

Counterfactual and Welfare Analysis with an Approximate Model*

Roy Allen

Department of Economics

University of Western Ontario

rallen46@uwo.ca

John Rehbeck

Department of Economics

The Ohio State University

rehbeck.7@osu.edu

October 1, 2020

Abstract

We propose a conceptual framework for counterfactual and welfare analysis for approximate models. Our key assumption is that model approximation error is the same magnitude at new choices as the observed data. Applying the framework to quasilinear utility, we obtain bounds on quantities at new prices using an approximate law of demand. We then bound utility differences between bundles and welfare differences between prices. All bounds are computable as linear programs. We provide detailed analytical results describing how the data map to the bounds including continuity and other shape restrictions. An application to gasoline demand illustrates the methodology.

*We thank Victor Aguiar, Victor Aguirregabiria, Lars Hansen, Nail Kashaev, Lance Lochner, Nirav Mehta, Magne Mogstad, Ismael Mourifié, Salvador Navarro, Andres Santos, and participants at the Banff Empirical Microeconomics Workshop, the Western Conference on Counterfactuals with Economic Restrictions, the University of Chicago, and the University of Toronto for helpful comments.

1 Introduction

Models are generally viewed as approximations. A common intuition in empirical work is that conclusions of a model are robust to “small” amounts of approximation error. Unfortunately, this intuition does not apply to many standard frameworks. For example, when performing a revealed preference analysis [Varian, 1982], if a dataset is inconsistent with a model, then counterfactual predictions described by certain inequalities cross. Thus, if there is *any* violation of the model (no matter how “small”), then the model fails to generate coherent counterfactual or welfare statements.¹

Alternatively, one can formally acknowledge approximation error throughout the analysis. Rather than treat approximation error as nonexistent, one can place restrictions on the magnitude of the approximation error. This paper does so for counterfactual and welfare analysis, with the following assumption on this magnitude.

Assumption 1. *When making counterfactual predictions or measuring welfare changes, we assume the approximation error of the model on the counterfactual predictions is the same as the approximation error of the model on the observed dataset.*

This assumption is a natural extension of the standard approach to generate counterfactual predictions that assumes that both the observed data and counterfactual predictions are consistent with a model. We present a framework in which a model can be used even though it is not exactly consistent with observed data. In particular, this paper assumes that the approximation error of the model on observed and unobserved situations has the same magnitude. We call the counterfactuals that are consistent with Assumption 1 *adaptive counterfactuals* because they adapt to approximation error present in the observed dataset. This assumption can be questioned, especially when the counterfactual setting is significantly different than observed data, yet provides a way to conduct counterfactual analysis taking approximation error seriously.

The conceptual framework of this paper is general and can be applied to different settings. In this paper, we formalize how to generate counterfactual predictions and

¹Related concerns have been raised in the econometric literature on partial identification [Ponomareva and Tamer, 2011, Müller and Norets, 2016].

measure welfare changes for the quasilinear utility model using the notion of approximation error from Allen and Rehbeck [2020]. In particular, we present bounds on counterfactual quantities at new prices, differences of utility over consumption bundles, and welfare differences involving a price change. The computation of all bounds is facilitated by linear programming and we demonstrate the methods in an illustrative empirical example using gasoline demand data from Blundell et al. [2012].

The quasilinear model is a suitable setting in which to study approximation error because it is implicitly viewed as an approximation.² The most common criticism is that the model does not allow income effects. In addition, the model neglects dynamics, limited consideration, and peer effects, among many other omitted features. We present a framework in which one does not need to pick a single story why the baseline model is imperfect when generating counterfactual predictions on quantities. However, to make welfare comparisons we take a stand on the interpretation of the approximation error: the individual ranks bundles according to a quasilinear utility function but for reasons we do not model explicitly, the choices do not exactly maximize the function. Overall, the framework we propose permits many reasons why the baseline model is wrong, provided the approximation error is the same magnitude in the counterfactual setting.

Despite often being viewed as an approximation, the quasilinear model is widely used. Examples include work in insurance choice [Einav et al., 2010, Bundorf et al., 2012, Tebaldi et al., 2018] and public health [Cohen et al., 2010]. In addition, the quasilinear structure is closely related to a large class of latent utility models (e.g. McFadden [1981], Allen and Rehbeck [2019a]), and so the insights of this paper are directly relevant beyond a setting with just prices and quantities. In particular, many latent utility models used in applied work involve characteristics other than prices that shift the desirability of goods but not the budget constraint.³

We now describe the framework in more detail. We begin by studying counterfactual bounds for the demand of goods at new prices. We construct the counterfactual bounds by looking for the maximal and minimal demand for each good in the pres-

²A notable exception that studies the approximation error explicitly is Willig [1976].

³Our analysis is also relevant for specifications in which latent utilities depend on a nonlinear function of prices. For example, Berry et al. [1995] specifies that the utility of alternative j depends on several observables including a term $\ln(m - p_j)$ where p_j is the price of good j and m is income. This is a quasilinear model in the variable $\tilde{p}_j = \ln(m - p_j)$.

ence of approximation error. We assume approximation error does not increase for the counterfactual predictions relative to the dataset we have seen. This leads to nontrivial restrictions on demand at new prices as long as the prices are not too low. For each counterfactual price, this procedure gives an interval for lower and upper demands for a given good.

For welfare analysis we view approximation error with a specific interpretation. In particular, we assume an individual cannot perfectly maximize their utility function because they are satisficers in the spirit of Simon [1947]. This means an individual chooses quantities that are “close” to the maximum utility possible, but not necessarily optimal. The interpretation of satisficing allows us to assign special welfare significance to the latent utility over bundles. This utility over consumption bundles is key for policy decisions involving the allocation of goods. In addition, we study welfare over price changes, which is key for tax policy and other policy thought to affect prices.

Two features limit the ability to measure welfare changes using data. First, the choices we see do not exactly maximize utility, which leads to a “measurement wedge” on the underlying utility function over bundles. Second, even if we knew the utility function over bundles exactly, we do not know which approximately-optimizing choices will be made at new prices. This leads to an additional “prediction wedge” when bounding utility differences obtained at different *prices*.⁴

Taking into account these wedges, we present bounds on differences in utility over consumption bundles and robust consumer surplus bounds involving price changes. We present computational results for both bounds, as well as analytical results designed to interpret specifically how the data are used to measure welfare changes. These bounds generalize existing work in several directions: first, and most importantly, they are valid with approximation error; second, they apply to finite datasets rather than requiring demand functions; third, the bounds apply to the (approximate) indirect utility at a new price without needing to first bound the quantity at that price. In particular, we show that the bounds on (approximate) indirect utility is a generalization of the standard integral definition of consumer surplus and we

⁴We thus complement the core analysis of Bernheim and Rangel [2009], which focuses on recovering ordinal information on preferences over *consumption bundles* and does not distinguish between these wedges.

establish a close connection between counterfactual quantities and welfare bounds.

In order to understand how the counterfactual/welfare bounds depend on data and the counterfactual price, we present several shape restrictions. We are not aware of any work in the tradition of Varian [1982] that discusses shape restrictions *of* the counterfactual/welfare bounds (viewed as functions of data or counterfactual prices).⁵ Studying these shape restrictions is important to understand how data are used in an empirical analysis.

For counterfactual quantities, we find that when there is a single good and the counterfactual price changes, the upper and lower bounds on counterfactual demands are each weakly decreasing in price. Thus, the upper and lower bounds on demand are functions that satisfy the law of demand. For utility differences, we establish monotonicity and continuity properties of the bounds as the quantities being compared change. For a price change, the approximate indirect utility bounds satisfy convexity and monotonicity conditions as prices vary; these are also key shape restrictions for the indirect utility function for quasilinear utility.⁶ Finally, we establish several convexity properties describing how quantities data map to the bounds. To our knowledge these results are all new even under correct specification for quasilinear utility.⁷

Our analysis also establishes that the bounds satisfy a key continuity property: as the degree of approximation error limits to 0, our analysis limits to the analysis under correct specification. In fact, we show a stronger property that the counterfactual and approximate indirect utility bounds are *jointly* continuous when viewed as a function of the quantities in the data and the degree of approximation error. This facilitates plug-in estimation of the bounds in which we replace true quantities with estimated quantities. This is needed to cover our empirical application in which we apply the framework with data on gasoline purchases used previously in Blundell

⁵There is a large literature using shape restrictions to produce bounds in new settings. We mention work in revealed preference below, but there is also work in other literatures such as discrete choice [Manski, 2007] and program evaluation [Mogstad et al., 2018]. Here we clarify that we study shape restrictions on the bounds themselves.

⁶More specifically, convexity holds for the money metric version of the utility function. In general one can only obtain quasiconvexity.

⁷The closest work appears to be a computational approach to describing bounds for models related to quasilinear utility, without describing detailed shape restrictions [Chiong et al., 2017, Tebaldi et al., 2018, Allen and Rehbeck, 2019a].

et al. [2012]. The data is a single cross section, and we pre-process the data as in Blundell et al. [2012] by kernel smoothing. We conduct a representative agent analysis with quantities (conditional means) estimated from the kernel-smoothed data. Like Blundell et al. [2012], we find that for several natural choices of the bandwidth, demand is not downward sloping. Thus, it is inconsistent with the exact quasilinear model. Nonetheless, the minimal degree of approximation error need to describe the kernel-smoothed data is small and welfare bounds are surprisingly narrow for all bandwidths we consider. In contrast, the informativeness of the counterfactual bounds depends on the bandwidth.

The rest of this paper is organized as follows. After reviewing the related literature, Section 2 describes the setup and conceptual framework of approximate counterfactuals for the quasilinear framework. Section 3 studies counterfactuals. Section 4 studies welfare. Section 5 presents additional shape restrictions and discusses plug-in estimation. Section 6 contains the application to gasoline demand. Section 7 concludes.

1.1 Literature Review

This paper is part of a long literature that uses the revealed preference approach to do counterfactual and welfare analysis. The primary model used is the general model of utility maximization subject to a budget constraint, whose empirical content has been characterized in Afriat [1967], Diewert [1973], and Varian [1982]. Recent econometric work considering counterfactual or welfare bounds includes Blundell et al. [2003], Blundell et al. [2008], Blundell et al. [2012], Blundell et al. [2014], Hoderlein and Stoye [2015], Kline and Tartari [2016], Blundell et al. [2017], Cosaert and Demuynck [2018], Aguiar and Kashaev [2018], Adams [2019], Cherchye et al. [2019], and Kitamura and Stoye [2019]. Several proposals have been made to assess the fit of a model using revealed preference tools, including Afriat [1973], Houtman and Maks [1985], Varian [1990], and Echenique, Lee, and Shum [2011].⁸ Other papers outside of the revealed preference literature that discuss fit of an approximate model include Kydland and Prescott [1982], Vuong [1989], and Hansen and Jagannathan [1997]. The primary way in which we differ from existing work is that we *use* a measure of fit to adjust bounds on counterfactuals and welfare. In addition, relative to the general model

⁸See Allen and Rehbeck [2019b] for additional references and discussion of units.

with income effects, which has been the focus of the revealed preference literature, we conduct counterfactual analysis fixing prices at a new value without also fixing expenditure.⁹

A growing econometric literature has studied sensitivity analysis and other ways in which a model can be used formally viewing it as an approximation. Examples include Imbens [2003], Conley et al. [2012], Kline and Santos [2013], Andrews et al. [2017], Manski and Pepper [2018], Masten and Poirier [2018a, 2019], Armstrong and Kolesár [2018], Bonhomme and Weidner [2018], Christensen and Connault [2018], d'Haultfoeuille et al. [2018], Fessler and Kasy [2019], Salanié and Wolak [2019], and Andrews and Kwon [2019]. See Masten and Poirier [2018b] for additional references and discussion. We differ from this work by focusing on a notion of approximation derived from approximate optimization. We complement the robustness approach [Hansen and Sargent, 2008] by focusing on the theoretical bounds for welfare/counterfactuals rather than focusing on an optimal decision.

This paper is naturally related to other frameworks for welfare analysis that go beyond the classic revealed preference tradition. Our approach differs substantially from an approach that posits individuals have consistent choices (that can be modeled as solutions to a “decision utility”), but whose consistent choices do not reveal the “true utility.” Subtleties with this approach have been discussed by Bernheim and Rangel [2009] among others. A framework of “behavioral welfare analysis” is presented in Bernheim [2016] and Bernheim and Taubinsky [2018], which provide further summaries of what has become a large literature. Broadly, we differ from welfare proposals in behavioral economics by addressing welfare questions given data (only) on prices and quantities, without observing decision frames or other variables thought to alter the choice process. In addition, when conducting welfare analysis concerning price changes, we introduce the “prediction wedge” because given prices, we do not know precisely what an individual would choose, even if utility were known *a priori*.

⁹Work that uses revealed preference techniques to go beyond measuring the fit of the model includes Varian [1990], Halevy et al. [2018], and Gauthier [2019], which study parameter recoverability. See also Chetty [2012] for a related approach.

2 Framework and Setting

The goal of this paper is to provide a framework where a researcher begins with a baseline model that is taken seriously as an approximation. Since the model is an approximation, the researcher does not expect *all* data to be consistent with the model. Nonetheless, the researcher may want to use the model for counterfactual and welfare analysis. We present an adaptive framework framework for this by operationalizing Assumption 1 for the quasilinear utility model. Recall that Assumption 1 maintains that a researcher considers counterfactuals that are “no worse” than the observed data. To do this, we enlarge the baseline model to fit the data, which introduces a measurement wedge and prediction wedge. The *measurement wedge* concerns limits on what an analyst can learn due to approximation error for objects defined in the existing dataset (e.g. utility functions). The *prediction wedge* describes limits on what can be said in new settings where the model may not be perfect (e.g. counterfactual quantities). We formalize Assumption 1 by making these wedges as small as possible while still fitting the observed data, using the notion of approximation error from Allen and Rehbeck [2020]. We describe the framework more below.

We formalize the *baseline model* of quasilinear utility. A consumption bundle $(x, y) \in \mathbb{R}_+^K \times \mathbb{R}$ is evaluated according to $u(x) + y$, where $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ is a utility function over bundles x . The numeraire good is given by y and has a price of one. Given prices $p \in \mathbb{R}_{++}^K$ and income $I \in \mathbb{R}$ decisions in a quasilinear utility model follow

$$\begin{aligned} \max_{x \in \mathbb{R}_+^K, y \in \mathbb{R}} \quad & u(x) + y \quad \iff \quad \max_{x \in \mathbb{R}_+^K} \quad u(x) + I - p \cdot x \\ \text{s.t. } & p \cdot x + y \leq I \end{aligned}$$

where consumption of the numeraire good is allowed to be negative for unobserved borrowing.¹⁰ We study the quasilinear utility model since it is regularly used in applications described in the Introduction, including adaptations to handle non-price characteristics.¹¹ In addition, it has a tractable notion of welfare in terms of units of the numeraire.

¹⁰Allowing negative expenditure avoids boundary issues for chosen consumption bundles.

¹¹Allen and Rehbeck [2019a] show many applications including the additive random utility model [McFadden, 1981] are quasilinear models with utility indices playing the role of prices.

Now we present an *enlargement* of the baseline model that relaxes the assumption of exact maximization to a notion of approximate optimization. Because we focus on the empirical analysis, we define the enlargement in terms of a finite datasets of the form $\{(x^t, p^t)\}_{t=1}^T$.¹² There are T observations, quantities are weakly positive $x^t \in \mathbb{R}_+^K$, and prices are strictly positive $p^t \in \mathbb{R}_{++}^K$. Importantly, quantities can be discrete or continuous, and 0 quantities are permitted in this framework.

Definition 1. A dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε -rationalized by quasilinear utility for $\varepsilon \geq 0$ if there exists a utility function $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ such that for all $t \in \{1, \dots, T\}$ and for all $x \in \mathbb{R}_+^K$, the following inequality holds:

$$u(x^t) - p^t \cdot x^t \geq u(x) - p^t \cdot x - \varepsilon.$$

When ε equals zero, we say the dataset is quasilinear rationalized.

The value ε is in the same units as the price of the numeraire good, e.g. dollars per time period. When $\varepsilon > 0$, the observed bundles are within ε dollars of the maximum utility possible at a given price. We treat ε as a single value here for all observations, though in principle it can depend on other variables. One interpretation is that ε captures “unstructured” deviations from the quasilinear utility model (cf. Chetty [2012], Hansen and Sargent [2018]), without a single interpretation of the nature of the deviations. Instead, the *magnitude* of the deviations is controlled. This interpretation can be pursued when making counterfactual predictions. In contrast, to make welfare predictions an interpretation of the model is crucial. When discussing welfare, we follow Allen and Rehbeck [2020] and interpret the value ε as a level of satisficing in the spirit of Simon [1947]. In this case, a higher value of ε means there is a larger set of consumption bundles that are “good enough” to be chosen.

When using the model for counterfactual or welfare analysis, a *measurement wedge* and a *prediction wedge* arise. These concepts will become more clear when we turn to specific analysis below, but we first provide an overview. When observed data is not exactly consistent with a quasilinear utility model, the econometrician knows

¹²Here we follow the revealed preference literature, where a finite dataset is the primitive empirical object. In econometric analysis, the data generating process may instead be the primitive empirical object. For interpretation in econometric analysis, interpret (x^t, p^t) as an identifiable structural quantity-price tuple, identified over the finite set $t \in \{1, \dots, T\}$. For example, x^t could be mean structural quantity for a population. See Section 5 for analysis when structural quantities are estimated.

there is no utility function that rationalizes the entire dataset, so at some observation the quantity is not optimal. Thus, if a researcher still wants to use the quasilinear model even when data is inconsistent with the baseline quasilinear model, then there is a *measurement wedge* when trying to recover information about candidate utility functions and indirect utility. Second, even after the econometrician has a set of candidate utility functions that match the original dataset, the econometrician cannot know that counterfactual choices will exactly maximize a candidate utility function. Thus, there is a *prediction wedge* when forecasting even after recovering information on the utility function.

We now discuss how to formalize Assumption 1 for the approximate quasilinear utility model in relation to the measurement wedge and prediction wedge. Let ε_M denote the measurement wedge and ε_P denote the prediction wedge. In principle, the two wedges may not be the same, but Assumption 1 allows us to treat these as equal.¹³ To further formalize Assumption 1, we introduce ε^* as the smallest value of ε such that the dataset is ε^* -rationalized by quasilinear utility. We also refer to ε^* as the level of approximation error or approximation error of the quasilinear utility model for the observed dataset.

Proposition 1 (Allen and Rehbeck [2020]). *Let $\varepsilon^* \geq 0$ be the smallest value such that for all $\varepsilon \geq \varepsilon^*$ the dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε -rationalized by quasilinear utility. The value ε^* exists and is obtained by a linear program.*

We note that ε^* is a function of the dataset to a number, so for a dataset $D = \{(x^t, p^t)\}_{t=1}^T$ we can write $\varepsilon^*(D)$. When we discuss only a single dataset, we typically drop dependence on D . The value ε^* will be used in our framework to place restrictions on the magnitude of the measurement and prediction wedges. Setting $\varepsilon_M \geq \varepsilon^*$ formalizes that the measurement wedge is large enough to explain the data we have seen. Similarly, setting $\varepsilon_P \geq \varepsilon^*$ formalizes that the model is no better at predicting in new settings than the data we have seen. We make these bounds as tight as possible, and formalize Assumption 1 for this setting as follows.

Assumption 1'. *When performing counterfactual analysis, the measurement wedge,*

¹³See Appendix B for additional discussion on this case and a more formal treatment of the measurement and prediction wedges.

prediction wedge, and approximation error of the model are equal,

$$\varepsilon_M = \varepsilon_P = \varepsilon^*.$$

This is a direct generalization of the standard approach to counterfactual and welfare analysis, which sets $\varepsilon_M = \varepsilon_P = 0$. The conceptual framework of the standard approach only applies to models that perfectly fit the data, which translates to $\varepsilon^* = 0$ here. We later develop a framework for counterfactual and welfare analysis when the measurement and prediction wedges are equal, $\varepsilon_M = \varepsilon_P$. For notational convenience, we will let ε denote the common value. Assumption 1' is the special case where $\varepsilon = \varepsilon^*$.

3 Counterfactuals

For the quasilinear utility model with approximation error that does not exceed ε , we consider sharp counterfactual bounds. More formally, we consider when the measurement and prediction wedges are both equal to a single value ε . We impose $\varepsilon = \varepsilon^*$ to implement Assumption 1' and describe some properties on the counterfactual bounds. In Section 3.1, we provide graphical intuition for the bounds. In Section 3.2, we describe how to compute bounds on quantities fixing a price. In Section 3.3 we describe additional restrictions that can be imposed to tighten the bounds, such as *a priori* bounds on expenditure at a new price.

To that end, for notational convenience, let $D = \{(x^t, p^t)\}_{t=1}^T$ denote the observed dataset. Suppose we have a candidate quantity-price tuple (\tilde{x}, \tilde{p}) . We can add this to the original dataset to form an augmented dataset $D \cup (\tilde{x}, \tilde{p})$. We consider candidates such that the approximation error of the augmented dataset is bounded by ε . In particular, the set of *consistent demands and prices* for the level of approximation error ε is given by

$$C(D, \varepsilon) = \{(\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid \varepsilon^*(D \cup (\tilde{x}, \tilde{p})) \leq \varepsilon\}.$$

To check whether a candidate tuple (\tilde{x}, \tilde{p}) is in $C(D, \varepsilon)$, one can calculate ε^* for the augmented dataset using Proposition 1. If this measure of approximation error for the augmented dataset is below ε , then the candidate tuple is in the set $C(D, \varepsilon)$.

Our framework imposes Assumption 1' to generate counterfactual predictions assuming the level of approximation error does not get worse. This amounts to setting ε equal to the approximation error of the observed dataset. In particular, we focus on the *adaptive counterfactual set*

$$\begin{aligned} AC(D) &= \{(\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid \varepsilon^*(D \cup (\tilde{x}, \tilde{p})) \leq \varepsilon^*(D)\} \\ &= C(D, \varepsilon^*(D)). \end{aligned}$$

We collect some facts about $AC(\cdot)$ and $C(\cdot)$.

Fact 1 (Constant Approximation Error). *For any $(\tilde{x}, \tilde{p}) \in AC(D)$, we have*

$$\varepsilon^*(D \cup (\tilde{x}, \tilde{p})) = \varepsilon^*(D).$$

Thus, when a candidate observation in $AC(D)$ is added to D , the measure of approximation stays the same. This follows from the construction of $AC(D)$. This equality does not hold for all measures of model approximation error. For example, if we had chosen to take ε^* divided by the number of observations T as the measure of approximation error, then Fact 1 would not hold in general since the measure of approximation error for the augmented dataset would divide by $T + 1$.

Fact 2 (Monotonicity). *If $\varepsilon < \varepsilon'$, then $C(D, \varepsilon) \subseteq C(D, \varepsilon')$.*

Higher values of ε correspond to less informative counterfactual predictions. This follows from the fact that if a dataset is ε -rationalized by quasilinear utility, then it is also ε' -rationalized for $\varepsilon < \varepsilon'$.

Fact 3 (Nonemptiness). *$C(D, \varepsilon)$ is nonempty if and only if $\varepsilon \geq \varepsilon^*(D)$.*

This states that the observed data places a lower bound on the minimal amount of approximation error needed to conduct counterfactual analysis. If $\varepsilon \geq \varepsilon^*$, nonemptiness of $C(D, \varepsilon)$ is guaranteed by considering \tilde{p} sufficiently high along each dimension and $\tilde{x} = 0$. Alternatively, for the dataset D , when $\varepsilon < \varepsilon^*$ even observations within the dataset cannot be ε -quasilinear rationalized.

Fact 4 (Minimality). *$AC(D)$ is obtained from the smallest ε such that $C(D, \varepsilon)$ is nonempty.*

This formalizes that setting $\varepsilon = \varepsilon^*$ for counterfactual values obtains the sharpest

restrictions under Assumption 1' subject to the constraint that counterfactuals are nonempty. This follows from the previous facts. To perform a sensitivity analysis, one could examine any $\varepsilon > \varepsilon^*$ and use $C(D, \varepsilon)$ as the counterfactual set. Our framework allows this yet focuses on $\varepsilon = \varepsilon^*$.

3.1 Approximate Law of Demand

To gain intuition on the “shape” of the counterfactual sets $C(D, \varepsilon)$ and $AC(D)$, we present a graphical description of the restrictions on counterfactuals. For exposition we focus on *some* of the restrictions rather than all of them. First we describe a restriction that must hold for a dataset to be ε -rationalized. At price p^r we must have

$$u(x^r) - p^r \cdot x^r \geq u(x^s) - p^r \cdot x^s - \varepsilon$$

for some unknown function u . This states that x^s cannot be much better than x^r at price p^r . Flipping the role of observations r and s and basic algebra yields

$$\frac{1}{2}(p^s - p^r) \cdot (x^s - x^r) \leq \varepsilon. \quad (1)$$

This is a multivariate approximate law of demand. The usual multivariate law of demand obtains when $\varepsilon = 0$. For a given value $\varepsilon \geq 0$, this also places restrictions on counterfactual demand \tilde{x} at prices \tilde{p} since for any $r \in \{1, \dots, T\}$, a potential counterfactual tuple must satisfy

$$\frac{1}{2}(\tilde{p} - p^r) \cdot (\tilde{x} - x^r) \leq \varepsilon. \quad (2)$$

When we apply Assumption 1', we evaluate counterfactuals at $\varepsilon = \varepsilon^*$. This inequality places a restriction on candidate quantity-price tuples when compared with any observation in the dataset. In the one dimensional case ($K = 1$), this states that if the price increases from p^r to \tilde{p} , then demand cannot increase by too much. The bound on the increase in quantities is inversely related to the magnitude of the price increase. That is, when $\tilde{p} - p^r > 0$ we have $\tilde{x} \leq x^r + \frac{2\varepsilon}{\tilde{p} - p^r}$.

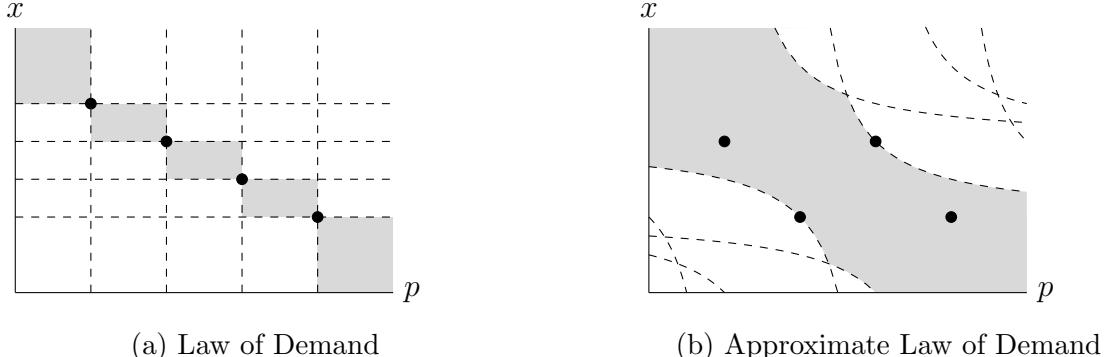


Figure 1: Restrictions of Law of Demand

We illustrate these bounds in two example datasets displayed in Figure 1. Each dataset has four observations represented as black dots. The gray area denotes the set of quantity-price tuples (\tilde{x}, \tilde{p}) that are consistent with the existing dataset with minimal level of approximation error $\varepsilon = \varepsilon^*$. In panel (a), the observed dataset is exactly consistent with quasilinear utility and the counterfactual set has $\varepsilon = 0$. In this case, the constructed bounds have the property that when price increases quantity cannot increase. This leads to the “rectangular” bounds in panel (a). Note that for low values of prices, quantity has a lower bound but not an upper bound. Similarly, when prices are higher than any observed data the lower bound on counterfactual demand is zero.

In panel (b), the dataset is not consistent with quasilinear utility, because there is an instance in which price goes up and quantity goes up. Here we graphically obtain the counterfactual restrictions using the approximate law of demand constructed from Equation 2, setting $\varepsilon = \varepsilon^*$.¹⁴ This approach leads to “hyperbolic” bounds, in contrast with the rectangular bounds in panel (a). The fact that $\varepsilon = \varepsilon^*$ is the minimal approximation error needed to rationalize the data is demonstrated on the graph by two points touching dashed hyperbolas.

¹⁴There are additional restrictions beyond Equation 2; here we provide a graphical illustration but the general framework uses additional inequalities.

3.2 Bounding Quantities

The sets $C(\cdot)$ and $AC(\cdot)$ completely describe counterfactuals. An analyst may not be interested in the entire set of counterfactual quantity-price tuples, but rather certain features of it. For example, an analyst may only be interested in quantities at a fixed counterfactual price $\tilde{p} \in \mathbb{R}_{++}^K$ allowing approximation error ε . This set may be written

$$X(\tilde{p}, D, \varepsilon) = \{\tilde{x} \in \mathbb{R}_+^K \mid (\tilde{x}, \tilde{p}) \in C(D, \varepsilon)\}.$$

Our first question is when this set is nonempty, i.e. when can we conduct counterfactual analysis.

Proposition 2. *For a dataset D and counterfactual price \tilde{p} , the set $X(\tilde{p}, D, \varepsilon)$ is nonempty if and only if $\varepsilon \geq \varepsilon^*$. Moreover, when $\varepsilon \geq \varepsilon^*$ there is a concave, strictly increasing, continuous utility function $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ that ε -rationalizes the dataset and has an exact maximizer for each $p \in \mathbb{R}_{++}^K$.*¹⁵

This shows that by allowing enough approximation error, we can find counterfactual quantities for any price. This is stronger than Fact 3 because it gives nonemptiness of the counterfactual quantity set for *any* price. Existing work has studied when observed datasets can be rationalized by quasilinear utility (Brown and Calsamiglia [2007] for $\varepsilon = 0$ or with certain random shocks) or an enlargement (Allen and Rehbeck [2020] for $\varepsilon \geq 0$). This strengthens those results by showing when we can conduct counterfactual analysis of quantities at a new price. The question is nontrivial because the utility functions constructed in Brown and Calsamiglia [2007] and Allen and Rehbeck [2020] to explain data have no approximate maximizer for low prices because the indirect utility is infinite.¹⁶

We now discuss additional properties of $X(\tilde{p}, D, \varepsilon)$.

Proposition 3. *For a dataset $D = \{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. The set $X(\tilde{p}, D, \varepsilon)$ is a*

¹⁵By strictly increasing we mean the usual definition, i.e. if each component of x is weakly greater than each component of z , then $u(x) \geq u(z)$, and if in addition some component of x is strictly greater than the corresponding component of z , then $u(x) > u(z)$.

¹⁶See the proof of Proposition 13. See also Aguiar et al. [2020] for recent work concerning emptiness of counterfactual sets when using the weak axiom of revealed preference.

closed, convex polyhedron. In particular, $\tilde{x} \in X(\tilde{p}, D, \varepsilon)$ if and only if the inequalities

$$(\tilde{p} - p^{t_M}) \cdot \tilde{x} \leq (M + 1)\varepsilon + \tilde{p} \cdot x^{t_1} - p^{t_M} \cdot x^{t_M} - \sum_{m=1}^{M-1} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \quad (3)$$

hold for all finite sequences $\{t_m\}_{m=1}^M$ without cycles where $t_m \in \{1, \dots, T\}$ and $M \geq 1$.

When $M = 1$, the inequalities in Equation 3 yield the approximate law of demand described in Equation 2. In this case we compare an observation in the dataset with a conjectured counterfactual tuple (\tilde{x}, \tilde{p}) , which leads to two instances of ε in Equation 3, just like the approximate law of demand. Proposition 3 shows there are other restrictions imposed on counterfactuals beyond the law of demand by considering more than one observation at a time ($M \geq 2$). These restrictions arise by adding up additional sequences of inequalities similar to constructing the approximate law of demand. By summing up appropriate sequences, the unknown utility function is removed so restrictions on counterfactual quantities is given only using observable data. We later use similar information to generate bounds on certain welfare objects.

Suppose now that we are only interested in bounding the quantity of the k -th good at a price \tilde{p} , allowing up to ε approximation error. These bounds are extrema of $X(\tilde{p}, D, \varepsilon)$ along the k -the dimension. That is, they are the extreme points of the set

$$X_k(\tilde{p}, D, \varepsilon) = \{x_k \in \mathbb{R}_+ \mid \text{There is some } \tilde{x} \in X(\tilde{p}, D, \varepsilon) \text{ with } x_k = \tilde{x}_k\}.$$

The following proposition discusses the bounds for the k -th good. In particular, when the bounds exist they can be computed by a linear program and the bounds satisfy monotonicity properties with respect to the approximation error ε .

Proposition 4. *For a dataset D , let $\varepsilon \geq \varepsilon^*$. The bounds*

$$\begin{aligned} \bar{x}_k(\tilde{p}, \varepsilon) &= \sup_{x_k \in X_k(\tilde{p}, D, \varepsilon)} x_k \\ \underline{x}_k(\tilde{p}, \varepsilon) &= \inf_{x_k \in X_k(\tilde{p}, D, \varepsilon)} x_k \end{aligned}$$

can each be computed as a linear program whenever they are finite.

Under Assumption 1' ($\varepsilon = \varepsilon^*$), these bounds cannot be improved.

The details on the linear program to compute bounds are found in Proposition A.1 of Appendix A. Recall $X_k(\tilde{p}, D, \varepsilon)$ is convex from Proposition 3. Thus, any quantity between $\underline{x}_k(\tilde{p}, \varepsilon)$ and $\bar{x}_k(\tilde{p}, \varepsilon)$ is a candidate counterfactual quantity for good k .

Next we elaborate on when these bounds are finite. We show upper bound is finite as long as prices are not too low. The lower bound is always finite. To formalize this define the upper comprehensive convex hull of a finite set $\{z^\ell\}_{\ell=1}^L \subseteq \mathbb{R}^K$ as

$$\text{CCo}(\{z^\ell\}_{\ell=1}^L) = \left\{ z \in \mathbb{R}^K \mid z \geq \sum_{\ell=1}^L \alpha_\ell z^\ell \text{ for some nonnegative } \alpha_1, \dots, \alpha_L \right. \\ \left. \text{such that } \sum_{\ell=1}^L \alpha_\ell = 1 \right\}.$$

The inequality in the definition here is componentwise. In addition, let $\text{int } A$ denote the interior of a set A .

Proposition 5. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. The upper bound $\bar{x}_k(\tilde{p}, \varepsilon)$ is finite if and only if $\tilde{p} \in \text{int CCo}(\{p^t\}_{t=1}^T)$. The lower bound of $\underline{x}_k(\tilde{p}, \varepsilon)$ is always finite. The upper bound $\bar{x}_k(\tilde{p}, \varepsilon)$ is weakly increasing in ε and the lower bound $\underline{x}_k(\tilde{p}, \varepsilon)$ is weakly decreasing in ε .*

Finally, we show that for one good ($K = 1$), the bounds on demand are downward sloping in own-price.

Proposition 6 (Univariate Monotonicity). *For a dataset D , let $\varepsilon \geq \varepsilon^*$ and suppose there is a single good ($K = 1$). For any pair of prices $\tilde{p}^1, \tilde{p}^0 \in \mathbb{R}_{++}$, it follows that*

$$(\bar{x}(\tilde{p}^1, \varepsilon) - \bar{x}(\tilde{p}^0, \varepsilon))(\tilde{p}^1 - \tilde{p}^0) \leq 0$$

and

$$(\underline{x}(\tilde{p}^1, \varepsilon) - \underline{x}(\tilde{p}^0, \varepsilon))(\tilde{p}^1 - \tilde{p}^0) \leq 0.$$

When $\varepsilon = 0$, the dataset satisfies the exact law of demand. When ε^* (and $K = 1$), there is some pair of observations $r, s \in \{1, \dots, T\}$ that violates the law of demand so

$$(x^r - x^s)(p^r - p^s) > 0.$$

Proposition 6 shows that while such violations can occur in the data, the bounds themselves satisfy the exact law of demand.

Remark 1 (Sensitivity Analysis). If an analyst is unsure what is a sensible choice of ε (other than the requirement $\varepsilon \geq \varepsilon^*$), then it is possible to perform sensitivity analysis of $\bar{x}_k(\tilde{p}, \varepsilon)$ and $\underline{x}_k(\tilde{p}, \varepsilon)$ as ε varies. A specific question is the largest amount of approximation error in which one can still bound the quantity of the k th good by a pre-specified value, e.g.

$$\sup\{\varepsilon \geq \varepsilon^* \mid \bar{x}_k(\tilde{p}, \varepsilon) \leq \bar{q}_k\}.$$

This bound is related the analysis of breakdown frontiers of Masten and Poirier [2019], which involve the weakest assumptions under which one can reach a conclusion. Here, weakest assumption translates to most approximation error.

Remark 2 (Other Bounds). It is straightforward to generalize Proposition 4 to bound certain linear combinations of the candidate demand vector \tilde{x} . Bounds on such linear combinations may be computed as the value of a linear programming problem. One interesting linear combination is $\tilde{p} \cdot \tilde{x}$, which is the expenditure on the K goods. Sharp bounds on general functionals $f(\tilde{x})$ can also be described as the value of a constrained optimization problem. For example, an upper bound is given by

$$\sup_{\tilde{x} \in X(\tilde{p}, D, \varepsilon)} f(\tilde{x}).$$

Recall that Proposition 3 states this constraint set is a closed convex polyhedron. This can facilitate computation though we do not formally study computation for general f .

3.3 Expenditure Bounds

Additional assumptions can tighten the bounds on quantities in Proposition 4. For example, one can assume that expenditure is the same at the counterfactual value as the last period of data, so $\tilde{p} \cdot \tilde{x} = p^T \cdot x^T$. Alternatively, one could place bounds on the expenditures so that $m \leq \tilde{p} \cdot \tilde{x} \leq \bar{m}$. One may also impose *a priori* bounds on the quantities of other goods. These bounds can considerably shrink the set of counterfactual bounds, especially when there are multiple goods. In addition,

computation with these additional restrictions is not challenging because these are inequality constraints that can be appended to the original linear program. When adding these additional constraints, however, it is possible that the counterfactual set can be empty.

We emphasize that in general, such expenditure bounds are not needed to deliver nontrivial counterfactual bounds. It is helpful to contrast our approach with the general model of utility maximization subject to a budget constraint, with preferences that need not be quasilinear. In the general model, even under correct specification the sharp bounds on quantities of each good at a given price are the trivial bounds $[0, \infty)$ unless the analyst places *a priori* bounds on expenditure at the new price.¹⁷ This is because the general model does not rule out expenditure of 0 or arbitrarily high values at counterfactuals when we only fix prices.¹⁸

4 Welfare

To study welfare, we must take a stand on the interpretation of approximation error. For this section, we follow Allen and Rehbeck [2020] and treat the approximation error as arising from satisficing in the spirit of Simon [1947]. In particular, an individual has a utility function that describes the ranking over goods, but satisfies by choosing bundles that are “good enough.”

We now discuss how satisficing relates to the measurement and prediction wedge. When trying to learn about utility from data, a measurement wedge arises since observed choices may not be optimal. When trying to predict welfare for a price change, the prediction wedge occurs since we only know the region of bundles that are “good enough.” Assumption 1' means that the measurement and prediction wedge are the same size as the smallest amount of satisficing needed to describe the data. We note that one can also apply the satisficing interpretation to counterfactual quantities, but it is not necessary. For this reason we did not distinguish between these wedges

¹⁷The bounds $[0, \infty)$ are for bounding one good at a time (similar to \underline{x}_k and \bar{x}_k above). There are nontrivial restrictions on the entire demand tuple.

¹⁸The results in Deb et al. [2018] can be used to show nontrivial bounds are possible in the general model when income is always the same value (inside and outside the dataset) and there is an unobserved good whose price is fixed.

in Section 3. Appendix B provides additional discussion.

Since we are studying quasilinear utility there are two natural welfare objects. We look at differences in utility over consumption bundles and differences in (approximate) indirect utility over prices. An important asymmetry arises because learning about differences in utility only involves the measurement wedge because it does not involve choices in new situations. In contrast, differences in (approximate) indirect utility over prices involves both the measurement wedge and prediction wedge because one must consider choices in new settings. We elaborate more below.

4.1 Recoverability of Utility

Our first goal is to learn about the unknown utility function over consumption bundles using data. This is helpful when considering policies involving the direct distribution of goods.

In general, there is a collection utility functions that can ε -rationalize a dataset $\{(x^t, p^t)\}_{t=1}^T$. We study bounds on utility differences between consumption bundles. Specifically, given two consumption bundles $\tilde{x}^1, \tilde{x}^0 \in \mathbb{R}_+^K$ we consider the upper and lower bounds

$$\begin{aligned}\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) &= \sup_{\{u|u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{u(\tilde{x}^1) - u(\tilde{x}^0)\} \\ \underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) &= \inf_{\{u|u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{u(\tilde{x}^1) - u(\tilde{x}^0)\}.\end{aligned}$$

Here, we consider all possible utility functions $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ without additional restrictions such as monotonicity or concavity. A utility function u is said to ε -rationalize the dataset $\{(x^t, p^t)\}_{t=1}^T$ when for every $t \in \{1, \dots, T\}$ the inequality

$$u(x^t) - p^t \cdot x^t \geq u(x) - p \cdot x - \varepsilon$$

holds for every $x \in \mathbb{R}_+^K$.

To interpret these bounds, suppose for example that $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) < 0$. We conclude that the individual ranks \tilde{x}^0 above \tilde{x}^1 , even when the individual's choices do not ex-

actly maximize utility. Thus, there is no ambiguity in the ranking of these bundles according to the unknown utility function u . If $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) > 0$, then it is possible that the individual ranks \tilde{x}^1 above \tilde{x}^0 . Lastly, if $\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) > 0$, then we conclude the individual ranks \tilde{x}^1 above \tilde{x}^0 . More broadly, these bounds provide *cardinal* information on utility differences, in units of the price of the numeraire.

To gain some intuition how bounds on differences of utility are informed by data, consider two bundles x^r and x^s in the dataset. Since x^r is approximately optimal given prices p^r , we have the restriction

$$u(x^r) - p^r \cdot x^r \geq u(x^s) - p^r \cdot x^s - \varepsilon,$$

which rearranges to

$$u(x^s) - u(x^r) \leq p^r \cdot (x^s - x^r) + \varepsilon. \quad (4)$$

Differences in utility are thus bounded by changes in expenditure. Here, price is fixed and a change in quantity determines the magnitude of the expenditure change. The inequality in (4) arises because the point in the data x^r was approximately optimal at prices p^r . Thus, ε here directly involves the observed data and is part of the measurement wedge. There is no prediction wedge because an analyst does not contemplate choices in new situations.

We first formalize computation of the bounds before providing additional interpretation. We show the bounds can be calculated as a linear program. An explicit description is relegated to Proposition A.2 in Appendix A.

Proposition 7. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. If \tilde{x}^0 is in the dataset, i.e. $\tilde{x}^0 = x^S$ for some $S \in \{1, \dots, T\}$, then $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon)$ is finite and can be calculated as a linear program. If \tilde{x}^1 is in the dataset, i.e. $\tilde{x}^1 = x^F$ for some $F \in \{1, \dots, T\}$, then $\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon)$ is finite and can be calculated as a linear program.*

Under Assumption 1' ($\varepsilon = \varepsilon^$), these bounds cannot be improved.*

Note that the set

$$\{u \mid u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}$$

is convex in the sense that if each u^a, u^b ε -rationalize the dataset, then $\alpha u^a + (1-\alpha)u^b$

does for $\alpha \in [0, 1]$. This follows from inspecting inequalities such as

$$u(x^t) - p^t \cdot x^t \geq u(x) - p \cdot x - \varepsilon$$

that define ε -rationalizability by a utility function u . This means that any value between $\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon)$ and $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon)$ can be attained.

To gain further intuition how data bound utility differences, we provide an analytical characterization. This characterization builds on inequalities such as (4) above, yet uses longer sequences (rather than just pairs) of observations to describe the tightest possible bounds. This parallels analysis of counterfactuals, where restrictions other than the law of demand arise by considering sequences of observations.

Proposition 8. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. If \tilde{x}^0 is in the dataset, i.e. $\tilde{x}^0 = x^S$ for some $S \in \{1, \dots, T\}$, then for any $\tilde{x}^1 \in \mathbb{R}_+^K$ with $\tilde{x}^1 \neq x^S$, the upper bound on utility differences is given by*

$$\bar{u}(\tilde{x}^1, x^S, \varepsilon) = \min_{\sigma \in \Sigma_S} \left\{ p^{\sigma(M)} \cdot (\tilde{x}^1 - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

where Σ_S is the set of sequences that start with $\sigma(1) = S$, have no cycles, and have length at least $M \geq 1$. Moreover, the function \bar{u} is strictly increasing and continuous in $(\tilde{x}^1, \varepsilon)$ over the region that satisfies $\varepsilon \geq \varepsilon^*$ and excludes $\tilde{x}^1 = x^S$.

The sums inside the minimum are closely related to sums discussed in Proposition 3 for counterfactuals. The sums differ because Proposition 3 constructs sequences making a cycle (to remove the unknown utility numbers). In contrast, Proposition 8 considers sequences that do not make a cycle because the goal is to examine differences of utility numbers.

Continuity and concavity fail at $\tilde{x}^1 = x^S$ (when $\varepsilon > 0$) because the difference in utility is zero when the quantity is the same. Since $\bar{u}(\tilde{x}^0, \tilde{x}^1, \varepsilon) = -\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon)$, analogous results hold for the lower bound $\underline{u}(x^F, \tilde{x}^1, \varepsilon)$ if the first argument x^F is in the dataset. See Proposition A.2 for formal results.

An important feature for practical application is that the bounds on utilities are trivial unless an appropriate quantity is in the dataset. We formalize this as follows.

Proposition 9. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. If \tilde{x}^0 is not in the dataset, i.e.*

$\tilde{x}^0 \neq x^r$ for every $r \in \{1, \dots, T\}$, then

$$\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = \infty.$$

If \tilde{x}^1 is not in the dataset, then

$$\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = -\infty.$$

Recall that Proposition 8 shows that \bar{u} is strictly increasing and continuous over a region. Thus, the upper bound on utility differences has some shape restrictions like a “nice” utility function. Despite this, the bound is not concave/continuous in the first argument at $\tilde{x}^1 = x^S$ (when $\varepsilon > 0$). This means that imposing concavity/continuity can potentially tighten the bounds.¹⁹ Imposing these (or other) shape restrictions is important if one wishes to bound utility when neither quantity is in the dataset, since from Proposition 9 we know the bounds are trivial without more structure.

4.2 Recoverability of Approximate Indirect Utility

We now turn to welfare analysis concerning price changes. Here both the measurement and prediction wedge play a role. Recall that a “measurement wedge” shows up for bounds on the utility over bundles as in Section 4.1 since observations may not *exactly* maximize a quasilinear utility function. Here the prediction wedge also arises when $\varepsilon^* > 0$ because even when we know the utility function, we do not know which approximately-optimal choice would be made at a new price.

We first discuss the indirect utility, which is the standard welfare object for the exact quasilinear model. Later we introduce the approximate indirect utility to account for the fact that an individual does not exactly optimize. The indirect utility function associated with the utility function $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ is given by

$$V_u(p) = \sup_{x \in \mathbb{R}_+^K} u(x) - x \cdot p.$$

¹⁹Continuity and concavity do not tighten the bounds when $\varepsilon = 0$ because the upper bound \bar{u} is then continuous and concave for all values of \tilde{x}^1 . See the proof of Proposition A.2 for more details.

Since the researcher does not know the individual's utility *a priori*, we consider indirect utility associated with candidate utility functions.

We show how indirect utility interacts with the measurement wedge. If x^t is within ε of the maximum utility possible at price p^t , then we can write

$$V_u(p^t) \leq u(x^t) - p^t \cdot x^t + \varepsilon.$$

The definition of the indirect utility yields for arbitrary $p \in \mathbb{R}_{++}^K$, the inequality

$$V_u(p) \geq u(x^t) - p \cdot x^t.$$

Differencing these, we obtain

$$V_u(p^t) - V_u(p) \leq x^t \cdot (p - p^t) + \varepsilon. \quad (5)$$

Here, ε arises because the observed choices need not be exact maximizers and thus is part of the measurement wedge. With a restriction on the magnitude of ε , we can use observations of x^t and p^t to bound differences in indirect utility.

We now introduce the prediction wedge. This wedge arises because *raw* differences in indirect utility are not the natural welfare object in our setting for a price change because we focus on ex ante policy evaluation. Instead, we take into account that when $\varepsilon > 0$, an individual may choose bundles with different utility when facing the same prices. This is because we assume an individual satisfies.

The utility the individual attains for a given price and choice of consumption bundle is the *approximate indirect utility*. For observation $t \in \{1, \dots, T\}$, the approximate indirect utility is

$$u(x^t) - p^t \cdot x^t.$$

At price p , the approximate indirect utility is restricted to be somewhere in the interval

$$[V_u(p) - \varepsilon, V_u(p)].$$

In fact, (weakly) further restrictions take into account that the approximate indirect

utility attained is bounded below by

$$\underline{V}_{u,A}(p, \varepsilon) = \inf_{x \in \mathbb{R}_+^K} u(x) - p \cdot x \quad \text{s.t.} \quad u(x) - p \cdot x \geq V_u(p) - \varepsilon,$$

while the upper bound is the indirect utility. The lower bound on approximate indirect utility is the lower bound $\underline{V}_{u,A}(p, \varepsilon)$, while the upper bound on approximate indirect utility is the upper bound $\overline{V}_{u,A}(p, \varepsilon) = V_u(p)$.

Now suppose we wish to bound the change in approximate indirect utility between prices \tilde{p}^0 and \tilde{p}^1 . If the utility u and level of satisficing ε were known, then the welfare bounds would be

$$[\underline{V}_{u,A}(\tilde{p}^1, \varepsilon) - \overline{V}_{u,A}(\tilde{p}^0, \varepsilon), \overline{V}_{u,A}(\tilde{p}^1, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon)].$$

Fixing u , this interval becomes wider when ε increases. In general, ε controls the prediction wedge, which arises even if we knew u because we would not know what is chosen by the satisficer.

Since we do not know the utility function *a priori*, we consider bounds involving the smallest and largest changes in approximate indirect utility among all utility functions that ε -rationalize the dataset:

$$\begin{aligned} \overline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) &= \sup_{\{u|u \text{ } \varepsilon-\text{rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{\overline{V}_{u,A}(\tilde{p}^1, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon)\} \\ \underline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) &= \inf_{\{u|u \text{ } \varepsilon-\text{rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{\underline{V}_{u,A}(\tilde{p}^1, \varepsilon) - \overline{V}_{u,A}(\tilde{p}^0, \varepsilon)\}.^{20} \end{aligned}$$

These bounds incorporate both the measurement and prediction wedges. The measurement wedge shows up when considering u that ε -rationalize the data, while the prediction wedge arises when defining the approximate indirect utility. We use the same value ε for both since we maintain Assumption 1'.

These bounds can inform a researcher about changes in welfare even in the presence of satisficing. If $\underline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) > 0$, then we can conclude that given a price change from \tilde{p}^0 to \tilde{p}^1 the individual is better off at \tilde{p}^1 . If $\overline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) < 0$, then the price

²⁰Formally, we take the supremum over u such that $\overline{V}_{u,A}(\tilde{p}^0, \varepsilon)$ is not ∞ , and the infimum over u such that $\overline{V}_{u,A}(\tilde{p}^1, \varepsilon)$ is not ∞ .

change from \tilde{p}^0 to \tilde{p}^1 makes the individual worse off. In contrast, ambiguity arises when $\underline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) < 0$ and $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) > 0$. In this case an individual may be better or worse given the price change, but the data alone are inconclusive.

We now state a computational result for the bounds. A specific description of the linear program is given in Proposition A.3 in Appendix A.

Proposition 10. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. The bounds on approximate indirect utility $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ and $\underline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ can each be computed as a linear program whenever they are finite.*

Under Assumption 1' ($\varepsilon = \varepsilon^$), these bounds cannot be improved.*

When $\varepsilon = \varepsilon^* = 0$, the approximate indirect utility equals the indirect utility, and these are the sharp bounds on consumer surplus with limited price variation. When we set $\varepsilon = \varepsilon^*$, these are the *adaptive consumer surplus bounds*. These bounds may be used for arbitrary prices \tilde{p}^1, \tilde{p}^0 , not only at prices in $\{(p^t)\}_{t=1}^T$. In particular, these bounds provide welfare bounds at new prices *without* needing to first provide bounds on the quantities at the prices.

Recall (5) established for p^t in the dataset,

$$V_u(p^t) - V_u(p) \leq x^t \cdot (p - p^t) + \varepsilon.$$

This states that differences in indirect utility are bounded by changes in expenditure. Here, the change in expenditure involves keeping the quantity fixed and changing prices. We present lower and upper bounds on \bar{V} that build on this inequality. To state the result, first suppose $\tilde{p}^1 = p^S$ is in the dataset. Define

$$h(\tilde{p}^0) = \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\},$$

where Σ_S is the set of sequences that start with $\sigma(1) = S$, have no cycles, and have length at least $M \geq 1$.

Proposition 11. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. If \tilde{p}^1 is in the dataset, i.e. $\tilde{p}^1 = p^S$ for some $S \in \{1, \dots, T\}$, then for $\tilde{p}^1 \neq \tilde{p}^0$,*

$$h(\tilde{p}^0) - \varepsilon \leq \bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) \leq h(\tilde{p}^0) + \varepsilon,$$

and for $\tilde{p}^1 = \tilde{p}^0$, $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = \varepsilon$.

This result is established by leveraging duality results we present in Appendix C.3. Analogous results exist for \underline{V} because $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = -\underline{V}(\tilde{p}^0, \tilde{p}^1, \varepsilon)$, and are omitted for brevity.

We reiterate that Proposition 10 describes that \bar{V} can be computed exactly as a linear program. The goal of Proposition 11 is to make this process less of a “black box.” Note that when $\varepsilon = 0$, the lower and upper bounds coincide and we characterize $\bar{V}(\tilde{p}^1, \tilde{p}^0, 0)$. We recognize h as a function closely related to the construction of the Riemann integral, since it computes the sum of the area of certain rectangles. We may view h as a “discrete” analogue of the consumer surplus formula, which states that differences in indirect utility are the area of a demand function. In fact, this integration intuition can be formalized in the special case of a single good ($K = 1$), when $\varepsilon = 0$.

Proposition 12. *Suppose there is a single good ($K = 1$), the dataset $\{(x^t, p^t)\}_{t=1}^T$ is exactly consistent with quasilinear utility ($\varepsilon^* = 0$), and we set $\varepsilon = 0$. If $\tilde{p}^1 > \min\{p_1, \dots, p_T, \tilde{p}^0\}$, then*

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, 0) = \int_0^1 \bar{x}_1(t\tilde{p}^1 + (1-t)\tilde{p}^0, 0)(\tilde{p}_1^0 - \tilde{p}_1^1)dt.$$

Proposition 12 shows that in a certain case, there is a tight connection between bounds on quantities and welfare bounds. Further relationships between welfare and counterfactual quantities are left for future work.

To close this section, we present shape restrictions on \bar{V} that hold for all $\varepsilon \geq 0$.

Proposition 13. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is convex, weakly decreasing, and lower semicontinuous in \tilde{p}^1 ²¹ and weakly increasing in ε . If $\tilde{p}^1 \in \text{CCo}(\{p^t\}_{t=1}^T)$, then $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is finite. If $\tilde{p}^1 \notin \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$, then $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = \infty$.*

The shape restrictions in Proposition 13 are those of an indirect utility function. We do not obtain global continuity here because the welfare bounds can be infinite. However, $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is continuous in \tilde{p}^1 over the relative interior of $\text{CCo}(\{p^t\}_{t=1}^T)$

²¹A function $f : H \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$ is lower semicontinuous if for any $a \in \mathbb{R}$ the set $\{x \in H \mid f(x) \leq a\}$ is closed in the topology on H .

because it is convex and finite over this set (Rockafellar [2015], Theorem 10.1).

Recall that Proposition 9 shows that bounds on utility differences are trivial when a quantity is not in the dataset. In contrast, Proposition 13 shows that the bound on approximate indirect utility $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is typically finite provided \tilde{p}^1 is not too low. In particular, neither \tilde{p}^1 nor \tilde{p}^0 need be in the dataset. The reason we obtain these contrasting results is that indirect utility functions must satisfy certain shape restrictions while we consider utility functions that need not satisfy shape restrictions such as concavity or monotonicity.

5 Continuity and Convexity in Quantities and Approximation Error

In classic revealed preference, a small amount of measurement error can lead to refutation of the model. In this case, there is no way to use the model for counterfactual or welfare analysis. Below we show continuity of the welfare/counterfactual bounds in both quantities and degree of approximation error. Thus, we provide a way to still conduct analysis when the model is not perfect, and do so in a way that is a continuous enlargement of the standard conceptual framework.²²

In more detail, here we study the joint mapping from quantities and approximation error to the bounds analyzed previously. One motivation for this is that in applications, an analyst may not observe a dataset of interest $\{(x^t, p^t)\}_{t=1}^T$ exactly, and may instead only have an estimate of the quantities. We show below that if we can consistently estimate quantities, then we can consistently estimate the bounds.

For concreteness, suppose an analyst is conducting a representative agent analysis, and quantities are mean quantities from a population at each time period. We examine the mean demand vector at period t so that $x^t = \mathbb{E}[X^{i,t}]$, where $X^{i,t}$ is demand for individual i at time t . Here $X^{i,t}$ is treated as a random variable that is identically distributed across individuals. An analyst estimates $\widehat{\mathbb{E}[X^{i,t}]}$ from a cross-sectional dataset in which individuals at each time period face the same prices. For

²²To be clear, results in this paper are also new under $\varepsilon = 0$ with a few exceptions that are noted.

example, the estimator could be the sample average of demands at time t across many individuals.

We now turn to the formal results. While we allow estimation error associated with quantities, here we take each price p^t as nonrandom and measured exactly. Recall that $\bar{x}_k(\tilde{p}, \varepsilon)$ is the maximal quantity of good k at counterfactual price \tilde{p} assuming approximation error is no greater than ε . The definition of \bar{x}_k is presented in Proposition 4. We now treat \bar{x}_k as a function of the dataset of quantities, and with minor abuse of notation we write $\bar{x}_k(d, \tilde{p}, \varepsilon)$, where $d = (d^1, \dots, d^T) \in \mathbb{R}_+^{K \times T}$ denotes quantities across all goods at time periods. This allows us to study how the bound depends on quantities in the dataset while keeping prices, $\{p^t\}_{t=1}^T$, fixed. Similarly, $\underline{x}_k(d, \tilde{p}, \varepsilon)$ denotes the lower bound. Finally, let $A \subseteq \mathbb{R}_+^{K \times T} \times \mathbb{R}_+$ denote combinations of quantities and approximation error such that the counterfactual/welfare objects are defined, i.e.

$$A = \{(d, \varepsilon) \in \mathbb{R}_+^{K \times T} \times \mathbb{R}_+ \mid \varepsilon \geq \varepsilon^* (\{(d^t, p^t)\}_{t=1}^T)\}.$$

Proposition 14. *Fix a price \tilde{p} where we wish to bound counterfactual quantities and assume the dataset of prices $\{p^t\}_{t=1}^T$ is fixed. The set A is convex. The mapping $\bar{x}_k(\cdot, \tilde{p}, \cdot) : A \rightarrow \mathbb{R}_+ \cup \{\infty\}$ is concave in (d, ε) , and is continuous in (d, ε) at any point where it is finite. The mapping $\underline{x}_k(\cdot, \tilde{p}, \cdot) : A \rightarrow \mathbb{R}_+$ is convex and continuous in (d, ε) .*

We obtain a similar result for the bounds on approximate indirect utility \bar{V} and \underline{V} when we view them as a function of the dataset of quantities. To formalize this, with minor abuse of notation let $\bar{V}(d, \tilde{p}^1, \tilde{p}^0, \varepsilon)$ describe the upper bound as a mapping of the quantities $d \in \mathbb{R}_+^{K \times T}$ in a dataset. Similarly, $\underline{V}(d, \tilde{p}^1, \tilde{p}^0, \varepsilon)$ denotes the lower bound.

Proposition 15. *Fix a price pair \tilde{p}^1 and \tilde{p}^0 where we wish to bound the difference in approximate indirect utility, and assume the dataset of prices $\{p^t\}_{t=1}^T$ is fixed. The mapping $\bar{V}(\cdot, \tilde{p}^1, \tilde{p}^0, \cdot) : A \rightarrow \mathbb{R} \cup \{\infty\}$ is concave in (d, ε) , and is continuous in (d, ε) at any point where it is finite. The mapping $\underline{V}(\cdot, \tilde{p}^1, \tilde{p}^0, \cdot) : A \rightarrow \mathbb{R} \cup \{-\infty\}$ is convex in (d, ε) , and is continuous in (d, ε) at any point where it is finite.*

Recall Proposition 13 shows that when $\tilde{p}^1 \in \text{CCo}(\{p^t\}_{t=1}^T)$, $\bar{V}(d, \tilde{p}^1, \tilde{p}^0, \varepsilon)$ is finite for any $(d, \varepsilon) \in A$.

We need one more result. Here, we interpret the minimal approximation error ε^* as a function of quantities for fixed prices $\{p^t\}_{t=1}^T$.

Proposition 16 (Allen and Rehbeck [2020]). *The mapping $\varepsilon^* : \mathbb{R}_+^{K \times T} \rightarrow \mathbb{R}_+$ is convex and continuous.*

The previous continuity results imply the following consistency results.

Corollary 1. *Suppose we have some estimator of the quantities that satisfies $\hat{d}^n \xrightarrow{p} d$. Then*

$$\begin{aligned}\bar{x}_k\left(\hat{d}^n, \tilde{p}, \varepsilon^*\left(\hat{d}^n\right)\right) &\xrightarrow{p} \bar{x}_k(d, \tilde{p}, \varepsilon^*(d)) \\ \underline{x}_k\left(\hat{d}^n, \tilde{p}, \varepsilon^*\left(\hat{d}^n\right)\right) &\xrightarrow{p} \underline{x}_k(d, \tilde{p}, \varepsilon^*(d)). \\ \bar{V}\left(\hat{d}^n, \tilde{p}^1, \tilde{p}^0, \varepsilon^*\left(\hat{d}^n\right)\right) &\xrightarrow{p} \bar{V}(d, \tilde{p}^1, \tilde{p}^0, \varepsilon^*(d)) \\ \underline{V}\left(\hat{d}^n, \tilde{p}^1, \tilde{p}^0, \varepsilon^*\left(\hat{d}^n\right)\right) &\xrightarrow{p} \underline{V}(d, \tilde{p}^1, \tilde{p}^0, \varepsilon^*(d)),\end{aligned}$$

where each result holds whenever the right hand side result is finite.

This provides a theoretical foundation for plug-in estimation. We omit a formal description of the sampling scheme since the result applies to any collection of random variables $\{\hat{d}^n\}$ that converges in probability to d . For example, if we have panel data and the quantities $(X^{i,t})_{t=1}^T$ are independent and identically distributed across individuals, one can use sample averages so that $\hat{d}^n = (\frac{1}{n} \sum_{i=1}^n X^{i,t})_{t=1}^T$ when estimating $d = (\mathbb{E}[X^{i,t}])_{t=1}^T$. This requires that individuals face the same prices, which are treated as nonrandom parameters here for the consistency result.

Finally, we consider shape restrictions for the bounds on utility differences \bar{u} and \underline{u} , viewed as functions of quantities. As before, fixing prices $\{p^t\}_{t=1}^T$, we study dependence on the quantities $d \in \mathbb{R}_+^{K \times T}$. With minor abuse of notation write $\bar{u}(d, \tilde{x}^0, \tilde{x}^1, \varepsilon)$ as a function of quantities and approximation error, and similarly for \underline{u} .

Recall that Proposition 8 shows that when \tilde{x}^0 is in the dataset of quantities d , $\bar{u}(d, \tilde{x}^1, \tilde{x}^0, \varepsilon)$ is finite provided $\varepsilon \geq \varepsilon^*$. In contrast, Proposition 9 shows that whenever \tilde{x}^0 is outside the dataset, $\bar{u}(d, \tilde{x}^1, \tilde{x}^0, \varepsilon) = \infty$. We conclude that when viewed as a mapping of quantities, \bar{u} is no longer continuous. It is, however, continuous over a certain subset of A . To describe this, for a vector $\tilde{x} \in \mathbb{R}_+^K$ let $A(\tilde{x}) = \{(d, \varepsilon) \in A \mid d^1 = \tilde{x}\}$. This restricts attention to quantities datasets that all contain a certain vector \tilde{x} as

the first component. We formalize continuity and concavity results as follows.

Proposition 17. *Fix a quantity pair $\tilde{x}^1 \neq \tilde{x}^0$ where we wish to bound the difference in utility, and assume the dataset of prices $\{p^t\}_{t=1}^T$ is fixed. The set $A(\tilde{x})$ is convex for any $\tilde{x} \in \mathbb{R}_+^K$. The mapping $\bar{u}(\cdot, \tilde{x}^1, \tilde{x}^0, \cdot) : A \rightarrow \mathbb{R} \cup \{\infty\}$ is concave and continuous in (d, ε) over the region $A(\tilde{x}^0)$. The mapping $\underline{u}(\cdot, \tilde{x}^1, \tilde{x}^0, \cdot) : A \rightarrow \mathbb{R} \cup \{-\infty\}$ is convex and continuous in (d, ε) over the region $A(\tilde{x}^1)$.*

Continuity over all of A does not hold and so one cannot directly apply the continuous mapping theorem to establish a consistency result like Corollary 1. If some quantity vector d^1 is measured without error, however, then it is possible to consistently estimate the bounds on utility differences between d^1 and other bundles though we omit details for brevity.

5.1 Aggregation

The previous analysis in this section presents several results indicating how bounds are convex or concave in quantities. Building on this, we show that the framework we present has nontrivial trade-offs concerning the level of aggregation.

To illustrate this, suppose the analyst has a collection of individual-level datasets $\{(x^{i,t}, p^t)\}_{t=1}^T$, for $i \in \{1, \dots, n\}$, where $d^i = (x^{i,t})_{t=1}^T$ collects the quantities. Here, each individual faces the same prices. We abstract from sampling variability or measurement error for simplicity. The analyst can construct individual-level bounds on objects of interest. For example, the analyst may consider an upper bound on quantity at price \tilde{p} , denoted $\bar{x}_k(d^i, \tilde{p}, \varepsilon^*(d^i))$. The analyst can then construct bounds on aggregate quantities by averaging over the individual bounds, forming

$$\frac{1}{n} \sum_{i=1}^N \bar{x}_k(d^i, \tilde{p}, \varepsilon^*(d^i)).$$

As an alternative, the analyst can directly calculate a bound applied to average quantities, i.e. a representative agent analysis. That is, letting $d = \frac{1}{n} \sum_{i=1}^n d^i$, we can bound quantities via

$$\bar{x}_k(d, \tilde{p}, \varepsilon^*(d)).$$

When $\varepsilon^*(d^i) = 0$ for each individual i , there is no approximation error. In this

case, $\varepsilon^*(d) = 0$ by convexity of ε^* . Because \bar{x}_k is a convex mapping in quantities (Proposition 14), we see

$$\frac{1}{n} \sum_{i=1}^N \bar{x}_k(d^i, \tilde{p}, 0) \leq \bar{x}_k(d, \tilde{p}, 0),$$

which we recall is only stated assuming $\varepsilon^*(d^i) = 0$ for each i . This result states that it is weakly more informative to use the individual-level bounds.

An important distinction emerges when $\varepsilon^*(d^i) > 0$ for some individuals. In this case, ε^* is convex from Proposition 16, which yields $\varepsilon^*(d) \leq \frac{1}{n} \sum_{i=1}^n \varepsilon^*(d^i)$. In other worse, aggregation decreases the approximation error relative to the average approximation error. This can lead to instances in which

$$\frac{1}{n} \sum_{i=1}^N \bar{x}_k(d^i, \tilde{p}, \varepsilon^*(d^i)) \geq \bar{x}_k(d, \tilde{p}, \varepsilon^*(d)),$$

which states that aggregation can lead to more informative bounds when approximation error is taken into account. The ability of aggregation to reduce approximation error is a classic point emphasized by e.g. Grunfeld and Griliches [1960]. As one example, Allen and Rehbeck [2020] find using panel data that $\varepsilon^*(d^i) > 0$ for each household, but the aggregated data satisfies $\varepsilon^*(d) = 0$.²³

The analysis is similar with the other bounds because they satisfy concavity/convexity restrictions. Aggregation can either help or hurt in the present framework.

6 Application

We now illustrate the results in the paper with data on the demand for gasoline. Data are from the 2001 United States National Household Travel Survey, and have previously been used in Blundell et al. [2012]. The data are from a single cross-section. For brevity we refer to Blundell et al. [2012] for additional details, including construction of the particular sample.

The primary observables of interest are quantities and prices. Quantities are annual

²³That paper does not study counterfactuals or welfare and is a complement to the present paper.

gasoline consumption, which is constructed from odometer readings and an estimate of fuel efficiency. Prices are the average tax-inclusive price per gallon, in the county where the individual lives.

First note it is possible to map the raw quantities and prices to a (cross-sectional) dataset $\{(X^i, P^i)\}_{i=1}^n$, and then apply our previous analysis. Here i denotes the individual and n denotes the sample size. We use this notation rather than t and T to emphasize we have a cross-section. We use upper case X^i and P^i to denote that these are random variables.

We do not use the raw dataset, and instead “pre-process it” to map to our framework. We do so because we have a cross-section of individuals. We wish to both to diminish the impact of sampling variability as well as incorporate heterogeneity along observable variables. As in, Blundell et al. [2012] we pre-process by first considering a partially linear model given by

$$X = g(P, Y) + \beta' W + U,$$

where P is price, Y is income, W are observed covariates, and U is unobservable heterogeneity. While Blundell et al. [2012] interpret g as a demand curve for a representative agent for the general model of utility maximization subject to a budget constraint, here we have a different interpretation. We interpret $g(\cdot, Y)$ as the demand curve for the representative agent with income level Y ; Y thus serves as a demographic characteristic that alters the shape of the demand curve. We close the model with the restriction

$$\mathbb{E}[U | P = p, Y = y, W = w] = 0.$$

This specification allows price sensitivity to depend on the level of income of an individual. If this assumption fails because e.g. prices are endogenous, our analysis can be adapted provided we have an estimator of an average structural demand function [Blundell and Powell, 2003]. In this case, our analysis applies to the representative agent corresponding to the average structural demand.²⁴ Because we consider this an illustrative application, we do not pursue this here. The *starting point* for our

²⁴Recall the analysis in Section 5.1 about aggregation. Allen and Rehbeck [2020] studies how stochastic shocks and approximation error can be studied in a common framework.

framework here is a consistent estimator of structural quantities, which complements a large literature studying how to estimate structural quantities.

For each level of income \tilde{y} , we consider a dataset (in the sense used previously in the paper) of the form

$$D(\tilde{y}) = \{\hat{g}(P^i, \tilde{y}), P^i\}_{i=1}^{\tilde{n}},$$

where \hat{g} is an estimator of g described below. Thus, $g(P^i, \tilde{y})$ is akin to the structural quantity x^t in the previous notation, and P^i is akin to p^t . Recall Corollary 1 shows that a consistent estimator of quantities produces a consistent estimator of bounds. Like Blundell et al. [2012], we consider prices between the 5-th and 95-th quantile to mitigate endpoint issues, so \tilde{n} enumerates these observations.

The estimator \hat{g} is constructed similar to Blundell et al. [2012]. We first estimate $\hat{\beta}$ by a double residual regression as in Robinson [1988].²⁵ Then we set

$$\hat{g}(P^i, \tilde{y}) = \frac{\sum_{j=1}^N \left(X^j - \hat{\beta}' W_j \right) K_{h_p}(P^j - P^i) K_{h_y}(Y^j - \tilde{y})}{\sum_{j=1}^N K_{h_p}(P^j - P^i) K_{h_y}(Y^j - \tilde{y})},$$

where K_h is a kernel with bandwidth h . Following Blundell et al. [2012] we use the biweight kernel. Throughout, the bandwidths h_p and h_y are chosen so that

$$\frac{h_p}{h_y} = \frac{\hat{\sigma}_P}{\hat{\sigma}_Y},$$

where $\hat{\sigma}_P^2 = \frac{1}{n} \sum_{i=1}^n (P^i - \bar{P})^2$, $\bar{P} = \frac{1}{n} \sum_{i=1}^n P_i$, $\hat{\sigma}_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y^i - \bar{Y})^2$, and $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. Note that these are all constructed with all observations.

Figure 2 presents analysis with two choices of bandwidths. These correspond to the ad hoc choices .75 and 1 after standardizing. The top two panels display the kernel-smoothed “dataset”

$$D(\bar{Y}) = \{\hat{g}(P^i, \bar{Y}), P^i\}_{i=1}^{\tilde{n}}$$

as well as counterfactual bounds, where $\hat{g}(P^i, \bar{Y})$ is interpreted as a quantity for observation i facing prices P^i . Recall \tilde{n} denotes the middle 90% of observations in terms of price, where we drop the lower and upper 5% to mitigate endpoint issues. Income is evaluated at the sample mean \bar{Y} . The welfare bounds for approximate indirect

²⁵We use the biweight kernel with ad hoc bandwidth .75 after standardizing the data.

utility are displayed in the middle panels. The bounds are evaluated relative to the mean price $\bar{P} = 1.334$. The bounds for differences in utility for certain quantities are displayed in the lower panels. The utility bounds are relative to the median quantity in the dataset. There are 101 comparisons, which for computational reasons are made between 101 of the points in the dataset of quantities $\{\hat{g}(P^i, \bar{Y})\}_{i=1}^{\tilde{n}}$. Recall from Proposition 9, comparisons in utility when one quantity is not in the dataset will have at least one trivial bound. This is why we restrict attention to comparisons in which both quantities are in the dataset. It is important to note that in practice, bounding quantities over a grid that is not a subset of the observed quantities will lead to trivial bounds.

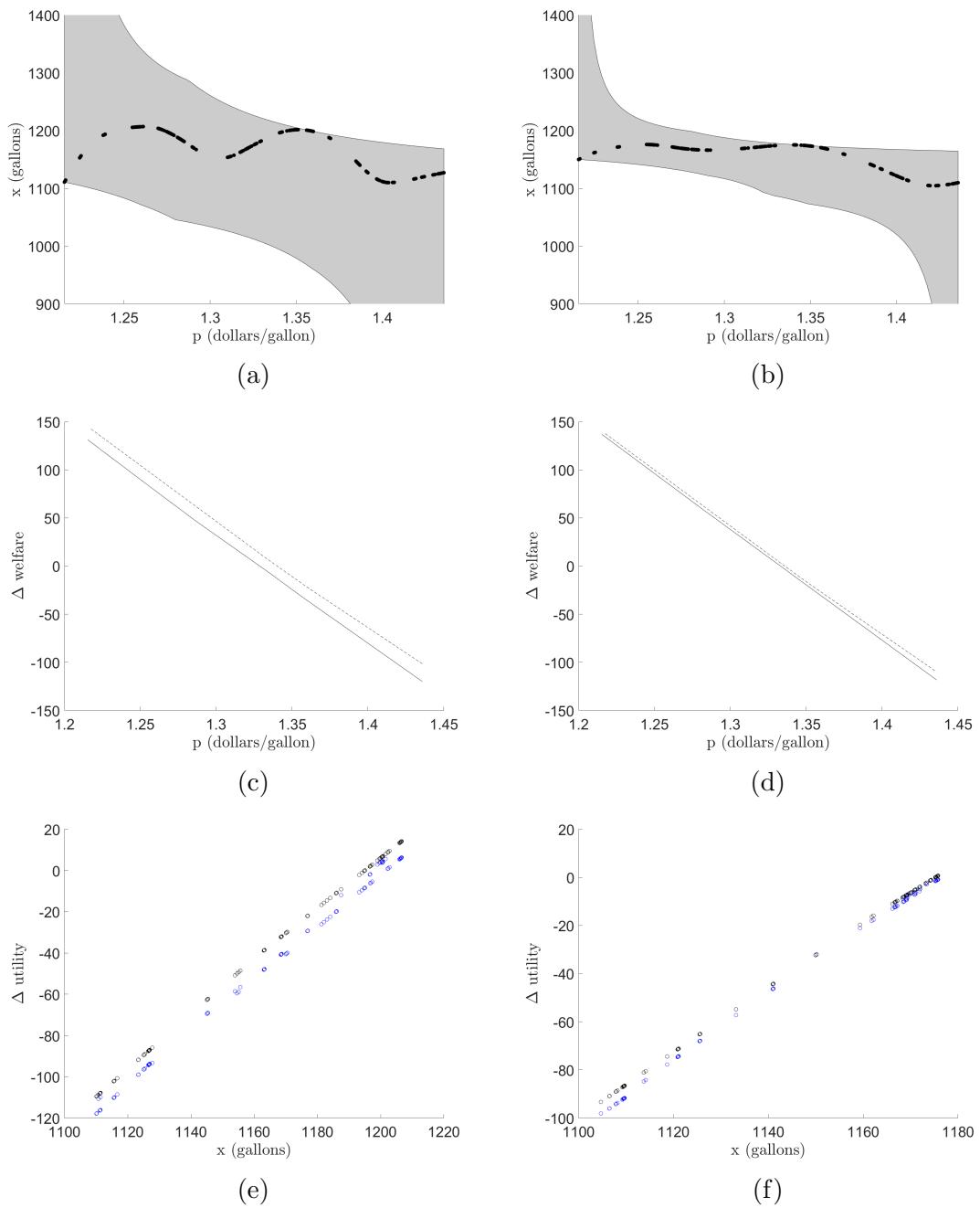


Figure 2: Quantity, Approximate Indirect Utility, and Utility Bounds at Two Bandwidths

Notes: The top panels depict quantity bounds at new prices. The middle panels depict bounds on approximate indirect utility relative to the mean price. The lower panels depict bounds on utility at certain quantities, relative to the median quantity in the dataset.

As can be seen from the figures, the choice of bandwidth noticeably alters the informativeness of the counterfactual bounds (upper panels). In contrast, the bounds on approximate indirect utility (middle panels) and utility (lower panels) are relatively narrow for both bandwidths. Similar results obtain for alternative bandwidths.

In the lower panels, a contrast emerges between the lower and upper bounds on utility. Recall that these bounds are for $u(\tilde{x}^1) - u(\tilde{x}^0)$, where \tilde{x}^0 is the median quantity, i.e. in the dataset. Because the second argument is in the dataset, Proposition 8 applies to the upper bounds and establishes monotonicity in the argument \tilde{x}^1 . The graphs are consistent with this, since the upper bounds are monotone in quantities. In contrast, the lower bounds are not monotone. In order to make lower bounds on utilities monotone, it would be necessary to have the *first argument* be in the dataset and fixed (\tilde{x}^1). See Proposition A.2(vi) in Appendix A.

7 Conclusion

This paper provides a conceptual framework for counterfactual and welfare analysis for approximate models. Our main conceptual assumption is that model approximation error has the same magnitude in new settings as the data we have seen. We formalize this for the quasilinear utility model. This assumption is portable to other settings, and generalizes the standard approach that requires correct specification in both the data we have seen and at hypothetical values.

Engaging with the possibility that a model may not perfectly match data is especially important using the nonparametric revealed preference approach. Indeed, a natural intuition is that if approximation error is “small,” then it is second order and we can ignore it for certain questions. Unfortunately, this intuition is false in the standard approach used in the revealed preference literature, since small violations of the model mean it cannot be used for counterfactual or welfare analysis. This paper presents an adaptive approach allowing the analyst to use the model formally viewing it as an approximation. Moreover, our counterfactual/welfare bounds are continuous in the degree of approximation error, and so they continuously transition to the standard framework when approximation error is negligible.

Appendix A Proofs of Main Results

This appendix provides proofs of the results in the main text. It also provides explicit descriptions of the linear programs mentioned in the main text. Some of the proofs require additional lemmas contained in Supplemental Appendix C.

A.1 Proofs for Section 3

Proof of Facts 1-4. The proofs are in the main text. \square

Proof of Proposition 2. Emptiness of $X(\tilde{p}, D, \varepsilon)$ when $\varepsilon < \varepsilon^*$ is immediate from Fact 3. It remains to show that when $\varepsilon \geq \varepsilon^*$, the set $X(\tilde{p}, D, \varepsilon)$ is nonempty.

First, fix $(x^1, p^1) \in \{(x^t, p^t)\}_{t=1}^T$ and let Σ_1 denote the set of finite sequences of $t \in \{1, \dots, T\}$ with no cycles that begins at (x^1, p^1) . Define

$$U(x) = \min_{\sigma \in \Sigma_1} \left\{ p^{\sigma(M)} \cdot (x - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

where $\sigma \in \Sigma_1$ is a sequence of length M , for all $m \in \{1, \dots, M\}$ it follows that $\sigma(m) \in \{1, \dots, T\}$, and $\sigma(1) = 1$. Allen and Rehbeck [2020] have shown that for $\varepsilon \geq \varepsilon^*$, this function ε -rationalizes the data in the sense that for each $t \in \{1, \dots, T\}$ and each $x \in \mathbb{R}_+^K$,

$$U(x^t) - p^t \cdot x^t \geq U(x) - p^t \cdot x - \varepsilon.$$

The function $U(x)$ need not induce an ε -maximizer when prices take low values. However, the constructed utility can be modified to guarantee maximizers exist.

To that end, let

$$\bar{U} = \sup_{x \in \text{Co}(\{x^t\}_{t=1}^T)} U(x)$$

where $\text{Co}(\{x^t\}_{t=1}^T)$ denotes the convex hull, i.e. the smallest convex set containing $\{x^t\}_{t=1}^T$. We see $\bar{U} < \infty$ since U is continuous and $\text{Co}(\{x^t\}_{t=1}^T)$ is compact. Define $f : \mathbb{R}_+^K \rightarrow \mathbb{R}$ by $f(x) = \frac{\sum_{k=1}^K x_k}{1 + \sum_{k=1}^K x_k} + \bar{U}$, which is bounded and concave. To see this, note the function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ given by $h(z) = z/(1+z)$ is concave by inspecting

derivatives. Since f is a composition of a concave function and an affine and strictly increasing function it is concave.

Now construct $\tilde{U}(x) = \min\{U(x), f(x)\}$. This function rationalizes the data since for $t \in \{1, \dots, T\}$ and $x \in \mathbb{R}_+^K$ we have

$$\tilde{U}(x^t) - p^t \cdot x^t = U(x^t) - p^t \cdot x^t \geq U(x) - p^t \cdot x - \varepsilon \geq \tilde{U}(x) - p^t \cdot x - \varepsilon.$$

In addition, \tilde{U} is concave since it is the minimum of concave functions. Similarly, \tilde{U} is continuous and strictly increasing as it is the minimum of finitely many continuous and strictly increasing functions. It remains to show this utility admits an ε -maximizer for all prices $p \in \mathbb{R}_{++}^K$.

To that end, note the indirect utility of f , denoted

$$V_f(p) = \sup_{x \in \mathbb{R}_+^K} f(x) - p \cdot x,$$

is everywhere finite over the region $p \in \mathbb{R}_{++}^K$ because f is bounded between \bar{U} and $\bar{U} + 1$. Moreover, since $\tilde{U} \leq f$ pointwise, we also have $V_{\tilde{U}}(p) \leq V_f(p)$, so that $V_{\tilde{U}}$ is finite for any $p \in \mathbb{R}_{++}^K$. Since \mathbb{R}_{++}^K is open, from Lemma C.2 we conclude that

$$\tilde{U}(x) - p \cdot x$$

admits an exact maximizer in x for any $p \in \mathbb{R}_{++}^K$. In particular, it admits ε -maximizers, completing the proof. \square

Proof of Proposition 3. An equivalent definition of $X(\tilde{p}, D, \varepsilon)$ is $\tilde{x} \in X(\tilde{p}, D, \varepsilon)$ if and only if the augmented dataset $D \cup (\tilde{x}, \tilde{p})$ is ε -rationalized by quasilinear utility. From the characterization in Lemma C.1(iii), this is equivalent to showing that certain sequences satisfy an inequality. For each sequence involving the augmented dataset there are two cases. If the sequence does not contain (\tilde{x}, \tilde{p}) , then the inequality in Lemma C.1(iii) is satisfied because we assume $\varepsilon \geq \varepsilon^*(D)$. (Note that $\varepsilon^*(D)$ is constructed to have this property.) It remains to check sequences involving (\tilde{x}, \tilde{p}) .

Rearranging the inequality of Lemma C.1(iii), we see that

$$(\tilde{p} - p^{t_M}) \cdot \tilde{x} \leq (M+1)\varepsilon + \tilde{p} \cdot x^{t_1} - p^{t_M} \cdot x^{t_M} - \sum_{m=1}^{M-1} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \quad (6)$$

must hold for all finite sequences $\{t_m\}_{m=1}^M$ without cycles where $t_m \in \{1, \dots, T\}$ and $M \geq 1$. Here we have the $M+1$ coefficient on ε since the sequences include the counterfactual observation and a length M sequence. This characterizes the set $X(\tilde{p}, D, \varepsilon)$ as an intersection of finitely many half-spaces. Thus, $X(\tilde{p}, D, \varepsilon)$ is a closed, convex polyhedron. \square

To prove Proposition 4, we prove a stronger result that explicitly describes the linear program.

Proposition A.1. *For a dataset $D = \{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. Then whenever $X_k(\tilde{p}, D, \varepsilon)$ is bounded above, its maximum is given by the linear program*

$$\begin{aligned} \bar{x}_k(\tilde{p}, \varepsilon) = & \max_{\substack{\tilde{x} \in \mathbb{R}_+^K \\ u^1, \dots, u^T, \tilde{u} \in \mathbb{R}_+}} \tilde{x}_k \\ \text{s.t. } & u^s \leq u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\ & \tilde{u} \leq u^r + p^r \cdot (\tilde{x} - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\ & u^r \leq \tilde{u} + \tilde{p} \cdot (x^r - \tilde{x}) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\}. \end{aligned}$$

The upper bound $\bar{x}_k(\tilde{p}, \varepsilon)$ may be equivalently calculated as

$$\begin{aligned} \bar{x}_k(\tilde{p}, \varepsilon) = & \max_{\tilde{x} \in \mathbb{R}_+^K} \tilde{x}_k \\ \text{s.t. } & (\tilde{p} - p^{t_M}) \cdot \tilde{x} \leq (M+1)\varepsilon + \tilde{p} \cdot x^{t_1} - p^{t_M} \cdot x^{t_M} - \sum_{m=1}^{M-1} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}), \end{aligned}$$

where this inequality must hold for all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \dots, T\}$ and $M \geq 1$. The value of $\underline{x}_k(\tilde{p}, \varepsilon)$ is calculated as the minimum of the objective with either constraint set of the above linear programs. The value $\underline{x}_k(\tilde{p}, \varepsilon)$ is weakly increasing in ε over the region $\varepsilon \geq \varepsilon^*$, and $\underline{x}_k(\tilde{p}, \varepsilon)$ is weakly decreasing in ε over the region $\varepsilon \geq \varepsilon^*$.

The first linear program is easy to implement as it has order $(T+1)^2$ constraints

and $T + 1 + K$ unknowns. The second linear program is useful to understand the mapping from data to bounds. However, directly operationalizing the second linear program would require enumerating all finite sequences of the dataset that do not contain cycles, which is computationally costly.

Related bounds have appeared in Chiong et al. [2017] and Allen and Rehbeck [2019a], which focus on latent utility models with observable characteristics of goods other than prices. The result here differs since ε can be nonzero and the first set of bounds directly describes a convenient linear program used to compute bounds.²⁶ We take \bar{x}_k to be positive infinity when there is no finite upper bound.

Proof of Proposition A.1 (and Proposition 4). The linear programming formulations are immediate from Lemma C.1 and the proof of Proposition 3. From Lemma C.3, the maximum is attained because the linear program has a bounded value function by construction. Recall that for the second formulation, we only need to consider cycles involving the counterfactual quantity-price tuple because we have assumed $\varepsilon \geq \varepsilon^*$. Recall that all cycles that do not involve the counterfactual quantity are necessarily less than ε^* and will not bind. We leveraged these properties in Proposition 3 already. The proof for x_k are analogous and are omitted.

Now we argue that these bounds cannot be improved under Assumption 1'. To see this, note from Proposition 5 that the bounds are weakly monotone in ε . Moreover, when $\varepsilon < \varepsilon^*$ we know the set $X(\tilde{p}, D, \varepsilon)$ (and hence $X_k(\tilde{p}, D, \varepsilon)$) is empty from Proposition 2. \square

Proof of Proposition 5. We begin by showing that $\bar{x}_k(\tilde{p}, \varepsilon)$ is finite if and only if $\tilde{p} \in \text{intCCo}(\{p^t\}_{t=1}^T)$. First, let $\tilde{p} \in \text{intCCo}(\{p^t\}_{t=1}^T)$ so that $\tilde{p} > \sum_{t=1}^T \alpha_t p^t$ for some nonnegative α_t such that $\sum_t \alpha_t = 1$. Note that for each $t \in \{1, \dots, T\}$, the approximate law of demand yields

$$(\tilde{p} - p^t) \cdot (\tilde{x} - x^t) \leq 2\varepsilon.$$

²⁶Chiong et al. [2017] essentially start with the second formulation of the bounds (in terms of cycles) and show that while there are many cycles, only a certain number are effectively binding. The first formulation of Proposition A.1 complements their analysis by describing an explicit linear program with order $(T + 1)^2$ scalar inequalities. Allen and Rehbeck [2019a] describe bounds in certain models with characteristics in place of prices, and use a characterization similar to the cycles condition, but do not study extreme points or describe computations.

Multiplying by α^t and summing up the inequalities gives that

$$\left(\tilde{p} - \sum_{t=1}^T \alpha_t p^t \right) \cdot \left(\tilde{x} - \sum_{t=1}^L \alpha_t x^t \right) \leq 2\varepsilon.$$

Thus,

$$\left(\tilde{p} - \sum_{t=1}^T \alpha_t p^t \right) \cdot \tilde{x} \leq 2\varepsilon + \left(\tilde{p} - \sum_{t=1}^T \alpha_t p^t \right) \cdot \sum_{t=1}^L (\alpha_t x^t).$$

Since $\left(\tilde{p} - \sum_{t=1}^T \alpha_t p^t \right) > 0$ and $\tilde{x} \in \mathbb{R}_+^K$ one can bound the values on each dimension of \tilde{x} so that

$$\tilde{x} \in \prod_{k=1}^K \left[0, \frac{2\varepsilon + \left(\tilde{p} - \sum_{t=1}^T \alpha_t p^t \right) \cdot \sum_{t=1}^L \alpha_t x^t}{\tilde{p}_k - \sum_{t=1}^T \alpha_t p_k^t} \right].$$

This shows that $\bar{x}_k(\tilde{p}, \varepsilon)$ is finite when $\tilde{p} \in \text{intCCo}(\{p^t\}_{t=1}^T)$.

Next, we show when $\tilde{p} \notin \text{intCCo}(\{p^t\}_{t=1}^T)$ that $\bar{x}_k(\tilde{p}, \varepsilon)$ is unbounded. Suppose that $\tilde{p} \notin \text{intCCo}(\{p^t\}_{t=1}^T)$. This means for all $t \in \{1, \dots, T\}$ that $\tilde{p} \leq p^t$. From Proposition 3 we know $\tilde{x} \in X(\tilde{p}, D, \varepsilon)$ if and only if for any sequence $\{t_m\}_{m=1}^M$ with $M \geq 1$,

$$(\tilde{p} - p^{t_M}) \cdot (\tilde{x} - x^{t_1}) \leq (M+1)\varepsilon + p^{t_M} \cdot x^{t_1} - p^{t_M} \cdot x^{t_M} - \sum_{m=1}^{M-1} p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}).$$

Note that the right hand side of the expression is always weakly positive. Moreover, $\tilde{p} - p^{t_M} \leq 0$ for every dimension. If all terms are zero, this places no restrictions on \tilde{x} so that one can choose arbitrarily positive amounts of every good. If there is a dimension k such that $\tilde{p}_k - p_k^{t_M} < 0$, then one can choose arbitrarily high amounts of \tilde{x}_k to satisfy all such inequalities. This establishes that $\bar{x}_k(\tilde{p}, \varepsilon)$ is unbounded above.

Note that the lower bound $\underline{x}_k(\tilde{p}, \varepsilon)$ is always finite because it is bounded below by 0.

To show monotonicity in ε note that the feasibility region is weakly increasing (with regard to set inclusion) as ε increases. Thus, \bar{x}_k is weakly increasing in ε , and \underline{x}_k is weakly decreasing in ε . \square

Proof of Proposition 6. First, let Σ be the set of sequences $\{t_m\}_{m=1}^M$ that contain no cycles where $t_m \in \{1, \dots, T\}$. From Proposition 3, the counterfactual bounds on

demand are given by inequalities of the form

$$(\tilde{p} - p^{t_M})\tilde{x} \leq (M + 1)\varepsilon + \tilde{p}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}}), \quad (7)$$

which much hold for every sequence in Σ . Dot products are removed since all objects are one-dimensional.

Whether a certain inequality of the form (7) provides an upper bound or lower bound on \tilde{x} depends on the sign of $\tilde{p} - p^{t_M}$. To see this, note that rearranging Equation 7 when $\tilde{p} > p^{t_M}$ yields

$$\begin{aligned} \tilde{x} &\leq \frac{(M + 1)\varepsilon + \tilde{p}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{\tilde{p} - p^{t_M}} \\ &= x^{t_1} + \frac{(M + 1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{\tilde{p} - p^{t_M}}. \end{aligned} \quad (8)$$

Note the the expression in Equation 8 that is divided by $\tilde{p} - p^{t_m}$ is positive since the terms above are those for a cycle of length M . To see this, note that

$$p^{t_M}(x^{t_M} - x^{t_1}) + \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}}) \leq M\varepsilon \leq (M + 1)\varepsilon,$$

where the first inequality holds because $\varepsilon \geq \varepsilon^*$ and the left hand side is a sequence of length M . Thus, such sequences constitute upper bounds.

When instead $\tilde{p} < p^{t_M}$, the sequence yields a lower bound since one is dividing by a negative number. Since the sign of the difference matters, we partition the set of sequences in Σ as follows. We consider the counterfactual prices where $\tilde{p}^0 < \tilde{p}^1$ without loss of generality. Let $\{t_m\}_{m=1}^M = \sigma \in \Sigma^1$ when $p^{t_M} \leq \tilde{p}^0 < \tilde{p}^1$. Let $\{t_m\}_{m=1}^M = \sigma \in \Sigma^0$ when $\tilde{p}^0 < p^{t_M} < \tilde{p}^1$. Lastly, let $\{t_m\}_{m=1}^M = \sigma \in \Sigma^{-1}$ when $\tilde{p}^0 < \tilde{p}^1 \leq p^{t_M}$.

Upper bounds on counterfactual demand for the price \tilde{p}^0 involve sequences where $p^{t_M} \leq \tilde{p}^0$ (i.e. sequences in Σ^1). Upper bounds on counterfactual demand for the price \tilde{p}^1 involve sequences where $p^{t_M} \leq \tilde{p}^1$ (i.e. sequences in $\Sigma^1 \cup \Sigma^0$). We denote the upper bound inequalities by

$$UB(\tilde{p}^0) = \{\tilde{x} \in \mathbb{R}_+ \mid \text{Equation 7 holds for sequences } \sigma \in \Sigma^1 \text{ with } \tilde{p} = \tilde{p}^0\}$$

and

$$UB(\tilde{p}^1) = \{\tilde{x} \in \mathbb{R}_+ \mid \text{Equation 7 holds for sequences } \sigma \in \Sigma^1 \cup \Sigma^0 \text{ with } \tilde{p} = \tilde{p}^1\}.$$

We use Equation (8) to show that $UB(\tilde{p}^1) \subseteq UB(\tilde{p}^0)$. First, if $\tilde{x} \in UB(\tilde{p}^1)$ then for every sequence $\{t_m\}_{m=1}^M = \sigma \in \Sigma^1$ with $p^{t_M} < \tilde{p}^0 < \tilde{p}^1$ it follows that

$$\begin{aligned}\tilde{x} &\leq x^{t_1} + \frac{(M+1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{\tilde{p}^1 - p^{t_M}} \\ &\leq x^{t_1} + \frac{(M+1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{\tilde{p}^0 - p^{t_M}}\end{aligned}$$

where the second inequality holds since the numerator is positive and $0 < \tilde{p}^0 - p^{t_M} < \tilde{p}^1 - p^{t_M}$. If $\tilde{p}^0 = p^{t_M}$ for a sequence $\sigma \in \Sigma^1$, then there is no restriction on the counterfactual demands. Since $UB(\tilde{p}^0)$ only is restricted by sequences in Σ^1 while $UB(\tilde{p}^1)$ is restricted by sequences in Σ^1 and Σ^0 , this shows $UB(\tilde{p}^1) \subseteq UB(\tilde{p}^0)$. This proves that $\bar{x}(\tilde{p}^1, \varepsilon) \leq \bar{x}(\tilde{p}^0, \varepsilon)$ since \bar{x} is the maximum, the upper bounds satisfy $UB(\tilde{p}^1) \subseteq UB(\tilde{p}^0)$, and a maximum over a larger set is weakly larger.

Next note that the lower bounds on counterfactual demand \tilde{x} are given by the following

$$LB(\tilde{p}^0) = \{\tilde{x} \in \mathbb{R}_+ \mid \text{Equation 7 holds for sequences } \sigma \in \Sigma^{-1} \cup \Sigma^0 \text{ with } \tilde{p} = \tilde{p}^0\}$$

and

$$LB(\tilde{p}^1) = \{\tilde{x} \in \mathbb{R}_+ \mid \text{Equation 7 holds for sequences } \sigma \in \Sigma^{-1} \text{ with } \tilde{p} = \tilde{p}^1\}.$$

To see this, note that rearranging Equation 7 when $\tilde{p} < p^{t_M}$ yields

$$\tilde{x} \geq x^{t_1} - \frac{(M+1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{p^{t_M} - \tilde{p}}.$$

We now show that $LB(\tilde{p}^0) \subseteq LB(\tilde{p}^1)$. If $\tilde{x} \in LB(\tilde{p}^0)$, then for every sequence

$\{t_m\}_{m=1}^M = \sigma \in \Sigma^{-1}$ with $\tilde{p}^0 < \tilde{p}^1 < p^{t_M}$ it follows that

$$\begin{aligned}\tilde{x} &\geq x^{t_1} - \frac{(M+1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{p^{t_M} - \tilde{p}^0} \\ &\geq x^{t_1} - \frac{(M+1)\varepsilon + p^{t_M}x^{t_1} - p^{t_M}x^{t_M} - \sum_{m=1}^{M-1} p^{t_m}(x^{t_m} - x^{t_{m+1}})}{p^{t_M} - \tilde{p}^1}\end{aligned}$$

since the term being subtracted weakly increases when dividing by a smaller difference since $0 < p^{t_M} - \tilde{p}^0 < p^{t_M} - \tilde{p}^1$. (Recall the numerator in each fraction is positive.) When the sequence $\sigma \in \Sigma^{-1}$ has $\tilde{p}^1 = p^{t_M}$ there is no restriction on counterfactual demands. Since $LB(\tilde{p}^1)$ only is restricted from sequences in Σ^{-1} while $LB(\tilde{p}^0)$ is restricted by sequences in Σ^{-1} and Σ^0 , this shows $LB(\tilde{p}^0) \subseteq LB(\tilde{p}^1)$. This also shows that $\underline{x}(\tilde{p}^1, \varepsilon) \leq \underline{x}(\tilde{p}^0, \varepsilon)$ since \underline{x} is a minimum, the constraint set on the lower bounds $LB(\tilde{p}^0) \subseteq LB(\tilde{p}^1)$, and a minimum over a smaller set is weakly larger.

□

A.2 Proofs for Section 4

A.2.1 Proofs for Section 4.1

Propositions 7 and 8 are proven together in the following result.

Proposition A.2. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$.*

i. *If \tilde{x}^0 is in the dataset, i.e. $\tilde{x}^0 = x^S$ for some $S \in \{1, \dots, T\}$, and $\tilde{x}^1 \neq \tilde{x}^0$, then*

$$\begin{aligned}\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) &= \max_{u^1, \dots, u^T, \tilde{u}^1 \in \mathbb{R}_+} \tilde{u}^1 - u^S \\ \text{s.t. } u^s &\leq u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\ \tilde{u}^1 &\leq u^r + p^r \cdot (\tilde{x}^1 - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\ \tilde{u}^1 &= u^r \quad \text{for all } r \in \{1, \dots, T\} \quad \text{with } \tilde{x}^1 = x^r.\end{aligned}$$

ii. *If $\tilde{x}^0 = x^S$ is in the dataset and $\tilde{x}^1 \neq \tilde{x}^0$, then the upper bound is equivalently*

given by

$$\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = \min_{\sigma \in \Sigma_S} \left\{ p^{\sigma(M)} \cdot (\tilde{x}^1 - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

where Σ_S is the set of sequences that start with $\sigma(1) = S$, have no cycles, and have length at least $M \geq 1$.

iii. If $\tilde{x}^0 = x^S$ is in the dataset, the function \bar{u} is strictly increasing and continuous in $(\tilde{x}^1, \varepsilon)$ over the region that excludes $\tilde{x}^1 = \tilde{x}^0$. In particular, under Assumption 1 the bound cannot be improved.

iv. If \tilde{x}^1 is in the dataset, i.e. $\tilde{x}^1 = x^F$ for some $F \in \{1, \dots, T\}$, and $\tilde{x}^1 \neq \tilde{x}^0$, then

$$\begin{aligned} \bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) &= \min_{u^1, \dots, u^T, \tilde{u}^0 \in \mathbb{R}_+} \quad u^F - \tilde{u}^0 \\ \text{s.t.} \quad u^s &\leq u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\ \tilde{u}^0 &\leq u^r + p^r \cdot (\tilde{x}^0 - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\ \tilde{u}^0 &= u^r \quad \text{for all } r \in \{1, \dots, T\} \text{ with } \tilde{x}^0 = x^r. \end{aligned}$$

v. If $\tilde{x}^1 = x^F$ is in the dataset and $\tilde{x}^1 \neq \tilde{x}^0$, then the lower bound is equivalently given by

$$\underline{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = \max_{\sigma \in \Sigma_F} \left\{ p^{\sigma(M)} \cdot (x^{\sigma(M)} - \tilde{x}^0) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m)} - x^{\sigma(m+1)}) - M\varepsilon \right\}.$$

vi. If $\tilde{x}^1 = x^F$ is in the dataset, the function \underline{u} is strictly decreasing and continuous in $(\tilde{x}^0, \varepsilon)$ over the region that excludes $\tilde{x}^0 = \tilde{x}^1$. In particular, under Assumption 1 the bound cannot be improved.

Parts (i) and (iv) describe the linear programs used for computation and stated as Proposition 7 in the main text. Note that parts (ii) and (v) show that the bounds are finite, as claimed in Proposition 7. The other parts cover Proposition 8 stated in the main text. Parts (ii) and (v) provide analytical characterizations of the bounds on utility differences. Parts (iii) and (vi) describe shape restrictions of the bounds.

Proof of Proposition A.2. We first prove parts (i) and (ii).

The definition of ε -rationalizability yields

$$\begin{aligned} \bar{u}(\tilde{x}^1, x^S, \varepsilon) &\leqslant \sup_{u^1, \dots, u^T, \tilde{u}^1 \in \mathbb{R}_+} \tilde{u}^1 - u^S \\ \text{s.t. } u^s &\leqslant u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\ \tilde{u}^1 &\leqslant u^r + p^r \cdot (\tilde{x}^1 - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\ \tilde{u}^1 &= u^r \quad \text{for all } r \in \{1, \dots, T\} \text{ with } \tilde{x}^1 = x^r. \end{aligned}$$

We shall show the opposite inequality holds to prove (i), and in doing so characterize the maximum as stated in part (ii). First, note that the problem on the right hand side is feasible since for the dataset $D = \{(x^t, p^t)\}_{t=1}^T$, we assumed $\varepsilon \geqslant \varepsilon^*(D)$. We show that there is a utility function \tilde{u} such that for any $u^1, \dots, u^T, \tilde{u}^1 \in \mathbb{R}_+$ that are feasible,

$$\tilde{u}^1 - u^S \leqslant \tilde{u}(\tilde{x}^1) - \tilde{u}(x^S).$$

To that end, first consider feasible values $u^1, \dots, u^T, \tilde{u}^1$. For any sequence that begins at $\sigma(1) = S$, we can sum up the inequalities in the program to obtain

$$\tilde{u}^1 - u^S \leqslant p^{\sigma(M)} \cdot (\tilde{x}^1 - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon.$$

Thus,

$$\tilde{u}^1 - u^S \leqslant \min_{\sigma \in \Sigma_S} \left\{ p^{\sigma(M)} \cdot (\tilde{x}^1 - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

where Σ_S is the set of sequences with $\sigma(1) = S$, have no cycles, and have length at least $M \geqslant 1$. We show in particular that provided $\tilde{x}^1 \neq x^S$, the upper bound on the right hand side can be attained by the utility function \tilde{u} , defined for $x \neq x^S$ by

$$\tilde{u}(x) = \min_{\sigma \in \Sigma_S} \left\{ p^{\sigma(M)} \cdot (x - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

and defined for x^S by $\tilde{u}(x^S) = 0$. Note that the summation on the right side defining $\tilde{u}(x)$ is zero whenever $M = 1$ because it is a summation over an empty set of indices.

Note that \tilde{u} is not continuous at x^S , which is key for our arguments.

We show that this utility function rationalizes the data. For any $x \in \mathbb{R}_+^K$, it follows that for any $t \in \{1, \dots, T\}$ such that $x^t \neq x^S$,

$$\begin{aligned}\tilde{u}(x) - p^t \cdot x &\leqslant p^t \cdot (x - x^t) + p^{\sigma^{*,t}(M^{*,t})} \cdot (x^t - x^{\sigma^{*,t}(M^{*,t})}) + \\ &\quad \sum_{m=1}^{M^{*,t}-1} p^{\sigma^{*,t}(m)} \cdot (x^{\sigma^{*,t}(m+1)} - x^{\sigma^{*,t}(m)}) + (M^{*,t} + 1)\varepsilon - p^t \cdot x \\ &= p^{\sigma^{*,t}(M^{*,t})} \cdot (x^t - x^{\sigma^{*,t}(M^{*,t})}) + \\ &\quad \sum_{m=1}^{M^{*,t}-1} p^{\sigma^{*,t}(m)} \cdot (x^{\sigma^{*,t}(m+1)} - x^{\sigma^{*,t}(m)}) + (M^{*,t} + 1)\varepsilon - p^t \cdot x^t \\ &= \tilde{u}(x^t) - p^t \cdot x^t + \varepsilon\end{aligned}$$

where $\sigma^{*,t} \in \Sigma_S$ is a sequence that obtains the minimum of $\tilde{u}(x^t)$ and $M^{*,t}$ is the length of that sequence.

Lastly, consider the observation $S \in \{1, \dots, T\}$. For any $x \in \mathbb{R}_+^K$, it follows that

$$\begin{aligned}\tilde{u}(x) - p^S \cdot x &\leqslant p^S \cdot (x - x^S) + \varepsilon - p^S \cdot x \\ &= \tilde{u}(x^S) - p^S \cdot x^S + \varepsilon\end{aligned}$$

where the inequality follows by looking at the sequence length one which only has observation S and the equality follows since $\tilde{u}(x^S) = 0$.

This utility function gives

$$\begin{aligned}\tilde{u}(\tilde{x}^1) - \tilde{u}(x^S) &= \min_{\sigma \in \Sigma_S} \left\{ p^{\sigma(M)} \cdot (\tilde{x}^1 - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\} \\ &\leqslant \bar{u}(\tilde{x}^1, x^S, \varepsilon).\end{aligned}$$

The inequality holds because \tilde{u} ε -rationalizes the dataset. The first part of the proof of the proposition established $\bar{u}(\tilde{x}^1, x^S, \varepsilon) \leqslant \tilde{u}(\tilde{x}^1) - \tilde{u}(x^S)$. This proves part (ii).

To prove part (i), note that we can use the function $\tilde{u}(x)$ to generate utility numbers that satisfy the inequality and equality conditions in the linear programming formulation. Indeed, set $u^t = \tilde{u}(x^t) - \min_{x \in \{x^s\}_{s=1}^T \cup \{\tilde{x}^1\}} \{\tilde{u}(x)\}$ for $t \in \{1, \dots, T\}$ and

$\tilde{u}^1 = \tilde{u}(\tilde{x}^1) - \min_{x \in \{x^s\}_{s=1}^T \cup \tilde{x}^1} \{\tilde{u}(x)\}$. We subtract the minimum because $\tilde{u}(x)$ can be negative. Here the values u^t and \tilde{u}^1 are weakly positive and satisfy the inequalities.

To prove (iii), recall that the upper bound is the minimum of finitely many affine functions as shown in (ii), except at $\tilde{x}^0 = x^S$ when $\varepsilon > 0$. Each such function is strictly increasing and continuous in $(\tilde{x}^0, \varepsilon)$. From this we conclude that \bar{u} is strictly increasing and continuous in $(\tilde{x}^0, \varepsilon)$ (except at $\tilde{x}^0 = x^S$ when $\varepsilon > 0$). Thus, taking the minimum over ε subject to the constraint that the bound is defined, we see that under Assumption 1' the bound cannot be tightened. Note that here we use that for the special case $\tilde{x}^1 = \tilde{x}^0$, $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = 0$, which is also the tightest possible under Assumption 1'.

The proofs for (iv)-(vi) are analogous since $\bar{u}(\tilde{x}^1, \tilde{x}^0, \varepsilon) = -\underline{u}(\tilde{x}^0, \tilde{x}^1, \varepsilon)$, and are omitted. \square

Proof of Proposition 9. First suppose \tilde{x}^0 is not in the dataset. Since $\varepsilon \geq \varepsilon^*$ there is some utility function u that ε -rationalizes the dataset. If we modify the utility function to make $u(\tilde{x}^0)$ arbitrarily negative, the modified function still rationalizes the dataset. Note that this modified function satisfies local nonsatiation, but is not (globally) strictly increasing or concave. This proves $\bar{u}(\tilde{x}^0, \tilde{x}^1, \varepsilon) = \infty$.

Now instead suppose \tilde{x}^1 is not in the dataset. For any utility function that ε -rationalizes the dataset, we can modify $u(\tilde{x}^1)$ to be arbitrarily negative. Such a modified utility function still rationalizes the dataset, and so $\underline{u}(\tilde{x}^0, \tilde{x}^1, \varepsilon) = -\infty$. \square

A.3 Proofs for Section 4.2

We prove a stronger and more formal version of Proposition 10, explicitly describing a tractable linear program. To state this result, relative to the main text we use argument \tilde{p}^{T+2} in place of \tilde{p}^1 and \tilde{p}^{T+1} in place of \tilde{p}^0 . We use this notation because in the proofs it is helpful to think of these as extra observations relative to a dataset of T observations.

Proposition A.3. *For a dataset $\{(x^t, p^t)\}_{t=1}^T$, let $\varepsilon \geq \varepsilon^*$. Whenever $\bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon)$ is finite, the change in the approximate indirect utility can be bounded by the following*

linear program:

$$\begin{aligned}
\bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon) = & \max_{\substack{\tilde{x}^{T+1}, \tilde{x}^{T+2} \in \mathbb{R}_+^K \\ u^1, \dots, u^T, \tilde{u}^{T+1}, \tilde{u}^{T+2} \in \mathbb{R}_+}} \tilde{u}^{T+2} - \tilde{p}^{T+2} \cdot \tilde{x}^{T+2} - \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot \tilde{x}^{T+1} \\
\text{s.t. } & u^s \leq u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\
& \tilde{u}^{T+1} \leq u^r + p^r \cdot (\tilde{x}^{T+1} - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
& \tilde{u}^{T+2} \leq u^r + p^r \cdot (\tilde{x}^{T+2} - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
& u^r \leq \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot (x^r - \tilde{x}^{T+1}) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
& u^r \leq \tilde{u}^{T+2} + \tilde{p}^{T+2} \cdot (x^r - \tilde{x}^{T+2}) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
& \tilde{u}^{T+1} \leq \tilde{u}^{T+2} + \tilde{p}^{T+2} \cdot (\tilde{x}^{T+1} - \tilde{x}^{T+2}) + \varepsilon \\
& \tilde{u}^{T+2} \leq \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot (\tilde{x}^{T+2} - \tilde{x}^{T+1}) + \varepsilon.
\end{aligned}$$

Moreover, when $\underline{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon)$ is finite it is the minimum of the same problem.

Under Assumption 1' ($\varepsilon = \varepsilon^*$), these bounds cannot be improved.

Note that we do not impose the constraint that if $\tilde{p}^{T+2} = p^t$ for some t , then $\tilde{x}^{T+2} = x^t$. This is because the observed demand x^t is not known to *exactly* maximize utility at the price p^t so we must account for the fact that \tilde{x}^{T+2} can differ.

Proof of Proposition A.3 (and Proposition 10). Recall

$$\bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon) = \sup_{\{u|u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{\bar{V}_{u,A}(\tilde{p}^{T+2}, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^{T+1}, \varepsilon)\}.$$

For a utility function u , price \tilde{p} , and bound on approximate optimization given by ε , let the set of approximate optimizers be given by

$$AO_u(\tilde{p}, \varepsilon) = \{x \in \mathbb{R}_+^K \mid u(x) - p \cdot x \geq V_u(p) - \varepsilon\}.$$

We can write

$$\overline{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon) = \sup_{\{u|u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \left\{ \begin{array}{l} \sup_{\tilde{x}^{T+2} \in AO_u(\tilde{p}^{T+2}, \varepsilon)} (u(\tilde{x}^{T+2}) - \tilde{p}^{T+2} \cdot \tilde{x}^{T+2}) \\ - \inf_{\tilde{x}^{T+1} \in AO_u(\tilde{p}^{T+1}, \varepsilon)} (u(\tilde{x}^{T+1}) - \tilde{p}^{T+1} \cdot \tilde{x}^{T+1}) \end{array} \right\}.$$

We can write the difference as

$$\begin{aligned} \overline{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon) &\leqslant \sup_{\substack{u \\ \tilde{x}^{T+1}, \tilde{x}^{T+2} \in \mathbb{R}_+^K}} u(\tilde{x}^{T+2}) - \tilde{p}^{T+2} \cdot \tilde{x}^{T+2} - u(\tilde{x}^{T+1}) + \tilde{p}^{T+1} \cdot \tilde{x}^{T+1} \\ &\quad \text{s.t. } u(x^t) - p^t \cdot x^t \geqslant \sup_{x \in \mathbb{R}_+^K} u(x) - p^t \cdot x - \varepsilon \quad \forall t \in \{1, \dots, T\} \\ &\quad u(\tilde{x}^{T+2}) - \tilde{p}^{T+2} \cdot \tilde{x}^{T+2} \geqslant \sup_{x \in \mathbb{R}_+^K} u(x) - \tilde{p}^{T+2} \cdot x - \varepsilon \\ &\quad u(\tilde{x}^{T+1}) - \tilde{p}^{T+1} \cdot \tilde{x}^{T+1} \geqslant \sup_{x \in \mathbb{R}_+^K} u(x) - \tilde{p}^{T+1} \cdot x - \varepsilon. \end{aligned}$$

The first inequality constraint imposes the requirement that u ε -rationalizes the dataset. The second inequality constraint only involves the variables \tilde{x}^{T+2} , and so when we take a supremum this is the upper approximate indirect utility $\overline{V}_{u,A}(\tilde{p}^{T+2}, \varepsilon) = V_u(\tilde{p}^{T+2})$. The third inequality constraint has infimum (over \tilde{x}^{T+1}) at the lower approximate indirect utility $\underline{V}_{u,A}(\tilde{p}^{T+1}, \varepsilon)$.

Consider the feasibility region of this problem. Checking the inequalities for all x is weakly more restrictive than checking for $x \in \{x^1, \dots, x^T, \tilde{x}^{T+1}, \tilde{x}^{T+2}\}$. Thus, we will replace the suprema over all $x \in \mathbb{R}_+^K$ with a finite collection of inequalities involving $\{x^1, \dots, x^T, \tilde{x}^{T+1}, \tilde{x}^{T+2}\}$. In addition, searching over all utility functions to satisfy these inequalities is weakly more restrictive than searching over all utility numbers. From these two monotonicity observations, and the fact that the value function is

monotone in its feasibility region (with regard to set inclusion), we obtain

$$\begin{aligned}
\overline{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) &\leq \sup_{\substack{\tilde{x}^{T+1}, \tilde{x}^{T+2} \in \mathbb{R}_+^K \\ u^1, \dots, u^T, \tilde{u}^{T+1}, \tilde{u}^{T+2} \in \mathbb{R}_+}} \tilde{u}^{T+2} - \tilde{p}^{T+2} \cdot \tilde{x}^{T+2} - \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot \tilde{x}^{T+1} \\
\text{s.t. } u^s &\leq u^r + p^r \cdot (x^s - x^r) + \varepsilon \quad \text{for all } r, s \in \{1, \dots, T\} \\
\tilde{u}^{T+1} &\leq u^r + p^r \cdot (\tilde{x}^{T+1} - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
\tilde{u}^{T+2} &\leq u^r + p^r \cdot (\tilde{x}^{T+2} - x^r) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
u^r &\leq \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot (x^r - \tilde{x}^{T+1}) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
u^r &\leq \tilde{u}^{T+2} + \tilde{p}^{T+2} \cdot (x^r - \tilde{x}^{T+2}) + \varepsilon \quad \text{for all } r \in \{1, \dots, T\} \\
\tilde{u}^{T+1} &\leq \tilde{u}^{T+2} + \tilde{p}^{T+2} \cdot (\tilde{x}^{T+1} - \tilde{x}^{T+2}) + \varepsilon \\
\tilde{u}^{T+2} &\leq \tilde{u}^{T+1} + \tilde{p}^{T+1} \cdot (\tilde{x}^{T+2} - \tilde{x}^{T+1}) + \varepsilon.
\end{aligned}$$

We will now show the opposite inequality holds. First, recall that Proposition 2 shows that this program is feasible provided $\varepsilon \geq \varepsilon^*$. Let $u^1, \dots, u^T, \tilde{u}^{T+1}, \tilde{u}^{T+2}, \tilde{x}^{T+1}$, and \tilde{x}^{T+2} denote some values that are feasible. Construct the augmented dataset $\{(x^t, p^t)\}_{t=1}^{T+2}$ that has $(x^{T+1}, p^{T+1}) = (\tilde{x}^{T+1}, \tilde{p}^{T+1})$ and $(x^{T+2}, p^{T+2}) = (\tilde{x}^{T+2}, \tilde{p}^{T+2})$. Construct a utility function as

$$\tilde{u}(x) = \min_{\sigma \in \Sigma_{T+1}} \left\{ p^{\sigma(M)} \cdot (x - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

for $x \neq x^{T+1}$, where Σ_{T+1} is the set of sequences in the augmented dataset that start with $\sigma(1) = T+1$, have no cycles, and have length at least $M \geq 1$. Finally, set $\tilde{u}(x^{T+1}) = 0$. The proof of Proposition A.2 shows that this function ε -rationalizes the augmented dataset $\{(x^t, p^t)\}_{t=1}^{T+2}$. Moreover, Proposition A.2 also established

$$\tilde{u}^{T+2} - \tilde{u}^{T+1} \leq \tilde{u}(x^{T+2}) - \tilde{u}(x^{T+1}). \tag{9}$$

Recall we set $(x^{T+1}, p^{T+1}) = (\tilde{x}^{T+1}, \tilde{p}^{T+1})$ and $(x^{T+2}, p^{T+2}) = (\tilde{x}^{T+2}, \tilde{p}^{T+2})$ for hypo-

theoreticals. We conclude

$$\begin{aligned}
u^{T+2} - p^{T+2} \cdot x^{T+2} - (u^{T+1} - p^{T+1} \cdot x^{T+1}) \\
&\leq \tilde{u}(x^{T+2}) - p^{T+2} \cdot x^{T+2} - (\tilde{u}(x^{T+1}) - p^{T+1} \cdot x^{T+1}) \\
&\leq \bar{V}_{\tilde{u},A}(p^{T+2}, \varepsilon) - \underline{V}_{\tilde{u},A}(p^{T+1}, \varepsilon) \\
&\leq \bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon).
\end{aligned}$$

The first inequality uses (9). The second inequality holds because for \tilde{u} , x^{T+2} is an approximate optimizer given p^{T+2} and similarly for x^{T+1} . The third inequality is the definition of the bounds on approximate indirect utility. Since this is true for any feasible values, we conclude that $\bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon)$ is obtained by the linear program described in the proposition. Recall that while we have used suprema throughout, in this last step since we have established $\bar{V}(\tilde{p}^{T+2}, \tilde{p}^{T+1}, \varepsilon)$ as the (bounded) value of a linear program, we know the supremum is attained by Lemma C.4.

Finally, we note that the bounds are weakly monotone in ε by Proposition 13. For $\varepsilon < \varepsilon^*$ we know the bounds are not defined because no utility function ε -rationalizes the dataset. Thus, the bounds are the tightest possible under Assumption 1' ($\varepsilon = \varepsilon^*$). \square

Proof of Proposition 11. Step 1 provides the upper bound on \bar{V} . Step 2 provides the lower bound on \bar{V} . Recall that $\tilde{p}^1 = P^S$ for some $S \in \{1, \dots, T\}$.

Step 1. Recall from Equation 5 that for any $r \in \{1, \dots, T\}$ and $p \in \mathbb{R}_{++}^K$,

$$V_u(p^r) - V_u(p) \leq x^r \cdot (p - p^r) + \varepsilon.$$

By summing up such inequalities over sequences, we obtain the upper bound

$$V_u(p^S) - V_u(\tilde{p}^0) \leq \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\}.$$

Since

$$\bar{V}_{u,A}(p^S, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon) \leq V_u(p^S) - V_u(\tilde{p}^0) + \varepsilon$$

by construction, we prove that

$$\bar{V}(p^S, \tilde{p}^0, \varepsilon) \leq \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\} + \varepsilon.$$

Step 2. We now establish the lower bound for $\bar{V}(p^S, \tilde{p}^0, \varepsilon)$. Define

$$V'(p) = -\min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (p - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\}.$$

Let \underline{V}' denote the minimum of $V'(p)$ over the convex hull of $\{p^t\}_{t=1}^T$, which is attained and finite because V' is continuous and the convex hull here is compact. Define $V(p) = \max\{V'(p), \underline{V}'\}$.

First we show that V satisfies a set of inequalities that are a dual version of ε -rationalizability. Duality is considered in more detail in the Supplemental Appendix C.3, and we use several results from that Appendix.

For some $t \in \{1, \dots, T\}$, let $\sigma^{*,t} \in \Sigma_S$ be a sequence that obtains the minimum of $V'(p^t)$, and let $M^{*,t}$ be the length of that sequence. We have

$$\begin{aligned} -V'(p) - x^t \cdot p &\leq x^t \cdot (p - p^t) + x^{\sigma^{*,t}(M^{*,t})} \cdot (p^t - p^{\sigma^{*,t}(M^{*,t})}) + \\ &\quad \sum_{m=1}^{M^{*,t}-1} p^{\sigma^{*,t}(m)} \cdot (x^{\sigma^{*,t}(m+1)} - x^{\sigma^{*,t}(m)}) + (M^{*,t} + 1)\varepsilon - x^t \cdot p \\ &= -V'(p^t) - x^t \cdot p^t + \varepsilon. \end{aligned}$$

Recall that $V(p) \geq V'(p)$ and $V(p^t) = V'(p^t)$ for $t \in \{1, \dots, T\}$. This implies

$$V(p) \geq V(p^t) - x^t \cdot (p - p^t) - \varepsilon \tag{10}$$

for any $t \in \{1, \dots, T\}$ and any $p \in \mathbb{R}^K$.

Define $u_V : \mathbb{R}^K \rightarrow \mathbb{R} \cup \{-\infty\}$ by $u_V(x) = \inf_{p \in \mathbb{R}_+^K} V(p) + p \cdot x$. Proposition C.1(ii) shows that u_V ε -rationalizes the dataset. Note that since $u_V(x) \geq \underline{V}'$, u_V is everywhere finite.

The function V is the maximum of finitely many affine functions, each weakly de-

creasing in p , and is hence continuous, weakly decreasing, and convex. We conclude from Lemma C.7 that $V = V_{u_V}$. In particular, V is the indirect utility function for u_V and we have established that u_V ε -rationalizes the dataset. We conclude that

$$V(p^S) - V(\tilde{p}^0) \leq \bar{V}_{u_V, A}(p^S, \varepsilon) - \underline{V}_{u_V, A}(\tilde{p}^0, \varepsilon) \leq \bar{V}(p^S, \tilde{p}^0, \varepsilon). \quad (11)$$

We now characterize $V(p^S) - V(\tilde{p}^0)$ to state the lower bound on the proposition. Since p^S is in the convex hull of prices, $V(p^S) = V'(p^S)$. Note that $V'(p^S) \leq 0$ because the dataset is ε -rationalized by quasilinear utility (see Lemma C.1); this relies on the fact that the sum of each sequence defining V' makes a cycle because it begins and ends at p^S . In addition, by considering a sequence of length 1, $V'(p^S) \geq -\left(x^S \cdot (p^S - p^S) + \varepsilon\right) = -\varepsilon$. Thus,

$$\begin{aligned} V(p^S) - V(\tilde{p}^0) &\geq -\varepsilon - V(\tilde{p}^0) \\ &= \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\} - \varepsilon. \end{aligned}$$

So from (11),

$$\begin{aligned} \bar{V}(p^S, \tilde{p}^0, \varepsilon) &\geq \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\} - \varepsilon \\ &= h(\tilde{p}^0) - \varepsilon, \end{aligned}$$

establishing the lower bound.

It is worth noting that the arguments above rely on the inequality $-\varepsilon \leq V(p^S) \leq 0$. If we instead rely on the true value for $V(p^S)$ we actually prove the stronger result

here that

$$\begin{aligned}
\bar{V}(p^S, \tilde{p}^0, \varepsilon) &\geq \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (\tilde{p}^0 - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\} \\
&\quad - \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (p^S - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\} \\
&= h(\tilde{p}^0) \\
&\quad - \min_{\sigma \in \Sigma_S} \left\{ x^{\sigma(M)} \cdot (p^S - p^{\sigma(M)}) + \sum_{m=1}^{M-1} x^{\sigma(m)} \cdot (p^{\sigma(m+1)} - p^{\sigma(m)}) + M\varepsilon \right\}.
\end{aligned}$$

Thus, the lower bound $h(\tilde{p}^0)$ stated in the proposition can be tightened a bit. \square

Proof of Proposition 12. Recall for this proof we consider when $K = 1$ so quantities and prices are scalar. We suppress dot product notation for this proof. We first show that when $\tilde{p}^1 > \min\{p^1, \dots, p^T, \tilde{p}^0\}$,

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, 0) \leq \int_0^1 \bar{x}_1(t\tilde{p}^1 + (1-t)\tilde{p}^0, 0) (\tilde{p}^0 - \tilde{p}^1) dt.$$

For any utility u function that admits maximizers over $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$, write

$$x_u(p) \in \operatorname{argmax}_{x \in \mathbb{R}_+} u(x) - px$$

for some selector from the argmax correspondence. We first argue

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, 0) = \sup_{\{u \in \mathcal{U} \mid u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \int x_u(t\tilde{p}^1 + (1-t)\tilde{p}^0) (\tilde{p}^0 - \tilde{p}^1) dt$$

where the supremum is over the set of utility functions \mathcal{U} that admit a maximizer over $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$.

To that end, note that $\tilde{p}^1 \in \operatorname{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$. From Proposition 13, $\bar{V}(\tilde{p}^1, \tilde{p}^0, 0)$ can be written as a supremum of differences in indirect utility, where each indirect utility V_u is finite for $p \geq \{p_1, \dots, p_T, \tilde{p}^0\}$. Each V_u is convex, and for such functions, the subdifferential

$$\partial V_u(p) = \{x \in \mathbb{R}_+ \mid V_u(\tilde{p}) \geq V_u(p) + x(\tilde{p} - p) \quad \forall \tilde{p} \in \mathbb{R}_{++}\}$$

is nonempty for any $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$. See the proof of Lemma C.2. We can construct a function \tilde{x}_{V_u} by selecting from the subdifferential, so that for each $p > \min\{p_1, \dots, p_T, \tilde{p}^0\}$

$$\tilde{x}_{V_u}(p) \in \partial V_u(p).$$

Any such function satisfies the formula

$$V_u(\tilde{p}^1) - V_u(\tilde{p}^0) = \int_0^1 \tilde{x}_{V_u}(t\tilde{p}^1 + (1-t)\tilde{p}^0) (\tilde{p}^0 - \tilde{p}^1) dt. \quad (12)$$

See for example Rockafellar [2015], Corollary 24.2.1, or Chambers and Echenique [2017], Theorem 2. (The selector $\tilde{x}_{V_u}(p)$ is always Reimann integrable.)

We note that for any V_u such that u ε -rationalizes the dataset, the utility function $u_{V_u}(x) = \inf_{p \in \mathbb{R}^K} V(p) + px$ also ε -rationalizes the dataset from Lemma C.1. Moreover, u_{V_u} is concave, weakly increasing, and satisfies $V_{u_{V_u}} = V_u$ from Lemmas C.6 and C.7. From this and the proof of Lemma C.2 we conclude that \tilde{x}_{V_u} is a maximizer of u_{V_u} . Summing up, we conclude that it is without loss of generality to consider utility functions \mathcal{U} that induce a maximizer for all $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$.

Putting these arguments together, we conclude that

$$\begin{aligned} \bar{V}(\tilde{p}^1, \tilde{p}^0, 0) &= \sup_{\{u|u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{\bar{V}_{u,A}(\tilde{p}^1, 0) - \underline{V}_{u,A}(\tilde{p}^0, 0)\} \\ &= \sup_{\{u|u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{V_u(\tilde{p}^1) - V_u(\tilde{p}^0)\} \\ &= \sup_{\{u \in \mathcal{U}|u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{V_u(\tilde{p}^1) - V_u(\tilde{p}^0)\} \\ &= \sup_{\{u \in \mathcal{U}|u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \int_0^1 x_u(t\tilde{p}^1 + (1-t)\tilde{p}^0) (\tilde{p}^0 - \tilde{p}^1) dt \\ &\leq \int_0^1 \bar{x}_1(t\tilde{p}^1 + (1-t)\tilde{p}^0, 0) (\tilde{p}^0 - \tilde{p}^1) dt. \end{aligned}$$

The first equality is the definition. The second equality uses the fact that when $\varepsilon = 0$, the approximate indirect utilities equal the indirect utility. The third equality uses the arguments above to conclude it is without loss of generality to consider utility functions that induce a maximizer for $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$. The fourth equality uses (12). The first inequality uses the fact that \bar{x}_1 is a pointwise maximizer of

demand induced by u that are ε -rationalized by the dataset.

Now it remains to show the opposite inequality. We first show that \bar{x}_1 is the demand induced by some quasilinear utility function. When $K = 1$ and $\varepsilon = 0$, for $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$, it follows that $\bar{x}_1(p)$ is finite from Proposition 5. Consider the sets $E^1 = \{(\bar{x}_1(p), p)\}_{p > \min\{p^1, \dots, p^T, \tilde{p}^0\}}$ and $E^2 = \{(x^t, p^t)\}_{t=1}^T$. We argue that for $e = (e_x, e_p) \in \{E^1 \cup E^2\}$ and $e' = (e'_x, e'_p) \in \{E^1 \cup E^2\}$, the inequality

$$(e_x - e'_x)(e_p - e'_p) \leq 0$$

holds. We argue by cases. Indeed, when $e, e' \in E^1$ this follows from Proposition 6. When $e, e' \in E^2$, the inequality holds by Lemma C.1 and the fact that $\varepsilon = 0$. When $e \in E^1$ and $e' \in E^2$, the result follows from Equation 3. This covers all cases.

Now consider the set $E = E^1 \cup E^2$. From Rockafellar [2015], p. 240 the set E is the graph of a multivalued mapping that is cyclically monotonically decreasing. From Rockafellar [2015], Theorem 24.3 there is some lower semicontinuous convex function f such that for any $e = (e_x, e_p) \in E$,

$$-e_x \in \partial f(e_p),$$

where ∂f denotes the subdifferential of f .²⁷ Let f^* denote the convex conjugate of f , which is defined in Appendix C.3. We conclude from Rockafellar [2015], Theorem 23.5 that for each $e = (e_x, e_p) \in E$,

$$-e_x \in \operatorname{argmax}_x xe_p - f^*(x).$$

Since $e_x \geq 0$, we conclude via a change in variables that

$$e_x \in \operatorname{argmax}_{x \geq 0} -xe_p - f^*(-x).$$

We conclude that by setting $\tilde{u}(x) = -f^*(-x)$, we have for each $e \in E$, the price e_p induces the demand e_x as some exact maximizer of a quasilinear utility function.

²⁷Note that we differ from the statement of Theorem 24.3 in Rockafellar [2015] because we consider cyclically monotonically decreasing mappings while that result considers increasing mappings; this is why we need to take a negative involving e_x .

In particular, recalling $E = E^1 \cup E^2$ and using that E^1 is the graph of \bar{x}_1 over $p > \min\{p^1, \dots, p^T, \tilde{p}^0\}$, we conclude that \bar{x}_1 is a demand function generated by quasilinear utility with \tilde{u} . Moreover, recall \tilde{u} rationalizes the dataset. Thus,

$$\begin{aligned} & \int \bar{x}_1(t\tilde{p}^1 + (1-t)\tilde{p}^0) dt (\tilde{p}^0 - \tilde{p}^1) \\ & \leq \sup_{\{u \in \mathcal{U} | u \text{ rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \int_0^1 x_u(t\tilde{p}^1 + (1-t)\tilde{p}^0) (\tilde{p}^0 - \tilde{p}^1) dt \\ & = \bar{V}(\tilde{p}^1, \tilde{p}^0, 0). \end{aligned}$$

□

Proof of Proposition 13. To show shape restrictions on $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$, recall

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = \sup_{\{u | u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{\bar{V}_{u,A}(\tilde{p}^1, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon)\},$$

where the supremum is over u such that $\underline{V}_{u,A}(\tilde{p}^0, \varepsilon) \neq \infty$. For any u that ε -rationalizes the data, we can add or subtract a constant to u and the new utility function rationalizes the data as well. In addition, if we let $u+a$ be the utility u plus the constant a , then

$$\bar{V}_{u,A}(\tilde{p}^1, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon) = \bar{V}_{u+a,A}(\tilde{p}^1, \varepsilon) - \underline{V}_{u+a,A}(\tilde{p}^0, \varepsilon).$$

Thus, it is thus without loss of generality to restrict u such that $\underline{V}_{u,A}(\tilde{p}^0, \varepsilon) = 0$. Now recall the upper approximate indirect utility satisfies $\bar{V}_{u,A}(\tilde{p}^1, \varepsilon) = V_u(\tilde{p}^1)$. Combining these arguments we can write

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = \sup_{\{u | u \text{ } \varepsilon\text{-rationalizes } \{(x^t, p^t)\}_{t=1}^T\}} \{V_u(\tilde{p}^1) - 0\},$$

where the supremum is over u such that $\underline{V}_{u,A}(\tilde{p}^0, \varepsilon) = 0$. We know that each V_u is convex, weakly decreasing, and lower semicontinuous by Lemma C.6. Note that this is true for any u , regardless of whether it is concave or upper semicontinuous.

We conclude that when viewing $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ only as a function of \tilde{p}^1 , it is the supremum (over u) of convex, weakly decreasing, lower semicontinuous functions. It is therefore convex, weakly decreasing, and lower semicontinuous in \tilde{p}^1 from Rockafellar [2015],

Theorems 5.5 and 9.4.

To see that $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is weakly increasing in ε , we recall the characterization as a linear program in Proposition A.3. The feasibility region of the linear program is weakly increasing (with regard to set inclusion) in ε , and so the value function of the problem is weakly increasing in ε .

We now establish the finiteness properties in the proposition. Let $\tilde{p}^1 \in \text{CCo}(\{p^t\}_{t=1}^T)$. We show $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon)$ is finite. We can write

$$\tilde{p}^1 \geq \sum_{t=1}^T \alpha_t p^t$$

for some nonnegative $\alpha_1, \dots, \alpha_T$ that sum to 1 where the inequality holds componentwise. We have

$$\begin{aligned} V_u(\tilde{p}^1) &\leq V_u \left(\sum_{t=1}^T \alpha_t p^t \right) \\ &\leq \sum_{t=1}^T \alpha_t V_u(p^t), \end{aligned}$$

where the first inequality follows because V_u is weakly decreasing, and the second inequality follows because V_u is convex. Recall from (5) in the main text that for any u that ε -rationalizes the dataset, we have

$$\bar{V}_{u,A}(p^t, \varepsilon) - \underline{V}_{u,A}(\tilde{p}^0, \varepsilon) \leq V_u(p^t) - V_u(\tilde{p}^0) + \varepsilon \leq (x^t \cdot (\tilde{p}^0 - p^t) + \varepsilon) + \varepsilon$$

for any $t \in \{1, \dots, T\}$. Combining the previous steps we obtain

$$\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) \leq \sum_{t=1}^T \alpha_t x^t \cdot (\tilde{p}^0 - p^t) + 2\varepsilon < \infty.$$

Now we show that if $\tilde{p}^1 \notin \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$, then $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) = \infty$. First note that from Proposition 2, there is some utility function that ε -rationalizes the dataset and has a maximizer at the price \tilde{p}^0 . Let \tilde{x}^0 denote such a maximizer. Now construct the augmented dataset $\{(x^t, p^t)\}_{t=1}^{T+1}$, where $(x^{T+1}, p^{T+1}) = (\tilde{x}^0, \tilde{p}^0)$. Note that by construction, there is some utility function u that ε -rationalizes the augmented dataset

$\{(x^t, p^t)\}_{t=1}^{T+1}$. For any such function u , the indirect utility satisfies $V_u(p) < \infty$ for any $p \in \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$ from the finiteness arguments above. Thus, it remains to show that when $p \notin \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$, there is some u that ε -rationalizes the augmented dataset and satisfies $V_u(p) = \infty$. That end, fix $(x^1, p^1) \in \{(x^t, p^t)\}_{t=1}^{T+1}$ and let Σ_1 denote the set of finite sequences of $t \in \{1, \dots, T\}$ with no cycles that begins at $\sigma(1) = 1$. Define

$$U(x) = \min_{\sigma \in \Sigma_1} \left\{ p^{\sigma(M)} \cdot (x - x^{\sigma(M)}) + \sum_{m=1}^{M-1} p^{\sigma(m)} \cdot (x^{\sigma(m+1)} - x^{\sigma(m)}) + M\varepsilon \right\},$$

where M corresponds to the length of a particular sequence. Allen and Rehbeck [2020] have shown that for $\varepsilon \geq \varepsilon^*$, this function ε -rationalizes the augmented dataset $\{(x^t, p^t)\}_{t=1}^{T+1}$.

Since $\tilde{p}^1 \notin \text{CCo}(\{p^t\}_{t=1}^T, \tilde{p}^0)$ from the separating hyperplane theorem, there is some $x \in \mathbb{R}^K$ with $x \neq 0$ such that

$$(p - \tilde{p}^1) \cdot x > 0 \quad \text{for all } p \in \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0).$$

We argue by contradiction that x contains no negative components. Indeed, suppose it does so that $x_k < 0$ for some k . Since $\tilde{p}^1 \in \mathbb{R}_{++}^K$ and $\text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$ is upper comprehensive, we can find some $p \in \text{CCo}(\{p^t\}_{t=1}^T \cup \tilde{p}^0)$ with p_k high enough so that $(p - \tilde{p}^1) \cdot x < 0$. We reach a contradiction and conclude $x \in \mathbb{R}_+^K$ and $x \neq 0$.

Note that in the definition of $U(x)$, the minimum is taken over certain functions that involve $p^t \cdot x$ for some t , plus a constant. Thus, for shorthand write

$$U(x) = \min_{t \in \{1, \dots, T+1\}} \min_{a \in A^t} \{p^t \cdot x + a\}$$

for certain finite sets A^t corresponding to the sums in the construction of $U(x)$. We conclude that

$$\lim_{\lambda \rightarrow \infty} U(\lambda x) - \tilde{p}^1 \cdot \lambda x = \lim_{\lambda \rightarrow \infty} \min_{t \in \{1, \dots, T+1\}} \min_{a \in A^t} \{(p^t - \tilde{p}^1) \cdot \lambda x + a\} = \infty.$$

This establishes that $V_U(\tilde{p}^1) = \infty$ and so since $V_U(\tilde{p}^0) < \infty$ we conclude $\bar{V}(\tilde{p}^1, \tilde{p}^0, \varepsilon) \geq V_U(\tilde{p}^1) - V_U(\tilde{p}^0) = \infty$.

□

A.4 Proofs for Section 5

In order to prove Proposition 14, we first prove a lemma. The lemma establishes a convexity property of the set of counterfactual quantities at a given price,

$$X(\tilde{p}, D, \varepsilon) = \{\tilde{x} \mid (\tilde{x}, \tilde{p}) \in C(D, \varepsilon)\}.$$

Lemma A.1. *Let $D^0 = \{(d^{0,t}, p^t)\}_{t=1}^T$ and $D^1 = \{(d^{1,t}, p^t)\}_{t=1}^T$ differ only for quantities. If $\tilde{d}^j \in X(\tilde{p}, D^j, \varepsilon^j)$ for $j \in \{0, 1\}$, then for any $\alpha \in [0, 1]$,*

$$\alpha\tilde{d}^0 + (1 - \alpha)\tilde{d}^1 \in X(\tilde{p}, \alpha D^0 + (1 - \alpha)D^1, \alpha\varepsilon^0 + (1 - \alpha)\varepsilon^1).$$

In addition, the set A described in Proposition 14 is convex.

Proof. The sets $C(D, \varepsilon)$ and $X(\tilde{p}, D, \varepsilon)$ can be characterized by using any of the equivalent statements of Lemma C.1, applied to the counterfactual-augmented dataset. In particular, for an arbitrary (hypothetical) dataset $D^j = \{(d^{j,t}, p^t)\}_{t=1}^T$, $\tilde{d}^j \in X(\tilde{p}, D^j, \varepsilon^j)$ means there are numbers $u^{j,1}, \dots, u^{j,K}, \tilde{u}^j \in \mathbb{R}_+$ such that

$$\begin{aligned} u^{j,s} &\leq u^{j,r} + p^r \cdot (d^{j,s} - d^{j,r}) + \varepsilon && \text{for all } r, s \in \{1, \dots, T\} \\ u^j &\leq u^{j,r} + p^r \cdot (\tilde{d}^j - d^{j,r}) + \varepsilon && \text{for all } r \in \{1, \dots, T\} \\ u^{j,r} &\leq \tilde{u}^j + \tilde{p} \cdot (d^{j,r} - \tilde{d}^j) + \varepsilon && \text{for all } r \in \{1, \dots, T\}. \end{aligned}$$

In addition, \tilde{d}^j must be non-negative. We can take a convex combination of the values for $j = 0$ and $j = 1$ and the inequalities are preserved. For example, considering the first inequalities that involve a pair r, s , we have

$$\begin{aligned} \alpha u^{0,s} + (1 - \alpha)u^{1,s} &\leq \alpha u^{0,r} + (1 - \alpha)u^{1,r} + \\ &p^r \cdot (\alpha d^{0,s} + (1 - \alpha)d^{1,s} - (\alpha d^{0,r} + (1 - \alpha)d^{1,r})) + \alpha\varepsilon^0 + (1 - \alpha)\varepsilon^1. \end{aligned}$$

Since by Lemma C.1 the inequalities displayed above are the only ones we need to

check, we obtain that

$$\alpha\tilde{d}^0 + (1 - \alpha)\tilde{d}^1 \in X(\tilde{p}, \alpha D^0 + (1 - \alpha)D^1, \alpha\varepsilon^0 + (1 - \alpha)\varepsilon^1).$$

Finally, convexity of A follows from similar averaging of the inequalities characterizing ε -rationalizability in Lemma C.1(iii). \square

Proof of Proposition 14. Recall we now include quantities as arguments of the bounds. In more detail, write

$$\bar{x}_k(d, \tilde{p}, \varepsilon) = \max_{\tilde{x} \in X(\tilde{p}, D, \varepsilon)} \tilde{x}_k,$$

where $D = \{(d^t, p^t)\}_{t=1}^T$ and we work in the extended reals so that $\bar{x}_k(d, \tilde{p}, \varepsilon)$ may be ∞ . Addition is defined as $\infty + a = \infty$ provided a is not $-\infty$.

Let $\tilde{d}^0, \tilde{d}^1 \in \mathbb{R}^{K \times T}$ be arbitrary quantities datasets. Since \bar{x}_k is a maximum, we obtain

$$\bar{x}_k\left(\alpha\tilde{d}^0 + (1 - \alpha)\tilde{d}^1, \tilde{p}, \alpha\varepsilon^0 + (1 - \alpha)\varepsilon^1\right) \geq \alpha\bar{x}_k\left(\tilde{d}^0, \tilde{p}, \varepsilon^0\right) + (1 - \alpha)\bar{x}_k\left(\tilde{d}^1, \tilde{p}, \varepsilon^1\right),$$

because from Lemma A.1 the weighted average of the values is feasible. This establishes concavity of \bar{x}_k in its non-price arguments. Since \underline{x}_k is a minimum we obtain

$$\underline{x}_k\left(\alpha\tilde{d}^0 + (1 - \alpha)\tilde{d}^1, \tilde{p}, \alpha\varepsilon^0 + (1 - \alpha)\varepsilon^1\right) \leq \alpha\underline{x}_k\left(\tilde{d}^0, \tilde{p}, \varepsilon^0\right) + (1 - \alpha)\underline{x}_k\left(\tilde{d}^1, \tilde{p}, \varepsilon^1\right),$$

and so the lower bound is convex in its non-price arguments.

To establish continuity of $\bar{x}_k(\cdot, \tilde{p}, \cdot)$ as stated in the proposition, recall the linear programming formulation in A.1. We see that quantities d and approximation error ε enter additively (relative to the choice variables) in the inequalities describing feasibility region. That is, they are part of the “ b ” in the canonical linear programming formulation from Appendix C.2. Continuity then follows from Lemma C.4. \square

Proof of Proposition 15. The proof is analogous to the proof of Proposition 14 and so we only outline it. Recall that \bar{V} and \underline{V} are described by a linear program in Proposition 10. By inspecting the feasibility region of this program we see convexity holds similar to Lemma A.1. From this, we conclude that \bar{V} satisfies the concavity

property in the proposition because it is a maximum, and \bar{V} satisfies the convexity property in the proposition because it is a minimum.

Continuity of \bar{V} and \underline{V} in (d, ε) over the region where these bounds are finite follows from the linear programming formulation in Proposition A.3 and Lemma C.4. \square

Proof of Corollary 1. By assumption $\bar{x}_k(d, \tilde{p}, \varepsilon^*(d))$ is finite. By construction, $(\hat{d}^n, \varepsilon(\hat{d}^n)) \in A$ for each n . Then from Propositions 14, 16, and the continuous mapping theorem, $\bar{x}(\hat{d}^n, \tilde{p}, \varepsilon(\hat{d}^n)) \xrightarrow{p} \bar{x}_k(d, \tilde{p}, \varepsilon^*(d))$. The arguments for \bar{x} , \underline{V} , and \bar{V} are analogous.

\square

Proof of Proposition 17. First note that the set $A(\tilde{x})$ is convex for any $\tilde{x} \in \mathbb{R}_+^K$ because it is a projection of the set A , which is convex by Lemma A.1. The proof is analogous to the proofs of Propositions 14 and 15. The feasibility region of the program describing \bar{u} and \underline{u} is given in Proposition A.2. This feasibility regions of \bar{u} and \underline{u} are convex in quantities and degree of approximation error over the sets $A(\tilde{x}^0)$ and $A(\tilde{x}^1)$ respectively, similar to Lemma A.1. Since \bar{u} is a supremum it is concave, and since \underline{u} is an infimum it is convex.

Continuity of \bar{u} and \underline{u} in (d, ε) over the region stated in the proposition follows from the characterization of the bounds in Proposition A.2(ii) and (iii). Indeed, \bar{u} is the maximum of finitely many functions that are each affine in (d, ε) , and \underline{u} is the minimum. \square

Appendix B Alternative Approaches

Assumption 1 is the key conceptual assumption for this paper, which posits that approximation error is the same in new settings as the data we have seen. We have operationalized this for counterfactual and welfare analysis with a number controlling approximation error as in Allen and Rehbeck [2020]. We now describe other potential ways to conduct counterfactual or welfare analysis. We also elaborate on the measurement and prediction wedges.

This paper focuses on approximation error being controlled by a single scalar. An alternative approach is to consider a multidimensional notion along the lines of Afriat [1972], Varian [1990], Varian [1991], Halevy et al. [2018], and Masten and Poirier [2018b]. We pursue this by allowing each observation to have its own value ε_V^t of approximation error relative to exact optimization.

Definition B.1. A dataset $\{(x^t, p^t)\}_{t=1}^T$ is ε_V -rationalized by quasilinear utility for $\varepsilon_V = (\varepsilon_V^1, \dots, \varepsilon_V^T) \in \mathbb{R}_+^T$ if there exists a utility function $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ such that for all $t \in \{1, \dots, T\}$ and for all $x \in \mathbb{R}_+^K$, the following inequality holds:

$$u(x^t) - p^t \cdot x^t + \geq u(x) - p^t \cdot x - \varepsilon_V^t.$$

We also refer to the above by saying a dataset is ε_V -quasilinear rationalized.

We can apply this concept to counterfactual analysis, as in the main text, by considering datasets in which the last observation is the hypothetical. That is, for a dataset $\{(x^t, p^t)\}_{t=1}^{T+1}$, interpret the first T observations as data we have seen and the last $T + 1$ observation as be the hypothetical. In this case, the *measurement wedges* are controlled by the collection $(\varepsilon_V^1, \dots, \varepsilon_V^T)$ of values for the observed data, while the *prediction wedge* is the scalar ε_V^{T+1} . In principle we can consider counterfactuals involving several observations such as $T + 1$ and $T + 2$. We focus on the case of a single counterfactual for brevity.

In the main text, we set a single number controlling the measurement wedge and the prediction wedge. The full strength of Assumption 1' also imposes that these are equal to the minimal approximation error needed to explain the data we have seen. In this appendix we drop the assumptions that these wedges are the same, which shows how to generalize our framework when Assumption 1' is relaxed.

B.1 Counterfactuals with Multidimensional Approximation Error

Because ε_V -rationalization is multidimensional, in general there is no single “smallest” vector ε_V such that the dataset is ε_V -rationalized by quasilinear utility. Nonetheless,

we can define a set of such rationalizing vectors via

$$E(\tilde{D}) = \left\{ \varepsilon_V \in \mathbb{R}_+^T \mid \tilde{D} \text{ is } \varepsilon_V\text{-quasilinear rationalized} \right\},$$

where we let $\tilde{D} = ((x^1, p^1), \dots, (x^T, p^T))$ be the observed dataset written as an ordered tuple. Note that we switch from an unordered dataset to an ordered tuple. The reason we care about the order of observations now is that the t -th dimension of $E(\tilde{D})$ corresponds to approximation error associated with the t -th observation. We can conduct counterfactual analysis as before by considering the set of quantity-price tuples that do not make approximation error worse.

Formalizing worse here leads to some ambiguity since we do not have a total order on vectors. We consider two possibilities. To formalize these, let $\pi_T : \mathbb{R}^{T+1} \rightarrow \mathbb{R}^T$ denote the projection onto the first T components. We can then define the sets of lower and upper approximate counterfactuals by

$$\begin{aligned} \underline{AC}(\tilde{D}) &= \left\{ (\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid \pi_T(E(\tilde{D} \times (\tilde{x}, \tilde{p}))) = E(\tilde{D}) \right\} \\ \overline{AC}(\tilde{D}) &= \left\{ (\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid \pi_T(E(\tilde{D} \times (\tilde{x}, \tilde{p}))) \cap E(\tilde{D}) \neq \emptyset \right\}. \end{aligned}$$

With minor abuse of notation we define $E(\cdot)$ in the obvious way for datasets of different dimensions. Clearly, $\underline{AC}(\tilde{D}) \subseteq \overline{AC}(\tilde{D})$. The smaller set formalizes that counterfactuals do not change the potential ε_V vectors that rationalize the data we see when we add an existing observation. This smaller set is conceptually closer to the original adaptive counterfactual set $AC(\cdot)$. The larger set formalizes that there is *some* ε_V vector that ε_V -rationalizes both the original dataset and the counterfactual-augmented dataset $\tilde{D} \times (\tilde{x}, \tilde{p})$.

While the sets \underline{AC} and \overline{AC} may appear to be intuitive alternatives to the adaptive approach presented in the main text, unfortunately these sets are trivial. To see this, note that the set $\underline{AC}(\tilde{D})$ allows the prediction wedge ε_V^{T+1} for the counterfactual value to be unbounded. This leads to trivial restrictions for both $\underline{AC}(\tilde{D})$ and the larger set $\overline{AC}(\tilde{D})$. However, nontriviality can be restored if we modify the sets by placing an *a priori* bound on approximation error at the new observation. To formalize this, let $\bar{\pi} : \mathbb{R}^{T+1} \rightarrow \mathbb{R}$ denote the projection of the last component. Then

we can modify the smaller set via

$$\underline{AC}'\left(\tilde{D}, \varepsilon_V^{T+1}\right) = \left\{(\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid \pi_T\left(E\left(\tilde{D} \times (\tilde{x}, \tilde{p})\right)\right) = E\left(\tilde{D}\right), \right. \\ \left. \bar{\pi}\left(E\left(\tilde{D} \times (\tilde{x}, \tilde{p})\right)\right) \leq \varepsilon_V^{T+1}\right\}.$$

Thus, the prediction wedge is restricted by the number ε_V^{T+1} . This set is not data-adaptive in the sense that ε_V^{T+1} needs to be chosen by the researcher. However, one can make this data adaptive by using information obtained from other measures of approximation error discussed below.

B.2 Other Measures of Approximation Error

It is natural to wonder for the multidimensional vector of approximation errors, ε_V , whether other intuitive one-dimensional summaries can be used for counterfactual analysis. In fact, there can be many ways to do this depending on how one aggregates the approximation error. We consider general aggregators of the elements of multidimensional approximation error ε_V that turn it into a one-dimensional measure of approximation error. Formally, an aggregator can be written $e^T : \mathbb{R}_+^T \rightarrow [0, \infty)$. Higher values of the aggregator can be interpreted as more approximation error. Varian [1990] and Halevy et al. [2018] consider a related notion in the standard consumer problem for general utility maximization.²⁸

Given a dataset $\tilde{D} = ((x^1, p^1), \dots, (x^T, p^T))$ and an aggregator e^T , we can define a measure of approximation error via

$$e^{T*}\left(\tilde{D}\right) = \inf_{\varepsilon_V \in E(\tilde{D})} e^T(\varepsilon_V).$$

One such aggregator is the max aggregator of $e_M^T(\varepsilon_V) = \max_{t \in T} \varepsilon_V^t$. The measure of approximation error for the max aggregator agrees with the one presented in the main text, i.e. $e_M^{T*} = \varepsilon^*$. In general, an aggregator can depend on the sample size. For example, consider the average approximation error aggregator $e_A^T(\varepsilon_V) = \frac{1}{T} \sum_{t=1}^T \varepsilon_V^t$. Our leading measure, ε^* , does not depend on T .

²⁸Many of the convenient shape restrictions we obtain in this paper do not hold for the case of general utility maximization since the constraint set of consistent utility indices is non-convex.

For an arbitrary aggregator, similar to how AC was constructed with the measure ε^* , we can define a set of counterfactuals such that approximation error does not get worse:

$$\left\{ (\tilde{x}, \tilde{p}) \in \mathbb{R}_+^K \times \mathbb{R}_{++}^K \mid e^{(T+1)*} \left(\tilde{D} \times (\tilde{x}, \tilde{p}) \right) \leq e^{T*} \left(\tilde{D} \right) \right\}. \quad (13)$$

This construction does not separately control the prediction wedge and approximation wedge as in \underline{AC}' described at the end of the previous subsection. Instead, it lumps together both prediction and approximation wedges via the aggregators $e^{(T+1)*}$ and e^{T*} .

To understand properties of this set, consider an aggregator that sums up the observation-specific bounds on approximate optimization, $e_S^T(\varepsilon_V) = \sum_{t=1}^T \varepsilon_V^t$. With this choice of aggregator, each conjectured observation in (13) must be perfectly consistent with the model, i.e. $\varepsilon_V^{T+1} = 0$, for approximation error to not be made worse. In other words, for each element of (13), there must exist some utility function that approximately explains the existing dataset \tilde{D} , but *exactly* explains the counterfactual. Thus, there is no prediction wedge. This property may be desirable when one thinks the observed dataset \tilde{D} comes from a “true” dataset that is generated by the quasilinear model but has been measured with error. Using the previous terminology, in this case we may wish to conduct counterfactual analysis without a prediction wedge. If instead we think approximation error propagates to new settings, then we may wish to allow the prediction wedge. This is one motivation for ε^* and the adaptive set AC .

One potential way to address this limitation of the sum-type aggregator e_S^T is to adjust it by dividing by the sample size to obtain the average approximation error aggregator so that $\tilde{e}_A^{T*} = \frac{1}{T} e_S^{T*}$. This division allows one to construct a set analogously to (13) that allows a prediction wedge when generating counterfactual information.

Appendix C Supplemental Appendix

This appendix contains additional results needed for proofs of the main results. Section C.1 contains miscellaneous lemmas, Section C.2 presents lemmas specifically for linear programming results, and Section C.3 presents duality results used in proofs

for approximate indirect utility in Section 4.2.

C.1 Miscellaneous Lemmas

Lemma C.1 (Allen and Rehbeck [2020]). *For any dataset $\{(x^t, p^t)\}_{t=1}^T$ and $\varepsilon \geq 0$, the following are equivalent:*

(i) $\{(x^t, p^t)\}_{t=1}^T$ is ε -rationalized by quasilinear utility.

(ii) There exist numbers $\{u^t\}_{t=1}^T$ that satisfy the following inequalities for all $r, s \in \{1, \dots, T\}$:

$$u^s \leq u^r + p^r \cdot (x^s - x^r) + \varepsilon.$$

(iii) For all finite sequences $\{t_m\}_{m=1}^M$ with $t_m \in \{1, \dots, T\}$ and $M \geq 2$, the inequality

$$\frac{1}{M} \sum_{m=1}^M p^{t_m} \cdot (x^{t_m} - x^{t_{m+1}}) \leq \varepsilon$$

holds, where $x^{t_{M+1}} = x^{t_1}$.

We require a lemma that will be used in the proof of Proposition 2 to ensure a maximizer exists. In contrast with models with compact budget constraints, continuity of the utility function u is *not* enough to ensure a maximizer exists, which is why we require the following lemma. To state the lemma, recall that for a utility function u , the indirect utility is defined as

$$V_u(p) = \sup_{x \in \mathbb{R}_+^K} u(x) - p \cdot x.$$

Lemma C.2. *Suppose $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ is concave, monotonically increasing, and continuous. Moreover, suppose $V_u(p)$ is finite over some open set $O \subseteq \mathbb{R}^K$. It follows that for any price $p \in \text{ri}(\text{Co}(O))$,*²⁹

$$u(x) - p \cdot x$$

admits a maximizer for $x \in \mathbb{R}_+$.

²⁹For a set $S \subseteq \mathbb{R}^K$, $\text{ri}(S)$ gives the relative interior of the set S as defined in Rockafellar [2015].

Proof of Lemma C.2. Since $V_u(p)$ is convex and finite over O , then $V_u(p)$ is finite on the $\text{ri}(\text{Co}(O))$. Thus, the subdifferential

$$\partial V_u(p) = \{x \mid V_u(\tilde{p}) \geq V_u(p) + x \cdot (\tilde{p} - p) \quad \forall \tilde{p} \in \mathbb{R}^K\}$$

is nonempty for any $p \in \text{ri}(\text{Co}(O))$ by Rockafellar [2015], Theorem 23.4. Extend u to all of \mathbb{R}^K by setting $u(x) = -\infty$ for any $x \in \mathbb{R}^K \setminus \mathbb{R}_+^K$. Recall that the original u defined on \mathbb{R}_+^K is continuous, so that it is upper semicontinuous and $\{x \mid u(x) \geq a\}$ is closed for any $a \in \mathbb{R}$ by Theorem 7.1 in Rockafellar [2015]. Note that the extension is also upper semicontinuous because it does not change the topological properties of the upper contour sets for all $a \in \mathbb{R}$. Since u is upper semicontinuous and concave, we conclude from Rockafellar [2015] Theorem 23.5 parts (b) and (a*) that for any $p \in \text{ri}(\text{Co}(O))$ there is some $x^* \in \mathbb{R}^K$ such that

$$u(x) - p \cdot x$$

is maximized over $x \in \mathbb{R}$ at x^* since $\partial V_u(p)$ is nonempty. Since u is $-\infty$ outside of \mathbb{R}_+^K , we conclude $x^* \in \mathbb{R}_+^K$. This establishes existence of an exact maximizer for the utility function u over the region $p \in \text{ri}(\text{Co}(O))$, which completes the proof. \square

C.2 Linear Programming Lemmas

We require some existing results from the theory of linear programming. In canonical form, a linear program is written as

$$\begin{aligned} & \max_{x \in \mathbb{R}^{J_1}} c \cdot x \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0. \end{aligned}$$

Here, $c, x \in \mathbb{R}^{J_1}$, and $b \in \mathbb{R}^{J_2}$ are vectors and $A \in \mathbb{R}^{J_2 \times J_1}$ is a matrix.

This is written as a maximum rather than a supremum because provided the supremum is finite, the maximum is attained as we formalize now.

Lemma C.3. *If the value function of a linear program is finite, then the maximum*

is attained.

Proof. See e.g. Bertsekas [2009], Proposition 1.4.12. \square

Fixing all other variables, let $B \subseteq \mathbb{R}^{J_2}$ be the set of b where the linear program is bounded. Write the value function as a function of b so that

$$\begin{aligned} G(b) = \sup_{x \in \mathbb{R}^{J_1}} c \cdot x \\ \text{s.t.} \quad Ax \leq b \\ x \geq 0. \end{aligned}$$

Lemma C.4. Let $b_m \rightarrow b^*$ where $b^* \in B$ and for each b_m , the set

$$\{x \mid Ax \leq b_m, x \geq 0\}$$

is nonempty. It follows that $G(b_m) \rightarrow G(b^*)$.

Proof. Let $\eta > 0$ and define

$$\bar{G}(b) = \min\{G(b), G(b^*) + \eta\}.$$

The value $\bar{G}(b)$ is the value function of a linear program defined by $G(b)$ that appends the inequality constraint $c \cdot x \leq G(b^*) + \eta$. Because $G(b^*) + \eta$ is finite, $\bar{G}(b)$ is finite for any feasible b . Böhm [1975], Theorem 1 states that $\bar{G}(b)$ is continuous over the set of b such that the feasibility region is nonempty. Since $b_m \rightarrow b^*$ and each b_m and $b^* \in B$ are feasible, we then obtain $\bar{G}(b_m) \rightarrow \bar{G}(b^*) = G(b^*)$. \square

C.3 Duality

The focus of the paper is on counterfactuals with approximate utility maximization. In other words, we consider utility functions u such that the inequality

$$u(x^t) - p^t \cdot x^t \geq u(x) - p^t \cdot x - \varepsilon$$

holds for every $t \in \{1, \dots, T\}$ and $x \in \mathbb{R}_+^K$. In this supplement, we consider a dual approach involving functions V such that

$$V(p) \geq V(p^t) - x^t \cdot (p - p^t) - \varepsilon$$

holds for every $t \in \{1, \dots, T\}$ and $p \in \mathbb{R}_{++}^K$. We also mention some results from convex analysis. Results from this section are used to prove several results in the main text.

Our first result formalizes that finding a utility function u that satisfies the first set of inequalities is equivalent to finding a V function that satisfies the second set of inequalities.

To state the result, recall the indirect utility function of u is given by $V_u : \mathbb{R}_+^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$

$$V_u(p) = \sup_{x \in \mathbb{R}_+^K} u(x) - p \cdot x.$$

We make use of a “dual” utility function $u_V : \mathbb{R}_+^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$ constructed via

$$u_V(x) = \inf_{p \in \mathbb{R}_+^K} V(x) + p \cdot x.$$

These operations can be defined for any extended real-valued functions $u : \mathbb{R}^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$ and $V : \mathbb{R}^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$.

Proposition C.1. *Let $\varepsilon \geq 0$ and let $\{(x^t, p^t)\}_{t=1}^T$ be an arbitrary dataset.*

i. Suppose $u : \mathbb{R}_+^K \rightarrow \mathbb{R}$ satisfies

$$u(x^t) - p^t \cdot x^t \geq u(x) - p^t \cdot x - \varepsilon$$

for every $t \in \{1, \dots, T\}$ and every $x \in \mathbb{R}_+^K$. It follows that V_u satisfies

$$V_u(p) \geq V_u(p^t) - x^t \cdot (p - p^t) - \varepsilon$$

for every $t \in \{1, \dots, T\}$ and every $p \in \mathbb{R}_+^K$.

ii. Suppose $V : \mathbb{R}_+^K \rightarrow \mathbb{R}$ satisfies

$$V(p) \geq V(p^t) - x^t \cdot (p - p^t) - \varepsilon$$

for every $t \in \{1, \dots, T\}$ and every $p \in \mathbb{R}_+^K$. It follows that u_V satisfies

$$u_V(x^t) - p^t \cdot x^t \geq u_V(x) - p^t \cdot x - \varepsilon$$

for every $t \in \{1, \dots, T\}$ and every $x \in \mathbb{R}_+^K$.

Proof. First we show (i). For arbitrary $t \in \{1, \dots, T\}$, write

$$V_u(p^t) = u(x^t) - p^t \cdot x^t + \delta^t \quad (14)$$

where $\delta^t = V_u(p^t) - u(x^t) + p^t \cdot x_t \geq 0$ and $\delta^t \leq \varepsilon$ since the observed quantities are only approximately optimal. For arbitrary $p \in \mathbb{R}_+^K$ we have

$$V_u(p) \geq u(x^t) - p \cdot x^t.$$

Differencing yields

$$V_u(p) - V_u(p^t) \geq -x^t \cdot (p - p^t) - \delta^t.$$

The term $V_u(p)$ may equal ∞ , in which case we define $\infty - a = \infty$ for any finite a . Here, $V_u(p^t)$ is finite from (14) and the fact that u is finite. We know for all $t \in \{1, \dots, T\}$ that $0 \leq \delta^t \leq \varepsilon$ by assumption, so (i) is established.

Now we show (ii). As before, write

$$u_V(x^t) = V(p^t) + p^t \cdot x^t - \delta^t$$

where $\delta^t = u_V(x^t) - V(p^t) - p^t \cdot x^t \geq 0$ and $\delta \leq \varepsilon$ since the observed quantities and prices are only supposed to satisfy the inequality in (ii) for V . For arbitrary $x \in \mathbb{R}_+^K$ we have

$$u_V(x) \leq V(p^t) + p^t \cdot x,$$

and so

$$u_V(x^t) - p^t \cdot x^t \geq u_V(x) - p^t \cdot x - \delta^t.$$

As before, $u_V(x)$ can equal $-\infty$, but $u_V(x^t)$ is always finite. Recall, for all $t \in \{1, \dots, T\}$ that $0 \leq \delta^t \leq \varepsilon$ by assumption, and so (ii) is established. \square

The mappings $u \rightarrow V_u$ and $V \rightarrow u_V$ are closely related to convex conjugates, and we

can adapt existing results from convex analysis. Recall that for a function $f : \mathbb{R}^K \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$, the *convex conjugate* is given by

$$f^*(p) = \sup_{x \in \mathbb{R}^K} p \cdot x - f(x).$$

The *monotone conjugate* is given by

$$f^+(p) = \sup_{x \in \mathbb{R}_+^K} p \cdot x - f(x).$$

Let the function \tilde{f} equal f over $x \in \mathbb{R}_+^K$, and ∞ otherwise. It follows that $\tilde{f}^*(p) = f^+(p)$.

We now formalize the relationships between u_V and V_u and monotone conjugates. Following this, we present some immediate consequences.

Lemma C.5.

$$u_V(x) = -V^+(-x)$$

and

$$V_u(p) = (-u)^+(-p).$$

Proof.

$$u_V(x) = \inf_{p \in \mathbb{R}_+^K} V(p) + p \cdot x = -\sup_{p \in \mathbb{R}_+^K} -V(p) - p \cdot x = -V^+(-x).$$

and

$$V_u(p) = \sup_{x \in \mathbb{R}_+^K} u(x) - p \cdot x = \sup_{x \in \mathbb{R}_+^K} -(-u(x)) + (-p) \cdot x = (-u)^+(-p).$$

□

Lemma C.6. *The function u_V is concave, weakly increasing, and upper semicontinuous. The function V_u is convex, weakly decreasing, and lower semicontinuous.*

Proof. To see that V_u is weakly decreasing, consider p^a, p^b with $p^a \geq p^b$. For $x \in \mathbb{R}_+^K$ we have $p^a \cdot x \geq p^b \cdot x$ and so

$$V_u(p^a) = \sup_{x \in \mathbb{R}_+^K} u(x) - p^a \cdot x \leq \sup_{x \in \mathbb{R}_+^K} u(x) - p^b \cdot x \leq V_u(p^b).$$

Also, V_u is convex and lower semicontinuous from Bertsekas [2009], p. 83. The arguments for u_V are analogous by applying Lemma C.5. \square

Lemma C.7. *i. Suppose $u : \mathbb{R}_+^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$ is concave, weakly increasing, upper semicontinuous, and finite at 0. Then $u_{V_u} = u$.*

ii. Suppose $V : \mathbb{R}_+^K \rightarrow \mathbb{R}^K \cup \{-\infty, \infty\}$ is concave, weakly decreasing, lower semicontinuous, and finite at 0. Then $V_{u_V} = V$.

Proof. Write f^{++} as the monotone conjugate of f^+ . Rockafellar [2015], Theorem 12.4 states $V^{++} = V$ and $(-u)^{++} = -u$. From Lemma C.5 we conclude

$$-u_{V_u}(x) = (V_u)^+(-x) = (-u)^{++}(x) = -u(x)$$

and

$$V_{u_V}(p) = (-u_V)^+(-p) = V^{++}(p) = V(p).$$

\square

References

- Abi Adams. Mutually consistent revealed preference demand predictions. *American Economic Journal: Microeconomics*, 2019. Forthcoming.
- Sidney N Afriat. The construction of utility functions from expenditure data. *International economic review*, 8(1):67–77, 1967.
- Sidney N Afriat. Efficiency estimation of production functions. *International economic review*, pages 568–598, 1972.
- Sidney N Afriat. On a system of inequalities in demand analysis: an extension of the classical method. *International Economic Review*, pages 460–472, 1973.
- Victor Aguiar, Per Hjertstrand, and Roberto Serrano. A rationalization of the weak axiom of revealed preference. 2020.
- Victor H. Aguiar and Nail Kashaev. Stochastic revealed preferences with measurement error. 2018. Working Paper.

- Roy Allen and John Rehbeck. Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054, 2019a.
- Roy Allen and John Rehbeck. Measuring rationality: Percentages vs expenditures. Available at SSRN 3399065, 2019b.
- Roy Allen and John Rehbeck. Satisficing, aggregation, and quasilinear utility. 2020.
- Donald WK Andrews and Soonwoo Kwon. Inference in moment inequality models that is robust to spurious precision under model misspecification. 2019.
- Isaiah Andrews, Matthew Gentzkow, and Jesse M Shapiro. Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592, 2017.
- Timothy Armstrong and Michal Kolesár. Sensitivity analysis using approximate moment condition models. 2018.
- B Douglas Bernheim. The good, the bad, and the ugly: a unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis*, 7(1):12–68, 2016.
- B Douglas Bernheim and Antonio Rangel. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104, 2009.
- B Douglas Bernheim and Dmitry Taubinsky. Behavioral public economics. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 381–516. Elsevier, 2018.
- Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.
- Dimitri P Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, 2009.
- Richard Blundell and James L Powell. Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357, 2003.
- Richard Blundell, Martin Browning, and Ian Crawford. Best nonparametric bounds on demand responses. *Econometrica*, 76(6):1227–1262, 2008.

- Richard Blundell, Joel L Horowitz, and Matthias Parey. Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, 3(1):29–51, 2012.
- Richard Blundell, Dennis Kristensen, and Rosa Matzkin. Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics*, 179(2):112–127, 2014.
- Richard W Blundell, Martin Browning, and Ian A Crawford. Nonparametric engel curves and revealed preference. *Econometrica*, 71(1):205–240, 2003.
- Richard W Blundell, Dennis Kristensen, and Rosa Liliana Matzkin. Individual counterfactuals with multidimensional unobserved heterogeneity. Technical report, cemmap working paper, 2017.
- Volker Böhm. On the continuity of the optimal policy set for linear programs. *SIAM Journal on Applied Mathematics*, 28(2):303–306, 1975.
- Stéphane Bonhomme and Martin Weidner. Minimizing sensitivity to model misspecification. *arXiv preprint arXiv:1807.02161*, 2018.
- Donald J Brown and Caterina Calsamiglia. The nonparametric approach to applied welfare analysis. *Economic Theory*, 31(1):183–188, 2007.
- M Kate Bundorf, Jonathan Levin, and Neale Mahoney. Pricing and welfare in health plan choice. *American Economic Review*, 102(7):3214–48, 2012.
- Christopher P Chambers and Federico Echenique. A characterization of combinatorial demand. *Mathematics of Operations Research*, 2017.
- Laurens Cherchye, Thomas Demuynck, and Bram De Rock. Bounding counterfactual demand with unobserved heterogeneity and endogenous expenditures. *Journal of Econometrics*, 2019.
- Raj Chetty. Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018, 2012.
- K Chiong, YW Hsieh, and Matthew Shum. Counterfactual estimation in semiparametric discrete choice models. URL <https://ssrn.com/abstract=2979446>, 2017.

- Timothy M Christensen and Benjamin Connault. Counterfactual sensitivity and robustness. *arXiv preprint arXiv:1807.02161*, 2018.
- Jessica Cohen, Pascaline Dupas, et al. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Quarterly journal of Economics*, 125(1):1, 2010.
- Timothy G Conley, Christian B Hansen, and Peter E Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012.
- Sam Cosaert and Thomas Demuynck. Nonparametric welfare and demand analysis with unobserved individual heterogeneity. *Review of economics and statistics*, 100(2):349–361, 2018.
- Rahul Deb, Yuichi Kitamura, John Quah, and Jörg Stoye. Revealed price preference: Theory and stochastic testing. 2018. Working Paper.
- Xavier d’Haultfoeuille, Christophe Gaillac, and Arnaud Maurel. Rationalizing rational expectations? tests and deviations. Technical report, National Bureau of Economic Research, 2018.
- W Erwin Diewert. Afriat and revealed preference theory. *The Review of Economic Studies*, 40(3):419–425, 1973.
- Federico Echenique, Sangmok Lee, and Matthew Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223, 2011.
- Liran Einav, Amy Finkelstein, and Mark R Cullen. Estimating welfare in insurance markets using variation in prices. *The quarterly journal of economics*, 125(3):877–921, 2010.
- Pirmin Fessler and Maximilian Kasy. How to use economic theory to improve estimators: Shrinking toward theoretical restrictions. *Review of Economics and Statistics*, 101(4):681–698, 2019.
- Charles Gauthier. Nonparametric identification of discount factors under partial efficiency. 2019. Working Paper.

Yehuda Grunfeld and Zvi Griliches. Is aggregation necessarily bad? *The Review of Economics and Statistics*, pages 1–13, 1960.

Yoram Halevy, Dotan Persitz, and Lanny Zrill. Parametric recoverability of preferences. *Journal of Political Economy*, 126(4):1558–1593, 2018.

Lars Peter Hansen and Ravi Jagannathan. Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590, 1997.

Lars Peter Hansen and Thomas J Sargent. *Robustness*. Princeton university press, 2008.

Lars Peter Hansen and Thomas J Sargent. Structured uncertainty and model misspecification. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2018-77), 2018.

Stefan Hoderlein and Jörg Stoye. Testing stochastic rationality and predicting stochastic demand: the case of two goods. *Economic Theory Bulletin*, 3(2):313–328, 2015.

Martijn Houtman and J Maks. Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden*, 19(1):89–104, 1985.

Guido W Imbens. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2):126–132, 2003.

Yuichi Kitamura and Jörg Stoye. Nonparametric counterfactuals in random utility models. *arXiv preprint arXiv:1902.08350*, 2019.

Patrick Kline and Andres Santos. Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics*, 4(2):231–267, 2013.

Patrick Kline and Melissa Tartari. Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *American Economic Review*, 106(4):972–1014, 2016.

Finn E Kydland and Edward C Prescott. Time to build and aggregate fluctuations. *Econometrica*, pages 1345–1370, 1982.

- Charles F Manski. Partial identification of counterfactual choice probabilities. *International Economic Review*, 48(4):1393–1410, 2007.
- Charles F Manski and John V Pepper. How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics*, 100(2):232–244, 2018.
- Matthew Masten and Alexandre Poirier. Inference on breakdown frontiers. 2019.
- Matthew A Masten and Alexandre Poirier. Identification of treatment effects under conditional partial independence. *Econometrica*, 86(1):317–351, 2018a.
- Matthew A Masten and Alexandre Poirier. Salvaging falsified instrumental variable models. *arXiv preprint arXiv:1812.11598*, 2018b.
- Daniel McFadden. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272, 1981.
- Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, 86(5):1589–1619, 2018.
- Ulrich K Müller and Andriy Norets. Credibility of confidence sets in nonstandard econometric problems. *Econometrica*, 84(6):2183–2213, 2016.
- Maria Ponomareva and Elie Tamer. Misspecification in moment inequality models: Back to moment equalities? *The Econometrics Journal*, 14(2):186–203, 2011.
- Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.
- Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.
- Bernard Salanié and Frank A Wolak. Fast, “robust,” and approximately correct: estimating mixed demand systems. Technical report, National Bureau of Economic Research, 2019.
- Herbert A Simon. *Administrative Behavior*. Macmillan, 1947.
- Pietro Tebaldi, Alexander Torgovitsky, and Hanbin Yang. Nonparametric estimates

of demand in the california health insurance exchange. Technical report, Working Paper, 2018.

Hal R Varian. The nonparametric approach to demand analysis. *Econometrica*, pages 945–973, 1982.

Hal R Varian. Goodness-of-fit in optimizing models. *Journal of Econometrics*, 46(1-2):125–140, 1990.

Hal R Varian. Goodness-of-fit for revealed preference tests. Technical report, 1991.

Quang H Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

Robert D Willig. Consumer's surplus without apology. 1976.