_0/q_1)}$	$\frac{(p_1 - p_0)\,q_1(1-r)}{\log(q_1/q_0)}\left[1 - \left(\frac{q_0}{q_1}\right)^{\frac{1}{2-r}}\right] + \frac{(p_1 - p_0)\,q_0}{\log(q_1/q_0)(1-r)}\left[\left(\frac{q_1}{q_0}\right)^{\frac{1-r}{2-r}} - 1\right]$	$\frac{(p_1 - p_0)\,q_0(1-r)}{\log(q_1/q_0)}\left[\left(\frac{q_1}{q_0}\right)^{\frac{1}{2-r}} - 1\right] + \frac{(p_1 - p_0)\,q_1}{\log(q_1/q_0)(1-r)}\left[1 - \left(\frac{q_0}{q_1}\right)^{\frac{1-r}{2-r}}\right]$
logarithmic $A(q) = q,\ B(p) = \log p$ $\Delta\text{CS} = p_1 q_1 - p_0 q_0 + \frac{(p_1 - p_0)(q_0 - q_1)}{\log(p_1/p_0)}$	$p_1 q_1 + \frac{p_0(q_0 - q_1)}{\log(p_1/p_0)}\left[\left(\frac{p_1}{p_0}\right)^{\frac{1}{2-r}} - 1\right] + \frac{p_1(q_0 - q_1)}{\log(p_1/p_0)(1-r)}\left[1 - \left(\frac{p_0}{p_1}\right)^{\frac{1-r}{2-r}}\right] - p_0 q_0$	$p_1 q_1 + \frac{p_1(q_0 - q_1)(1-r)}{\log(p_1/p_0)}\left[1 - \left(\frac{p_0}{p_1}\right)^{\frac{1}{2-r}}\right] + \frac{p_0(q_0 - q_1)}{\log(p_1/p_0)(1-r)}\left[\left(\frac{p_1}{p_0}\right)^{\frac{1-r}{2-r}} - 1\right] - p_0 q_0$
monomial $A(q) = q^n,\ B(p) = p,\ n > 0$ $\Delta\text{CS} = \frac{n}{n+1}\frac{(p_1 - p_0)\left(q_0^{n+1} - q_1^{n+1}\right)}{q_0^n - q_1^n}$	$\frac{n(p_1 - p_0)(1-r)\left[\left(\frac{q_0^n + q_1^n(1-r)}{2-r}\right)^{1+1/n} - q_1^{1+n}\right]}{(1+n)(q_0^n - q_1^n)} + \frac{n(p_1 - p_0)\left[q_0^{1+n} - \left(\frac{q_0^n + q_1^n(1-r)}{2-r}\right)^{1+1/n}\right]}{(1+n)(q_0^n - q_1^n)(1-r)}$	$\frac{n(p_1 - p_0)(1-r)\left[q_0^{1+n} - \left(\frac{q_0^n(1-r) + q_1^n}{2-r}\right)^{1+1/n}\right]}{(1+n)(q_0^n - q_1^n)} + \frac{n(p_1 - p_0)\left[\left(\frac{q_0^n(1-r) + q_1^n}{2-r}\right)^{1+1/n} - q_1^{1+n}\right]}{(1+n)(q_0^n - q_1^n)(1-r)}$

Table 1: $\overline{\Delta \text{CS}}_1(r)$ and $\underline{\Delta \text{CS}}_1(r)$ for common functional forms.

Theorem 2. *Define the auxiliary functions* $h^*, h_* : [B(p_0), B(p_1)] \to \mathbb{R}$ *as follows:*

$$h^*(s) = \begin{cases} -\frac{A(q_0)-A(q_1)}{B(p_1)-B(p_0)} - \frac{\underline{\gamma}}{2}\left[B(p_0)+B(p_1)\right] & \text{if } \underline{\gamma} \geq -\frac{2[A(q_0)-A(q_1)]}{[B(p_1)-B(p_0)]^2}, \\ \begin{cases} -\underline{\gamma}\left[B(p_1) - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}\right] & \text{if } s > B(p_1) - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}, \\ -\underline{\gamma} s & \text{if } s \leq B(p_1) - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}, \end{cases} & \text{if } \underline{\gamma} < -\frac{2[A(q_0)-A(q_1)]}{[B(p_1)-B(p_0)]^2}; \end{cases}$$

$$h_*(s) = \begin{cases} \begin{cases} -\overline{\gamma} s & \text{if } s > B(p_0) + \sqrt{\frac{2[A(q_0)-A(q_1)]}{\overline{\gamma}}}, \\ -\overline{\gamma}\left[B(p_0) + \sqrt{\frac{2[A(q_0)-A(q_1)]}{\overline{\gamma}}}\right] & \text{if } s \leq B(p_0) + \sqrt{\frac{2[A(q_0)-A(q_1)]}{\overline{\gamma}}}, \end{cases} & \text{if } \overline{\gamma} \geq \frac{2[A(q_0)-A(q_1)]}{[B(p_1)-B(p_0)]^2}, \\ -\frac{A(q_0)-A(q_1)}{B(p_1)-B(p_0)} - \frac{\overline{\gamma}}{2}\left[B(p_0)+B(p_1)\right] & \text{if } \overline{\gamma} < \frac{2[A(q_0)-A(q_1)]}{[B(p_1)-B(p_0)]^2}. \end{cases}$$

Under Assumption 2, the largest and smallest possible losses in consumer surplus between p_0 *and* p_1*,* $\overline{\Delta \text{CS}}$ *and* $\underline{\Delta \text{CS}}$*, are respectively:*

$$\begin{cases} \overline{\Delta \text{CS}} = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \left[h^*(s) + \underline{\gamma} s\right] \, \mathrm{d}s\right) \, \mathrm{d}p, \\ \underline{\Delta \text{CS}} = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \left[h_*(s) + \overline{\gamma} s\right] \, \mathrm{d}s\right) \, \mathrm{d}p. \end{cases}$$

Like Theorem 1, Theorem 2 derives the largest and smallest possible losses in consumer surplus by characterizing the extremal demand curves implied by Assumption 2. These extremal curves are now given by

$$\begin{cases} E^*(p) := A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \left[h^*(s) + \underline{\gamma} s\right] \, \mathrm{d}s\right), \\ E_*(p) := A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \left[h_*(s) + \overline{\gamma} s\right] \, \mathrm{d}s\right). \end{cases}$$

Although the computation is slightly more involved than in Theorem 1, it can similarly be verified that E^* and E_* satisfy Assumption 2. As we show in Appendix A.1, the largest and smallest possible losses in consumer surplus under Assumption 2 are attained at E^* and E_* respectively.

Unlike Theorem 1, Theorem 2 shows that the one-dimensional parametrization that we use is, in fact, without loss of generality. Although Assumption 2 allows $\underline{\gamma}$ and $\overline{\gamma}$ to co-vary in an arbitrary way, Theorem 2 shows that the largest possible loss in consumer surplus depends only on $\underline{\gamma}$ and the smallest possible loss in consumer surplus depends only on $\overline{\gamma}$. As such, based on whether $\Delta\text{CS} < G$ or $\Delta\text{CS} > G$, the robustness measure κ^* depends only on $\underline{\gamma}$ or $\overline{\gamma}$, but not both.

Finally, Theorem 2 can also be used to derive $\overline{\Delta \mathrm{CS}}_2$ and $\underline{\Delta \mathrm{CS}}_2$ as functions of κ. Indeed, our one-dimensional parametrization implies that the auxiliary functions h^* and h_* in Theorem 2 can be written as

$$h^*(s;\kappa) = \begin{cases} \hat{\beta}\left[1-\kappa\cdot\frac{B(p_0)+B(p_1)}{B(p_1)-B(p_0)}\right] & \text{if } \kappa \leq 1, \\ \begin{cases} 2\hat{\beta}\sqrt{\kappa}\left[1-\sqrt{\kappa}\cdot\frac{B(p_1)}{B(p_1)-B(p_0)}\right] & \text{if } s > B(p_1)-\frac{B(p_1)-B(p_0)}{\sqrt{\kappa}}, \\ -2\hat{\beta}\kappa\cdot\frac{s}{B(p_1)-B(p_0)} & \text{if } s \leq B(p_1)-\frac{B(p_1)-B(p_0)}{\sqrt{\kappa}}, \end{cases} & \text{if } \kappa > 1; \end{cases}$$

$$h_*(s;\kappa) = \begin{cases} \begin{cases} 2\hat{\beta}\kappa\cdot\frac{s}{B(p_1)-B(p_0)} & \text{if } s > B(p_0)+\frac{B(p_1)-B(p_0)}{\sqrt{\kappa}}, \\ 2\hat{\beta}\sqrt{\kappa}\left[1+\sqrt{\kappa}\cdot\frac{B(p_0)}{B(p_1)-B(p_0)}\right] & \text{if } s \leq B(p_0)+\frac{B(p_1)-B(p_0)}{\sqrt{\kappa}}, \end{cases} & \text{if } \kappa > 1, \\ \hat{\beta}\left[1+\kappa\cdot\frac{B(p_0)+B(p_1)}{B(p_1)-B(p_0)}\right] & \text{if } \kappa \leq 1. \end{cases}$$

In turn, Theorem 2 implies that $\overline{\Delta \mathrm{CS}}_2$ and $\underline{\Delta \mathrm{CS}}_2$ can be expressed as

$$\begin{cases} \overline{\Delta \mathrm{CS}}_2(\kappa) = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)}\left[h^*(s;\kappa) + \frac{2\hat{\beta}\kappa s}{B(p_1)-B(p_0)}\right]\,\mathrm{d}s\right)\,\mathrm{d}p, \\ \underline{\Delta \mathrm{CS}}_2(\kappa) = \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)}\left[h_*(s;\kappa) - \frac{2\hat{\beta}\kappa s}{B(p_1)-B(p_0)}\right]\,\mathrm{d}s\right)\,\mathrm{d}p. \end{cases} \tag{5}$$

These expressions allow $\overline{\Delta \mathrm{CS}}_2$ and $\underline{\Delta \mathrm{CS}}_2$ to be computed straightforwardly, such as by numerical integration. Closed-form expressions are generally more difficult to obtain for $\overline{\Delta \mathrm{CS}}_2$ and $\underline{\Delta \mathrm{CS}}_2$ because of the different cases in $h^*(s;\kappa)$ and $h_*(s;\kappa)$. As we discuss below in Section 4.4, these cases arise because the monotonicity constraint—that the demand curve must be downward-sloping—might potentially bind under Assumption 2, whereas it never binds under Assumption 1.

4.2 Interpretation of Robustness Measures

Whereas Theorems 1 and 2 can be used to compute our robustness measures, we now explain how they can be interpreted. For simplicity, we focus on r^* and provide three possible interpretations.

Comparison with institutional knowledge. First, our robustness measures can be directly interpreted through comparison with institutional knowledge and surveys of related studies. For example, consider a researcher studying the welfare impact of a sugar tax who measures an elasticity of $\hat{\varepsilon}$ for sugary foods and beverages (e.g., by running a log-log regression in the spirit of our motivating example in Section 2). A robustness measure of r^* would therefore imply that, in order to overturn her welfare conclusion, the elasticities along the demand curve would have to exceed the interval $[\hat{\varepsilon}/(1-r^*), \hat{\varepsilon}(1-r^*)]$. To determine if this is likely, the researcher could compare this range to known ranges of elasticities in the literature for comparable products. For instance, Andreyeva, Long and Brownell (2010) summarize elasticities for food and beverage in the U.S. from 160 empirical studies and determine that all lie between -3.18 and -0.01; they also provide narrower ranges for each distinct food category.

Implications on counterfactual demand. Next, our robustness measures can be interpreted through their implications on counterfactual demand. We do so with the following proposition:

Proposition 1. *Let the extremal demand curves D^*, D_*, E^*, and E_* be as defined in the discussions following Theorems 1 and 2.*

(a) Under Assumption 1, the demand at any price $p \in [p_0, p_1]$ satisfies $D(p) \in [D_*(p), D^*(p)]$.

(b) Under Assumption 2, the demand at any price $p \in [p_0, p_1]$ satisfies $D(p) \in [E_*(p), E^*(p)]$.

Proposition 1 shows that the demand curves D^*, D_*, E^*, and E_* are extremal not just for the loss in consumer surplus (as shown in Theorems 1 and 2), but also for quantities demanded at any intermediate price. As we show in Appendix A.2, this is not a coincidence: the demand curve that maximizes (or minimizes) the loss in consumer surplus must be the pointwise highest demand curve under either Assumption 1 or Assumption 2, and hence it must also maximize (or minimize) the quantity demanded at any intermediate price.

Proposition 1 therefore allows us to interpret our robustness measures by reasoning about the potential range of quantities demanded at a given intermediate price. Continuing the example of a researcher studying the welfare impact of a sugar tax from above, a robustness measure of r^* would imply that the counterfactual demand at any intermediate price can lie only within a restricted range. Proposition 1 implies that this range increases with r^*. Consequently, if the researcher reasons that this implied range of quantities is too restrictive, then r^* is too low and her welfare conclusion is not robust. In particular, if she can conduct a new experiment and sample an additional point on the demand curve at an intermediate price, she can test the null hypothesis (that the quantity demanded at that price lies within this range) implied by r^*.

Comparison with higher-order approximations. Finally, our robustness measures can also be interpreted through higher-order approximations. For example, rather than interpret a linear demand curve as the true demand curve, it can instead be viewed as a first-order Taylor approximation (cf. Kleven, 2021).

Higher-order Taylor approximations provide natural benchmarks for r^*. To illustrate, suppose that the true demand curve $D(p)$ has a second-order term:

$$D(p) = q_0 + \beta (p - p_0) + \frac{\delta}{2} (p - p_0)^2 ,$$

where β and δ are constants. Because only two points, (p_0, q_0) and (p_1, q_1), are observed, β and δ are not uniquely pinned down. However, the possible values that β and δ can jointly take are restricted by the requirements that the demand curve passes through (p_1, q_1) and that the demand curve is downward-sloping (note that the demand curve passes through (p_0, q_0) by construction). These restrictions imply that, when the true demand curve $D(p)$ is a second-order polynomial in p, the loss in consumer surplus satisfies

$$\Delta \text{CS}_{\text{second-order}} \in \left[\frac{1}{3} (p_1 - p_0) (q_0 + 2q_1) , \frac{1}{3} (p_1 - p_0) (2q_0 + q_1) \right] .$$

Under our framework, such a demand curve would imply a robustness measure of at most 1/2. In turn, this provides a natural benchmark for r^*. For instance, if $\Delta \text{CS} > G$ and $r^* > 1/2$, then the welfare conclusion cannot be reversed by using only a second-order approximation of demand: higher-order terms are required. Higher-order benchmarks for r^* can be similarly computed.

4.3 Implications for Shape Constraints

Next, we discuss the implications of our robustness measures for shape constraints on demand curves. Shape constraints have a long history of precedence in economics. For instance, Marshall (1890) went so far as to define a demand curve as a decreasing function whose elasticity also decreases with price, while Robinson (1933) suggested that demand curves, ought to be convex lest the monopoly output rises when third-degree price discrimination causes prices to rise. These intuitions underlie the standard textbook depiction of a demand curve as a convex function.

Shape constraints are imposed for a variety of reasons, arguably the most important of which is to ensure that comparative statics predicted by models are consistent with observed data. For example, Marshall's assumption (now more commonly known as Marshall's second law) was maintained "without apology" by Krugman (1979) so that his model would produce "reasonable results." Melitz (2018) argues that Marshall's second law—which he also imposes in his model—is "equivalent to the property that more productive firms (or alternatively lower cost) set higher markups," and that violations would "directly contradict the [empirical] evidence on markups and pass-through."

Different literatures in economics employ a variety of shape constraints on demand curves that capture other intuitions pertaining to their fields of interest. Examples include Marshall's second law, decreasing marginal revenue, ρ-concavity (including concavity and log-concavity), and ρ-convexity (including convexity and log-convexity). We defer formal definitions and discussions of these shape constraints to Appendix C.

Our robustness measures imply that common functional form assumptions sometimes lead to maximally robust welfare conclusions under common shape constraints. In particular, Theorems 1 and 2 imply that:

Corollary 1.

(a) For a welfare conclusion obtained with isoelastic demand, $r^* = 1$ and $\kappa^* = +\infty$ if $\Delta\text{CS} > G$ and the true demand curve satisfies Marshall's second law.

(b) For a welfare conclusion obtained with constant marginal revenue demand, $r^* = 1$ and $\kappa^* = +\infty$ if $\Delta\text{CS} > G$ and the true demand curve exhibits decreasing marginal revenue.

(c) For a welfare conclusion obtained with ρ-linear demand, $r^* = 1$ and $\kappa^* = +\infty$ if $\Delta\text{CS} > G$ and the true demand curve is ρ-concave, or if $\Delta\text{CS} < G$ and the true demand curve is ρ-convex.

Corollary 1 thus provides some justification for the use of common functional form assumptions when shape constraints are imposed. For instance, the isoelastic demand curve can be interpreted as a "conservative" functional form assumption in environments that impose Marshall's second law because it is maximally robust (i.e., $r^* = 1$ and $\kappa^* = +\infty$) if the estimated loss in consumer surplus exceeds the benchmark gain.

4.4 Proofs of Theorem 1

Next, we present two approaches to prove Theorem 1. The first is simple and intuitive due to its geometric nature, but cannot be easily adapted to prove Theorem 2. The second is more technically complex and relies on a fortuitous connection between our problem—of determining extremal welfare impacts subject to feasibility constraints—and Bayesian persuasion problems that have been considered by the theoretical literature stemming from Gentzkow and Kamenica

(2016). While our second approach is less straightforward, it generalizes more easily to other environments, such as Theorem 2. We discuss how this approach generalizes to Theorem 2 below but defer its proof to Appendix A.1.

4.4.1 Geometric Proof

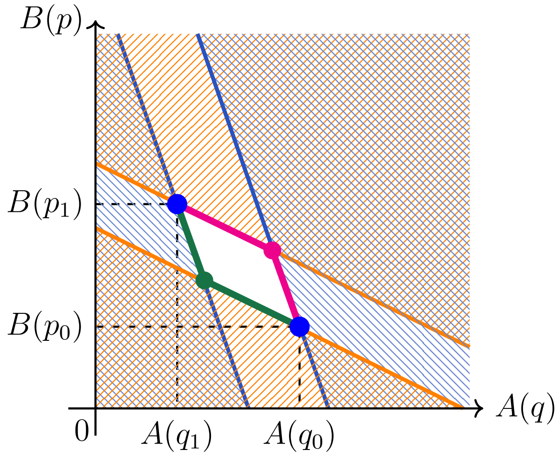


Figure 5: Sketch of the proof of Theorem 1.

We now explain the geometric proof of Theorem 1 with the help of Figure 5. We begin with a change of variables: rather than plot prices p against quantities q, we depict a demand curve by plotting transformed prices $B(p)$ against transformed quantities $A(q)$. The key insight is that, because A and B are monotone transformations, this change of variables does not qualitatively alter our problem: as before, our goal is to find decreasing curves that pass through two points, $(B(p_0), A(q_0))$ and $(B(p_1), A(q_1))$, that respectively maximize and minimize the areas under the curves that are bounded between $B(p_0)$ and $B(p_1)$.

Although it does not qualitatively alter the problem, this change of variables enables a more natural interpretation of Assumption 1. Notice that demand curves with constant gradient of $A(q)$ with respect to $B(p)$ correspond to *linear* curves when we plot $B(p)$ against $A(q)$. To rule out transformed demand curves with gradients higher than $\overline{\beta}$, we draw two (blue) straight lines such that one passes through $(B(p_0), A(q_0))$ and the other, $(B(p_1), A(q_1))$; each of these lines has a gradient of $\overline{\beta}$. Any demand curve satisfying Assumption 1 must therefore correspond to a curve that lies between these two lines—and not in the (blue) shaded regions. Similarly, to rule out transformed demand curves with gradients lower than $\underline{\beta}$, we draw two (orange) straight lines with gradients equal to $\underline{\beta}$. Any demand curve satisfying Assumption 1 must also correspond to a curve that lies between these two lines—and not in the (orange) shaded regions.

The next step is to find the curve within the unshaded parallelogram-shaped region that maximizes the area under it between $B(p_0)$ and $B(p_1)$, since the loss in consumer surplus must vary monotonically with this area. This curve can be read directly off the diagram: it must be the top boundary of the parallelogram (i.e., the red curve). Likewise, the bottom boundary of the parallelogram (i.e., the green curve) minimizes the area under it between $B(p_0)$ and $B(p_1)$.

It follows that the demand curves corresponding to each of these curves must respectively maximize and minimize the loss in consumer surplus. Reversing the change of variables, the top (red) curve and the bottom (green) curve correspond respectively to the demand curves

$$\begin{cases} D^*(p) := A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta^*(s)\ \mathrm{d}s\right), \\ D_*(p) := A^{-1}\left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta_*(s)\ \mathrm{d}s\right), \end{cases}$$

where β^* and β_* are as defined in Theorem 1.

One minor technicality is that, while any demand curve that satisfies Assumption 1 must correspond to a curve in the unshaded parallelogram-shaped region, the converse is not true: not any curve in this region can be mapped back to a demand curve that satisfies Assumption 1. Nevertheless, it is easy to see that D^* and D_* both satisfy Assumption 1.

Although minor, this technicality highlights the somewhat "coincidental" simplicity of this proof. This geometric argument relies on finding the constraints implied by Assumption 1 that are not only necessary, but also sufficient *a posteriori*. In more complicated environments, the binding constraints are not as easily determined, nor do they necessarily take on such a simple form. In turn, this explains why this geometric argument fails to generalize to Theorem 2, necessitating the information design approach that we undertake below.

4.4.2 Information Design Proof

We now provide an alternative proof of Theorem 1 that exploits a fortuitous connection between our problem and Bayesian persuasion problems. We divide the proof into three steps: *(i)* employing a change of variables to map the problem into an appropriate functional space; *(ii)* endowing this space with a partial order and characterizing its extremal functions; and *(iii)* mapping the solution back to the original problem.

Step 1: Changing variables. Similar to our geometric approach above, we begin by employing a change of variables. Instead of choosing a demand curve to maximize or minimize the loss in consumer surplus, we choose the function $\beta : [B(p_0), B(p_1)] \to \mathbb{R}_-$ that represents the gradient of $A(q)$ against $B(p)$. To this end, let $\pi = B(p)$ be defined on $[\pi_0, \pi_1] = [B(p_0), B(p_1)]$, and consider $\widetilde{D} : [\pi_0, \pi_1] \to \mathbb{R}$ be defined by $\widetilde{D}(B(p)) = A(D(p))$. Then we define $\beta(\pi) \coloneqq \widetilde{D}'(\pi)$. Given β, $\widetilde{D}$ is completely determined, and vice versa:

$$\widetilde{D}(\pi) = A(q_0) + \int_{\pi_0}^{\pi} \beta(s)\ \mathrm{d}s \qquad \text{for } \pi \in [\pi_0, \pi_1].$$

This is obtained via integration by parts. Next, we define the set of feasible gradient functions that are consistent with Assumption 1:

$$\mathcal{B} \coloneqq \left\{ \beta : [\pi_0, \pi_1] \to [\underline{\beta}, \overline{\beta}] \text{ s.t. } \int_{\pi_0}^{\pi_1} \beta(s)\ \mathrm{d}s = A(q_1) - A(q_0) \right\}.$$

Thus we arrive at the equivalent problem:

$$\begin{cases} \overline{\Delta \mathrm{CS}} = \sup\limits_{\beta \in \mathcal{B}} \displaystyle\int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(p)} \beta(s)\ \mathrm{d}s \right)\ \mathrm{d}p, \\ \underline{\Delta \mathrm{CS}} = \inf\limits_{\beta \in \mathcal{B}} \displaystyle\int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(p)} \beta(s)\ \mathrm{d}s \right)\ \mathrm{d}p. \end{cases} \tag{6}$$

Step 2: Characterizing the set $\mathcal{B}$. We now endow the set $\mathcal{B}$ with a partial order. Formally, for any two functions $\beta_1, \beta_2 \in \mathcal{B}$, we write

$$\beta_1 \succeq \beta_2 \iff \int_{\pi_0}^{\pi} \beta_1(s)\ \mathrm{d}s \geq \int_{\pi_0}^{\pi} \beta_2(s)\ \mathrm{d}s \qquad \text{for } \pi \in [\pi_0, \pi_1].$$

This partial order is motivated by the definition of second-order stochastic dominance, but with a few differences: β is not necessarily a monotone function, nor is $\beta(\pi_0)$ or $\beta(\pi_1)$ fixed. For these reasons, β cannot be interpreted as a cumulative distribution function (CDF), making the above definition slightly different from second-order stochastic dominance.

Nevertheless, a familiar mathematical property of second-order stochastic dominance holds in this environment. Just as the second-order stochastic dominance order defines a lattice structure on the set of all CDFs with the same mean, the partial order $\succeq$ defines a lattice structure on the set $\mathcal{B}$.

Lemma 1. *Any $\beta \in \mathcal{B}$ satisfies $\beta^* \succeq \beta \succeq \beta_*$, where β^* and β_* are as defined in Theorem 1.*

Proof. To see that $\beta^* \succeq \beta$ for any $\beta \in \mathcal{B}$, observe that

$$\int_{\pi_0}^{\pi} \beta^*(s) \, \mathrm{d}s = \begin{cases} A(q_1) - A(q_0) - (\pi_1 - \pi) \cdot \underline{\beta} & \text{for } \frac{\overline{\beta}\pi_0 - \underline{\beta}\pi_1 - A(q_0) + A(q_1)}{\overline{\beta} - \underline{\beta}} < \pi \leq \pi_1, \\ (\pi - \pi_0) \cdot \overline{\beta} & \text{for } \pi_0 \leq \pi \leq \frac{\overline{\beta}\pi_0 - \underline{\beta}\pi_1 - A(q_0) - A(q_1)}{\overline{\beta} - \underline{\beta}}. \end{cases}$$

Since $\operatorname{im} \beta \subset [\underline{\beta}, \overline{\beta}]$, we conclude that, in either case,

$$\int_{\pi_0}^{\pi} \beta^*(s) \, \mathrm{d}s \geq \int_{\pi_0}^{\pi} \beta(s) \, \mathrm{d}s.$$

A similar argument shows that $\beta \succeq \beta_*$ for any $\beta \in \mathcal{B}$. □

It is easy to check that $\beta^*, \beta_* \in \mathcal{B}$. Therefore, Lemma 1 characterizes the largest and smallest elements of the partially ordered set $(\mathcal{B}, \succeq)$. With more work, one can show that $(\mathcal{B}, \succeq)$ is a lattice (cf. Theorem 3.3 of Müller and Scarsini, 2006); however, as the lattice property is not important for our purposes, we do not pursue that here.

Step 3: Mapping back to the original problem. Having characterized the largest and smallest elements of $(\mathcal{B}, \succeq)$, it remains to map these back to the original problem. To this end, we define the functional $\Delta\mathrm{CS} : \mathcal{B} \to \mathbb{R}$ by

$$\Delta\mathrm{CS}(\beta) := \int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(p)} \beta(s) \, \mathrm{d}s \right) \, \mathrm{d}p.$$

Our problem (6) is equivalent to maximizing and minimizing this functional over the family $\mathcal{B}$. The following lemma shows that this can be done with the aid of the partial order $\succeq$ defined in our previous step:

Lemma 2. *The functional $\Delta\mathrm{CS}(\cdot)$ is increasing in the partial order $\succeq$; that is, for any $\beta_1 \succeq \beta_2$,*

$$\int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(p)} \beta_1(s) \, \mathrm{d}s \right) \, \mathrm{d}p \geq \int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(p)} \beta_2(s) \, \mathrm{d}s \right) \, \mathrm{d}p.$$

Proof. The result follows straightforwardly from the definition of the partial order $\succeq$, the fact that A (and hence A^{-1}) is increasing, and a pointwise comparison of the two integrands. □

Together, Lemmas 1 and 2 imply that the functional $\Delta\text{CS}(\cdot)$ is maximized at β^* and minimized at β_*; that is, $\overline{\Delta\text{CS}} = \Delta\text{CS}(\beta^*)$ and $\underline{\Delta\text{CS}} = \Delta\text{CS}(\beta_*)$. This completes the proof of Theorem 1.

4.5 Discussion

We conclude this section by: *(i)* clarifying the conceptual contributions in our paper; *(ii)* discussing how the information design approach compares with the geometric approach; and *(iii)* describing various extensions of our robustness measures.

Clarification of contributions. As foreshadowed in Section 2, our paper makes two different, but closely related, conceptual contributions. First, we derive bounds on changes in consumer surplus for demands in general constraint sets (namely, under Assumptions 1 and 2). These bounds are stated in Theorems 1 and 2 respectively. Second, to provide measures of robustness, we focus on bounds obtained under specific one-dimensional parametrizations of these constraints. These bounds are given by equations (4) and (5) respectively and allow us to graphically represent the sensitivity of welfare conclusions with respect to the assumed functional form.

Comparison of approaches. While our information design approach is more complicated than our geometric approach, it has the advantage of being easily generalizable—in particular, to Theorem 2. Unlike in Theorem 1, where the monotonicity constraint (i.e., the demand curve must be downward-sloping) never binds, the monotonicity constraint could bind in Theorem 2 if the bounds on curvature are sufficiently permissive. Whereas the geometric approach is unable to easily determine when the monotonicity constraint binds, the monotonicity constraint can be easily accommodated in the functional space analog of $\mathcal{B}$. Notice also that Lemma 2 does not depend on how $\mathcal{B}$ is defined. Therefore, in the information design approach, the additional monotonicity constraint only requires determining the analog of Lemma 1 for the constrained problem—that is, finding the largest and smallest elements of the partially ordered set $(\mathcal{B}, \succeq)$.

It is worth pointing out that the structure of $\mathcal{B}$ is reminiscent of Bayesian persuasion problems stemming from the work of Gentzkow and Kamenica (2016). If $-\beta$ could be interpreted as a distribution of posterior means, then the mean constraint

$$\int_{B(p_0)}^{B(p_1)} \beta(s)\ \mathrm{d}s = A(q_1) - A(q_0)$$

could be interpreted as a Bayes plausibility constraint, where $A(q_0) - A(q_1)$ is the mean of the prior belief. This analogy breaks down for the sole reason that $-\beta$ *cannot* be interpreted as a distribution of posterior means: $-\beta$ is not monotone and hence cannot be interpreted as a CDF.

However, as we show in Appendix A.1, there is a closer equivalence between such Bayesian persuasion problems and the problem that we solve in Theorem 2. For example, to focus on the bounds implied by $\widetilde{D}''(\pi) \geq \underline{\gamma}$ in our proof of Theorem 2, we construct the analog of β by defining $h(\pi) := \widetilde{D}'(\pi) - \underline{\gamma}\pi$. Then h must be non-decreasing since $\widetilde{D}''(\pi) \geq \underline{\gamma}$ and, with appropriate rescaling, *can* be interpreted as a CDF representing the distribution of posterior means.

The fortuitous connection between our problem of bounding welfare and Bayesian persuasion problems implies that tools developed for *constrained* information design problems can potentially also be used to evaluate robust welfare bounds. From a technical point of view, our approach (in the alternative proof presented above) is based on the proof strategy of Kang and Vondrák (2019), who solve an infinite-dimensional optimization problem by showing that the objective functional is monotone with respect to the convex partial order. For the convex partial order in particular, Kleiner, Moldovanu and Strack (2021) recently develop an approach based on a characterization of extreme points, which yields a general solution to similar problems—even when the objective function is not monotone with respect to the convex partial order. While their method applies to problems in information and mechanism design, our discussion here suggests potential applications also to problems of robustness with respect to distributions and functional form assumptions.

Extensions of robustness measures. Our robustness measures can be extended when more complexity is allowed for. In Appendix B, we consider four extensions that are motivated by our applications in Section 5: *(i)* counterfactual exercises; *(ii)* more observations; *(iii)* measurement error; and *(iv)* other welfare measures. In each of these extensions, we show that analogs of Theorems 1 and 2 continue to hold, thereby allowing us to compute robustness measures.

5 Empirical Applications

Our approach to constructing robustness measures can be applied to a number of settings. In this section, we present three applications drawn from different fields.

5.1 Trade Tariffs

Between 2018 and 2019, the United States imposed an unprecedented wave of escalating import tariffs on a large set of product sectors and major trading partners. This "return to protectionism" inspired numerous academic studies to assess the welfare impact of the new tariffs (Amiti, Redding and Weinstein, 2019; Fajgelbaum, Goldberg, Kennedy and Khandelwal, 2020; Cavallo, Gopinath, Neiman and Tang, 2021). All of these studies document the same patterns: *(i)* quantities consumed fell in sectors targeted by the tariffs; *(ii)* foreign producer prices did not change significantly in the short run; and *(iii)* the net domestic impact of the tariffs was ultimately negative.[8] But the modeling choices and empirical techniques employed in each study are slightly different—for instance, Amiti et al. assume linear demand curves, while Fajgelbaum et al. and Cavallo et al. assume isoelastic demand curves. As a result, while all of the estimates are similar, it is difficult to discern the extent to which their differences stem from substantive modeling choices (e.g., accounting for substitution across product sectors) rather than different parametrizations—and the extent to which other specifications would have led to still larger differences in results.

In this subsection, we examine the robustness of conclusions regarding the welfare impact of the 2018–2019 import tariffs based on the approach taken by Amiti et al. (2019). In their paper, Amiti et al. consider the deadweight loss incurred due to the tariffs across the range of affected products over the course of 2018, and compare it to several benchmarks of gains that might be attributed to the tariff policy. To do this, they consider the price and quantity imported for each affected product in each calendar month in 2017—before the tariffs were implemented—and in 2018. They then perform three exercises: First, they use a regression approach to argue that foreign exporter prices did not change in response to the tariffs,[9] which means that the change in producer surplus can be ignored in the deadweight loss computation. Second, they calculate an average treatment effect of tariffs on quantities through a log-log regression with product and country-time fixed effects. Third, they impute a linear demand curve for all products and calendar months using a transformation of the regression results and compute the deadweight loss as the area of the resulting triangle, depicted by region B of Figure 2(a).

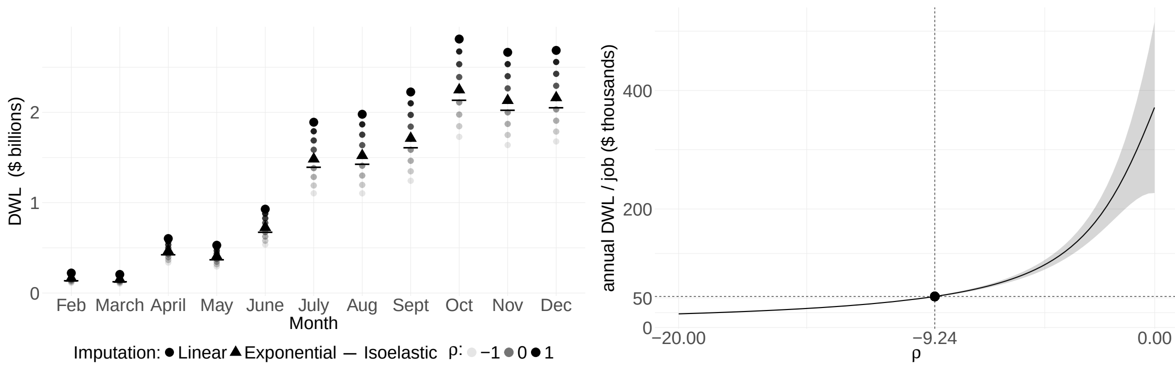
What if a different functional form had been chosen instead of linear demand? Figure 6(a) plots the deadweight loss estimate that would have been obtained in each month of 2018 under Amiti et al.'s framework, had they employed other possible imputations, such as ρ-linear curves (cf. Appendix C)—including linear ($\rho = 0$) and exponential ($\rho = 1$) curves as special cases—and an isoelastic curve.[10]

---

8 By 2020, the Wall Street Journal editorial board had written about the "piling" evidence of net economic harm from tariffs in an article titled "How Many Tariff Studies Are Enough?"

9 Specifically, they find that a log-log regression of the change in exporter prices on tariffs yields a precise zero.

10 In each case, we follow Amiti et al. by considering each product-month pair as an independent market (in 2017 and 2018) and aggregating the deadweight loss estimates across all of the products that were affected by tariffs in each month.



(a) DWL for common demand curves. *(b) DWL per manufacturing job under ρ-linearity.*

Figure 6: Comparison of DWL bounds for different imputations.

As Figure 6(a) shows, there is substantial variation in the magnitude of deadweight loss over time, and across imputations. In the early months of the tariffs—when relatively few products were taxed and tariffs were relatively low—the monthly deadweight loss estimates under all imputations are between \$0.25 and \$0.5 billion. However, by July—after a wave of 25% tariffs on \$34 billion worth of imports from China was added—the deadweight loss estimates are much higher and more dispersed. For instance, while the deadweight loss would be estimated at about \$1.5 billion under an exponential or isoelastic imputation, it would be estimated at nearly \$2 billion under a linear imputation. This difference grows larger in the later months of 2018 when two more waves of tariffs targeted at China were added.

This exercise demonstrates that the linear demand assumption of Amiti et al. is not without loss of generality: while linear demand is a good first-order approximation when price changes are small, it would generate a substantial bias if the true demand curve was of one of the other imputations considered. Note, however, that while the magnitudes of the differences between imputations vary from month to month—reflecting baseline differences in the quantities demanded at different times of the year along with changes to prices and quantities due to the gradual addition of tariffs—their relative ordering does not change. This actually follows from our Corollary 1. These imputations are the most conservative for different shape constraints, but as we show in Appendix C, the shape constraints often nest one another. As such, the relative ordering of the deadweight loss estimates that they imply is guaranteed to be preserved no matter how big the change in prices.

Next, we evaluate the robustness of Amiti et al.'s welfare conclusion with respect to their linear demand assumption. In their paper, Amiti et al. offer several benchmarks to assess the deadweight loss from the trade war in the context of potential benefits. For instance, if the trade war resulted in recouping the 35,400 manufacturing jobs that were lost in the steel and aluminum industry over the 2010s, then under Amiti et al.'s estimate, this would result in \$1.86 billion in annual wages (assuming an average annual wage of \$52,500 per job). Comparing this figure to the annual deadweight loss estimated under a linear imputation, one would conclude that the deadweight loss estimated at \$473,125 greatly exceeds the value of wages from job creation.

A simple way to quantify the robustness of Amiti et al.'s welfare conclusion is to extend the linear demand assumption to a slightly more general—but nonetheless parametric—functional form family. To illustrate, Figure 6(b) plots the lower bound on deadweight loss per manufacturing job added under a ρ-linear functional form assumption, where ρ is allowed to vary. The point at $\rho = 1$ corresponds to Amiti et al.'s estimate, while the point at $\rho = 0$ corresponds to the estimate that would have been obtained with an exponential imputation—that is, the sum of the triangles in Figure 6(a) divided by 35,400. The shaded area around the curve corresponds to the 95% confidence interval on the bound at each ρ with respect to the treatment effect estimate in Amiti et al.'s regression of log-quantities on log-tariffs. As Figure 6(b) shows, ρ would have to be at most -9.24 in order for a ρ-linear demand curve to rationalize a per-job deadweight loss as low as \$52,500. This suggests that Amiti et al.'s welfare conclusion is very robust, at least when one restricts attention to only ρ-linear imputations.

Using our robustness measures, we can quantify the robustness of Amiti et al.'s welfare conclusion more generally, without restricting attention to a particular family of functional forms. Because Amiti et al. estimate a log-log regression of prices on quantities, we use an isoelastic demand curve as our benchmark, and consider relaxations around the estimated elasticity parameter.[11] Figure 7 plots the lower bound on deadweight loss per manufacturing job added against gradient relaxations when $A(q) = \log(q)$ and $B(p) = \log(p)$ based on Theorem 1. As the figure shows, relative to an isoelastic demand benchmark, the smallest r required for the deadweight loss to break even with the manufacturing wage gains is $r^* = 0.89$. That is, we would need to allow for elasticities 89% smaller in magnitude than Amiti et al.'s estimate. As this is quite extreme, we conclude that Amiti et al.'s negative assessment of the tariffs policy is indeed robust to a large set of demand specifications given the observed demand response.

[11] We could have instead used a linear benchmark fitted using the projections of tariff treatment effects based on the log-log regression. However, as our measures of robustness account for a much broader class of possible demand curves in any case, we prefer to use the benchmark that is most consistent with the estimated parameters.

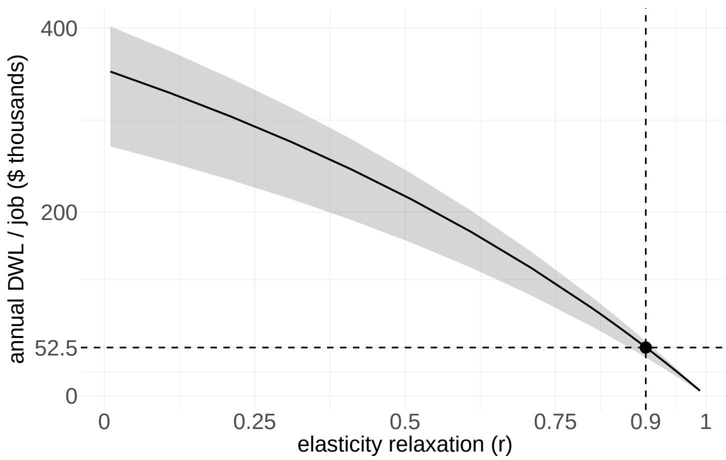


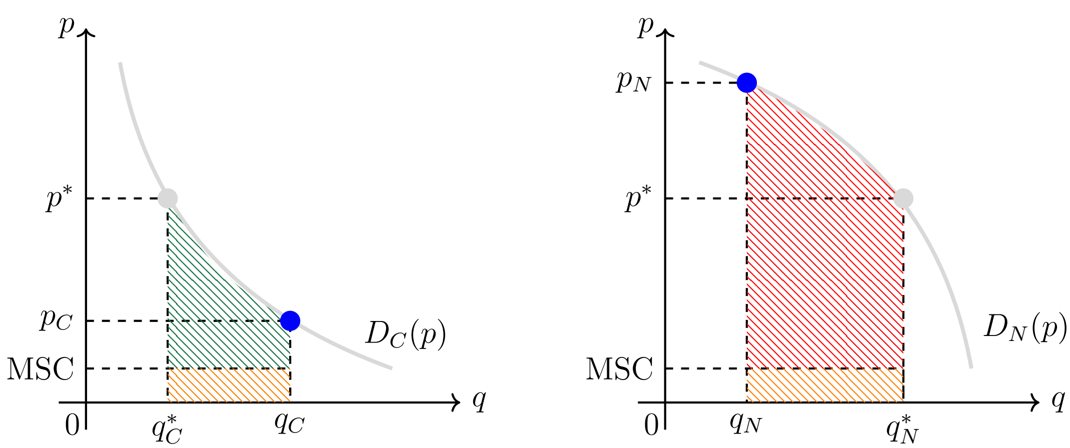
Figure 7: DWL per manufacturing job: lower bound by elasticity relaxation

5.2 Energy Subsidies

Some policies benefit one group while harming another. To assess the net welfare impacts of such policies, researchers must therefore estimate the impacts on several different market segments and compare them. In this section, we consider an application evaluating the energy subsidies offered through the California Alternate Rates for Energy (CARE) program, following the study by Hahn and Metcalfe (2021). The CARE program offers wholesale discounts on unit prices for gas and electricity to eligible low-income households. In Hahn and Metcalfe's sample, CARE households receive a 20% discount on marginal rates, from an average price of \$0.95 to \$0.75, per therm of gas. However, the CARE program comes at a cost through several channels. First, discounts for eligible households are subsidized by higher-income households, who shoulder a higher cost to compensate for the difference in revenues. Second, lower gas prices encourage higher gas consumption, which harms the environment. Finally, the CARE program incurs \$7 million in administrative cost. In their paper, Hahn and Metcalfe estimate that on net, the CARE program results in a welfare loss of \$4.8 million.

To reach their welfare conclusion, Hahn and Metcalfe estimate the difference in welfare between the status quo and a hypothetical in which all consumers face a uniform price determined by a fixed revenue formula. For CARE households, they estimate a local elasticity of consumption at the subsidized price using a LATE research design with randomized nudges for eligible households to sign up and receive the discounted rate. For non-CARE households, they adopt an estimate of the local elasticity of consumption from Auffhammer and Rubin (2018). In each case, they impute counterfactual demand from the relevant local elasticity estimate by assuming that the demand curve is linear. Under this functional form assumption, the observed price-quantity pair and local elasticity estimate pin down the entire demand curve for each type of household, allowing Hahn and Metcalfe to project a counterfactual quantity at every price point between the status quo and the uniform price and to compute the changes in total surplus.

The mechanics of Hahn and Metcalfe's welfare computation are depicted in Figure 8. For CARE households, the counterfactual unit price p^* is higher than the discounted CARE price p_C, and so the counterfactual quantity q_C^* is lower than the observed quantity q_C.[12] For non-CARE households, the opposite is true: $p_N > p^*$ and $q_N < q_N^*$. To compute the change in total surplus, Hahn and Metcalfe integrate under the inverse demand curve for each group. In addition, they account for environmental costs by subtracting the change in quantities consumed multiplied by the marginal social cost (MSC), assessed at \$0.68 per therm. Net of environmental costs (in orange), the gain in total surplus (in green) for a representative CARE household is shown in Figure 8(a), while the loss in total surplus (in red) for a representative non-CARE household is shown in Figure 8(b). The net change in total surplus is thus given by the difference between the green area, multiplied by the number of CARE households, and the red area, multiplied by the number of non-CARE households, minus the fixed administrative cost for the program.



(a) CARE households *(b) Non-CARE households*

Figure 8: The change in total surplus (excluding the fixed administrative cost) from the CARE program based on Hahn and Metcalfe (2021).

Note: *The demand curves $D_C(\cdot)$ and $D_N(\cdot)$, and counterfactual quantities q_C^* and q_N^*, are unknown to the researcher and must be inferred.*

[12] Hahn and Metcalfe derive p^* using an accounting identity that equalizes status quo transfers under CARE. See their Section 4.1.2 for a detailed discussion on the derivation, its robustness to alternative specifications, and its relationship with existing policy.

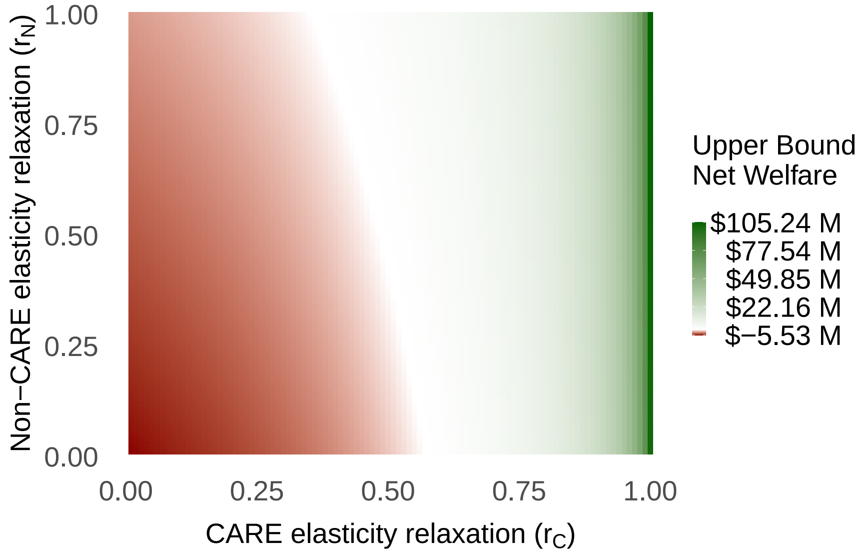


Figure 9: Upper bounds on net welfare under CARE and non-CARE elasticity relaxations.

Hahn and Metcalfe find that CARE households ($\hat{\varepsilon}_C = -0.35$) are substantially more elastic than non-CARE households ($\hat{\varepsilon}_N = -0.14$). This suggests that the more price-sensitive CARE households may benefit more from the subsidy than non-CARE households are harmed by it. Indeed, under their linear imputation, Hahn and Metcalfe estimate a total surplus gain of \$5.1 million for CARE households, which outweighs a total surplus loss of \$3.1 million for non-CARE households. However, the net change in total surplus for the CARE program becomes negative once the \$7 million fixed administrative costs are taken into account.

How robust is this result? To answer this question it is useful to note that the choice of a functional form does more than interpolation in this exercise: it also determines the counterfactual quantities at p^*. To see why this may be restrictive, suppose that counterfactual elasticities were instead allowed to vary unboundedly outside of p_C and p_N, respectively. The welfare impact of the CARE program is maximized when recipients of the program are maximally elastic (i.e., $\varepsilon_C \to -\infty$) so that their counterfactual demand is $q_C^* = 0$, while non-recipients are inelastic (i.e., $\varepsilon_N = 0$) so that their counterfactual demand is the same as in the baseline $q_N^* = q_N$. It is not difficult to see that this extremal outcome would generate large net welfare gains. Indeed, the corresponding upper bound on the net welfare impact of CARE is over \$100M.

Of course, it may not be reasonable to assume that elasticities can be totally unrestricted outside of p_C and p_N. To provide a more informative metric of the robustness of Hahn and Metcalfe's conclusion, we can adapt our framework to find the threshold levels of variability—in both consumer types' demand curves—at which the welfare conclusion would reverse. Figure 9 plots the upper bound on net welfare across different levels of elasticity relaxations for CARE (on the horizontal axis) and non-CARE households (on the vertical axis).[13] As the figure shows, the upper bound on net welfare is negative (in red) for a large range of relaxations r_C and r_N. Indeed, if we only consider variability in CARE household demand, then the threshold r_C^* would be above 1/2. The welfare conclusion seems quite robust.

However, the figure also makes clear that the conclusion is not inevitable. The upper bound on net welfare is positive when r_C and r_N are 0.5, for instance—corresponding to allowing CARE elasticities to be as low as -0.7 and non-CARE elasticities to be as high as -0.07. To put these numbers in context, the standard errors for $\hat{\varepsilon}_C$ and $\hat{\varepsilon}_N$ are 0.159 and 0.068 respectively.[14] These standard errors are with respect to the local average elasticities estimated at p_C and p_N, while the thresholds r_C and r_N in the plot reflect the maximum (resp., minimum) of elasticities allowed at every price between p_C (resp., p_N) and p^*. Still, they provide a benchmark for how likely an elasticity above the threshold value might be. Under this interpretation, Figure 9 suggests that the robustness of Hahn and Metcalfe's conclusion may change substantially when taking the non-CARE households' demand into account: while the 95% confidence interval around $\hat{\varepsilon}_C$ does not include -0.7, the joint 95% confidence interval around $(\hat{\varepsilon}_C, \hat{\varepsilon}_N)$ does include other green points on the plot, such as $(0.45, 0.75)$.

5.3 Old-Age Pensions

Recent work in public finance has sought to compare the effectiveness of government policies by their marginal values of public funds, or MVPFs (Hendren and Sprung-Keyser, 2020). In this subsection we consider an example of this type of analysis, focused on the introduction of old-age pensions in the UK (Giesecke and Jäger, 2021). The 1908 Old-Age Pension Act (OPA) in the UK launched the first universal pension for low-income workers in the UK. To study its effects, Giesecke and Jäger collect individual level census data from 1891, 1901, and 1911. Using a regression discontinuity (RD) design around the minimum eligibility age of 70, Giesecke and Jäger find that labor participation for eligible workers dropped from 46% to 40% after the OPA was introduced—nearly all due to workers who retired.

[13] Although Hahn and Metcalfe assume that demand is linear, we prefer to use an isoelastic benchmark based on the two elasticity estimates at the center of their analysis. See Appendix D for a detailed discussion.

[14] The Auffhammer and Rubin SE is reported directly. To obtain the SE on ε_C, we applied the delta method to Hahn and Metcalfe's "arc elasticity" formula: $|\hat{\varepsilon}_C| = \frac{\frac{22.91-21}{(21+22.91)/2}}{\frac{0.9-0.7}{(0.7+0.9)/2}} = \frac{\frac{(q_0+\hat{\beta})-q_0}{(0.5)\times(2\times q_0+\hat{\beta})}}{\frac{\Delta p}{0.5\times(2\times p+\Delta p)}}$ with $\mathrm{SE}(\hat{\beta}) = 0.91$.

Giesecke and Jäger use this estimate to evaluate the MVPF of the OPA as follows. Nearly all pension recipients received the maximum pension of 260 shillings per year, but their willingness to pay—reflecting the relative value that they anticipated from staying in the workforce—may have been heterogeneous. As a conservative bound on this heterogeneity, Giesecke and Jäger assume that any individual who retired *because* of the OPA must have been on the margin of retiring in exchange for 260 shillings per year and received zero net benefit from the program. Anyone who was *inframarginal* in the sense that they would have retired anyway received the full pension value. The aggregate willingness to pay within the RD population is then $\frac{0.54}{0.6} \times 260$. The cost of supplying the OPA is the total cost of the pensions multiplied by a cost factor of 1.13: $0.6 \times 260 \times 1.13$. Dividing the total willingness to pay by the total cost yields an MVPF estimate of about 0.8.[15]

To connect their exercise to our framework, we start by interpreting the willingness to retire at a given pension amount as a point on a labor supply curve, as depicted in Figure 10(a).[16] We consider the population of workers that are eligible for the OPA. Each eligible worker i derives a utility from working w_i that includes his wage, taste for working, and distaste for retirement. Given a pension amount p, the worker retires if and only if $p \geq w_i$. As such, the share of eligible workers who retire at a pension p is the share of workers for whom $w_i \leq p$.

Denote the CDF of the aggregate distribution of w_i in the population of eligible workers by F. As Figure 10(a) demonstrates, the distribution F can be thought of as a supply curve: $F(p)$ is the proportion of workers who would retire at a pension of p. Thus, when the OPA increased pensions from $p_0 = 0$ shillings to $p_1 = 260$ shillings, worker surplus increased accordingly by

$$\Delta W = \int_{p_0}^{p_1} F(p)\ \mathrm{d}p.$$

While the full supply curve F is not observed, Giesecke and Jäger's RD estimates correspond to measurements of F at two points, namely, $q_0 = \hat{F}(0s) = 0.54$ and $q_1 = \hat{F}(260s) = 0.60$. The conservative extrapolation of these measurements used in Giesecke and Jäger's welfare calculation is equivalent to assuming that the willingness to pay for retirement is constant between $p_0 = 0s$ and $p_1 = 260s$, as depicted by the blue curve in Figure 10(a). As the figure shows, this assumption is quite extreme. How might the MVPF estimate change if a different assumption had been

[15] For their welfare analysis, Giesecke and Jäger extrapolate their RD results by assuming that the labor force participation rate would have continued to evolve after the cutoff age (71+), either: *(i)* at the same rate as for people aged 65–69 absent the OPA; or *(ii)* at a 10% higher rate. Since this is treated as a calibration exercise and standard errors are not provided, we focus on just the cutoff population directly studied by the RD.

[16] See Appendix D for a simple neoclassical model of labor-leisure that results in a labor supply curve of this type.

(a) Labor supply curve illustration. *(b) MVPF bounds under linear gradient relaxations.*

Figure 10: Welfare analysis of the OPA.

made? In Figure 10(b), we plot the upper and lower bounds on the MVPF of the OPA with respect to variability in the gradient of the labor supply curve between p_0 and p_1.[17] When $r = 0$, corresponding to a linear interpolation (with constant gradient), the MVPF is about 0.84. As r goes to 1, we recover Giesecke and Jäger's extreme lower bound and obtain the extreme upper bound, which attributes the full pension amount to all marginal types. Using this plot, we can compare the estimated MVPF of the OPA against external benchmarks or other programs much like the exercise in Section 2: for a given comparison point G, we can find the smallest level of variability r (if any) such that the relevant bound on the OPA MVPF crosses it.[18] We can then evaluate the plausibility of the labor supply curves implied by this threshold.

6 Concluding Remarks

The rapid growth of academic articles on welfare analysis in the last decade (cf. Kleven, 2021) is testament to its importance and relevance to policy. While the welfare effects of small policy changes can be well-approximated by any functional form due to Taylor's theorem, functional form assumptions necessarily entail some loss of generality when policy changes are large. It is thus important to know when welfare conclusions made on the basis of functional form assumptions are robust, and when they are not.

In this paper, we have developed measures to quantify how robust welfare conclusions are with respect to functional form assumptions. Our measures are flexible and simple to use, as we have

[17] The shaded areas around the gradient bounds reflect their 95% confidence intervals with respect to the standard error of the OPA treatment effect estimates from the RD performed by Giesecke and Jäger.

[18] See Hendren and Sprung-Keyser (2020) and the Policy Impacts Library for examples of MVPFs across programs.

demonstrated through empirical applications. Our measures are also easy to compute, as we have shown by exploiting a serendipitous connection between information design and welfare analysis in empirical work.

While we have focused on quantifying the robustness of conclusions under specific functional form assumptions rather than generating new conclusions under alternative, potentially more general, assumptions, there is a close connection between our work and the literature on partial identification for welfare conclusions (cf. Tamer, 2010). Recently, Tebaldi, Torgovitsky and Yang (2023) have shown that meaningful sharp bounds for welfare impacts can be estimated under weak assumptions (e.g., quasilinearity in preferences) when the underlying data has sufficiently rich variation in prices and choices. In the extreme, with infinite variation, Bhattacharya (2015) shows that precise welfare impacts can be estimated nonparametrically as well.

Our paper accommodates the other extreme, in which there might only be enough variation to estimate a single average treatment effect. In this case, even if the average treatment effect is well-identified, there may be substantial ambiguity about the shape of demand at unobserved intermediate price points. Our results show that bounds on the welfare impact in these settings can be easily computed under a range of assumptions that can be interpreted in economic terms. While our current results are limited to markets with homogeneous goods, we view further exploration into the potential for connections between information design and empirical welfare analysis to be a promising area for future research.

References

Aguirre, Inaki, Simon Cowan, and John Vickers, "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, 2010, *100* (4), 1601–15.

Amir, Rabah, Isabelle Maret, and Michael Troege, "On Taxation Pass-Through for a Monopoly Firm," *Annales d'Économie et de Statistique*, 2004, *Jul.–Dec.* (75/76), 155–172.

Amiti, Mary, Stephen J. Redding, and David E. Weinstein, "The Impact of the 2018 Tariffs on Prices and Welfare," *Journal of Economic Perspectives*, 2019, *33* (4), 187–210.

An, Mark Y., "Logconcavity versus Logconvexity: A Complete Characterization," *Journal of Economic Theory*, 1998, *80* (2), 350–369.

Andreyeva, Tatiana, Michael W. Long, and Kelly D. Brownell, "The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food," *American Journal of Public Health*, 2010, *100* (2), 216–222.

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

Auffhammer, Maximilian and Edward Rubin, "Natural Gas Price Elasticities and Optimal Cost Recovery Under Consumer Heterogeneity: Evidence From 300 Million Natural Gas Bills," *Working paper*, 2018.

Bagnoli, Mark and Ted Bergstrom, "Log-Concave Probability and Its Applications," *Economic Theory*, 2005, *26* (2), 445–469.

Bhattacharya, Debopam, "Nonparametric welfare analysis for discrete choice," *Econometrica*, 2015, *83* (2), 617–649.

Bishop, Robert L., "The Effects of Specific and Ad Valorem Taxes," *Quarterly Journal of Economics*, 1968, *82* (2), 198–218.

Caplin, Andrew and Barry Nalebuff, "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 1991, *59* (1), 25–59.

_ and _ , "Aggregation and Social Choice: A Mean Voter Theorem," *Econometrica*, 1991, *59* (1), 1–23.

Cavallo, Alberto, Gita Gopinath, Brent Neiman, and Jenny Tang, "Tariff Pass-Through at the Border and at the Store: Evidence from US Trade Policy," *American Economic Review: Insights*, 2021, *3* (1), 19–34.

Chetty, Raj, "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 2009, *1* (1), 451–488.

Diamond, Peter A., "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *88* (1), 83–95.

Dixit, Avinash, "Comparative Statics for Oligopoly," *International Economic Review*, 1986, *27* (1), 107–122.

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.

Gentzkow, Matthew and Emir Kamenica, "A Rothschild–Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 2016, *106* (5), 597–601.

Giesecke, Matthias and Philipp Jäger, "Pension Incentives and Labor Supply: Evidence from the Introduction of Universal Old-Age Assistance in the UK," *Journal of Public Economics*, 2021, *203*, 104516.

Hahn, Robert W. and Robert D. Metcalfe, "Efficiency and Equity Impacts of Energy Subsidies," *American Economic Review*, 2021, *111* (5), 1658–88.

Harberger, Arnold C., "The Measurement of Waste," *American Economic Review: Papers & Proceedings*, 1964, *54* (3), 58–76.

Hendren, Nathaniel and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

Johnson, Justin P., "The Agency Model and MFN Clauses," *Review of Economic Studies*, 2017, *84* (3), 1151–1185.

Kang, Zi Yang and Jan Vondrák, "Fixed-Price Approximations to Optimal Efficiency in Bilateral Trade," *Working paper*, 2019.

Kleiner, Andreas, Benny Moldovanu, and Philipp Strack, "Extreme Points and Majorization: Economic Applications," *Econometrica*, 2021, *89* (4), 1557–1593.

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

Krugman, Paul R., "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 1979, *9* (4), 469–479.

Malueg, David A., "Monopoly Output and Welfare: The Role of Curvature of the Demand Function," *Journal of Economic Education*, 1994, *25* (3), 235–250.

Marshall, Alfred, *Principles of Economics*, London, UK: Macmillan and Co., 1890.

Melitz, Marc J., "Competitive Effects of Trade: Theory and Measurement," *Review of World Economics*, 2018, *154* (1), 1–13.

Müller, Alfred and Marco Scarsini, "Stochastic Order Relations and Lattices of Probability Measures," *SIAM Journal on Optimization*, 2006, *16* (4), 1024–1043.

Prékopa, András, "On Logarithmic Concave Measures and Functions," *Acta Scientiarum Mathematicarum*, 1973, *34*, 335–343.

Robinson, Joan, *The Economics of Imperfect Competition*, London, UK: Macmillan, 1933.

Tamer, Elie, "Partial Identification in Econometrics," *Annual Review of Economics*, 2010, *2* (1), 167–195.

Tebaldi, Pietro, Alexander Torgovitsky, and Hanbin Yang, "Nonparametric estimates of demand in the california health insurance exchange," *Econometrica*, 2023, *91* (1), 107–146.

Wall Street Journal, "How Many Tariff Studies Are Enough?" 2020.

Weyl, E. Glen and Michal Fabinger, "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 2013, *121* (3), 528–583.

Willig, Robert D., "Consumer's Surplus Without Appology," *American Economic Review*, 1976, *66* (4), 589–597.

Appendix A Omitted Proofs

A.1 Proof of Theorem 2

We adapt our second proof of Theorem 1 and break down the proof of Theorem 2 into the same three steps: *(i)* employing a change of variables to map the problem into an appropriate functional space; *(ii)* endowing this space with a partial order and characterizing its extremal functions; and *(iii)* mapping the solution back to the original problem. We define $\widetilde{D}(\pi) = A(D(B^{-1}(\pi)))$ and $\pi = B(p)$, so that

$$\begin{aligned}\widetilde{D}''(\pi) &= \frac{\mathrm{d}}{\mathrm{d}\pi}\left[\frac{A'(D(B^{-1}(\pi)))D'(B^{-1}(\pi))}{B'(B^{-1}(\pi))}\right] \\ &= \frac{1}{B'(p)}\frac{\mathrm{d}}{\mathrm{d}p}\left[\frac{A'(D(p))D'(p)}{B'(p)}\right] \in [\underline{\gamma},\overline{\gamma}] \qquad \text{for } \pi \in [\pi_0,\pi_1].\end{aligned}$$

Throughout, we focus on the bounds implied by $\widetilde{D}''(\pi) \geq \underline{\gamma}$; the bounds implied by $\widetilde{D}''(\pi) \leq \overline{\gamma}$ can be similarly derived.

Step 1: Changing variables. Instead of choosing a demand curve to maximize or minimize the loss in consumer surplus, we choose the function $h : [B(p_0), B(p_1)] \to \mathbb{R}$ defined by

$$h(\pi) \coloneqq \widetilde{D}'(\pi) - \underline{\gamma}\pi \qquad \text{for } \pi \in [\pi_0,\pi_1].$$

Given h, $\widetilde{D}$ is completely determined, and vice versa:

$$\widetilde{D}(\pi) = A(q_0) + \int_{\pi_0}^{\pi}\left[h(s) + \underline{\gamma}s\right]\,\mathrm{d}s \qquad \text{for } \pi \in [\pi_0,\pi_1].$$

This is obtained via integration by parts. Next, we define the set of feasible functions h that are consistent with Assumption 2:

$$\mathcal{H} \coloneqq \left\{h \in \mathcal{H}_0 : h(\pi) \leq \underline{\gamma}\pi,\ \int_{\pi_0}^{\pi_1} h(s)\,\mathrm{d}s = A(q_1) - A(q_0) - \frac{1}{2}\underline{\gamma}\left(\pi_1^2 - \pi_0^2\right)\right\},$$

where $\mathcal{H}_0 = \left\{h : [\pi_0,\pi_0] \to [\underline{h},\overline{h}] \text{ is non-decreasing}\right\}$ for some given $\underline{h}$ and $\overline{h}$. Here, we assume that $\underline{h} \leq \min\left\{-\underline{\gamma}\pi_0, -\underline{\gamma}\pi_1, -\left[A(q_0) - A(q_1)\right]/\left[\pi_1 - \pi_0\right]\right\}$ and $\overline{h} = -\underline{\gamma}\pi_1$, with the goal of eventually taking the limit $\underline{h} \to -\infty$. Thus we arrive at the equivalent problem:

$$\begin{cases} \overline{\Delta\text{CS}} = \sup_{h\in\mathcal{H}} \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{\pi_0}^{B(p)} \left[h(s) + \underline{\gamma} s\right] \, \mathrm{d}s\right) \, \mathrm{d}p, \\ \underline{\Delta\text{CS}} = \inf_{h\in\mathcal{H}} \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{\pi_0}^{B(p)} \left[h(s) + \overline{\gamma} s\right] \, \mathrm{d}s\right) \, \mathrm{d}p. \end{cases} \tag{7}$$

Step 2: Characterizing the set $\mathcal{H}$. We now endow the set $\mathcal{H}$ with a partial order. Formally, for any two functions $h_1, h_2 \in \mathcal{H}$, we write

$$h_1 \succeq h_2 \iff \int_{\pi_0}^{\pi} h_1(s) \, \mathrm{d}s \geq \int_{\pi_0}^{\pi} h_2(s) \, \mathrm{d}s \qquad \text{for } \pi \in [\pi_0, \pi_1].$$

Analogous to Lemma 1, we show:

Lemma 3. *Any function $h \in \mathcal{H}$ satisfies $h^* \succeq h \succeq h_*$, where:*

(i) *if $0 \leq \underline{\gamma} \leq 2\left[A(q_0) - A(q_1)\right] / \left[A(p_1) - A(p_0)\right]^2$, then*

$$h^*(s) \coloneqq -\frac{A(q_0) - A(q_1)}{\pi_1 - \pi_0} - \frac{\gamma}{2}\left(\pi_0 + \pi_1\right),$$
$$h_*(s) \coloneqq \begin{cases} \overline{h} & \text{if } s > \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{\gamma}{2}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}}, \\ \underline{h} & \text{if } s \leq \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{\gamma}{2}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}}; \end{cases}$$

(ii) *if $-2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2 \leq \underline{\gamma} < 0$, then*

$$h^*(s) \coloneqq -\frac{A(q_0) - A(q_1)}{\pi_1 - \pi_0} - \frac{\gamma}{2}\left(\pi_0 + \pi_1\right),$$
$$h_*(s) \coloneqq \begin{cases} -\underline{\gamma} s & \text{if } s > -\frac{\underline{h} + \sqrt{\underline{h}^2 + \underline{\gamma}^2\pi_0^2 + 2\underline{\gamma}\left[\underline{h}\pi_0 - A(q_0) + A(q_1)\right]}}{\underline{\gamma}}, \\ \underline{h} & \text{if } s \leq -\frac{\underline{h} + \sqrt{\underline{h}^2 + \underline{\gamma}^2\pi_0^2 + 2\underline{\gamma}\left[\underline{h}\pi_0 - A(q_0) + A(q_1)\right]}}{\underline{\gamma}}; \end{cases}$$

(iii) *if* $\underline{\gamma} < -2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2$, *then*

$$h^*(s) := \begin{cases} -\underline{\gamma}\left[\pi_1 - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}\right] & \textit{if } s > \pi_1 - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}, \\ -\underline{\gamma}s & \textit{if } s \leq \pi_1 - \sqrt{\frac{2[A(q_1)-A(q_0)]}{\underline{\gamma}}}, \end{cases}$$

$$h_*(s) := \begin{cases} -\underline{\gamma}s & \textit{if } s > -\frac{\underline{h}+\sqrt{\underline{h}^2+\underline{\gamma}^2\pi_0^2+2\underline{\gamma}[\underline{h}\pi_0-A(q_0)+A(q_1)]}}{\underline{\gamma}}, \\ \underline{h} & \textit{if } s \leq -\frac{\underline{h}+\sqrt{\underline{h}^2+\underline{\gamma}^2\pi_0^2+2\underline{\gamma}[\underline{h}\pi_0-A(q_0)+A(q_1)]}}{\underline{\gamma}}; \end{cases}$$

Proof. When the constraint $h(\pi) \leq -\underline{\gamma}\pi$ is slack, results from the information design literature (e.g., Kang and Vondrák, 2019; Kleiner et al., 2021) imply that $h^* \succeq h \succeq h_*$ for any $h \in \mathcal{H}$, where

$$h^*(s) := -\frac{A(q_0) - A(q_1)}{\pi_1 - \pi_0} - \frac{\underline{\gamma}}{2}\left(\pi_0 + \pi_1\right),$$

$$h_*(s) := \begin{cases} \overline{h} & \text{if } s > \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{1}{2}\underline{\gamma}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}}, \\ \underline{h} & \text{if } s \leq \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{1}{2}\underline{\gamma}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}}. \end{cases}$$

To verify that the constraint $h(\pi) \leq -\underline{\gamma}\pi$ is slack, we require:

(a) $-\frac{A(q_0)-A(q_1)}{\pi_1-\pi_0} - \frac{\underline{\gamma}}{2}\left(\pi_0 + \pi_1\right) \leq \min\left\{-\underline{\gamma}\pi_0, -\underline{\gamma}\pi_1\right\}$ in order for $h^*(s)$ to be as stated above.

Equivalently,

$$\underline{\gamma}\left(\pi_1 - \pi_0\right) \geq -\frac{2\left[A(q_0) - A(q_1)\right]}{\pi_1 - \pi_0} \quad \text{and} \quad -\underline{\gamma}\left(\pi_1 - \pi_0\right) \geq -\frac{2\left[A(q_0) - A(q_1)\right]}{\pi_1 - \pi_0}.$$

Clearly, these inequalities hold when

$$-2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2 \leq \underline{\gamma} \leq 2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2.$$

We therefore conclude that

$$h^*(s) = -\frac{A(q_0) - A(q_1)}{\pi_1 - \pi_0} - \frac{\underline{\gamma}}{2}\left(\pi_0 + \pi_1\right) \qquad \text{for } -\frac{2\left[A(q_0) - A(q_1)\right]}{\left(\pi_1 - \pi_0\right)^2} \leq \underline{\gamma} \leq \frac{2\left[A(q_0) - A(q_1)\right]}{\left(\pi_1 - \pi_0\right)^2}.$$

(b) $\frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{\underline{\gamma}}{2}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}} \in [\pi_0, \pi_1]$ and $\overline{h} \leq -\underline{\gamma}s$ if $s > -\frac{\underline{h}+\sqrt{\underline{h}^2+\underline{\gamma}^2\pi_0^2+2\underline{\gamma}[\underline{h}\pi_0-A(q_0)+A(q_1)]}}{\underline{\gamma}}$ in order for $h_*(s)$ to be as stated above.

Equivalently,

$$\begin{cases} \underbrace{\overline{h}}_{=-\underline{\gamma}\pi_1} (\pi_1 - \pi_0) + A(q_0) - A(q_1) + \dfrac{\underline{\gamma}}{2}\left(\pi_1^2 - \pi_0^2\right) \geq 0, \\ \underbrace{\underline{h}}_{\leq -\underline{\gamma}\pi_0} (\pi_1 - \pi_0) + A(q_0) - A(q_1) + \dfrac{\underline{\gamma}}{2}\left(\pi_1^2 - \pi_0^2\right) \leq 0, \end{cases} \quad \text{and} \quad \underline{\gamma} \geq 0.$$

These inequalities hold when $0 \leq \underline{\gamma} \leq 2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2$. We therefore conclude that

$$h_*(s) := \begin{cases} \overline{h} & \text{if } s > \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{\underline{\gamma}}{2}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}}, \\ \underline{h} & \text{if } s \leq \frac{\overline{h}\pi_1 - \underline{h}\pi_0 + A(q_0) - A(q_1) + \frac{\underline{\gamma}}{2}\left(\pi_1^2 - \pi_0^2\right)}{\overline{h} - \underline{h}} \end{cases} \qquad \text{for } 0 \leq \underline{\gamma} \leq \frac{2\left[A(q_0) - A(q_1)\right]}{\left(\pi_1 - \pi_0\right)^2}.$$

The above argument thus proves part (i) of Lemma 3.

Given the form of h_* stated in parts (ii) and (iii) of Lemma 3, we next prove that $h \succeq h_*$ for any $h \in \mathcal{H}$. Let $\pi_* := -\left[\underline{h} + \sqrt{\underline{h}^2 + \underline{\gamma}^2\pi_0^2 + 2\underline{\gamma}\left[\underline{h}\pi_0 - A(q_0) + A(q_1)\right]}\right] / \underline{\gamma}$. Observe that $\underline{h} \leq \min\left\{-\underline{\gamma}\pi_0, -\underline{\gamma}\pi_1\right\}$ implies:

$$\begin{aligned} \pi_* \geq \pi_0 &\iff \underline{h} + \sqrt{\underline{h}^2 + \underline{\gamma}^2\pi_0^2 + 2\underline{\gamma}\left[\underline{h}\pi_0 - A(q_0) + A(q_1)\right]} \geq -\underline{\gamma}\pi_0 \\ &\iff A(q_0) - A(q_1) \geq 0, \\ \pi_* \leq \pi_1 &\iff \underline{h} + \sqrt{\underline{h}^2 + \underline{\gamma}^2\pi_0^2 + 2\underline{\gamma}\left[\underline{h}\pi_0 - A(q_0) + A(q_1)\right]} \leq -\underline{\gamma}\pi_1 \\ &\iff \underline{h} \leq -\frac{A(q_0) - A(q_1)}{\pi_1 - \pi_0}. \end{aligned}$$

These inequalities hold; hence $\pi_* \in [\pi_0, \pi_1]$. Then, to complete the proof of part (ii) of Lemma 3:

- If $\pi \in [\pi_0, \pi_*]$, then the inequality $\int_{\pi_0}^{\pi} h(s)\ \mathrm{d}s \geq \int_{\pi_0}^{\pi} h_*(s)\ \mathrm{d}s$ holds trivially from the fact that $h(s) \geq \underline{h} = h_*(s)$ for $s \leq \pi_*$.
- If $\pi \in [\pi_*, \pi_1]$, then the inequality $\int_{\pi}^{\pi_1} h(s)\ \mathrm{d}s \leq \int_{\pi}^{\pi_1} h_*(s)\ \mathrm{d}s$ holds from the fact that $h(s) \leq -\underline{\gamma}s = h_*(s)$ for $s \geq \pi_*$. Since $\int_{\pi_0}^{\pi_1} h(s)\ \mathrm{d}s = \int_{\pi_0}^{\pi_1} h_*(s)\ \mathrm{d}s$, we conclude that $\int_{\pi_0}^{\pi} h(s)\ \mathrm{d}s \geq \int_{\pi_0}^{\pi} h_*(s)\ \mathrm{d}s$.

Finally, given the form of h^* as stated above in part (iii) of Lemma 3, we assume that $\underline{\gamma} < -2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2$ and prove that $h^* \succeq h$ for any $h \in \mathcal{H}$. Now, because $\underline{\gamma} < -2\left[A(q_0) - A(q_1)\right] / \left(\pi_1 - \pi_0\right)^2$, we must have $\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right] / \underline{\gamma}} \in [\pi_0, \pi_1]$. Then:

- If $\pi \in [\pi_0, \pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}]$, then the inequality $\int_{\pi_0}^{\pi} h^*(s)\ \mathrm{d}s \geq \int_{\pi_0}^{\pi} h(s)\ \mathrm{d}s$ holds trivially from the fact that $h^*(s) = -\underline{\gamma}s \geq h(s)$ for $s \leq \pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}$.

- If $\pi \in [\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}, \pi_1]$, then suppose there exists $\hat{\pi}$ satisfying

$$\hat{\pi} \in (\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}, \pi_1) \quad \text{and} \quad \int_{\pi_0}^{\hat{\pi}} h(s)\ \mathrm{d}s > \int_{\pi_0}^{\hat{\pi}} h^*(s)\ \mathrm{d}s.$$

  Then $h(\hat{\pi}) > h^*(\hat{\pi}) = -\underline{\gamma}\left[\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}\right]$; otherwise, $h(s) \leq h^*(s)$ for every $s \in [\pi_0, \hat{\pi}]$, contradicting our assumption that $\int_{\pi_0}^{\hat{\pi}} h(s)\ \mathrm{d}s > \int_{\pi_0}^{\hat{\pi}} h^*(s)\ \mathrm{d}s$. Because h is non-decreasing, this implies that $h(s) \geq h(\hat{\pi}) > -\underline{\gamma}\left[\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}\right] = h^*(s)$ for every $s \in (\hat{\pi}, \pi_1)$. Then:

$$\begin{aligned} \int_{\pi_0}^{\pi_1} h(s)\ \mathrm{d}s &= \int_{\pi_0}^{\hat{\pi}} h(s)\ \mathrm{d}s + \int_{\hat{\pi}}^{\pi_1} h(s)\ \mathrm{d}s \\ &> \int_{\pi_0}^{\hat{\pi}} h^*(s)\ \mathrm{d}s + \int_{\hat{\pi}}^{\pi_1} h(s)\ \mathrm{d}s \\ &\geq \int_{\pi_0}^{\hat{\pi}} h^*(s)\ \mathrm{d}s + \int_{\hat{\pi}}^{\pi_1} h^*(s)\ \mathrm{d}s = \int_{\pi_0}^{\pi_1} h^*(s)\ \mathrm{d}s. \end{aligned}$$

  This contradicts the fact that $\int_{\pi_0}^{\pi_1} h(s)\ \mathrm{d}s = \int_{\pi_0}^{\pi_1} h^*(s)\ \mathrm{d}s = A(q_1) - A(q_0) - \underline{\gamma}\left(\pi_1^2 - \pi_0^2\right)/2$ since $h, h^* \in \mathcal{H}$. Here, the first inequality follows by the definition of $\hat{\pi}$, while the second inequality follows from our observation that $h(s) > h^*(s)$ for every $s \in (\hat{\pi}, \pi_1)$. Consequently, our initial supposition was wrong: no such $\hat{\pi}$ exists; hence $\int_{\pi_0}^{\pi} h^*(s)\ \mathrm{d}s \geq \int_{\pi_0}^{\pi} h(s)\ \mathrm{d}s$ for any $\pi \in [\pi_1 - \sqrt{2\left[A(q_1) - A(q_0)\right]/\underline{\gamma}}, \pi_1]$.

This completes the proof of part (iii) of Lemma 3. □

It is easy to check that $h^*, h_* \in \mathcal{H}$. Therefore, Lemma 3 characterizes the largest and smallest elements of the partially ordered set $(\mathcal{H}, \succeq)$.

Step 3: Mapping back to the original problem. Having characterized the largest and smallest elements of $(\mathcal{H}, \succeq)$, it remains to map these back to the original problem. To this end, we define the functional $\Delta\mathrm{CS} : \mathcal{H} \to \mathbb{R}$ by

$$\Delta\mathrm{CS}(h) := \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{\pi_0}^{B(p)} \left[h(s) + \underline{\gamma} s\right] \,\mathrm{d}s\right) \mathrm{d}p.$$

Our problem (7) is equivalent to maximizing and minimizing this functional over the family $\mathcal{H}$. The following lemma shows that this can be done with the aid of the partial order $\succeq$ defined in our previous step:

Lemma 4. *The functional $\Delta\mathrm{CS}(\cdot)$ is increasing in the partial order $\succeq$; that is, for any $h_1 \succeq h_2$,*

$$\int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{\pi_0}^{B(p)} \left[h_1(s) + \underline{\gamma} s\right] \,\mathrm{d}s\right) \mathrm{d}p \geq \int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \int_{\pi_0}^{B(p)} \left[h_2(s) + \underline{\gamma} s\right] \,\mathrm{d}s\right) \mathrm{d}p.$$

Proof. The result follows straightforwardly from the definition of the partial order $\succeq$, the fact that A (and hence A^{-1}) is increasing, and a pointwise comparison of the two integrands. □

Together, Lemmas 3 and 4 imply that the functional $\Delta\mathrm{CS}(\cdot)$ is maximized at h^* and minimized at h_*:

$$\overline{\Delta\mathrm{CS}} = \Delta\mathrm{CS}(h^*) \quad \text{and} \quad \underline{\Delta\mathrm{CS}} = \Delta\mathrm{CS}(h_*).$$

Through straightforward computation and taking the limit $\underline{h} \to -\infty$, we obtain the result of Theorem 2.

A.2 Proof of Proposition 1

Similar to the proofs of Theorems 1 and 2, we prove Proposition 1 in three steps. We focus on part (a) of Proposition 1.

Step 1: Changing variables. Let $\pi = B(p)$ be defined on $[\pi_0, \pi_1] = [B(p_0), B(p_1)]$, and consider $\widetilde{D} : [\pi_0, \pi_1] \to \mathbb{R}$ be defined by $\widetilde{D}(B(p)) = A(D(p))$. We choose the gradient function $\beta(\cdot)$:

$$\beta(\pi) := \widetilde{D}'(\pi) \qquad \text{for } \pi \in [\pi_0, \pi_1].$$

Given $\beta(\cdot)$, $\widetilde{D}(\cdot)$ is completely determined, and vice versa:

$$\widetilde{D}(\pi) = A(q_0) + \int_{\pi_0}^{\pi} \beta(s) \,\mathrm{d}s \qquad \text{for } \pi \in [\pi_0, \pi_1].$$

This is obtained via integration by parts, which assumes that $\widetilde{D}(\cdot)$ is absolutely continuous on $[\pi_0, \pi_1]$. Analogous to the family of demand curves $\mathcal{D}$, we define the set of feasible gradient functions:

$$\mathcal{B} \coloneqq \left\{ \beta : [\pi_0, \pi_1] \to [\underline{\beta}, \overline{\beta}] \text{ s.t. } \int_{\pi_0}^{\pi_1} \beta(s) \, \mathrm{d}s = A(q_1) - A(q_0) \right\}.$$

Thus we arrive at the equivalent problem:

$$\begin{cases} \overline{q} = \sup_{\beta \in \mathcal{B}} \; A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(\hat{p})} \beta(s) \, \mathrm{d}s \right), \\ \underline{q} = \inf_{\beta \in \mathcal{B}} \; A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(\hat{p})} \beta(s) \, \mathrm{d}s \right). \end{cases} \tag{8}$$

Step 2: Characterizing the set $\mathcal{B}$. Recall that, in Lemma 1, we showed that $\beta^* \succeq \beta \succeq \beta_*$ for any $\beta \in \mathcal{B}$, thereby characterizing the largest and smallest elements of the partially ordered set $(\mathcal{B}, \succeq)$. Here, β^* and β_* are as defined in the statement of Theorem 1.

Step 3: Mapping back to the original problem. Having characterized the largest and smallest elements of $(\mathcal{B}, \succeq)$, it remains to map these back to the original problem. To this end, we define the functional $\hat{q} : \mathcal{B} \to \mathbb{R}$ by

$$\hat{q}(\beta) \coloneqq A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(\hat{p})} \beta(s) \, \mathrm{d}s \right).$$

Our problem (8) is equivalent to maximizing and minimizing this functional over the family $\mathcal{B}$. The following lemma shows that this can be done with the partial order defined previously:

Lemma 5. *The functional $\hat{q}(\cdot)$ is increasing in the partial order $\succeq$:*

$$A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(\hat{p})} \beta_1(s) \, \mathrm{d}s \right) \geq A^{-1} \left(A(q_0) + \int_{\pi_0}^{B(\hat{p})} \beta_2(s) \, \mathrm{d}s \right) \quad \textit{for any } \beta_1 \succeq \beta_2.$$

Proof. The result follows straightforwardly from the definition of the partial order $\succeq$, the fact that A (and hence A^{-1}) is increasing, and a pointwise comparison of the two integrands. □

Together, Lemmas 1 and 5 imply that the functional $\hat{q}(\cdot)$ is maximized at β^* and minimized at β_*: $\overline{q} = \hat{q}(\beta^*)$ and $\underline{q} = \hat{q}(\beta_*)$. This completes the proof of part (a) of Proposition 1; part (b) of Proposition 1 can be proven very similarly and is therefore omitted.

Appendix B Extensions of Robustness Measures

In this appendix, we show how our robustness measures can be extended when more complexity is allowed for. Motivated by applications in Section 5, we focus on four extensions: *(i)* counterfactual exercises; *(ii)* more observations; *(iii)* measurement error; and *(iv)* other welfare measures.

B.1 Counterfactual Exercises

Our robustness measures can be extended to counterfactual exercises where only one point on the demand curve is observed. So far, we have assumed that two points on the demand curve are observed: (p_0, q_0) and (p_1, q_1). However, in counterfactual exercises such as our application in Section 5.2, the quantity that would be demanded at p_1 is not known.

To illustrate, we focus on extending our robustness measure r^* to this setting. Following our approach in Section 4, it suffices to establish the analog of Theorem 1:

Theorem 3. *Suppose that only (p_0, q_0) and p_1 are observed. Under Assumption 1, the largest and smallest possible losses in consumer surplus between p_0 and p_1, $\overline{\Delta\text{CS}}$ and $\underline{\Delta\text{CS}}$, are respectively:*

$$\begin{cases} \overline{\Delta\text{CS}} \coloneqq \displaystyle\int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \overline{\beta}\left[B(p) - B(p_0)\right]\right)\,\mathrm{d}p, \\ \underline{\Delta\text{CS}} \coloneqq \displaystyle\int_{p_0}^{p_1} A^{-1}\left(A(q_0) + \underline{\beta}\left[B(p) - B(p_0)\right]\right)\,\mathrm{d}p. \end{cases}$$

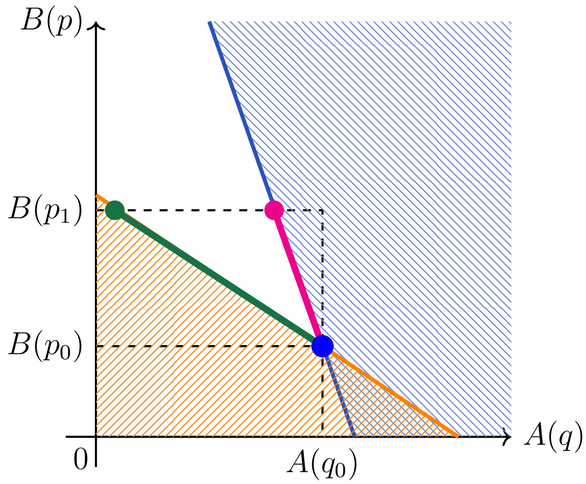


Figure B.1: Illustration of bounds when only (p_0, q_0) and p_1 are observed.

Theorem 3 can be shown using our earlier geometric argument for Theorem 1, illustrated in Figure B.1. The largest possible value of $A(q_1)$ that is consistent with Assumption 1 can be found by drawing the (blue) straight line with gradient $\overline{\beta}$ that passes through the point $(B(p_0), A(q_0))$, and then finding the (red) point on the line at $B(p_1)$. It is clear that this value of q_1 must also yield the maximal $\overline{\Delta \text{CS}}$; hence $\overline{\Delta \text{CS}}$ must be attained by the red curve. A symmetric argument shows that $\underline{\Delta \text{CS}}$ must be attained by the green curve.

B.2 More Observations

Our robustness measures can also be extended to settings where more than two points on the same demand curve are observed, as is the case in some empirical applications. Doing so requires a generalization of our robustness measures to an arbitrary (finite) number of observations, which we denote by $(p_0, q_0), \ldots, (p_{n-1}, q_{n-1})$.

To illustrate, we again focus on extending our robustness measure r^* to this setting. The generalization of Theorem 1 to this setting is:

Theorem 4. *Suppose that $(p_0, q_0), \ldots, (p_{n-1}, q_{n-1})$ are observed. Define the auxiliary functions $\beta^*, \beta_* : [B(p_0), B(p_1)] \to \mathbb{R}$ as follows: for each $j \in \{0, 1, \ldots, n-1\}$,*

$$\beta^*(s) := \begin{cases} \underline{\beta} & \text{if } \frac{\overline{\beta}B(p_j) - \underline{\beta}B(p_{j+1}) - A(q_j) + A(q_{j+1})}{\overline{\beta} - \underline{\beta}} < s \leq B(p_{j+1}), \\ \overline{\beta} & \text{if } B(p_j) < s \leq \frac{\overline{\beta}B(p_j) - \underline{\beta}B(p_{j+1}) - A(q_j) + A(q_{j+1})}{\overline{\beta} - \underline{\beta}}; \end{cases}$$

$$\beta_*(s) := \begin{cases} \overline{\beta} & \text{if } \frac{\overline{\beta}B(p_{j+1}) - \underline{\beta}B(p_j) + A(q_j) - A(q_{j+1})}{\overline{\beta} - \underline{\beta}} < s \leq B(p_{j+1}), \\ \underline{\beta} & \text{if } B(p_j) < s \leq \frac{\overline{\beta}B(p_{j+1}) - \underline{\beta}B(p_j) + A(q_j) - A(q_{j+1})}{\overline{\beta} - \underline{\beta}}. \end{cases}$$

Under Assumption 1, the largest and smallest possible losses in consumer surplus between p_0 and p_1, $\overline{\Delta \text{CS}}$ and $\underline{\Delta \text{CS}}$, are respectively:

$$\begin{cases} \overline{\Delta \text{CS}} = \displaystyle\int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta^*(s) \, \mathrm{d}s \right) \, \mathrm{d}p, \\ \underline{\Delta \text{CS}} = \displaystyle\int_{p_0}^{p_1} A^{-1} \left(A(q_0) + \int_{B(p_0)}^{B(p)} \beta_*(s) \, \mathrm{d}s \right) \, \mathrm{d}p. \end{cases}$$

Theorem 4 can be shown by applying Theorem 1 between every two adjacent points. Figure B.2 illustrates the geometric argument for the case of $n = 3$ observations, where both the largest (in

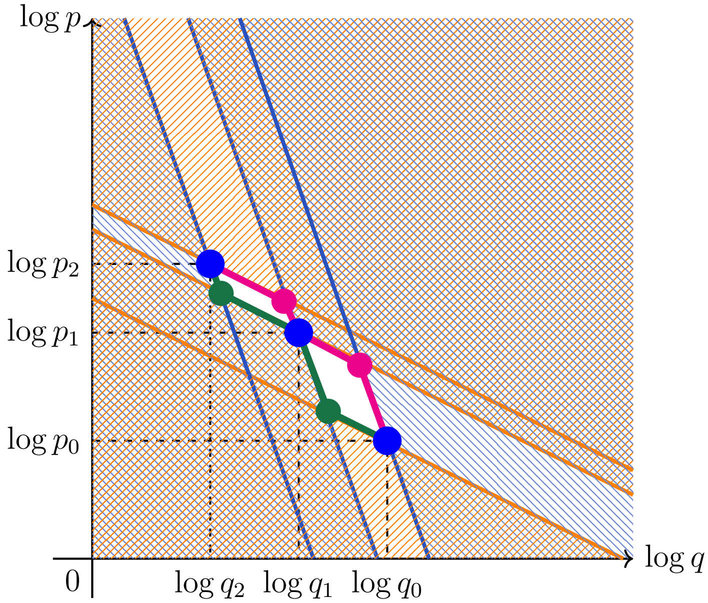


Figure B.2: Illustration of bounds with $n = 3$ observations.

red) and smallest (in green) possible losses in consumer surplus are depicted.

B.3 Measurement Error

Our robustness measures can be extended to account for uncertainty due to measurement error. Following our discussion in Section 2, we focus on the uncertainty in the treatment effect estimate, $\hat{\beta}$. By propagating this uncertainty into the bounds implied by Assumptions 1 and 2—for example, by applying a bootstrap procedure to equations (4) and (5)—we can extend Theorems 1 and 2 to account for measurement error in $\hat{\beta}$. This is straightforward because equations (4) and (5) are explicit expressions of $\hat{\beta}$. It can be readily verified that these expressions are monotone with respect to $\hat{\beta}$, which implies that more precise measurements of $\hat{\beta}$ would lead to narrower bounds. In turn, we can derive corresponding confidence intervals for our robustness measures.[19]

B.4 Other Welfare Measures

We have so far focused on measures of robustness for estimates of consumer surplus, defined as the integral of the Marshallian demand curve. To conclude this section, we discuss how our analysis

[19] In some empirical applications, it may be possible that measurement error leads to a confidence interval for $\hat{\beta}$ that includes 0. In the spirit of our framework, non-negative estimates of β correspond to $r^* = 1$ and $\kappa^* = \infty$: all curves that pass through the implied points $(p_0, \hat{q}_0)$ and $(p_1, \hat{q}_1)$ cannot overturn the welfare conclusion.

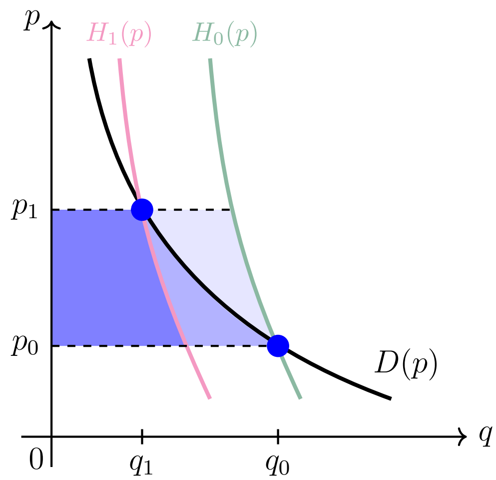


Figure B.3: Illustration of EV *and* CV *relative to* ΔCS *for a normal good.*

extends to various alternative welfare measures: *(i)* deadweight loss; *(ii)* equivalent variation (EV) and compensating variation (CV); and *(iii)* supply-side welfare measures such as producer surplus.

Deadweight loss. Under additional supply-side assumptions, our results can be extended when the welfare measure of interest is deadweight loss. For example, when the supply curve is flat and producers are price takers (see Section 5.1 for an empirical application), the change in deadweight loss due to a tariff $\tau = p_1 - p_0$ is

$$\Delta\text{DWL} = \int_{p_0}^{p_1} D(p)\ \mathrm{d}p - (p_1 - p_0)\, q_1 = \Delta\text{CS} - (p_1 - p_0)\, q_1.$$

Since (p_0, q_0) and (p_1, q_1) are observed, maximizing or minimizing ΔDWL is equivalent to maximizing or minimizing ΔCS. As such, analogs of Theorems 1 and 2 continue to hold.

EV and CV. When consumer utility is quasilinear in money, there are no income effects and the change in consumer surplus coincides exactly with the EV and CV. However, for markets in which income effects are significant, our framework can be adapted to examine the EV and CV directly. To see this, note that the EV and CV can be defined as follows for a normal good:

$$\text{EV} := \int_{p_0}^{p_1} H_1(p)\ \mathrm{d}p \quad \text{and} \quad \text{CV} := \int_{p_0}^{p_1} H_0(p)\ \mathrm{d}p,$$

where H_1 and H_0 respectively denote the Hicksian demand curves at the utility levels obtained at p_1 and p_0. Figure B.3 plots an illustration of the Hicksian demand curves relative to the Marshallian demand curve considered in Sections 3 and 4. The EV corresponds to the most darkly shaded area, left of H_1; the change in consumer surplus corresponds to the shaded area left of D as before; and the CV is the entire shaded area, left of H_0.

As noted by Willig (1976), the change in consumer surplus offers a one-sided bound to EV and CV. As Figure B.3 illustrates, when $p_1 > p_0$, EV $\geq \Delta$CS $\geq$ CV (as these welfare measures are negative). This suggests that the robustness measures for ΔCS discussed in Section 4 can apply as conservative measures of robustness for EV (if in the benchmark, $\Delta\text{CS} \geq G$) or CV (otherwise) as well. When this is not sufficient, Theorems 1 and 2 can be applied directly to the Hicksian demand curves H_0 and H_1 instead. Note, however, that since the counterfactual expenditures—$e(p_0, u_1)$ for EV and $e(p_1, u_0)$ for CV—are not observed, the points $(p_0, H_1(p_0))$ for CV and $(p_1, H_0(p_1))$ for EV must be treated as counterfactuals as in Appendix B.1.

Producer surplus. Our results also extend straightforwardly to supply-side welfare measures like producer surplus when producers are price takers. Section 5.3 considers such an empirical application where individuals supply labor. In this case, the relevant integrals are with respect to an upward-sloping supply curve, rather than a downward-sloping demand curve; but the remainder of the exercise is much the same.

Appendix C Shape Constraints

C.1 Common Shape Constraints

Different literatures in economics employ a variety of constraints on the shape of demand that capture other intuitions pertaining to their fields of interest. To be comprehensive, we consider a range of shape constraints that are considered standard in different fields. Each shape constraint (abbreviated by "SC") restricts ΔCS in a different way. We detail these assumptions below and provide some examples of how they are invoked in different fields.

(**SC1**) **Marshall's second law.** Demand is said to satisfy Marshall's second law if its price elasticity $\varepsilon(p) = pD'(p)/D(p)$ is decreasing in p. This was introduced by Marshall (1890) and is widely used in international trade, macroeconomics, and microeconomics, including by Krugman (1979), Bishop (1968), Johnson (2017), and Melitz (2018), who also provides some empirical justification for this shape constraint in the context of trade models.

(**SC2**) **Decreasing marginal revenue.** Let $P(q) \coloneqq D^{-1}(q)$ denote the inverse demand curve. Demand exhibits decreasing marginal revenue if marginal revenue $\text{MR}(q) \coloneqq P(q) + qP'(q)$ is decreasing in q. This shape constraint is standard in microeconomics (see Robinson, 1933, for example) and ensures that a profit-maximizing price exists for a monopolist who faces a convex cost function.

(**SC3**) **Log-concave demand.** Demand is log-concave if $D'(p)/D(p)$ is decreasing in p. The comprehensive surveys of Bagnoli and Bergstrom (2005) and An (1998) demonstrate that many common demand curves are log-concave. Log-concave demand also has a simple economic interpretation, as Amir, Maret and Troege (2004) show: the pass-through rate of a change in a monopolist's marginal cost is less than one if and only if demand is log-concave (see also Weyl and Fabinger, 2013). It is also well-known that log-concavity is a sufficient condition for a unique equilibrium to exist in common models of Cournot competition (Dixit, 1986) and differentiated products Bertrand competition (Caplin and Nalebuff, 1991a).

(**SC4**) **Concave demand.** Demand is concave if $D'(p)$ is decreasing in p. Robinson (1933) shows that concave demand has a simple economic interpretation: total output increases when third-degree price discrimination by a monopolist causes prices to rise in markets with concave demands (see also Malueg, 1994 and Aguirre, Cowan and Vickers, 2010 for variations and generalizations of this result).

(**SC5**) **ρ-concave demand.** For a given real number ρ, demand is ρ-concave if $D'(p)\,[D(p)]^{\rho-1}$ is decreasing in p. Based on the work of Prékopa (1973), this shape constraint was introduced to the economics literature by Caplin and Nalebuff (1991a,b) as a generalization of log-concavity ($\rho = 0$) and concavity ($\rho = 1$). Different values of ρ parametrize the restrictiveness of this constraint: a ρ'-concave demand curve is ρ''-concave for any $\rho'' < \rho'$.

(**SC6**) **Convex demand.** Demand is convex if $D'(p)$ is increasing in p. Similar to concave demand (**SC4**), Robinson (1933) shows that total output increases when third-degree price discrimination by a monopolist causes prices to fall in markets with convex demands (see also Malueg, 1994 and Aguirre, Cowan and Vickers, 2010 for variations and generalizations of this result).

(**SC7**) **Log-convex demand.** Demand is log-convex if $D'(p)/D(p)$ is increasing in p. Similar to log-concave demand (**SC3**), Amir et al. (2004) show that the pass-through rate of a change in a monopolist's marginal cost is more than one if and only if demand is log-convex.

(**SC8**) **ρ-convex demand.** For a given real number ρ, demand is ρ-convex if $D'(p)\,[D(p)]^{\rho-1}$ is increasing in p. Similar to ρ-concave demand (**SC5**), ρ-convexity generalizes convexity ($\rho = 1$) and log-convexity ($\rho = 0$); a ρ'-convex demand curve is ρ''-convex for any $\rho'' > \rho'$.

These shape constraints can be divided into two categories: *concave-like* shape constraints (**SC1**)–(**SC5**) and *convex-like* shape constraints (**SC6**)–(**SC8**). Concave-like and convex-like shape constraints respectively bound the curvature of the demand curve from above and from below.

These shape constraints are not mutually disjoint. For example, it is well-known that concave demand curves are log-concave, and that log-convex demand curves are convex. In fact:

$$\text{(SC4)} \Longrightarrow \text{(SC3)} \begin{array}{l} \nearrow\!\!\!\!\nearrow \text{(SC1)} \\ \searrow\!\!\!\!\searrow \text{(SC2)} \end{array} \qquad \text{and} \qquad \text{(SC7)} \Longrightarrow \text{(SC6)}.$$

For reference, these relationships are proven below in Appendix C.2. In Appendix C.3 where we also provide examples of common demand curves that satisfy each shape constraint.

C.2 Relationships between Assumptions

(SC4) $\implies$ (SC3)

Proof. Given a concave demand curve $D(\cdot)$, suppose on the contrary that there exist $p_H > p_L$ such that

$$\frac{D'(p_H)}{D(p_H)} > \frac{D'(p_L)}{D(p_L)} \implies D(p_L)D'(p_H) > D(p_H)D'(p_L).$$

Since $D(\cdot)$ is concave, $D'(p_H) \leq D'(p_L)$; since $D(\cdot)$ is decreasing, $D'(\cdot) \leq 0$ and $D(p_L) \geq D(p_H)$. Thus

$$D(p_L)D'(p_H) \leq D(p_H)D'(p_H) \leq D(p_H)D'(p_L).$$

This is a contradiction. Hence $D(\cdot)$ is log-concave. □

(SC3) $\implies$ (SC1)

Proof. For any $p_H > p_L$, log-concavity implies that

$$\frac{D'(p_H)}{D(p_H)} \leq \frac{D'(p_L)}{D(p_L)} \implies \frac{p_H D'(p_H)}{D(p_H)} \leq \frac{p_L D'(p_H)}{D(p_H)} \leq \frac{p_L D'(p_L)}{D(p_L)}.$$

Here, we have used the fact that $D'(\cdot) \leq 0$ as $D(\cdot)$ is decreasing. Since the above inequalities hold for any $p_H > p_L$, it follows that $D(\cdot)$ satisfies Marshall's second law. □

(SC3) $\implies$ (SC2)

Proof. For any $p_H > p_L$, log-concavity implies that

$$\frac{D'(p_H)}{D(p_H)} \leq \frac{D'(p_L)}{D(p_L)} \implies p_H + \frac{D(p_H)}{D'(p_H)} \geq p_L + \frac{D(p_L)}{D'(p_L)}.$$

Since this holds for any $p_H > p_L$, it follows that $D(\cdot)$ has a decreasing marginal revenue curve. □

(SC7) $\implies$ (SC6)

Proof. For any $p_H > p_L$, log-convexity implies that

$$\frac{D'(p_H)}{D(p_H)} \geq \frac{D'(p_L)}{D(p_L)} \implies D(p_L)D'(p_H) \geq D(p_H)D'(p_L).$$

Since $D(\cdot)$ is decreasing, $D'(\cdot) \leq 0$ and $D(p_L) \geq D(p_H)$. Thus

$$D(p_H)D'(p_H) \geq D(p_L)D'(p_H) \geq D(p_H)D'(p_L) \implies D'(p_H) \geq D'(p_L).$$

Since this holds for any $p_H > p_L$, it follows that $D(\cdot)$ is convex. □

C.3 Common Demand Curves

We now review some common demand curves that satisfy these shape constraints.

(i) *Isoelastic demand curves.* Each isoelastic demand curve is parametrized by its elasticity $\varepsilon \leq 0$:

$$D(p) = q_0 \left(\frac{p}{p_0}\right)^{\varepsilon}.$$

Because elasticity is constant, it must also be trivially decreasing. Hence any isoelastic demand curve satisfies Marshall's second law (**SC1**).

(ii) *Constant marginal revenue demand curve.* Analogous to a CES demand curve, each constant marginal revenue demand curve is parametrized by its marginal revenue $0 \leq \mu < p_0$:

$$D(p) = \frac{q_0 \left(p_0 - \mu\right)}{p - \mu}.$$

Because marginal revenue is constant, it must also be trivially decreasing. Hence each constant marginal revenue demand curve exhibits decreasing marginal revenue (**SC2**).

(iii) *Exponential demand curves.* Each exponential demand curve is parametrized by $\lambda \geq 0$:

$$D(p) = q_0 \exp\left[-\lambda \left(p - p_0\right)\right].$$

Observe that the logarithm of any exponential demand curve is linear in p:

$$\log D(p) = \log q_0 - \lambda \left(p - p_0\right).$$

Hence each exponential demand curve is both log-concave (**SC3**) and log-convex (**SC7**).

(iv) *Linear demand curves.* Each linear demand curve is parametrized by $\lambda \geq 0$:

$$D(p) = q_0 - \lambda \left(p - p_0\right).$$

Each linear demand curve is both concave (**SC4**) and convex (**SC6**).

(v) *ρ-linear demand curves.* Each ρ-linear demand curve is parametrized by $\lambda \geq 0$:

$$D(p) = [q_0 - \lambda (p - p_0)]^{1/\rho} .$$

Each ρ-linear demand curve is both ρ-concave (**SC5**) and ρ-convex (**SC8**).

Appendix D Empirical Application Details

This appendix provides additional details on our empirical applications in Section 5.

D.1 Trade Tariffs

In this section, we provide the technical details behind our application to the deadweight loss of trade tariffs. To obtain the data for our exercise, we follow Amiti et al.'s (2019) data appendix to obtain a comprehensive dataset of products hit by new tariffs during 2018. Products are denoted by a ten-digit Harmonized Tariff Schedule (HTS10) product code and by country or origin. The dataset contains a unit quantity and total import value for each product, along with a tariff amount for each month in 2017 and 2018.

As the first step of our exercise, we replicate Amiti et al.'s log-log regression used to estimate the relationship between prices and quantities, assuming that prices change proportionally to tariffs within the same market (product-calendar month) between 2017 and 2018. Following Amiti et al., we estimate the regression:

$$\log\left(\frac{q_{ijt}}{q_{ij(t-12)}}\right) = \Delta\log(1+\tau_{ijt}) + \mathrm{FE}_i + \mathrm{FE}_j + \eta_{ijt}, \tag{9}$$

where i denotes an HTS10 product code, j denotes a country-year, t denotes a month and $\Delta\log(1+\tau_{ijt})$ denotes the log change in the relevant tariff. This yields the elasticity estimate $\hat{\varepsilon} = -5.89$ with standard error 0.59, as reported in column (3) of Table 1 in their paper. We then follow Amiti et al. in imputing the potential outcome for each q_0 based on an isoelastic curve:

$$\log(\hat{q}_{ij(t-12)}) = \log(q_{ijt}) - \hat{\varepsilon}\Delta\log((1+\tau_{ijt})). \tag{10}$$

To compute the deadweight loss under the linear, exponential, and isoelastic curves in Figure 6, we compute the change in consumer surplus directly using the formulas in Table 1 and subtract $q_1 \times (p_1 - p_0)$ as discussed in Appendix B.4.[20] In each case, we treat q_1, p_1 for each good using

[20] Note that Amiti et al. apply an additional approximation argument before imputing a linear demand curve for each market. As they explain in footnote 9 (pp. 199–200), they make use of a second Taylor approximation in computing deadweight loss:

$$-\log(m_1/m_0) \approx (m_0 - m_1)/m_1,$$

where m_t is the total import value of a product in year t. In general, it can be shown that this approximation will underestimate deadweight loss: $-\log z \leq \frac{1}{z} - 1$ for any $z \in \mathbb{R}$. As the magnitudes of the tariffs are substantial, we find that this approximation shrinks the deadweight loss estimates substantially and makes the comparison the calendar month in 2018 as period 1 and the same month in 2017 as period 0. We then impute $p_0 = \frac{p_1}{1+\tau}$ and $\hat{q}_0$ based on equation (10) and plug these values into the formulas directly. Following Amiti et al., we compute the deadweight loss for each market separately and then aggregate across all markets in our sample. To compute the deadweight loss under a ρ-linear demand curve, we integrate over the curve $D(p) = [q_0 - \lambda(p - p_0)]^{1/\rho}$ for each value of ρ. The formula for this is given by:

$$\text{DWL}_\rho = \frac{\rho\,(p_1 - p_0)\left(q_0^{1+\rho} - q_1^{1+\rho}\right)}{(1+\rho)\,(q_0^\rho - q_1^\rho)} - q_1\,(p_1 - p_0)\,.$$

To compute standard confidence bands, we apply the delta method with respect to the standard error in $\hat{q}_0$ due to $\hat{\varepsilon}$ as in the motivating example. Finally, to calculate the bounds with respect to elasticity relaxations in Figure 7, we compute the lower bound on the change in consumer surplus for the isoelastic benchmark, as in the third column of Table 1 and subtract $q_1 \times (p_1 - p_0)$. Because these bounds vary at large magnitudes for some values of r, the Taylor approximation assumed in the delta method may not apply. As such, we compute standard errors by bootstrapping over the distribution of $\hat{q}_0$.

D.2 Energy Subsidies

In this subsection, we provide a detailed derivation of our robustness exercise with respect to the welfare conclusion in Hahn and Metcalfe (2021). Following the description of the setting in Section 5.2,[21] the welfare effect of the CARE program is given by:

$$\Delta W = N_C \int_{q_C^*}^{q_C} \left[P_C(q) - \text{MSC}\right]\,\mathrm{d}q + N_N \int_{q_N^*}^{q_N} \left[P_N(q) - \text{MSC}\right]\,\mathrm{d}q - A.$$

Here N_C and N_N are the numbers of CARE and non-CARE consumers, respectively; $P_C(\cdot)$ and $P_N(\cdot)$ are their respective inverse demand curves; and A is the administrative cost of the program. We assume that q_N and q_C are observed and that $p^* = P_N(q_N) = P_C(q_C)$. Moreover, we use Hahn and Metcalfe's equation (5) to relate p^* to the observed prices p_N and q_N and the observed quantities q_N and q_C:

$$p^* = \frac{N_N p_N q_N + N_C p_C q_C - A}{N_N q_N + N_C q_C}.$$

across assumptions more difficult to interpret. As such, we skip this approximation step in our calculations and instead present the deadweight loss estimates from linear (and other) interpolations using just the quantities and prices produced in their first step.

[21] See also equation (A3) in Hahn and Metcalfe's online appendix.

Summarizing:

1. We observe p^* and the points (p_N, q_N) and (p_C, q_C).

2. Hahn and Metcalfe estimate $\hat{\varepsilon}_C$ and take $\hat{\varepsilon}_N$ from Auffhammer and Rubin (2018).

3. We wish to examine how robust ΔW is to the functional form assumptions imposed by Hahn and Metcalfe and Auffhammer and Rubin. We therefore introduce two parameters, r_C and r_N, and consider

$$\begin{cases} \dfrac{\hat{\varepsilon}_C}{1-r_C} \leq \varepsilon_C \leq (1-r_C)\,\hat{\varepsilon}_C, \\ \dfrac{\hat{\varepsilon}_N}{1-r_N} \leq \varepsilon_N \leq (1-r_N)\,\hat{\varepsilon}_N. \end{cases}$$

Note that because Hahn and Metcalfe assume linear demand, it would be natural to consider relaxations of gradient variability, rather than elasticity variability. This would allow us to test the robustness of their linear benchmark directly. However, Hahn and Metcalfe provide elasticity estimates, not gradients. As such, we must decide which benchmark to use: (i) a linear benchmark using gradients inferred from elasticity estimates through the linear function or (ii) an isoelastic benchmark using the estimated elasticities directly. For our application, we choose the latter option. Our reasoning is that an isoelastic benchmark prioritizes the decisions that the authors made with respect to their exposition of price treatment effects. Hahn and Metcalfe chose to present their results in terms of an elasticity; although they could have extrapolated to a linear curve directly, interpreting their LATE estimate as a gradient, they did not do so. This decision may reflect important considerations that we would like our robustness measures to preserve.[22]

The largest possible ΔW is attained when the welfare gains from CARE households are maximized and the welfare losses from non-CARE households are minimized. Symmetrically, the smallest possible ΔW is attained when the welfare gains from CARE households are minimized and the welfare losses from non-CARE households are maximized. Importantly, we can consider these welfare effects separately since the counterfactual quantities q_C^* and q_N^* are independent of each other (as they lie on separate demand curves). Notice also that the additional costs, MSC and A, do not change our earlier analysis. This is because: *(i)* instead of prices p^*, p_C, and p_N,

[22] For instance, the LATE estimator in Hahn and Metcalfe's example uses different baseline usage numbers for CARE consumers (22.9 therms/month) than their counterfactual welfare exercise (310 therms/year), which considers a different program duration and accounts for the full CARE consumer base. Using an elasticity estimate, which is unitless, may therefore reflect an intention to accommodate the difference in magnitudes between the two quantities.

we can perform our earlier analysis on *net* prices $p^* - \text{MSC}$, $p_C - \text{MSC}$, and $p_N - \text{MSC}$; and *(ii)* the administrative cost A is simply an additive constant.

CARE Households. Let the welfare gains for CARE households be denoted by

$$\Delta W_C = N_C \int_{q_C^*}^{q_C} \left[P_C(q) - \text{MSC}\right] \, \mathrm{d}q.$$

We want to find the largest possible values of welfare gains for CARE households, ΔW_C. We proceed in two steps: (1) we consider the problem for a given q_C^*; and (2) we optimize over the possible values of q_C^*. Throughout, we maintain the assumption that $\underline{\varepsilon} \leq \varepsilon(\cdot) \leq \overline{\varepsilon}$ on $p \in [p_C, p^*]$.

Step #1: Fixing q_C^*. For a given q_C^*, the upper bound of ΔW_C is given by:

$$\overline{\Delta W_C}(q_C^*) = N_C \max_{P \in \mathcal{P}} \int_{q_C^*}^{q_C} \left[P_C(q) - \text{MSC}\right] \, \mathrm{d}q.$$

Our previous results (cf. Theorem 1) imply that the extremal demand curves are 2-piecewise isoelastic, with elasticities equal to $\underline{\varepsilon}$ and $\overline{\varepsilon}$ and an average elasticity of $\log(q_C/q_C^*)/\log(p_C/p^*)$. The upper bound is attained by the inverse demand curve:

$$\overline{P_C}(q; q_C^*) = \begin{cases} p^* \left(\dfrac{q}{q_C^*}\right)^{1/\underline{\varepsilon}} & \text{for } q_C^* \leq q \leq \hat{q}, \\ p_C \left(\dfrac{q}{q_C}\right)^{1/\overline{\varepsilon}} & \text{for } \hat{q} \leq q \leq q_C, \end{cases} \quad \text{where } \hat{q} = \exp\left[\frac{\overline{\varepsilon}\underline{\varepsilon} \log\left(p_C/p^*\right) + \overline{\varepsilon} \log q_C^* - \underline{\varepsilon} \log q_C}{\overline{\varepsilon} - \underline{\varepsilon}}\right].$$

To obtain $\Delta W_C(q_C^*)$ at a given level of r_C, we can integrate $\overline{P_C}(q; q_C^*)$ over $q \in [q_C^*, q_C]$, substituting in $\overline{\varepsilon} = \hat{\varepsilon}_C \left(1 - r_C\right)$ and $\underline{\varepsilon} = \hat{\varepsilon}_C / \left(1 - r_C\right)$.

Step #2: Optimizing over q_C^*. To optimize over q_C^*, we first determine the range of values $[\underline{q_C^*}, \overline{q_C^*}]$ that q_C^* can take, given p^*, p_C, q_C, $\hat{\varepsilon}_C$, and r_C:

$$\begin{cases} \overline{q_C^*} = q_C \left(\dfrac{p^*}{p_C}\right)^{\hat{\varepsilon}_C/(1-r_C)}, \\ \underline{q_C^*} = q_C \left(\dfrac{p^*}{p_C}\right)^{\hat{\varepsilon}_C(1-r_C)}. \end{cases}$$

The largest value of welfare losses for non-CARE households is attained by maximizing $\overline{\Delta W_C}(q_C^*)$ over $q_C^* \in [\underline{q_C^*}, \overline{q_C^*}]$. Notice, however, that the objective function is concave in q_C^*; hence $q_C^* = \underline{q_C^*}$ or $q_C^* = \overline{q_C^*}$. The largest value of welfare gains for CARE households is therefore attained by maximizing $\underline{\Delta W_C}(q_C^*)$ over $q_C^* \in [\underline{q_C^*}, \overline{q_C^*}]$. For each value of r_C, we can find the maximizing value of $\underline{\Delta W_C}(q_C^*)$ through a standard numerical optimization procedure (e.g., through a grid search).

Non-CARE Households. Let the welfare losses for non-CARE households be denoted by

$$\Delta W_N = N_N \int_{q_N^*}^{q_N} \left[P_N(q) - \text{MSC}\right] \, \mathrm{d}q.$$

We want to find the smallest possible values of welfare losses for non-CARE households, ΔW_N. We proceed in two steps: (1) we consider the problem for a given q_N^*; and (2) we optimize over the possible values of q_N^*. Throughout, we maintain the assumption that $\underline{\varepsilon} \leq \varepsilon(\cdot) \leq \overline{\varepsilon}$ on $p \in [p_C, p^*]$.

Step #1: Fixing q_N^*. For a given q_N^*, the lower bound of ΔW_N is given by

$$\underline{\Delta W_N}(q_N^*) = N_N \min_{P \in \mathcal{P}} \int_{q_N}^{q_N^*} \left[P_N(q) - \text{MSC}\right] \, \mathrm{d}q.$$

Again, our previous results (cf. Theorem 1) imply that the extremal demand curves are 2-piecewise isoelastic, with elasticities equal to $\underline{\varepsilon}$ and $\overline{\varepsilon}$ and an average elasticity of $\log(q_N/q_N^*)/\log(p_N/p^*)$. The lower bound is attained by the inverse demand curve:

$$\underline{P_N}(q; q_N^*) = \begin{cases} p_N \left(\dfrac{q}{q_N}\right)^{1/\overline{\varepsilon}} & \text{for } q_N \leq q \leq \hat{q}, \\ p^* \left(\dfrac{q}{q_N^*}\right)^{1/\underline{\varepsilon}} & \text{for } \hat{q} \leq q \leq q_N^*, \end{cases} \quad \text{where } \hat{q} = \exp\left[\frac{\overline{\varepsilon}\underline{\varepsilon} \log\left(p_N/p^*\right) - \underline{\varepsilon} \log q_N + \overline{\varepsilon} \log q_N^*}{\overline{\varepsilon} - \underline{\varepsilon}}\right].$$

As in the CARE case, we can obtain $\Delta W_N(q_N^*)$ by integrating $\underline{P_N}(q; q_N^*)$ over $q \in [q_N, q_N^*]$ and substituting in $\overline{\varepsilon} = \hat{\varepsilon}_N \left(1 - r_N\right)$ and $\underline{\varepsilon} = \hat{\varepsilon}_N / \left(1 - r_N\right)$.

Step #2: Optimizing over q_N^*. To optimize over q_N^*, we first determine the range of values $[\underline{q_N^*}, \overline{q_N^*}]$ that q_N^* can take, given p^*, p_N, q_N, $\hat{\varepsilon}_N$, and r_N:

$$\begin{cases} \overline{q_N^*} = q_N \left(\frac{p^*}{p_N}\right)^{\hat{\varepsilon}_N/(1-r_N)}, \\ \underline{q_N^*} = q_N \left(\frac{p^*}{p_N}\right)^{\hat{\varepsilon}_N(1-r_N)}. \end{cases}$$

The largest value of welfare losses for non-CARE households is attained by maximizing $\overline{\Delta W_N}(q_N^*)$ over $q_N^* \in [\underline{q_N^*}, \overline{q_N^*}]$. Notice, however, that the objective function is convex in q_N^*; hence $q_N^* = \underline{q_N^*}$ or $q_N^* = \overline{q_N^*}$. The smallest value of welfare losses for non-CARE households is there attained by minimizing $\underline{\Delta W_N}(q_N^*)$ over $q_N^* \in [\underline{q_N^*}, \overline{q_N^*}]$, and we can solve for the lower bound at each r_N through a standard bounded numerical optimization procedure like grid search, as well.

Combining Bounds for Analysis. In order to create Figure 9, we compute the upper bound on welfare gains for CARE consumers $\overline{\Delta W_C}$ and the lower bound of welfare losses for non-CARE consumers $\underline{\Delta W_N}$ for each pair of indices $(r_C, r_N) \in [0,1] \times [0,1]$. We then plot $\Delta W = N_C \cdot \overline{\Delta W_C} + N_N \cdot \underline{\Delta W_N} - A$. In each case, we use the numbers from Online Appendix B in Hahn and Metcalfe: $N_N = 3.85$M, $N_C = 1.6$M, $q_N = 490, q_C = 310, p_N = 0.95, p_C = 0.75$, and $p^* = 0.90$.

D.3 Old-Age Pensions

In this subsection, we provide the technical details behind our application to the MVPF of the Old-Age Pensions Act based on Giesecke and Jäger (2021). The key empirical result underlying our analysis is the marginal propensity to retire early, estimated by Giesecke and Jäger through a regression discontinuity design. As a first step to our analysis, we derive a micro-foundation for interpreting this as a casual response to an increase in the pension amount through the supply curves of the eligible population.

Neoclassical Labor-Leisure Model. We begin by considering the neoclassical labor-leisure model. Suppose that an individual i has utility over C and L, where C is consumption of goods (measured in dollars) and L is hours of leisure. We assume that utility is quasilinear with respect to consumption (as in Diamond, 1998):

$$u_i(L) + C.$$

The individual's budget constraint is

$$C \leq w_i (T_i - L) + V_i,$$

where T_i is total hours available, w_i is the wage rate, and V_i is other income for that individual.

To allow for the possibility of retiring in exchange for a pension, we augment this model by assuming that each individual can choose either to work ($y_i = 1$) or not ($y_i = 0$), but (for simplicity) cannot choose how much time they work. Therefore, if an individual chooses to work ($y_i = 1$), they work for $T_i - L_i$ hours. However, if an individual chooses not to work ($y_i = 0$), they receive a pension p. This changes the individual's budget constraint:

$$C \leq \begin{cases} w_i (T_i - L_i) + V_i & \text{if } y_i = 1, \\ p + V_i & \text{if } y_i = 0. \end{cases}$$

Summarizing, each individual faces the utility maximization problem:

$$\max \{u_i(L_i) + w_i (T_i - L_i) + V_i, u_i(T_i) + p + V_i\}.$$

Labor Supply. Under this model, an individual chooses to retire ($y_i = 0$) if and only if

$$u_i(T_i) + p \geq u_i(L_i) + w_i (T_i - L_i) \iff p \geq \underbrace{u_i(L_i) - u_i(T_i) + w_i (T_i - L_i)}_{=:\varepsilon_i}.$$

Let the aggregate distribution of ε_i in the population be denoted by F. In theory, a fraction q_t of people retires when the pension is p_t, where

$$q_t = \mathbf{E}\left[\mathbf{1}_{p_t \geq \varepsilon_i}\right] = F(p_t).$$

We interpret F as a supply curve for retirement: ε_i is the pension that individual i would have to be paid in order for him to be indifferent to retiring.

Welfare Impact of a Pension Increase. Suppose that the pension increases from p_0 to p_1. The change in each individual's surplus is then given by

$$\begin{aligned} \Delta W_i &= \max \{u_i(L_i) + w_i (T_i - L_i), u_i(T_i) + p_1\} - \max \{u_i(L_i) + w_i (T_i - L_i), u_i(T_i) + p_0\} \\ &= \max \{0, p_1 - \varepsilon_i\} - \max \{0, p_0 - \varepsilon_i\}. \end{aligned}$$

Integrating over the population with Q_0 individuals yields

$$\begin{aligned}\Delta W &= Q_0 \int_{-\infty}^{\infty} \left[\max\{0, p_1 - \varepsilon_i\} - \max\{0, p_0 - \varepsilon_i\}\right] \, \mathrm{d}F(\varepsilon_i) \\ &= Q_0 \int_{-\infty}^{p_1} (p_1 - \varepsilon_i) \, \mathrm{d}F(\varepsilon_i) - \int_{-\infty}^{p_0} (p_0 - \varepsilon_i) \, \mathrm{d}F(\varepsilon_i) \\ &= Q_0 \int_{-\infty}^{p_1} F(\varepsilon_i) \, \mathrm{d}\varepsilon_i - \int_{-\infty}^{p_0} F(\varepsilon_i) \, \mathrm{d}\varepsilon_i \implies \Delta W = Q_0 \int_{p_0}^{p_1} F(\varepsilon_i) \, \mathrm{d}\varepsilon_i.\end{aligned}$$

Robustness With Respect to Gradients. Our analysis builds on Giesecke and Jäger's baseline result (Section 4.1) that labor supply dropped by 6 percentage points, from 46% to 40%, at the eligibility cutoff age upon the introduction of old-age pensions in the U.K. Mapping this result to our framework, these estimates correspond to measurements of $F(\cdot)$ at two points, namely, $q_0 = \hat{F}(0s) = 0.54$ and $q_1 = \hat{F}(260s) = 0.60$. Because Q_0 is a constant (and cancels out in the MVPF calculation), we focus only on the supply curve F.

The authors include a welfare analysis in their online appendix, in which they take the extreme stance that any worker who was willing to retire at the observed pension would have been willing to retire at *any* non-zero pension. This assumption corresponds to the lower bound of any welfare measure in our framework at the limiting measure of variability (e.g., $r = 1$). As such, we conduct our robustness analysis with respect to a benchmark that is calibrated to their empirical exercise.

Because the treatment effects estimated by the regression discontinuity design are in levels-space, we focus on a linear benchmark and consider variability in gradients. Theorem 1 allows us to derive the upper and lower bounds at each r; hence we can apply the formulas from Table 1 to compute the bounds. Finally, to compute the MVPF at each r, we follow Giesecke and Jäger's Online Appendix D and divide the welfare gain at r by 1.13 to account for the net government cost of supplying each pension. To obtain confidence bands, we apply the delta method to each MVPF bound with respect to the standard error on the 6 percentage point treatment effect estimate as provided by Giesecke and Jäger.