An LLM Workflow for Functional Form Sensitivity Analysis

Isaiah Andrews, *MIT and NBER*
Simon Essig Aberg, *Harvard University*
Jesse M. Shapiro, *Harvard University and NBER**

October 2026

Abstract

We introduce an LLM workflow for functional form sensitivity analysis. A researcher estimates the value of an object of interest by maximizing the fit of a working model to some data. The working model assumes an arbitrary form for some function. We define sensitivity by how much it is possible to change the estimate of the object of interest, without compromising fit, by selecting a different functional form. The workflow first looks for an analytic solution. If it fails to find one, it approximates the solution via numerical methods that respect the fit constraint at each step of a solver path. The workflow yields a verifiable lower bound on the in-sample sensitivity of the researcher's estimate, and is fast in many practical situations. We illustrate with an application to the welfare cost of adverse selection in health insurance.

1 Introduction

It seems unlikely that demand functions are linear or logistic, that risk preferences are CRRA, or that unobservables are Gaussian. Yet, whether because of convenience or convention, such assumed functional forms pervade modern structural models. As estimated structural models are often intended to guide theory, policy, or both, it is important to know whether their implications would be substantively different under other, similarly plausible, functional forms.

In many cases the answer is unknown. Numerically checking sensitivity to all functional forms in a rich class can be prohibitively difficult. Analytic results characterizing objects of interest across all forms in a class are not always available. Perhaps for these reasons, the most common approach to functional form sensitivity analysis for estimated structural models is to report estimates under a small number of alternative forms, themselves often convenient or conventional.[1]

*We are grateful to Scott Kominers for comments and suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2140743. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. LLMs were used in the preparation of this article and its accompanying materials. E-mail: iandrews@mit.edu, sessigaberg@g.harvard.edu, jesse_shapiro@fas.harvard.edu.

[1]Of the 33 articles published in the *American Economic Review* in 2025 that report estimates of a structural model and use it to compute counterfactuals, 32 impose some parametric form on an unknown function or distribution, with

In this paper, we introduce an LLM workflow to assess the sensitivity of a given conclusion to the form of a given function. The workflow yields a verifiable lower bound on the in-sample sensitivity of the researcher's estimate, and is fast in many practical situations. We illustrate with an application to the welfare cost of adverse selection in health insurance.

Our setup envisions a researcher interested in a scalar *object of interest* (e.g., consumer surplus) whose value depends, in part, on the shape of an unknown function (e.g., a demand function). The researcher specifies a *working model* that restricts the unknown function to some class (e.g., linear functions) and under which the object of interest is known up to a finite-dimensional parameter (e.g., a slope and intercept). The researcher estimates the parameter (and hence the object of interest) by minimizing a data-dependent *criterion function* (e.g., sum of squared residuals) that describes how well the model fits the data.

The researcher wants to report how much their conclusions about the object of interest depend on their assumption about the form of the unknown function. To this end the researcher specifies a *nesting model* that broadens the function class, ideally to include all economically reasonable forms (e.g., all smooth, downward-sloping functions). We define the *sensitivity* of the researcher's estimate as the extent to which it is possible to increase or decrease the researcher's estimate of the object of interest, without increasing the criterion function, by choosing an alternative function from the nesting model.

We introduce an LLM-assisted workflow to assess sensitivity. The workflow consists of a generic LLM prompt and an associated Python library, together called `DirectedSensitivity`. The workflow takes advantage of modern LLMs' capabilities in mathematics and numerical methods.

To use the workflow, the researcher declares the object of interest, the nesting model (including any constraints on the unknown function), the criterion function, and the working model (which restricts the nesting model). In our default implementation, the researcher makes these declarations via mathematical statements in a Markdown file; alternatively they may be declared by (and then inferred from) the researcher's code. Having declared these elements, the researcher invokes the workflow.

The workflow begins by searching for an exact solution. For some problems, such solutions are known within the economics literature, as for the case of extremal demand curves considered by, for example, Kremer and Snyder (2018) or Kang and Vasserman (2025). For other problems, solutions are not known within the economics literature, but are available using techniques adapted from other fields. In our experience, even when the workflow does not find an exact solution, it is often able to find a functional form implying meaningful sensitivity.

If the workflow does not find an exact solution, it searches for an approximate solution. It does so by parameterizing a finite-dimensional *basis model* that relaxes the working model but remains

an average of 3.4 such forms per paper. In 32 of these articles at least one assumed form is "conventional" (logistic, Gaussian, linear, CRRA, etc.); in 17 of these articles every assumed form is conventional. Of the 32 articles, 14 report some kind of sensitivity analysis, showing sensitivity to an average of 1.8 alternative forms, of which 0.7 are themselves conventional, and which vary on average 1.5 of the 3.6 forms assumed in the model. None of these articles reports estimates of bounds on an object of interest that hold over a rich class of functional forms. See Online Appendix Table 1 for additional details.

consistent with the nesting model. Parameterizing a finite-dimensional relaxation enables the LLM to take advantage of the associated Python library, which hands the problem to a numerical solver. The solver attempts to solve the sensitivity problem across the portion of the nesting model spanned by the basis model.

We design the numerical search to be computationally frugal. The researcher declares a maximum number of calls to the object of interest or criterion functions. When the solver reaches the maximum tolerated number of function calls, the library terminates. The library prioritizes solvers that keep the criterion function at or below that of the estimated working model across all major solver steps.[2] As a result, as long as the solver moves away from its starting parameters, early termination yields a verifiable lower bound on sensitivity.

In our experience, the two steps of the workflow are complementary. In the first step, the LLM learns the mathematical structure of the problem, and locates related knowledge in other disciplines. In the second step, the LLM uses this understanding to parameterize a basis function that is amenable to numerical optimization.

We illustrate the workflow with an application to Einav et al. (2010). Einav et al. (2010) estimate the welfare cost from underinsurance due to adverse selection in an employer-sponsored health insurance plan. Under their given assumptions, this welfare cost is fully determined by demand and cost functions. Einav et al. (2010) assume that demand and cost functions are linear and estimate them via linear regression. In an appendix, Einav et al. (2010) report sensitivity to particular alternative forms for the demand function.

In our application, we instead explore sensitivity to the set of all demand functions consistent with the economic restrictions that Einav et al. (2010, Section II) describe, plus an additional (unstated) restriction that the marginal cost of an insurance plan is nonnegative. The resulting set of demand functions is rich in many respects, including that it is infinite-dimensional. Although related to other problems with a known solution, we are not aware of a published solution to the problem we study within the economics literature.

The workflow shows that the welfare cost is sensitive to the shape of the demand function. In the first step, the workflow solves the problem analytically, and shows that the solution can change the estimated welfare cost by a factor of around 2.8 in either direction without compromising fit. The analytic solution proceeds by a change of variables that rewrites the area between demand and marginal cost (a curve that itself depends on demand) as the area between demand and a reference curve that does not depend on demand. The solution then turns the economic restrictions into monotonicity conditions and collapses the infinite-dimensional problem into a one-dimensional problem. The workflow verifies its analytic solution using Lean (De Moura and Ullrich, 2021), an open-source proof verification library.

Although in this application the second step is made redundant by the first, we still apply the second step of the workflow for illustration. The second step proposes a basis model expansion

[2]Such solvers are used in engineering applications, when respecting feasibility constraints may be necessary for an early termination to yield valuable knowledge (Parkinson et al., 2013, Ch. 8). As a default that is practical in many applications, we implement a variant of the Generalized Reduced Gradient algorithm (Lasdon et al., 1978).

around the linear working model. The expansion is parameterized as a relaxation of the form of the semi-elasticity implied by the working model. Parameterizing the relaxation directly in terms of the semi-elasticity facilitates enforcing properties such as that demand slopes downward. Searching over the parameters of the basis model, a numerical solver is able to achieve sensitivity 99.9 percent as great as the analytic solution.

The extremal demand functions that lead to the greatest increases or decreases in the estimated welfare cost are quite dramatically curved, and appear to us (and, we suspect, many economists) as implausible. We view this as a feature of the workflow because it exposes the fact, in this and other applications, that there are likely *implicit* economic restrictions—restrictions that are believed by researchers, but not stated explicitly in research. By exposing extremal functions that violate these implicit restrictions, the workflow can help a researcher to make these restrictions explicit. We illustrate by showing how the sensitivity of the estimated welfare cost changes if we adopt the form of restriction proposed by Kang and Vasserman (2025).

Our main contribution is a practical, often fast procedure for computing (a lower bound on) the sensitivity of structural estimates to functional form choices. We emphasize practicality because structural estimation often imposes a large computational burden, and we do not believe researchers will adopt procedures that greatly increase this burden. We adopt an approach that searches within a large function class because we think that in many economic applications, it is more reasonable to require functions to be in such a class than to suppose that they take on one of a small number of conventional parametric forms.

Several existing papers propose procedures for calculating sensitivity (or close analogues) exactly over some class of unknown functions (e.g., Bonhomme and Weidner, 2022; Kang and Vasserman, 2025; Ansari et al., 2026). Schennach (2014) considers the goal of bounding an object of interest under a nesting model that is defined through moment conditions involving a latent variable. Christensen and Connault (2023) consider the goal of allowing for the set of all latent-variable distributions in a divergence neighborhood of a working model. Both papers use convex duality techniques to collapse an infinite-dimensional problem to a finite-dimensional problem. Our approach is not restricted to latent-variable distributions and leverages modern LLMs unavailable at the time these articles were published. Perhaps reflecting the computational burden these methods impose, few applied articles in peer-reviewed journals appear to have implemented either approach; for an exception, see De Silva et al. (2023).

A large literature (Leamer, 1983; Manski, 2003; Tamer, 2010; Matzkin, 2013) argues for exposing the role of arbitrary shape and other restrictions by peeling them back in layers (see esp. Tamer 2010). Despite (what we consider to be) a compelling logic, these recommendations are not typically followed in empirical practice. We hope that introducing a workflow that leverages modern LLMs can allow more researchers to take advantage of this perspective without the technical expertise of its prominent (human) advocates.

The remainder of the paper proceeds as follows. Section 2 sets up the sensitivity problem and defines our workflow. Section 3 shows our application to Einav et al. (2010).

2 Setup and Workflow

2.1 Sensitivity

A researcher is interested in a scalar-valued *object of interest* $c \in \mathbb{R}$. Under a nesting model specified by the researcher, $c = c(\nu)$, for $\nu \in \mathcal{N}$ an unknown and possibly infinite-dimensional parameter. The space $\mathcal{N}$ encodes economic restrictions on the unknown parameter, for instance shape restrictions like monotonicity or convexity.

The researcher observes some data. Given the data and any candidate parameter $\nu \in \mathcal{N}$, the researcher can compute a data-dependent criterion function $\hat{q}(\nu)$ that describes the fit of the nesting model to the data at the candidate parameter ν.

The researcher specifies a working model $\mathcal{A} \subset \mathcal{N}$ that restricts the unknown parameter to a finite-dimensional subset of the nesting model. We denote a generic value of the working model's parameters by $\alpha \in \mathcal{A} \subset \mathcal{N}$. The researcher estimates the working model by minimizing the criterion function, obtaining an estimate $c(\hat{\alpha})$ of the object of interest, where

$$\hat{\alpha} \in \arg\min_{\alpha \in \mathcal{A}} \hat{q}(\alpha)$$

is a minimizer of the criterion function within the working model. We assume for ease of exposition that such a minimizer exists, but do not assume it is unique.

We define the sensitivity under the nesting model as

$$\sup_{\{\nu \in \mathcal{N} : \hat{q}(\nu) \leq \hat{q}(\hat{\alpha})\}} |c(\nu) - c(\hat{\alpha})| . \tag{1}$$

Sensitivity is the extent to which it is possible to change the estimate for the object of interest without compromising fit to the data relative to a best-fitting parameterization of the working model.

Depending on the economic setting, we may be especially interested in changes that decrease or increase the object of interest relative to the estimate under the working model. We define the lower sensitivity under the nesting model as

$$\sup_{\{\nu \in \mathcal{N} : \hat{q}(\nu) \leq \hat{q}(\hat{\alpha})\}} c(\hat{\alpha}) - c(\nu) \tag{2}$$

and the upper sensitivity under the nesting model as

$$\sup_{\{\nu \in \mathcal{N} : \hat{q}(\nu) \leq \hat{q}(\hat{\alpha})\}} c(\nu) - c(\hat{\alpha}) . \tag{3}$$

Sensitivity (1) is the maximum of lower sensitivity (2) and upper sensitivity (3).

2.2 Workflow

Algorithm 1 summarizes an LLM-assisted workflow for computing a lower bound on sensitivity under the nesting model (1). The steps for the directional counterparts, (2) and (3), are analogous.

In its first step, the workflow attempts an analytic solution to the sensitivity problem. If it solves the problem analytically, the workflow stops. Otherwise, the workflow attempts to provide a problem reduction (such as from an infinite-dimensional to a finite-dimensional problem) or informative bounds (by proposing specific parameters in the nesting model that obtain large deviations in the object of interest without worsening fit) using analytic methods. Regardless of its success in these efforts, if it has not fully solved the problem analytically, the workflow proceeds to a second step.

Algorithm 1 Lower bound on sensitivity (LLM workflow)

- Inputs:
  - Nesting model $\mathcal{N}$
  - Working model $\mathcal{A}$
  - Object of interest $c(\cdot)$
  - Criterion function $\hat{q}(\cdot)$
- Attempt to solve (1) analytically.
  - If solved, terminate.
  - If not solved, attempt each of the below analytically and then move to the next step:
    - Reduce (1) to a finite-dimensional problem.
    - Provide informative bounds on (1).
- Develop a suitable basis model $\mathcal{B}$ using context from the preceding step.
- Apply Algorithm 2 to approximate the solution to (4).

In its second step, the workflow searches for an approximate solution to the sensitivity problem using numerical methods. The workflow proposes a finite-dimensional *basis model* $\mathcal{B}$ that relaxes the working model while respecting the nesting model: $\mathcal{A} \subset \mathcal{B} \subset \mathcal{N}$. We denote a generic element of the basis model by $\beta \in \mathcal{B} \subset \mathcal{N}$. We define the sensitivity under the basis model as

$$\sup_{\{\beta \in \mathcal{B}: \hat{q}(\beta) \leq \hat{q}(\hat{\alpha})\}} |c(\beta) - c(\hat{\alpha})| \tag{4}$$

with lower and upper sensitivity defined analogously to (2) and (3). Sensitivity under the basis model is a lower bound on sensitivity under the nesting model. The bound is tighter the better the basis model approximates the nesting model.[3]

[3]Indeed, (4) converges to (1) along any increasing sequence of basis models whose limit is dense in a topology under

The workflow uses numerical methods to arrive at a lower bound on sensitivity under the nesting model. Algorithm 2 summarizes the generic approach. The researcher specifies a maximum number of steps and a tolerance for convergence. For each problem, we start at the parameters $\beta = \hat{\alpha}$. From this starting point, we take steps in basis model space. At each step, we find parameters that increase or decrease the object of interest, while respecting the constraint that $\hat{q}(\beta) \leq \hat{q}(\hat{\alpha})$. The algorithm terminates when it either converges at the specified tolerance or reaches the maximum number of steps. As a leading example implementation, Online Appendix Algorithm 1 presents a variant of the Generalized Reduced Gradient algorithm.

Algorithm 2 Lower bound on sensitivity (numerical workflow)

- Inputs:
  - Maximum number of steps L.
  - Convergence tolerance ϵ.
- Initialize:
  - $\beta_0^- = \beta_0^+ = \hat{\alpha}$.
  - $c_0^- = c_0^+ = c(\hat{\alpha})$.
- For $\ell \in \{1, ..., L\}$ do:
  - Find β_ℓ^- such that $c_\ell^- := c\left(\beta_\ell^-\right) < c_{\ell-1}^-$ and $\hat{q}\left(\beta_\ell^-\right) \leq \hat{q}(\hat{\alpha})$, if one exists; otherwise set $\beta_\ell^- = \beta_{\ell-1}^-$ and $c_\ell^- = c_{\ell-1}^-$.
  - Find β_ℓ^+ such that $c_\ell^+ := c\left(\beta_\ell^+\right) > c_{\ell-1}^+$ and $\hat{q}\left(\beta_\ell^+\right) \leq \hat{q}(\hat{\alpha})$, if one exists; otherwise set $\beta_\ell^+ = \beta_{\ell-1}^+$ and $c_\ell^+ = c_{\ell-1}^+$.
  - If solutions have converged at tolerance ϵ, set $c_L^- = c_\ell^-$ and $c_L^+ = c_\ell^+$ and terminate.
- Return lower bound on lower sensitivity $c(\hat{\alpha}) - c_L^-$, upper sensitivity $c_L^+ - c(\hat{\alpha})$, and sensitivity $\max\left\{c(\hat{\alpha}) - c_L^-, c_L^+ - c(\hat{\alpha})\right\}$.

3 Application to Einav, Finkelstein, and Cullen (2010)

3.1 Problem Specification

Einav et al. (2010) estimate the welfare cost c from underinsurance due to adverse selection in an employer-sponsored health insurance plan. Under the assumptions in Einav et al. (2010, Section II), this welfare cost is fully determined by demand and cost functions.

Under their working model, Einav et al. (2010) assume that demand and (average) cost functions are linear and estimate them via linear regression, weighting by the number of employees purchasing each plan. To represent the linear regression estimator in the setup of Section 2, we treat the

which both $\hat{q}(\cdot)$ and $c(\cdot)$ are continuous, provided $\{\nu \in \mathcal{N} : \hat{q}(\nu) \leq \hat{q}(\hat{\alpha})\}$ is the closure of $\{\nu \in \mathcal{N} : \hat{q}(\nu) < \hat{q}(\hat{\alpha})\}$.

criterion function as the employee-weighted sum of squared deviations between the observed and predicted purchase shares at each of the six prices observed in the data.

Under the working model, Einav et al. (2010) estimate that the welfare cost c is equal to 9.55. We reproduce this value as 9.28 using the published (but rounded) data reported in Einav et al. (2010, Table II).

We specify a rich nesting model based on the assumptions in Einav et al. (2010, Section II). The nesting model permits all continuous, differentiable, and downward-sloping demand functions that are bounded to $[0, 1]$ and cross marginal cost at most once from above. We further impose the (unstated, but presumably implicit) requirement that the marginal cost is everywhere nonnegative. And, to guarantee coherence of the authors' working model, we suppose that all of these requirements must hold on a bounded interval of prices $\left[0, \overline{P}\right]$.[4] The nesting model maintains the working model's specification of a linear average cost function.

To invoke the workflow (Algorithm 1), we declare its inputs mathematically in a Markdown template. The nesting model is the set of demand functions defined above. The working model is the set of linear demand functions. The object of interest is the welfare cost c. The criterion function is the sum of squared deviations.

Figure 1 reproduces the key portions of the filled template. The template is filled in natural language, interspersing plain English, math, and references to data files. An LLM is then instructed to follow a prespecified workflow using the information in the template. Online Appendix Figure 1 reproduces the complete filled template, which includes additional definitions and illustrative examples that we omit from Figure 1 for brevity.

3.2 Analytic Solution

The first step of the workflow is to attempt to solve (1) analytically. In our execution (using Claude Opus 5.5), the workflow succeeds in this step. The analytic solution rests on a change of variables. The welfare cost is the area between demand and marginal cost, but marginal cost itself depends on demand. The change of variables re-expresses the welfare cost as the area between demand and a fixed reference curve. In these coordinates the economic restrictions in the nesting model can be represented as monotonicity conditions, and the infinite-dimensional search over demand curves reduces to a one-dimensional problem that the workflow solves.

The workflow returns that lower sensitivity is 5.91 and upper sensitivity is 18.10. These findings mean that, by deviating from the working model's shape restriction, it is possible to increase or decrease the estimated welfare cost c by a factor of 2.8 without worsening the fit to the data.

It is possible to verify the workflow's conclusions. First, for any proposed function, we can simply calculate the criterion and object of interest implied by the function, and verify that the criterion is no worse than under the working model. Second, for any analytic claim, the workflow returns a proof and, optionally, verifies the proof using Lean. These steps corroborate the workflow's findings.

[4] Following Einav et al. (2010, Online Appendix Table A2) we take $\overline{P} = 800$.

Figure 1: Abbreviated, filled problem-specification template for the Einav et al. (2010) application

Problem specification

Fill out this Markdown file. After completing the verification steps below, hand the filled Markdown file to an LLM instructed to follow the workflow in `WORKFLOW.md`. Provide precise mathematical definitions in each section unless instructed otherwise.

1. The nesting model

Define the nesting model $\mathcal{N}$, which encodes economic restrictions on unknown parameters, including unknown functions.

Your version.

Throughout, the primitive ν is the demand curve $D(p)$. The cost parameters (γ, δ) are fixed. They give the average cost curve $AC(p) = \gamma + \delta p$ and the marginal cost curve

$$MC(p) = AC(p) + \delta \frac{D(p)}{D'(p)}.$$

$\mathcal{N}$ is the set of demand curves $D : [0, 800] \to [0, 1]$ that satisfy the following properties:

- Continuous and differentiable.
- Strictly decreasing: $D'(p) < 0$ for all $p \in [0, 800]$.
- Bounded: $0 \leq D(p) \leq 1$ for all $p \in [0, 800]$.
- Single crossing of marginal cost and demand: if $MC(p) > p$ for some $p \in [0, 800]$, then $MC(p') > p'$ for all $p' \in [0, p)$.
- There exists $\bar{p} \in [0, 800]$ with $D(\bar{p}) > 0$ and $MC(p) < p$ for all $p \in (\bar{p}, 800]$.
- Nonnegative marginal cost: $MC(p) \geq 0$ for all $p \in [0, 800]$.

The cost parameters (γ, δ) are given in 'Location of data, code, and estimates'.

2. The working model

Define the working model $\mathcal{A} \subset \mathcal{N}$. Give its parameters, their domain, and the map from parameters to primitives.

Your version.

$\mathcal{A}$ is the linear-demand family

$$\mathcal{A} = \{D_{a,b}(p) = a + bp : D_{a,b} \in \mathcal{N}\},$$

with parameters (a, b) mapping to the primitive $D_{a,b}$.

3. The object of interest

Define the object of interest $c(\cdot) : \mathcal{N} \to \mathbb{R}$. It must be well-defined and finite for every $\nu \in \mathcal{N}$.

Your version.

The competitive price p_C is defined such that $p_C = AC(p_C)$, and the efficient price is the smallest price at which marginal cost is met from below:

$$p_E = \min\Big(\{p \in [0, 800] : p = MC(p)\}\Big).$$

$c(D)$ is the deadweight-loss area between demand and marginal cost, from the competitive price p_C to the efficient price p_E,

$$c(D) = \int_{D(p_C)}^{D(p_E)} \big[D^{-1}(Q) - MC\big(D^{-1}(Q)\big)\big] dQ.$$

Inverse demand $D^{-1}(Q)$ is well-defined for all $D \in \mathcal{N}$ because D is strictly decreasing. Inverse demand and marginal cost are bounded, so the integral exists and is finite for all $D \in \mathcal{N}$.

4. The criterion

Define the criterion $\hat{q}(\nu)$. The criterion is a data-dependent measure of fit, where smaller values indicate a better fit. The criterion must be finite and well-defined for every $\nu \in \mathcal{N}$.

Your version.

$\hat{q}(D)$ is the employee-weighted sum of squared share residuals at the six observed prices,

$$\hat{q}(D) = \sum_{i=1}^{6} n_i\big(s_i - D(p_i)\big)^2,$$

where (p_i, s_i, n_i) are the observed price, share, and employee count.

5. Computational constraints (optional)

State the largest number of evaluations of $c(\cdot)$ and $\hat{q}(\cdot)$ you are willing to spend, for when either is expensive. Fill in the blanks below.

At most 100,000 evaluations of $c(\cdot)$ and at most 100,000 evaluations of $\hat{q}(\cdot)$.

6. Location of data, code, and estimates

Provide the location of data inputs. Provide the location of code that implements the object of interest and criterion (or numerical approximations of these functions). Provide the values of any parameters referenced in the prior sections. Provide the location or values of the parameters estimated under the working model.

Your version.

- $\{(p_i, s_i, n_i)\}_{i=1}^{6}$ — `source/raw/EFC2010/data/EFC2010_table2.csv`.
- $c(D)$ and $\hat{q}(D)$ — `source/analysis/EFC2010/specification.py:{Target,Criterion}`.
- Demand estimates $(\hat{a}, \hat{b})$ and cost estimates (γ, δ) — `output/analysis/EFC2010/replicate.txt`.

7. Consistency and completeness checklist

After you complete the template, ask your LLM to verify the following prior to launching the workflow:

- ☒ 'The nesting model', 'The working model', 'The object of interest', and 'The criterion' are each filled with precise mathematical definitions.
- ☒ The nesting model $\mathcal{N}$ is non-empty.
- ☒ The working model is a genuine subset of the nesting model, $\mathcal{A} \subset \mathcal{N}$, and is finite-dimensional.
- ☒ The object of interest $c(\nu)$ is well-defined and finite for every $\nu \in \mathcal{N}$.
- ☒ The criterion $\hat{q}(\nu)$ is well-defined and finite for every $\nu \in \mathcal{N}$.
- ☒ Every data or code symbol used in 'The nesting model' through 'Computational constraints' is resolved in 'Location of data, code, and estimates'.

8. Workflow configuration

Should the LLM attempt a Lean formalization of mathematical claims? Formalization may add time and token cost, and is not required to verify numerical certificates.

- ☒ Yes
- ☐ No

Figure 2 illustrates the findings graphically. Following Figure V of Einav et al. (2010), each panel plots the proposed demand function, estimated average cost function, and implied marginal cost function. The area of the shaded region corresponds to the welfare cost c. Panel (a) uses the linear demand curve estimated under the working model. Panel (b) uses a demand curve that (nearly) attains the infimum of the welfare cost. Panel (c) uses a demand curve that (nearly) attains the supremum of the welfare cost.

Figure 2 provides an intuition for the sensitivity of the welfare cost. The welfare cost is the area between demand and marginal cost. Therefore, a demand curve that most decreases (increases) the welfare cost bends in such a way that it is as close to (far from) the implied marginal cost curve as possible. A demand curve that does so without worsening fit must bend sharply at specific points to thread the data precisely.

Figure 3 reproduces the authors' original sensitivity analyses, which include two alternative conventional functional forms—probit demand and quadratic demand—as well as two restrictions on the working model. Across these sensitivity analyses, the estimated welfare cost ranges from 3.27 to 11.28, as against 3.37 through 27.37 under the analytic bounds. Thus, relative to the workflow, the authors' sensitivity analysis achieves a lower lower bound (because it does not require fit as good as the working model), but also a lower upper bound.

Some economists (ourselves included) might regard the demand functions exhibited in panels (b) and (c) of Figure 2 as economically implausible, despite their being consistent with the stated assumptions in Einav et al. (2010, Section II). A researcher encountering these possibilities might be inclined to state (and, perhaps, defend) additional restrictions on the class of demand functions.

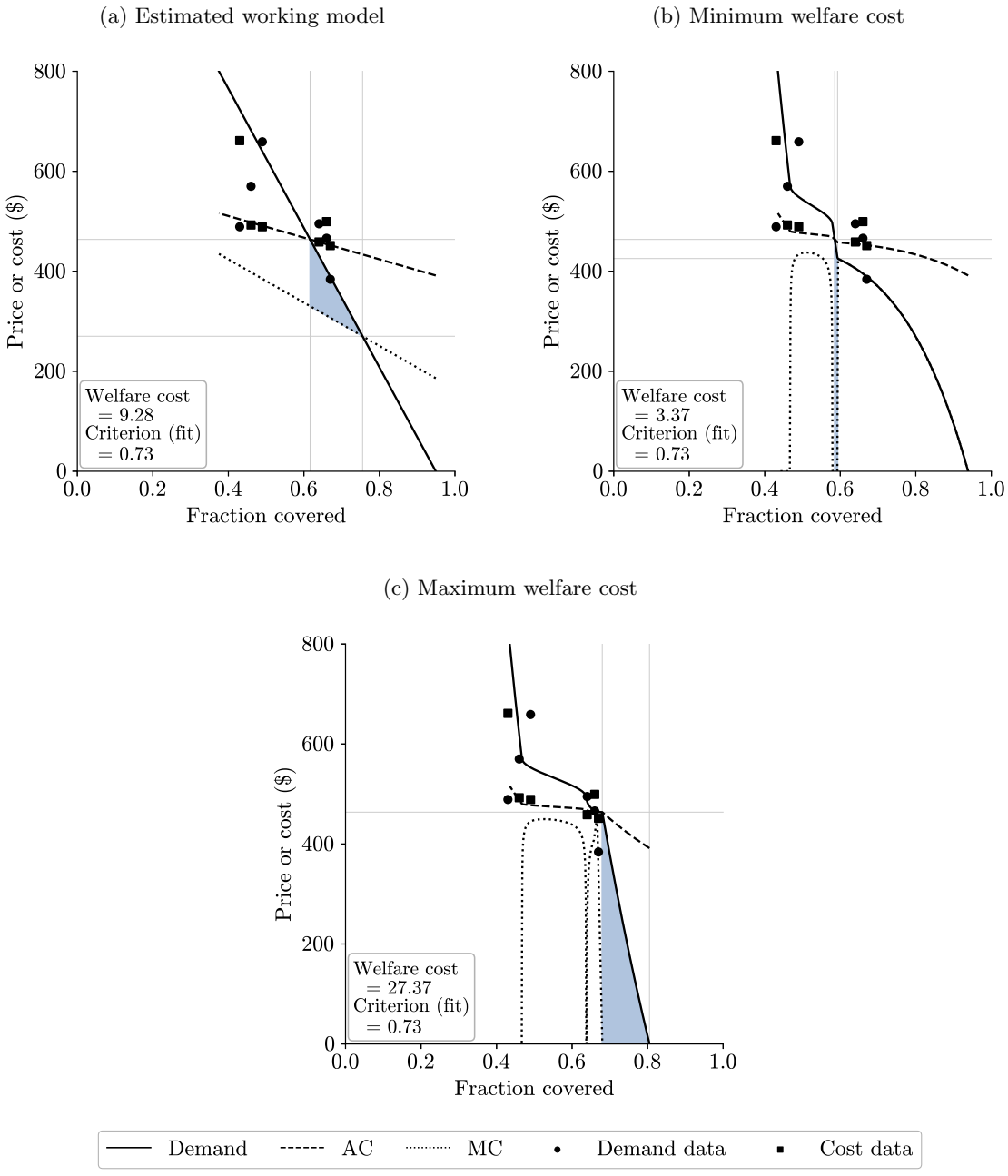
We regard this dialogue between assumptions and conclusions as a feature of the workflow, because we suspect that in many situations economists choose shape restrictions based on unstated assumptions that could be usefully surfaced for their audience. To illustrate this possibility, Figure 4 shows (nearly) extremal curves under restrictions on the curvature of the demand function in the spirit of those considered in Kang and Vasserman (2025). Each column corresponds to a successively looser bound on the curvature of the demand function. The top row corresponds to demand functions that (nearly) minimize the estimated welfare cost, the bottom to functions that (nearly) maximize it. As we loosen the bound on curvature, we move from the working model (left column) to the analytic solution (right column).

3.3 Numerical Bound

Although the workflow's first step succeeds, we proceed for illustration with the second step, which approximates sensitivity using a basis model expansion (Algorithm 2). In our execution, the workflow proposes a basis expansion (see Online Appendix C) that is parameterized as a relaxation of the functional form of the semi-elasticity implied by the working model. The workflow implements Online Appendix Algorithm 1 to find the parameters of the basis model that most change the welfare cost without sacrificing fit.

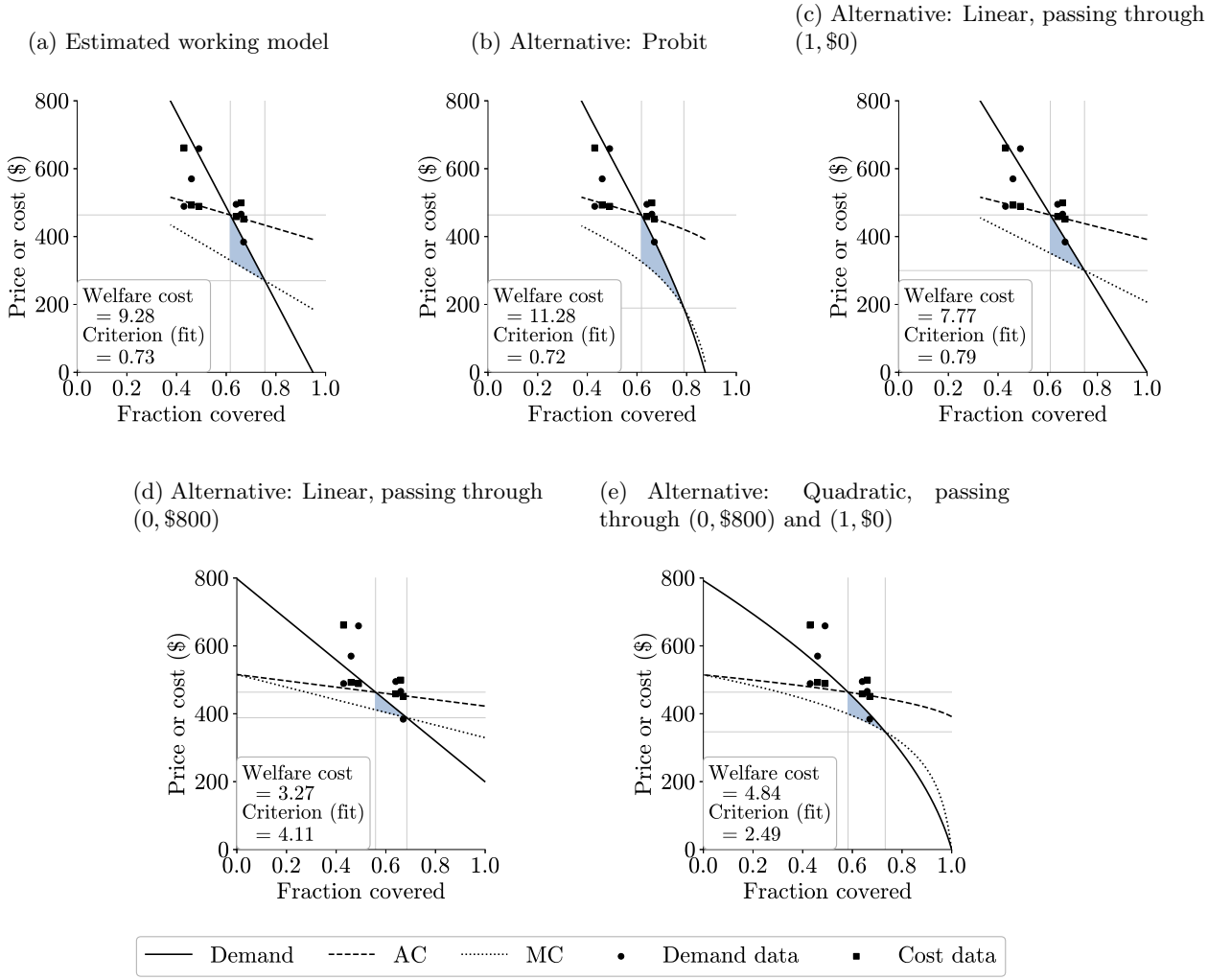
The numerical procedure performs well. Online Appendix Figure 2 shows that, against the exact

Figure 2: Sensitivity of the welfare cost in Einav et al. (2010) to alternative demand functions, analytic solution



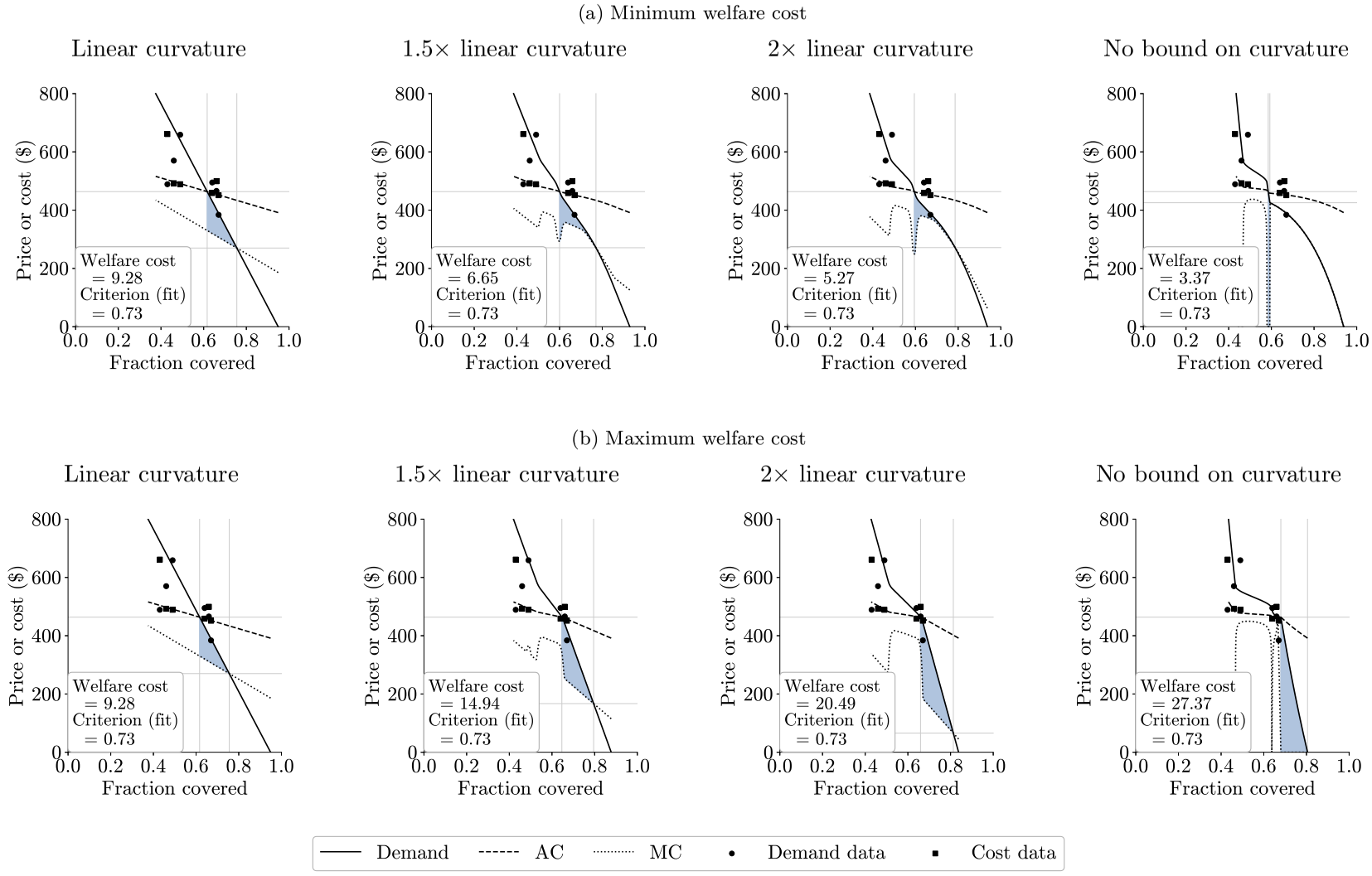
Notes: Each panel plots a demand function, the estimated average cost curve, and the implied marginal cost curve, as in Figure V of Einav et al. (2010), with the area of the shaded region denoting the implied welfare cost due to adverse selection. Panels differ in which demand function they use. Panel (a) uses the linear demand function implied by the estimated working model. Panel (b) uses a demand function that (nearly) attains the infimum of the welfare cost. Panel (c) uses a demand function that (nearly) attains the supremum of the welfare cost.

Figure 3: Sensitivity of the welfare cost in Einav et al. (2010) to alternative demand functions considered by the authors



Notes: Each panel plots a demand function, the estimated average cost curve, and the implied marginal cost curve, as in Figure V of Einav et al. (2010), with the area of the shaded region denoting the implied welfare cost due to adverse selection. Panels differ in which demand function they use. Panel (a) uses the linear demand function implied by the estimated working model. The remaining panels use demand functions corresponding to the four sensitivity analyses reported in Einav et al. (2010, Online Appendix Table A2). Values differ slightly from those reported in Einav et al. (2010) due to rounding.

Figure 4: Sensitivity of the welfare cost in Einav et al. (2010) to alternative demand functions, restrictions on curvature



Notes: Each panel plots a demand function, the estimated average cost curve, and the implied marginal cost curve, as in Figure V of Einav et al. (2010), with the area of the shaded region denoting the implied welfare cost due to adverse selection. Panels differ in which demand function they use. Each column corresponds to a restriction on curvature that requires that the slope of the demand function is uniformly within a given factor of that of the estimated linear working model, where the factor ranges from 1 (linear model) to infinity (unrestricted curvature). Panel (a) shows demand functions that (nearly) attain the infimum of the welfare cost given the restriction on curvature. Panel (b) shows demand functions that (nearly) attain the supremum of the welfare cost given the restriction on curvature. In the second and third columns, the demand functions are smoothed over a window of $50 using a triweight kernel and then shrunk towards the working model, so that they are differentiable and fit no worse than the working model.

analytic solution, the algorithm recovers 99.9 percent of the attainable upward movement and 96.1 percent of the downward movement, starting from the working model. Thus, in this application, it is possible to learn a lot about the sensitivity of the object of interest without relying entirely on the analytic solution. Although our execution uses a basis model with eight parameters, even a basis model with fewer parameters is able to certify a large portion of the true sensitivity (see Online Appendix Table 2).

When considering a generalization of their working model, it is common for researchers to select the best-fitting element of the more general model, by re-estimating their model under the more general specification.[5] Within our setup, such a procedure is akin to estimating the (parameters of the) basis model. Point-estimation under the basis model requires that the criterion be sufficiently informative for the minimizer to be unique (and thus, for instance, that the number of parameters in the present application be no larger than six), which our approach does not require. Moreover, estimating the basis model need not solve (1). Online Appendix Figure 2 shows that, in our application, a best-fitting basis model attains only 12.1 percent of the sensitivity under the full basis model.

The numerical procedure is not computationally demanding. It achieves its lower bound within 2,040 evaluations of the criterion function and 1,000 of the object of interest. Even if each function evaluation were expensive (which it is not in this application), the resulting computation would remain practical. Moreover, Online Appendix Figure 3 shows that the procedure certifies a large portion of the sensitivity before termination. Thus, in this application, it is possible to learn a lot about the sensitivity of the object of interest even with a smaller budget of function evaluations than we used in our execution.

References

Ansari, H. C., Donaldson, D., and Wiles, E. (2026). Quantifying the sensitivity of quantitative trade models. Working paper.

Bonhomme, S. and Weidner, M. (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics*, 13(3):907–954.

Christensen, T. and Connault, B. (2023). Counterfactual sensitivity and robustness. *Econometrica*, 91(1):263–298.

De Moura, L. and Ullrich, S. (2021). The Lean 4 theorem prover and programming language. In Platzer, A. and Sutcliffe, G., editors, *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer.

[5] Among the 14 articles in Online Appendix Table 1 that report results under an alternative functional form, eight report results under a generalization of one of their baseline forms, and seven of these re-estimate the model under the generalization.

De Silva, D. G., Hubbard, T. P., Schiller, A. R., and Tsionas, M. G. (2023). Estimating outcomes in the presence of endogeneity and measurement error with an application to R&D. *Quarterly Review of Economics and Finance*, 88:278–294.

Einav, L., Finkelstein, A., and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics*, 125(3):877–921.

Kang, Z. Y. and Vasserman, S. (2025). Robustness measures for welfare analysis. *American Economic Review*, 115(8):2449–2487.

Kremer, M. and Snyder, C. M. (2018). Worst-case bounds on R&D and pricing distortions: Theory with an application assuming consumer values follow the world income distribution. Working Paper 25119, National Bureau of Economic Research.

Lasdon, L. S., Waren, A. D., Jain, A., and Ratner, M. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Transactions on Mathematical Software*, 4(1):34–50.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73(1):31–43.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer, New York.

Matzkin, R. L. (2013). Nonparametric identification in structural economic models. *Annual Review of Economics*, 5(1):457–486.

Parkinson, A. R., Balling, R. J., and Hedengren, J. D. (2013). *Optimization Methods for Engineering Design*. Brigham Young University, Provo, UT, 1st edition.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica*, 82(1):345–385.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, 2(1):167–195.

Online Appendix for "An LLM Workflow for Functional Form Sensitivity Analysis"

Isaiah Andrews, *MIT and NBER*[1]
Simon Essig Aberg, *Harvard University*
Jesse M. Shapiro, *Harvard University and NBER*

A Functional Forms and Sensitivity Analyses in Structural Articles

Online Appendix Table 1: Functional forms and sensitivity analyses in structural articles

	Parametric forms		Sensitivity			
Article	# forms	# conventional	# forms	# alternatives	# conventional	Bound over rich class?
Aguirregabiria et al. (2025)	3	3	0	0	0	No
saver deposit-choice utility, logit with linear characteristics index.						
Allcott et al. (2025)	3	3	0	0	0	No
bias and nudge-effect heterogeneity, joint normal.						
Alvarez et al. (2025)	0	0	0	0	0	No
price-change hazard, nonparametric mixed proportional hazard.						
Amodio et al. (2025)	6	5	1	1	0	No
firm entry cost, square-root in entrant rank; alternative, Weibull across markets.						
Atkeson et al. (2025)	5	4	0	0	0	No
household utility, log utility.						
Balboni (2025)	4	4	2	1	0	No
worker utility, Cobb-Douglas goods-land; alternative, goods only.						
Benetton et al. (2025)	4	4	0	0	0	No
borrower taste shocks, nested logit.						
Berndt et al. (2025)	2	1	0	0	0	No
bank asset value, Brownian diffusion with Poisson exponentially distributed downward jumps.						
Bonfim et al. (2025)	3	3	0	0	0	No
firm credit demand across banks, Dixit-Stiglitz CES.						
Bornstein and Peter (2025)	3	1	1	1	0	No
consumption utility, CED shifted power; alternative, Kimball (Klenow-Willis).						
Chen et al. (2025)	4	4	1	1	1	No
consumer utility, CES; alternative, imperfect substitution.						
Choukhmane (2025)	3	2	2	2	0	No
switching cost, fixed additive utility cost; alternative, proportional to earnings.						
Chyn and Daruich (2025)	7	6	1	1	1	No
neighborhood spillover, power of mean resident income; alternative, CES power mean.						
Dhyne et al. (2025)	3	3	1	1	1	No
worker taste shocks, logit; alternative, nested logit.						
Doraszelski et al. (2025)	2	1	0	0	0	No
cash flow, additive components with Gaussian shock.						
Dube et al. (2025)	3	1	3	5	2	No
firm-level labor supply, constant elasticity with round-number jump; alternative, linear in wage.						

[1]E-mail: iandrews@mit.edu, sessigaberg@g.harvard.edu, jesse_shapiro@fas.harvard.edu.

	Parametric forms		Sensitivity			
Article	# forms	# conventional	# forms	# alternatives	# conventional	Bound over rich class?
Einav et al. (2025)	3	2	1	1	0	No
attention probability, constant Bernoulli; alternative, varying with months since sign-up.						
Gagliardone et al. (2025)	5	3	3	3	0	No
firm production technology, Cobb-Douglas; alternative, translog.						
Gaubert et al. (2025)	4	4	0	0	0	No
location preference distribution, logistic with skill-dependent mean.						
Gómez-Cram and Lawrence (2025)	2	2	0	0	0	No
cash flow growth process, Gaussian linear AR(1) with log-Gaussian stochastic volatility.						
Gourinchas et al. (2025)	4	3	0	0	0	No
bond investor demand, affine in log bond price with exponential maturity profile.						
Greenwald and Guren (2025)	6	5	0	0	0	No
landlord ownership-benefit heterogeneity, logistic.						
Gulesci et al. (2025)	2	2	0	0	0	No
taste shocks, logit.						
Haanwinckel (2025)	5	3	0	0	0	No
task-based production, exponential worker efficiency with Gamma task mix.						
Haddad et al. (2025)	2	2	0	0	0	No
investor stock demand, Koijen-Yogo log-linear logit.						
Jha and Leslie (2025)	1	1	0	0	0	No
unit fuel-use function, linear with start-up/running fixed terms.						
Kroft et al. (2025)	3	2	1	1	1	No
production function, Leontief over Cobb-Douglas composite; alternative, Cobb-Douglas.						
Li (2025)	6	6	0	0	0	No
household deposit-in-utility, power function.						
Mian et al. (2025)	3	2	1	3	1	No
saver bonds-in-utility, quadratic; alternative, log.						
Pellegrino (2025)	2	2	2	2	1	No
firm cost function, linear; alternative, quadratic.						
Shepard and Wagner (2025)	1	1	1	2	2	No
insurance demand, piecewise-linear; alternative, CARA utility.						
Stanton and Thomas (2025)	3	3	0	0	0	No
buyer consideration set size distribution, exponential.						
Starc and Wollmann (2025)	3	3	0	0	0	No
buyer demand for generic drugs, nested logit.						

Notes: Rows correspond to articles published in the 2025 *American Economic Review* that report estimates of the parameters of a model of economic agents' behavior fitted to data and use the estimated model to compute counterfactuals or welfare; articles whose parameters are only calibrated, and articles whose primary contribution is a method, are excluded. A parametric form is an unknown function or distribution that the baseline specification restricts to a finite-dimensional family and whose parameters the article estimates. A form nested inside another, such as an aggregator inside a utility function, counts once, and objects the article estimates nonparametrically are not counted. A form is conventional if it is linear, quadratic, logit (including nested and mixed logit and type-1 extreme value shocks), Gaussian (including probit), log-normal, CES, Cobb-Douglas, CRRA or isoelastic (including log utility), exponential (including Poisson arrivals), Pareto, or Fréchet. Under "Sensitivity" an alternative is a different functional form for one of the baseline forms, under which the article recomputes and reports its counterfactual or welfare quantity, in the article or in an online appendix that the article references. "# forms" under "Sensitivity" counts the baseline forms varied by any alternative, and "# conventional" counts the alternatives that are conventional. "Bound over rich class?" is Yes if the article reports bounds on a target quantity that hold over a rich or nonparametric class of functional forms. The second row for each article names one baseline form and, where the article reports one, the alternative form it is varied to. Alternatives that are special cases of the baseline form are excluded throughout. Each article was coded by Claude Fable 5.1 (high effort), which received the article PDF together with a fixed coding protocol and returned each classification with verbatim supporting quotes. We refined the protocol by comparing the model's coding with our own on a subset of articles, and spot-checked the final coding.

B Numerical Algorithm for Lower Bound on Sensitivity

Online Appendix Algorithm 1 Variant of Generalized Reduced Gradient

- Inputs:
  - Maximum number of steps L.
  - Convergence tolerance ϵ and stationarity tolerance ϵ_0.
  - Basis model $\mathcal{B} = \{\beta_\omega : \omega \in \Omega\}$ with coordinates $\omega \in \Omega \subseteq \mathbb{R}^m$, and $\hat{\omega}$ such that $\beta_{\hat{\omega}} = \hat{\alpha}$.
  - Constraint functions $g(\omega) \in \mathbb{R}^J$ that represent $\beta_\omega \in \mathcal{N}$ as $g(\omega) \geq 0$.
  - Positive definite matrix $M(\omega) \in \mathbb{R}^{\dim(\Omega)\times\dim(\Omega)}$ that may depend on $\omega \in \Omega$.
  - Initial step size $t_0 > 0$, growth factor $\eta > 1$, backtracking factor $\rho \in (0,1)$, caps K_t on trial steps and K_y on corrections, and correction margin $a \geq 0$.
- Notation: gradients ∇c, $\nabla \hat{q}$, and ∇g_j are with respect to ω and, like g and M, are evaluated at the current point. For margin $b \geq 0$, $\mathcal{F}_b(\omega)$ is the set of moves d that satisfy the fit and nesting constraints linearized at ω, each with slack b in the M-norm:

$$\hat{q}(\beta_\omega) + \nabla\hat{q}^\top d \leq \hat{q}(\hat{\alpha}) - b\left\|\nabla\hat{q}\right\|_{M^{-1}}, \qquad g_j(\omega) + \nabla g_j^\top d \geq b\left\|\nabla g_j\right\|_{M^{-1}} \text{ for all } j.$$

- For each direction $s \in \{+1, -1\}$ do:
  - Set $\omega_0 = \hat{\omega}$.
  - For $\ell \in \{1, ..., L\}$ do:
    - Form the predictor direction d, the move that best ascends sc within the linearized constraints,

$$d = \arg\max_{d\in\mathcal{F}_0(\omega_{\ell-1})} s\nabla c^\top d - \frac{1}{2}d^\top M d.$$

    - If no such d exists or $s\nabla c^\top d < \epsilon_0$, stop.
    - For each trial step $t \in \left\{t_0, \rho t_0, \ldots, \rho^{K_t-1}t_0\right\}$:
      - Set $y = \omega_{\ell-1} + td/\left\|d\right\|$.
      - Correct y: while $g(y) \geq 0$ or $\hat{q}(\beta_y) \leq \hat{q}(\hat{\alpha})$ fails, at most K_y times, set $y \leftarrow y + \Delta$ with

$$\Delta = \arg\min_{\Delta\in\mathcal{F}_a(y)} \frac{1}{2}\Delta^\top M \Delta.$$

      - If $g(y) \geq 0$, $\hat{q}(\beta_y) \leq \hat{q}(\hat{\alpha})$, and $sc(\beta_y) > sc\left(\beta_{\omega_{\ell-1}}\right)$, accept t and stop; set $t_0 \leftarrow \eta t$ if t was the first trial, else $t_0 \leftarrow t$. Reject y if any of these cannot be evaluated.
    - If no trial step is accepted, stop.
    - Set $\omega_\ell = y$. If $s\left(c(\beta_{\omega_\ell}) - c\left(\beta_{\omega_{\ell-1}}\right)\right) < \epsilon$, stop.
  - Set $c^s = c(\beta_\omega)$ at the last ω reached.
- Return lower sensitivity $c(\hat{\alpha}) - c^-$, upper sensitivity $c^+ - c(\hat{\alpha})$, and sensitivity $\max\{c(\hat{\alpha}) - c^-, c^+ - c(\hat{\alpha})\}$.

C Basis Model Expansion for the Application to Einav, Finkelstein, and Cullen (2010)

An LLM proposed the basis model below, including the numerical pegs and the dimension, based on information in Online Appendix Figure 1 and the results of its exploration in the first (analytic) step of Algorithm 1. Because in our execution the workflow succeeds in the first step, we used additional (ad hoc) prompts to request this basis model for illustration purposes.

- Throughout, the domain of all functions is $p \in \left[0, \overline{P}\right]$.
- The average cost curve is $AC(p) = \gamma + \delta p$. The competitive price p_C solves $p_C = AC(p_C)$.
- Marginal cost $MC(p)$ relates to average cost $AC(p)$ and demand $D(p)$ through $MC(p) = AC(p) + \delta D(p) / D'(p)$.
- Parameters are $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$ and $t = (t_1, \ldots, t_K) \in \mathbb{R}^K$ where $K \leq 6$. The basis model is $\mathcal{B} = \left\{D(\cdot; \theta, t) : \theta \in \mathbb{R}^2, t \in \mathbb{R}^K\right\} \cap \mathcal{N}$, with generic element $\beta = D(\cdot; \theta, t)$. Demand is linear when $t = 0$, and $\hat{\alpha} = D\left(\cdot; \hat{\theta}, 0\right)$.
  - $H_k(p)$ is defined as

$$H_k(p) = \begin{cases} S\left(\dfrac{p - P_{k-1}}{P_k - P_{k-1}}\right), & P_{k-1} \leq p \leq P_k, \\ 1 - S\left(\dfrac{p - P_k}{P_{k+1} - P_k}\right), & P_k \leq p \leq P_{k+1}, \\ 0, & \text{otherwise}, \end{cases} \qquad S(x) = x^2(3 - 2x),$$

    where the knot set $(P_1, \ldots, P_K)$ consists of the first K elements of the tuple $(p_C, p_2, \ldots, p_6)$, sorted so that $P_1 < \cdots < P_K$. At the ends, $H_1(p) = 1$ for $p \leq P_1$ and $H_K(p) = 1$ for $p \geq P_K$.
  - $\lambda(p; \theta, t)$ is defined as

$$\lambda(p; \theta, t) = \sigma\left(\sigma^{-1}\left(\lambda_{\text{lin}}(p; \theta_1)\right) + \sum_{k=1}^{K} t_k H_k(p)\right)$$

    where $\lambda_{\text{lin}}(p; \theta_1) = \delta\left(\overline{P} + \left(\gamma/\delta - \overline{P}\right)\sigma(\theta_1) - p\right) / AC(p)$ is the gap between average and marginal cost (expressed as a share of average cost) under the linear demand curve and $\sigma(x) = 1/(1 + e^{-x})$ is the logistic function.
  - Demand is given by

$$D(p; \theta, t) = e^{\theta_0} \exp\left(-\int_0^p \frac{\delta \, dz}{AC(z)\, \lambda(z; \theta, t)}\right).$$

- The parameters satisfy the following constraints:

- $\theta_0 \leq 0$.
- $\lambda(p;\theta,t) - (1 - p/AC(p))$ crosses zero at most once, from below, over $p \in [0, \overline{P}]$.

- The expansion uses the following numerical pegs:
  - $\overline{P} = 800$.
  - $(p_2, \ldots, p_6) = (384, 800, 0, 520, 425)$.

D Additional Results for the Application to Einav, Finkelstein, and Cullen (2010)

Online Appendix Table 2: Share of the analytic sensitivity of the welfare cost in Einav et al. (2010) recovered, by dimension of the basis model

	Number of parameters	Fraction of lower sensitivity	Fraction of upper sensitivity	Fraction of total sensitivity
(working model)	2	0.000	0.000	0.000
	4	0.170	0.986	0.986
	6	0.201	0.989	0.989
(proposed basis model)	8	0.961	0.999	0.999

Notes: The first row is the working model itself. Each subsequent row relaxes it to the basis model (see Online Appendix C) at the stated number of parameters. The fractions of lower and upper sensitivity are the decrease and increase in the welfare cost relative to the working model, divided by the lower and upper sensitivity from the analytic solution, respectively. The fraction of total sensitivity is the larger of the two deviations, divided by the sensitivity from the analytic solution. Deviations are computed by running Online Appendix Algorithm 1 from the working model in each direction. The algorithm uses $L = 1000$, $\epsilon = 10^{-6}$, $\epsilon_0 = 10^{-8}$, $t_0 = 0.5$, $\eta = 2$, $\rho = 0.5$, $K_t = 50$, $K_y = 25$, $a = 10^{-8}$, and $M = G + \lambda\Lambda$ with $\lambda = 10^{-2}$, where G is the Gauss–Newton matrix of $\hat{q}$ and Λ is its diagonal floored at $\varphi = 10^{-6}$ times its largest entry.

Online Appendix Figure 1: Complete, filled problem-specification template for the Einav et al. (2010) application

Problem specification

Fill out this Markdown file. After completing the verification steps below, hand the filled Markdown file to an LLM instructed to follow the workflow in `WORKFLOW.md`. Provide precise mathematical definitions in each section unless instructed otherwise.

The *object of interest* $c(\nu) \in \mathbb{R}$ depends on an (unknown) parameter $\nu \in \mathcal{N}$, where the space $\mathcal{N}$ is the *nesting model* and may be infinite-dimensional. The data-dependent *criterion* $\hat{q}(\nu) \in \mathbb{R}$ describes the fit of the nesting model to the data. The *working model* is a finite-dimensional subset $\mathcal{A}$ of $\mathcal{N}$. The working model's fit is $\min_{\alpha \in \mathcal{A}} \hat{q}(\alpha)$, and the baseline estimate $\hat{\alpha} \in \mathcal{A}$ is the minimizer.

The objectives of this template are:

1. Precisely define the following objects:
   - The nesting model $\mathcal{N}$,
   - The working model $\mathcal{A}$,
   - The object of interest $c(\cdot)$, and
   - The criterion $\hat{q}(\cdot)$.
2. Specify the location of data, code, and estimates.

1. The nesting model

Define the nesting model $\mathcal{N}$, which encodes economic restrictions on unknown parameters, including unknown functions. Define the domain and range of all unknown functions formally. Specify any restrictions on the function class, such as continuity and shape restrictions. The nesting model may also include a finite set of other known or unknown parameters. Define the domain or values of these parameters.

Example. The nesting model $\mathcal{N}$ is the set of all demand functions $D : [P_0, P_1] \to [0, 1]$ that are:

- Continuous, and
- Strictly decreasing.

The bounds P_0 and P_1 are given in 'Location of data, code, and estimates'.

Your version.

Throughout, the primitive ν is the demand curve $D(p)$. The cost parameters (γ, δ) are fixed. They give the average cost curve $AC(p) = \gamma + \delta p$ and the marginal cost curve

$$MC(p) = AC(p) + \delta \frac{D(p)}{D'(p)}.$$

$\mathcal{N}$ is the set of demand curves $D : [0, 800] \to [0, 1]$ that satisfy the following properties:

- Continuous and differentiable.
- Strictly decreasing: $D'(p) < 0$ for all $p \in [0, 800]$.
- Bounded: $0 \le D(p) \le 1$ for all $p \in [0, 800]$.
- Single crossing of marginal cost and demand: if $MC(p) > p$ for some $p \in [0, 800]$, then $MC(p') > p'$ for all $p' \in [0, p)$.
- There exists $\bar{p} \in [0, 800]$ with $D(\bar{p}) > 0$ and $MC(p) < p$ for all $p \in (\bar{p}, 800]$.
- Nonnegative marginal cost: $MC(p) \ge 0$ for all $p \in [0, 800]$.

The cost parameters (γ, δ) are given in 'Location of data, code, and estimates'.

2. The working model

Define the working model $\mathcal{A} \subset \mathcal{N}$. Give its parameters, their domain, and the map from parameters to primitives.

Example. The working model is the isoelastic family

$$\mathcal{A} = \{D_{a,b}(p) = \exp(a - b \log(p)) : b > 0, a \le b \log P_0\}.$$

Parameters a and b map to the primitive $\alpha = D_{a,b} \in \mathcal{A}$.

Your version.

$\mathcal{A}$ is the linear-demand family

$$\mathcal{A} = \{D_{a,b}(p) = a + bp : D_{a,b} \in \mathcal{N}\},$$

with parameters (a, b) mapping to the primitive $D_{a,b}$.

3. The object of interest

Define the object of interest $c(\cdot) : \mathcal{N} \to \mathbb{R}$. It must be well-defined and finite for every $\nu \in \mathcal{N}$.

Example. Writing $\nu = D$, the object is the loss in consumer surplus from a price increase $P_0 \to P_1$:

$$c(D) = \int_{P_0}^{P_1} D(p) \mathrm{d}p.$$

Since $D \in [0, 1]$ on the bounded interval $[P_0, P_1]$, the integral lies in $[0, P_1 - P_0]$ and is finite for every $D \in \mathcal{N}$.

Your version.

The competitive price p_C is defined such that $p_C = AC(p_C)$, and the efficient price is the smallest price at which marginal cost is met from below:

$$p_E = \min\Big(\{p \in [0, 800] : p = MC(p)\}\Big).$$

$c(D)$ is the deadweight-loss area between demand and marginal cost, from the competitive price p_C to the efficient price p_E,

$$c(D) = \int_{D(p_C)}^{D(p_E)} \big[D^{-1}(Q) - MC\big(D^{-1}(Q)\big)\big] dQ.$$

Inverse demand $D^{-1}(Q)$ is well-defined for all $D \in \mathcal{N}$ because D is strictly decreasing. Inverse demand and marginal cost are bounded, so the integral exists and is finite for all $D \in \mathcal{N}$.

4. The criterion

Define the criterion $\hat{q}(\nu)$. The criterion is a data-dependent measure of fit, where smaller values indicate a better fit. The criterion must be finite and well-defined for every $\nu \in \mathcal{N}$.

Example. With observed shares s_i at prices $p_i \in [P_0, P_1]$, for $i = 1, \ldots, n$,

$$\hat{q}(D) = \frac{1}{n} \sum_{i=1}^{n} \big(s_i - D(p_i)\big)^2.$$

This evaluates $D \in [0, 1]$ at finitely many points, so it is finite for every $D \in \mathcal{N}$. The shares s_i and prices p_i are given under 'Location of data, code, and estimates'.

Your version.

$\hat{q}(D)$ is the employee-weighted sum of squared share residuals at the six observed prices,

$$\hat{q}(D) = \sum_{i=1}^{6} n_i \big(s_i - D(p_i)\big)^2,$$

where (p_i, s_i, n_i) are the observed price, share, and employee count.

5. Computational constraints (optional)

State the largest number of evaluations of $c(\cdot)$ and $\hat{q}(\cdot)$ you are willing to spend, for when either is expensive. Fill in the blanks below.

At most 100,000 evaluations of $c(\cdot)$ and at most 100,000 evaluations of $\hat{q}(\cdot)$.

6. Location of data, code, and estimates

Provide the location of data inputs. Provide the location of code that implements the object of interest and criterion (or numerical approximations of these functions). Provide the values of any parameters referenced in the prior sections. Provide the location or values of the parameters estimated under the working model.

Example.

- $\{(p_i, s_i)\}_{i=1}^{n}$ — `demand.csv`, columns `price` and `share`.
- $c(D)$ — implemented in `consumer_surplus.py`.
- $\hat{q}(D)$ — implemented in `estimator.py`.
- P_0, P_1 — equal 100, 200.
- $\hat{a}, \hat{b}$ parameter estimates - `estimates.csv`, columns `intercept` and `slope`.

Your version.

- $\{(p_i, s_i, n_i)\}_{i=1}^{6}$ — `source/raw/EFC2010/data/EFC2010_table2.csv`.
- $c(D)$ and $\hat{q}(D)$ — `source/analysis/EFC2010/specification.py:{Target,Criterion}`.
- Demand estimates $(\hat{a}, \hat{b})$ and cost estimates (γ, δ) — `output/analysis/EFC2010/replicate.txt`.

7. Consistency and completeness checklist

After you complete the template, ask your LLM to verify the following prior to launching the workflow:

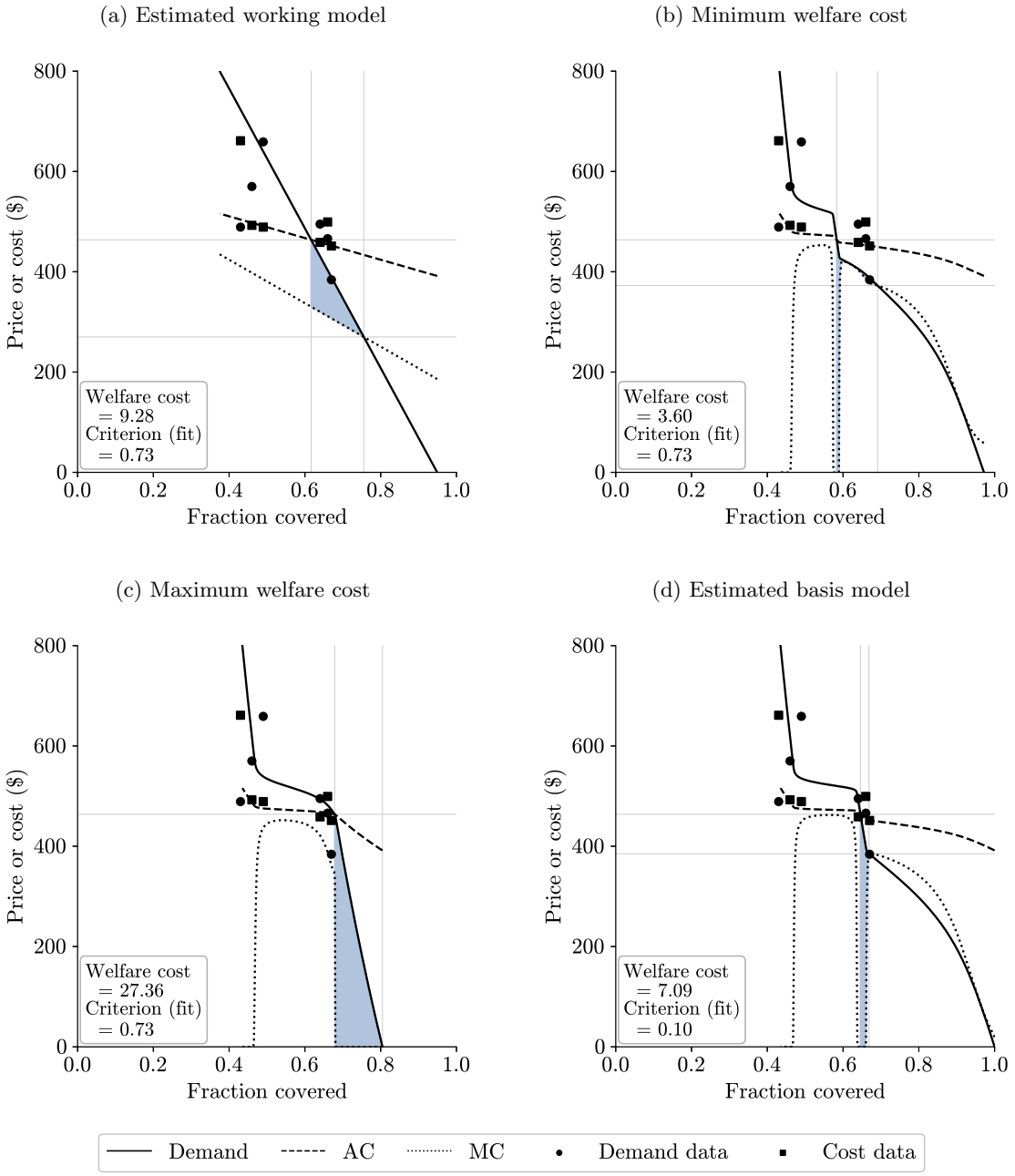
- ☒ 'The nesting model', 'The working model', 'The object of interest', and 'The criterion' are each filled with precise mathematical definitions.
- ☒ The nesting model $\mathcal{N}$ is non-empty.
- ☒ The working model is a genuine subset of the nesting model, $\mathcal{A} \subset \mathcal{N}$, and is finite-dimensional.
- ☒ The object of interest $c(\nu)$ is well-defined and finite for every $\nu \in \mathcal{N}$.
- ☒ The criterion $\hat{q}(\nu)$ is well-defined and finite for every $\nu \in \mathcal{N}$.
- ☒ Every data or code symbol used in 'The nesting model' through 'Computational constraints' is resolved in 'Location of data, code, and estimates'.

8. Workflow configuration

Should the LLM attempt a Lean formalization of mathematical claims? Formalization may add time and token cost, and is not required to verify numerical certificates.

- ☒ Yes
- ☐ No

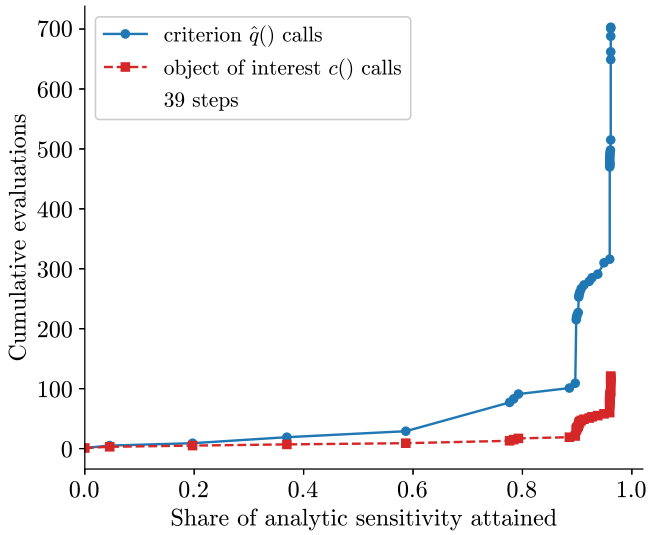
Online Appendix Figure 2: Sensitivity of the welfare cost in Einav et al. (2010) to alternative demand functions, numerical lower bound



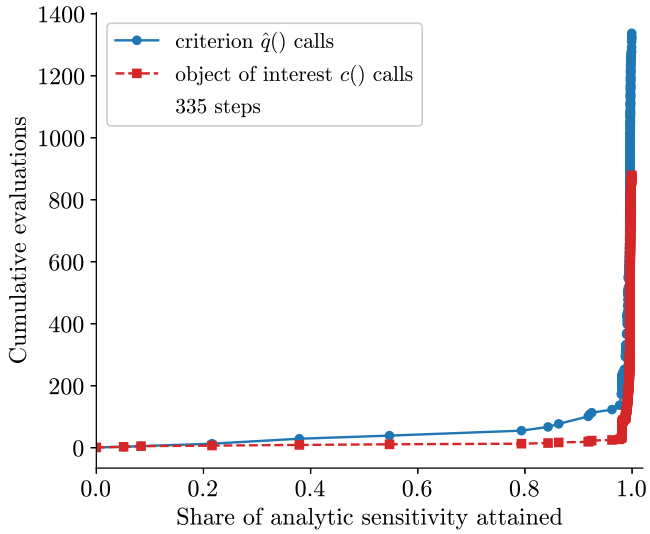
Notes: Each panel plots a demand function, the estimated average cost curve, and the implied marginal cost curve, as in Figure V of Einav et al. (2010), with the area of the shaded region denoting the implied welfare cost due to adverse selection. Panels differ in which demand function they use. Panel (a) uses the linear demand function implied by the estimated working model. Panels (b) and (c) use demand functions that result from applying Online Appendix Algorithm 1, under the eight-parameter basis model described in Online Appendix C. Panel (b) uses the demand function that attains the lowest welfare cost found. Panel (c) uses the demand function that attains the highest welfare cost found. Panel (d) uses a demand function that minimizes the criterion function under the basis model; the optimizer starts from the estimated working model. The implementation of Online Appendix Algorithm 1 uses $L = 1000$, $\epsilon = 10^{-6}$, $\epsilon_0 = 10^{-8}$, $t_0 = 0.5$, $\eta = 2$, $\rho = 0.5$, $K_t = 50$, $K_y = 25$, $a = 10^{-8}$, and $M = G + \lambda\Lambda$ with $\lambda = 10^{-2}$, where G is the Gauss–Newton matrix of $\hat{q}$ and Λ is its diagonal floored at $\varphi = 10^{-6}$ times its largest entry.

Online Appendix Figure 3: Computational cost of the numerical lower bound on sensitivity in Einav et al. (2010)

(a) Minimum welfare cost



(b) Maximum welfare cost



Notes: Each panel shows the cumulative number of evaluations of the criterion function and object of interest function used in applying Online Appendix Algorithm 1, under the eight-parameter basis model described in Online Appendix C, for the application to Einav et al. (2010). The horizontal axis is the welfare cost attained, as a share of the analytic bound in the same direction, so 0 is the working model and 1 is the analytic bound. Each point is a feasible step along the solver path. An evaluation of an analytic gradient or Jacobian counts as one call of the associated function. The implementation of Online Appendix Algorithm 1 uses $L = 1000$, $\epsilon = 10^{-6}$, $\epsilon_0 = 10^{-8}$, $t_0 = 0.5$, $\eta = 2$, $\rho = 0.5$, $K_t = 50$, $K_y = 25$, $a = 10^{-8}$, and $M = G + \lambda\Lambda$ with $\lambda = 10^{-2}$, where G is the Gauss–Newton matrix of $\hat{q}$ and Λ is its diagonal floored at $\varphi = 10^{-6}$ times its largest entry.

Appendix References

Aguirregabiria, V., Clark, R., and Wang, H. (2025). The geographic flow of bank funding and access to credit: Branch networks, synergies, and local competition. *American Economic Review*, 115(6):1818–1856.

Allcott, H., Cohen, D., Morrison, W., and Taubinsky, D. (2025). When do "nudges" increase welfare? *American Economic Review*, 115(5):1555–1596.

Alvarez, F., Borovičková, K., and Shimer, R. (2025). Consistent evidence on duration dependence of price changes. *American Economic Review*, 115(10):3322–3366.

Amodio, F., Medina, P., and Morlacco, M. (2025). Labor market power, self-employment, and development. *American Economic Review*, 115(9):3014–3057.

Atkeson, A., Heathcote, J., and Perri, F. (2025). The end of privilege: A reexamination of the net foreign asset position of the United States. *American Economic Review*, 115(7):2151–2206.

Balboni, C. (2025). In harm's way? Infrastructure investments and the persistence of coastal cities. *American Economic Review*, 115(1):77–116.

Benetton, M., Gavazza, A., and Surico, P. (2025). Mortgage pricing and monetary policy. *American Economic Review*, 115(3):823–863.

Berndt, A., Duffie, D., and Zhu, Y. (2025). The decline of too big to fail. *American Economic Review*, 115(3):945–974.

Bonfim, D., Ferreira, M. A., Queiró, F., and Zhao, S. E. (2025). Fiscal policy and credit supply in a crisis. *American Economic Review*, 115(6):1896–1935.

Bornstein, G. and Peter, A. (2025). Nonlinear pricing and misallocation. *American Economic Review*, 115(11):3853–3908.

Chen, Q., Chen, Z., Liu, Z., Suárez Serrato, J. C., and Xu, D. Y. (2025). Regulating conglomerates: Evidence from an energy conservation program in China. *American Economic Review*, 115(2):408–447.

Choukhmane, T. (2025). Default options and retirement saving dynamics. *American Economic Review*, 115(11):3749–3787.

Chyn, E. and Daruich, D. (2025). An equilibrium analysis of the effects of neighborhood-based interventions on children. *American Economic Review*, 115(12):4476–4522.

Dhyne, E., Kikkawa, A. K., Komatsu, T., Mogstad, M., and Tintelnot, F. (2025). Firm responses and wage effects of foreign demand shocks with fixed labor costs and monopsony. *American Economic Review*, 115(12):4328–4368.

Doraszelski, U., Seim, K., Sinkinson, M., and Wang, P. (2025). Ownership concentration and strategic supply reduction. *American Economic Review*, 115(3):903–944.

Dube, A., Manning, A., and Naidu, S. (2025). Monopsony and employer misoptimization explain why wages bunch at round numbers. *American Economic Review*, 115(8):2689–2721.

Einav, L., Finkelstein, A., and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics*, 125(3):877–921.

Einav, L., Klopack, B., and Mahoney, N. (2025). Selling subscriptions. *American Economic Review*, 115(5):1650–1671.

Gagliardone, L., Gertler, M., Lenzu, S., and Tielens, J. (2025). Anatomy of the Phillips curve: Micro evidence and macro implications. *American Economic Review*, 115(11):3941–3974.

Gaubert, C., Kline, P., Vergara, D., and Yagan, D. (2025). Place-based redistribution. *American Economic Review*, 115(10):3415–3450.

Gómez-Cram, R. and Lawrence, A. (2025). The value of software. *American Economic Review*, 115(10):3514–3558.

Gourinchas, P.-O., Ray, W., and Vayanos, D. (2025). A preferred-habitat model of term premia, exchange rates, and monetary policy spillovers. *American Economic Review*, 115(11):3788–3824.

Greenwald, D. L. and Guren, A. (2025). Do credit conditions move house prices? *American Economic Review*, 115(10):3559–3596.

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., and Young, P. (2025). A stepping stone approach to norm transitions. *American Economic Review*, 115(7):2237–2266.

Haanwinckel, D. (2025). Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution. *American Economic Review*, 115(12):4137–4182.

Haddad, V., Huebner, P., and Loualiche, E. (2025). How competitive is the stock market? Theory, evidence from portfolios, and implications for the rise of passive investing. *American Economic Review*, 115(3):975–1018.

Jha, A. and Leslie, G. (2025). Start-up costs and market power: Lessons from the renewable energy transition. *American Economic Review*, 115(2):690–724.

Kroft, K., Luo, Y., Mogstad, M., and Setzler, B. (2025). Imperfect competition and rents in labor and product markets: The case of the construction industry. *American Economic Review*, 115(9):2926–2969.

Li, Y. (2025). Fragile new economy: Intangible capital, corporate savings glut, and financial instability. *American Economic Review*, 115(4):1100–1141.

Mian, A., Straub, L., and Sufi, A. (2025). A Goldilocks theory of fiscal deficits. *American Economic Review*, 115(12):4253–4291.

Pellegrino, B. (2025). Product differentiation and oligopoly: A network approach. *American Economic Review*, 115(4):1170–1225.

Shepard, M. and Wagner, M. (2025). Do ordeals work for selection markets? Evidence from health insurance auto-enrollment. *American Economic Review*, 115(3):772–822.

Stanton, C. T. and Thomas, C. (2025). Who benefits from online gig economy platforms? *American Economic Review*, 115(6):1857–1895.

Starc, A. and Wollmann, T. G. (2025). Does entry remedy collusion? Evidence from the generic prescription drug cartel. *American Economic Review*, 115(5):1400–1438.