The Incredible Flexibility of Moment Matching

Isaiah Andrews and Bas Sanders*

Abstract

We ask how far the choice of which moments to match can push estimates in a misspecified structural model. The answer is: very far. Under mild conditions, an adversarial researcher informed about the data distribution can choose moments that render any parameter value the unique solution to the population moment-matching problem. Moreover, in many cases they can do so with little increase in model-implied standard errors relative to maximum likelihood. We illustrate both results in a menu-cost model, and discuss restricting, motivating, or pre-committing to the choice of moments as routes to reduce researcher degrees of freedom.

1 Introduction

Economists often view structural models as approximations rather than literal descriptions of the data-generating process. For this reason, many researchers estimate such models by matching a set of hand-selected moments rather than using more traditional statistical approaches like maximum likelihood.[1] This practice affords researchers the freedom to base estimation on the predictions of the model that they deem most economically important, and thus seems potentially appealing in settings where researchers think their models may be misspecified.

At the same time, the flexibility to choose moments introduces additional researcher degrees of freedom (Simmons et al., 2011) which could in principle be used to steer estimates

*This version: April 30, 2026. Andrews: MIT Department of Economics and NBER; iandrews@mit.edu. Sanders: SEO Amsterdam Economics; b.sanders@seo.nl. We thank Claude Opus 4.6/7 and GPT 5.4 Pro for outstanding research assistance.

[1]For instance, Bordalo et al. (2020) writes that "We prefer [moment matching] to maximum likelihood for two reasons. First, one advantage of our model is that it is simple and transparent. However, this simplicity comes at the cost of likely misspecification, and it is well known that with misspecification concerns moment estimators are often more reliable."

toward a desired conclusion. Moreover, the practical cost of searching over moment specifications has fallen sharply with the advent of larger research teams and, more recently, AI-assisted coding tools. In this paper, we study how much scope there is for such manipulation by asking what results an adversarial researcher could engineer through their choice of moments. Our main finding is an "anything-goes" result for moment matching estimation under misspecification: under mild conditions, for any pre-specified parameter value there exists a choice of moments which renders that parameter value the unique solution to the population moment-matching problem. Thus, an adversarial researcher who knows the true data-generating process can construct moments delivering any estimand they wish.[2]

One might wonder if such contrived moment choices will generate large standard errors and thus be statistically unconvincing, but this turns out not to be the case. Specifically, we characterize the choice of moments which minimizes model-implied asymptotic variance subject to delivering a given estimand, and show that in many cases the resulting standard errors are close to those from maximum likelihood, which in turn lower-bound the model-implied standard errors for any asymptotically normal and unbiased estimator.

To illustrate our results, following Cocci and Plagborg-Møller (2024) we estimate the model of Alvarez and Lippi (2014), who study menu-cost price setting in multiproduct firms. Fitting the Alvarez and Lippi (2014) model to the distribution of centered nonzero price changes in the scanner data used by Cocci and Plagborg-Møller (2024), we find that considering alternative, intuitively reasonable moments generates a broad range of point estimates, where the range of estimates is wide relative to the degree of statistical uncertainty (as reflected by the standard errors proposed in Cocci and Plagborg-Møller 2024). Going further, we apply our adversarial moment construction in these data, and show that an adversarial researcher could generate an even wider range of results, in many cases with standard errors little larger than those implied by maximum likelihood.

Our paper contributes to the literature on best practices for moment matching, a discussion that traces back at least to the calibration literature (Prescott, 1986; Hansen and Heckman, 1996; Cooley, 1997) and that has long recognized the importance of moment choice. For example, Wooldridge (2001) notes: "One problem is that the researcher must choose the additional moment conditions to be added in an ad hoc manner." In practice, researchers often argue that their moments capture the key empirical facts their model is designed to explain. Many papers begin by documenting what the authors view as the important pat-

[2]While we focus on ex-ante moment specification and the population estimand, as we discuss below a researcher who chooses the moments after observing the sample can likewise often engineer any estimate value they wish.

terns in the data, and then choose moments which reflect these patterns. This aims to make the recovered parameters economically meaningful in the sense that they reproduce what the researchers view as the central regularities in the data.[3]

In addition, many papers compare untargeted moments with their model-implied counterparts as a way to illustrate overall model fit.[4] It is also common for papers to include an "identification" section in which they discuss, formally or informally, what features of the data are important for a given estimate or conclusion.[5] Andrews et al. (2017, 2020) introduce diagnostics for this purpose and discuss the limitations of certain informal approaches.

Prior work has proposed methods for systematically constructing informative moments. For example, Gallant and Tauchen (1996) propose generating moment conditions from the score of an auxiliary model that approximates the data distribution. More recent work has focused on statistical inference and efficiency in calibration and moment-matching exercises. Cocci and Plagborg-Møller (2024) observe that calibration can be interpreted as minimum distance estimation and derive conservative standard errors when the covariance structure of the empirical moments is unknown. Kaji et al. (2023) introduce adversarial estimation, which estimates structural models by using a discriminator to learn features that best distinguish simulated from observed data. Our paper complements this literature by highlighting a different dimension of the moment selection problem: the scope for moment choice to alter the resulting estimands.

We close the paper with a discussion of what steps might be taken to mitigate the concerns we raise. For instance, one could restrict the class of estimators ex-ante (e.g. requiring that everyone report the MLE) but this ignores the context-specific misspecification concerns which motivate moment-matching in the first place. Alternatively, one could ask that researchers more explicitly articulate what misspecification concerns they have in mind and why this leads them to their particular choice of moments, or ask that (when possible) researchers commit to their choice of moments before they know the details of the data distribution. In our view, versions of these ideas are already present in some moment-matching applications. Our negative theoretical results suggest, however, it might be useful

[3]For example, Lagakos et al. (2023) write: "These moments help the model to jointly fit key aggregate facts from the Bangladeshi economy relevant for understanding rural-urban migration, plus the household responses to migration incentives, which are well identified through the experimental evidence."

[4]For example, Benhabib et al. (2019) write: "Recall that we only explicitly target the diagonal elements of the Markovian transition matrix. We plot the whole matrix in this figure in order to get a sense how well we do on the (non-targeted) off-diagonal cells."

[5]For example, Ahlfeldt et al. (2015) write: "We now consider each of the moment conditions in turn and show how they identify the parameters."

to think more systematically about the role of such norms and rules.

The rest of the paper is organized as follows. Section 2 illustrates the empirical consequences of moment choice in an example building on Alvarez and Lippi (2014) and Cocci and Plagborg-Møller (2024). Section 3 introduces our formal setting and characterizes the class of moments that yield a specific moment-matching estimand. Section 4 finds an efficient moment function in this class. Section 5 returns to the Alvarez and Lippi (2014) application to discuss implementation of the adversarial approach. Section 6 discusses possible ways to mitigate the issues suggested by our theoretical results. All proofs are collected in the Supplementary Appendix.

2 Illustration: Moment Choice in a Menu-Cost Model

To illustrate the empirical consequences of moment choice before turning to the theory, we follow Cocci and Plagborg-Møller (2024) and estimate Alvarez and Lippi (2014)'s model of menu-cost price setting in multiproduct firms. To obtain a simplified, two-parameter version of the model in which estimates can easily be plotted, we focus on the cross-sectional distribution of non-zero absolute centered log price changes, $X := |\Delta p - \overline{\Delta p}|$, conditional on $\Delta p \neq 0$.[6] Alvarez and Lippi (2014)'s model implies that the distribution of X depends on the model's structural primitives through two objects: the number of products N (n in the notation of Alvarez and Lippi 2014) and the threshold $\bar{y}$ at which the optimal policy resets prices.[7] We therefore parameterize the model by $\theta = (N, \bar{y})$, and study how the estimated value of θ varies with the choice of moments.

As a baseline, we consider matching the second and fourth moments of X, which is a just-identified subset of the moments considered in Cocci and Plagborg-Møller (2024). We compute the conservative standard errors of Cocci and Plagborg-Møller (2024) at the resulting estimate, and plot the estimate along with the marginal 95% confidence intervals for the two parameters in black in each of Figures 1-5.

We next explore the scope for obtaining different estimates of θ by matching alternative moments. Specifically, we consider a range of just-identified specifications, matching different

[6] Following Cocci and Plagborg-Møller (2024), we use demeaned log changes since Alvarez and Lippi (2014) abstract from inflation. We take absolute values since the absolute log change suffices to compute all of the estimation moments considered in Cocci and Plagborg-Møller (2024).

[7] In the original three-parameter formulation of Alvarez and Lippi (2014), the parameters are N, the volatility of desired prices σ, and a scaled menu-cost parameter s. The threshold in the optimal policy is then $\bar{y} \approx \sigma s\sqrt{2(N+2)}$. Proposition 6 in Alvarez and Lippi (2014) implies that the density of X depends on (N, σ, s) only through $(N, \bar{y})$.

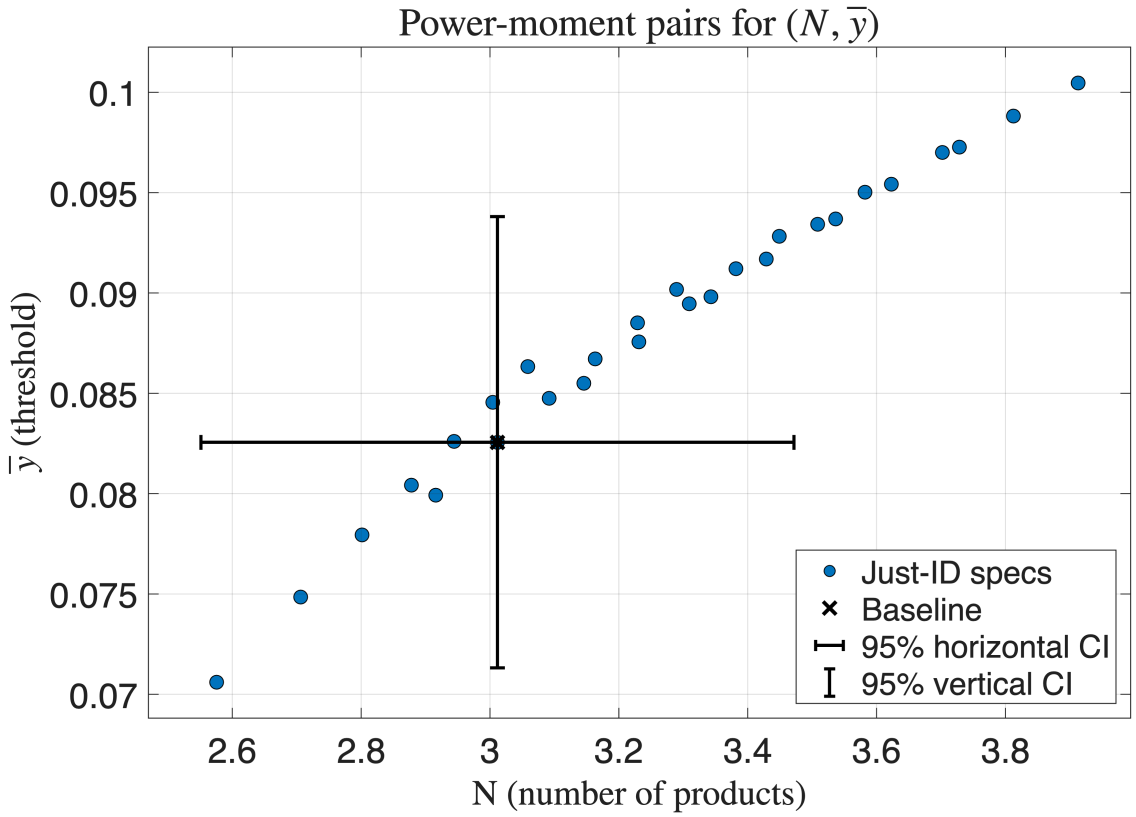

Figure 1: Point estimates for θ, considering powers of the distribution of non-zero absolute centered log price changes. The horizontal and vertical CIs are conservative confidence intervals based on Cocci and Plagborg-Møller (2024).

aspects of the distribution of log price changes. Figure 1 reports point estimates for θ obtained from all pairs of power moments $\mathrm{E}[X^k]$ with $k \in \{1, \ldots, 8\}$. As the figure highlights, the resulting range of estimates extends well beyond the range of statistical uncertainty captured by the Cocci and Plagborg-Møller (2024) standard errors.[8] This only becomes more apparent as we broaden the range of moments considered. Figure 2 shows results when we also match quantiles of the price change distribution, considering quantiles in the grid $\{0.10, 0.20, \ldots, 0.90\}$. Specifically, we plot results for all power-power pairs, all power-quantile pairs, and all quantile-quantile pairs (restricting to pairs with $|q_1 - q_2| \geq 0.25$ to exclude nearly redundant quantile pairs). The resulting range of estimates extends far beyond the range suggested by the Cocci and Plagborg-Møller (2024) standard errors.

We next consider a researcher who, rather than limiting attention to powers and quantiles, exploits the full flexibility of moment matching. For a given target parameter value

[8]This reflects no flaw in the Cocci and Plagborg-Møller (2024) standard errors, which are designed to upper bound statistical uncertainty given a set of moments, not variability induced by the choice of moments.

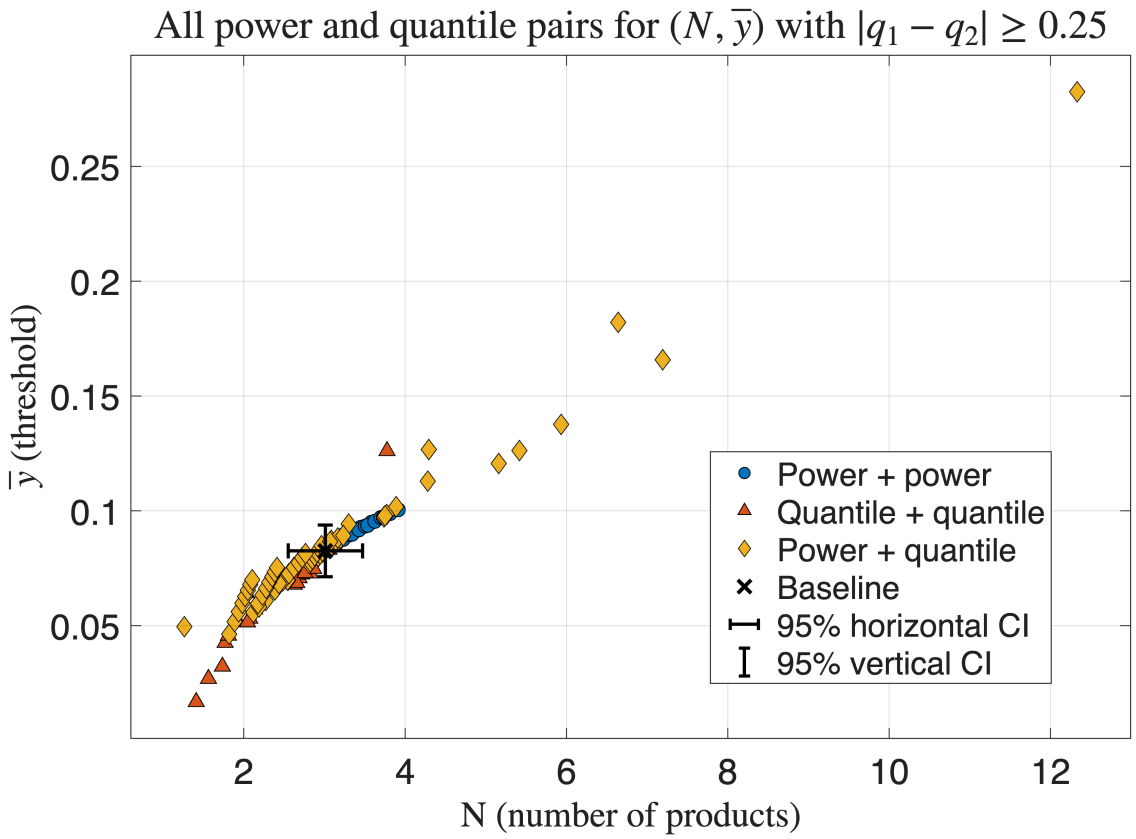

Figure 2: Point estimates for θ, considering powers and quantiles of the distribution of non-zero absolute centered log price changes. The horizontal and vertical CIs are conservative confidence intervals based on Cocci and Plagborg-Møller (2024).

$\bar{\theta} = (N, \bar{y})$, our results in Sections 3 and 4 below show how a researcher who knows the distribution of X can choose just-identified moments which (a) imply estimand $\bar{\theta}$ and (b) minimize the model-implied variance over moments with that estimand.[9] Figure 3 applies this construction to a grid of 12 target parameter values $\bar{\theta} = (N, \bar{y})$ with $N \in \{2, 3, 4, 5\}$ and $\bar{y} \in \{0.06, 0.12, 0.18\}$ and a distribution estimate based on the observed data. As expected given the theory, the resulting estimates track their targets up to sampling uncertainty, and cover an even broader range of parameter values than obtained using powers and quantiles. Moreover, these targeted estimates come with reasonable-looking standard errors: across the 12 sets of moments, the standard errors for N (evaluated at the estimates $\hat{\theta}$) range from 0.2% to 39% larger than the (model-implied) MLE standard error evaluated at the same point, with a median of 9.7%, while the standard errors for $\bar{y}$ range from 3.4% to 214.2% larger,

[9]For this exercise we censor the distribution of observed price changes to ensure that the support of X does not vary with θ, which would introduce even more researcher degrees of freedom. See Section 5 below for details.

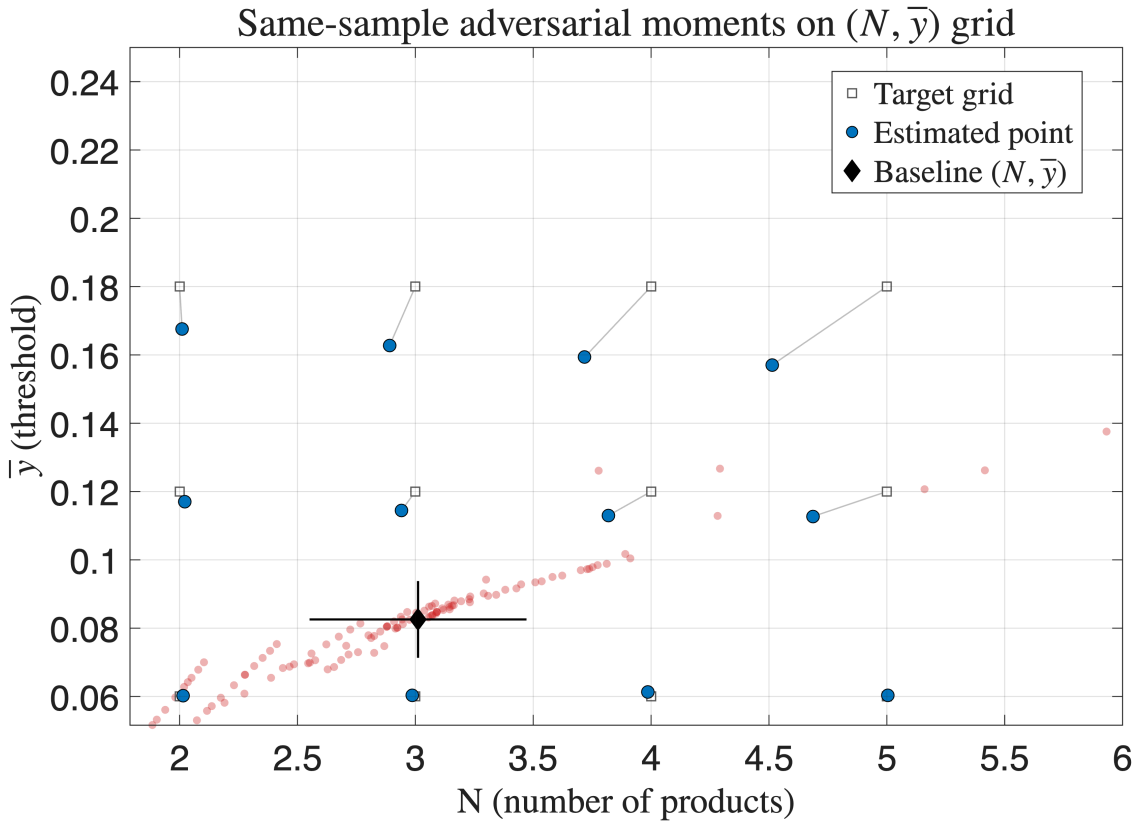

Figure 3: Preview of the adversarial construction in the Alvarez and Lippi (2014) application. Target grid (open squares) and point estimates (blue dots) for $\theta = (N, \bar{y})$ obtained from the target-specific adversarial moments. The gray lines connect each target to its estimate. The red scatter shows, for reference, the varying-moment estimates from Figure 2; the black diamond and bars are the baseline $(N, \bar{y})$ estimate and confidence intervals.

with a median of 17.2%. A full tabulation appears in Table 1 of the Online Appendix.

Overall, these results highlight that even in a simple model, the choice of moments can greatly influence the results obtained without necessarily incurring a large standard error cost. We next show that this is not a peculiarity of this specific example, but is instead a general property of moment matching applied to misspecified models.

3 The Flexibility of Moment Matching

3.1 Setting

Consider a researcher who observes i.i.d. draws $X_1, ..., X_n \in \mathcal{X}$ from a probability distribution P. The researcher postulates a parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with parameter space $\Theta \subseteq \mathbb{R}^K$. For each $\theta \in \Theta$, let p_θ denote the density of P_θ with respect to a common

dominating measure (e.g. Lebesgue measure for continuous data, or counting measure for discrete data). We do not assume correct specification: the model is statistically well-specified if and only if $P \in \mathcal{P}$, or equivalently if there exists $\theta \in \Theta$ such that $P = P_\theta$.

The researcher estimates their model by moment matching. Specifically, they choose target moments $g : \mathcal{X} \to \mathbb{R}^D$ with $D \geq K$ and a (possibly data-dependent) symmetric positive definite weighting matrix $W_n \in \mathbb{R}^{D \times D}$.[10] Define the *sample moment function* as the difference between the sample average of the target moments and their model-implied mean

$$m_n(g, \theta) := \frac{1}{n} \sum_{i=1}^{n} g(X_i) - \mathrm{E}_\theta[g(X)],$$

where E_θ denotes the expectation when $X \sim P_\theta$. The researcher estimates θ by minimizing the W_n-weighted distance of the sample moments from zero, yielding the set of minimizers

$$\hat{\Theta}_g(W_n) := \arg\min_{\vartheta \in \Theta} m_n(g, \vartheta)^\top W_n m_n(g, \vartheta). \tag{1}$$

As conventional, we define population analogs by replacing sample averages with expectations under the true data distribution P. Define the *population moment function* as

$$m(g, \theta) := \mathrm{E}_P[g(X)] - \mathrm{E}_\theta[g(X)].$$

Analogously, provided $W_n \xrightarrow{p} W = W(P)$ as $n \to \infty$ for some symmetric positive definite matrix $W \in \mathbb{R}^{D \times D}$, we define the population *pseudo-true set* as

$$\Theta_g(W) := \arg\min_{\vartheta \in \Theta} m(g, \vartheta)^\top W m(g, \vartheta).$$

Under misspecification, $\Theta_g(W)$ generally depends on both the choice of moments g and the weight matrix W.

In the special case where the moments g admit "perfect fit", meaning that there exists at least one $\theta \in \Theta$ such that $m(g, \theta) = 0$, define the set of population solutions to the moments as

$$\Theta_g := \{\vartheta \in \Theta : m(g, \vartheta) = 0\}.$$

When the perfect-fit condition holds, $\Theta_g(W) = \Theta_g$ for all positive-definite W, so the choice of weights no longer matters for the estimand. Since our interest in this paper is in the

[10] Note that g does not depend on θ; model dependence enters through $\mathrm{E}_\theta[g(X)]$.

choice of moments, rather than the choice of weights, we limit attention to g satisfying the perfect-fit condition.[11] This also ensures that the model's over-identifying restrictions hold so long as the researcher limits attention to moments in g.

3.2 Moment Engineering

Consider a motivated researcher who would like to report a particular parameter value $\bar{\theta}$. Since the moment-matching estimand Θ_g depends on the chosen moments g, presumably some choices of moments will deliver estimands closer to the target value than others. Absent other constraints (e.g. from scientific integrity, social conventions on the form of "acceptable" moments, etc.) how far could such a researcher get?

To answer this question, we consider the extreme case of a fully adversarial researcher who engineers their moments to make $\bar{\theta}$ a population solution of the moment conditions. For the time being we further assume that this researcher knows the true data generating process P, though we return to this point below. Restricting attention to moments which are square-integrable under the target parameter value (i.e. $g \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right)$), define

$$\mathcal{C}(\bar{\theta}) := \left\{ g \in L^2(P_{\bar{\theta}}; \mathbb{R}^D) : \mathrm{E}_P[g(X)] = \mathrm{E}_{P_{\bar{\theta}}}[g(X)] \right\}.$$

In words, $\mathcal{C}\left(\bar{\theta}\right)$ collects all $\mathbb{R}^D$-valued moment functions whose population mean under the true distribution P matches their mean under the model-implied distribution $P_{\bar{\theta}}$.

The next lemma characterizes the function class $\mathcal{C}\left(\bar{\theta}\right)$. It uses the chi-squared divergence, which for distributions Q_1 and Q_2 with Q_1 absolutely continuous with respect to Q_2 (that is, where all sets assigned probability zero by Q_2 are also assigned probability zero by Q_1), is defined as

$$\chi^2\left(Q_1 || Q_2\right) := \mathrm{E}_{Q_2}\left[\left(\frac{dQ_1}{dQ_2}(X) - 1\right)^2\right]. \tag{2}$$

When Q_1 is not absolutely continuous with respect to Q_2, the χ^2 divergence is defined to be infinite.

[11] Since this constrains the choice set, relaxing this constraint only increases researcher degrees of freedom. Even if the perfect fit condition fails, any interior minimizer of the weighted problem will satisfy the perfect-fit condition for new moments formed based on the first order condition $\frac{\partial m(g,\vartheta)}{\partial \vartheta}^\top W m(g,\vartheta) = 0$. See e.g. Andrews et al. (2025) for a complementary discussion of how weight choices shape estimates and estimands for a fixed set of moments.

Lemma 1 (Characterizing $\mathcal{C}\left(\bar{\theta}\right)$)**.** *If* $\chi^2\left(P||P_{\bar{\theta}}\right) \in (0,\infty)$, *then*

$$\mathcal{C}\left(\bar{\theta}\right) = \left\{ f - \frac{\mathrm{E}_{\bar{\theta}}\left[f\left(X\right)\left(\frac{dP}{dP_{\bar{\theta}}}\left(X\right)-1\right)\right]}{\chi^2\left(P||P_{\bar{\theta}}\right)}\left(\frac{dP}{dP_{\bar{\theta}}}(X)-1\right) : f \in L^2\left(P_{\bar{\theta}};\mathbb{R}^D\right)\right\}.$$

Note that $\chi^2\left(P||P_{\bar{\theta}}\right) \in (0,\infty)$ implies that (i) $P \neq P_{\bar{\theta}}$, so the model-implied distribution with parameter $\bar{\theta}$ is not the true DGP, and (ii) P is absolutely continuous with respect to $P_{\bar{\theta}}$.

Intuitively, $\mathcal{C}\left(\bar{\theta}\right)$ is constructed by taking all square-integrable functions f and linearly residualizing them against the recentered likelihood-ratio $\frac{dP}{dP_{\bar{\theta}}} - 1$. The recentered likelihood ratio has mean zero under $\bar{\theta}$ by construction, while the residualization corrects for the mean of f under P. If $f\left(x\right) = 1$, the construction yields $g\left(x\right) = 0$, which has no identifying power: every $\theta \in \Theta$ satisfies the resulting population moment equation. If $f\left(x\right) = x$, the resulting moment is

$$g\left(x\right) = x - \frac{\mathrm{E}_P\left[X\right] - \mathrm{E}_{\bar{\theta}}\left[X\right]}{\chi^2\left(P||P_{\bar{\theta}}\right)}\left(\frac{dP}{dP_{\bar{\theta}}}\left(x\right)-1\right),$$

which corrects for the difference in the mean of X under P and $P_{\bar{\theta}}$.

As the example with $f(x) = 1$, $g\left(x\right) = 0$ illustrates, a given moment function in $\mathcal{C}\left(\bar{\theta}\right)$ might fail to pin down $\bar{\theta}$ uniquely, in the sense that $\left\{\bar{\theta}\right\} \subset \Theta_g$. We next show, however, that if the model is misspecified then under mild conditions an adversarial researcher can eliminate such competing solutions by adding finitely many moments in $\mathcal{C}\left(\bar{\theta}\right)$, yielding a finite collection of moment conditions whose unique population solution is $\bar{\theta}$.

To state this result, let

$$L := \frac{dP}{dP_{\bar{\theta}}},\ L_\theta := \frac{dP_\theta}{dP_{\bar{\theta}}},$$

denote the likelihood ratio of the true distribution P relative to the target distribution $P_{\bar{\theta}}$, and of a generic model-implied distribution P_θ relative to $P_{\bar{\theta}}$, respectively. Furthermore, let $r := L - 1$ and $r_\theta := L_\theta - 1$ denote the corresponding recentered likelihood-ratios, and define $s_\theta\left(x\right) := \frac{\partial}{\partial\theta}\log p_\theta\left(x\right)$ as the score function (i.e. gradient of the log density) at θ.

Assumption 1 (Uniqueness of estimand)**.** *Suppose* $0 < \mathrm{E}_{P_{\bar{\theta}}}[r(X)^2] < \infty$, *and that* P_θ, $P_{\bar{\theta}}$ *are mutually absolutely continuous for all* $\theta \in \Theta$. *Further assume:*

(i) Θ *is compact.*

(ii) The map $\theta \mapsto r_\theta$ *is continuous from* Θ *into* $L^2(P_{\bar{\theta}})$.

(iii) *The functions* $\theta \mapsto \mathrm{E}_{P_\theta}[s_{\bar{\theta}}(X)]$ *and* $\theta \mapsto \mathrm{E}_{P_\theta}[r(X)]$ *are continuously differentiable in a neighborhood of* $\bar{\theta}$*, with derivative obtained by differentiating under the integral sign.*

(iv) *The score* $s_{\bar{\theta}} \in L^2(P_{\bar{\theta}}; \mathbb{R}^K)$ *satisfies rank* $\left(\mathrm{E}_{\bar{\theta}}\left[s_{\bar{\theta}}(X) s_{\bar{\theta}}(X)^\top\right]\right) = K$ *and*

$$\mathrm{E}_{P_{\bar{\theta}}}\left[\left(v^\top s_{\bar{\theta}}(X) - r(X)\right)^2\right] > 0 \qquad \text{for every } v \neq 0.$$

(v) *For every* $\theta \neq \bar{\theta}$*, none of* $P_{\bar{\theta}}$*,* P_θ*, and* P *is a (possibly degenerate) mixture of the other two.*

Condition (i) requires compactness of the parameter space, which is a standard regularity condition for nonlinear estimation results. Condition (ii) is a continuity condition on the model-implied density, and is sufficient for $\theta' \to \theta$ to imply $E_{\theta'}[f(X)] \to E_\theta[f(X)]$ for all $f \in L^2(P_{\bar{\theta}})$. In a similar spirit, condition (iii) is a differentiability condition which says, for the specific moments $E_\theta[s_{\bar{\theta}}] = \int s_{\bar{\theta}}(x) dP_\theta(x)$ and $E_\theta[r] = \int r(x) dP_\theta(x)$, we can pass the derivative with respect to θ through the integral. Condition (iv) requires that the recentered likelihood ratio r not lie in the K-dimensional linear subspace spanned by the score, and that the Fisher information have full rank at $\bar{\theta}$. Loosely speaking, the span requirement means that there is not some direction in which we can move, local to $\bar{\theta}$, so that the change in r_θ is proportional to r. Finally, condition (v) is a global condition on the model and implies that (a) the model is misspecified (in the sense that $P \notin \mathcal{P}$) and (b) the model is identified at $\bar{\theta}$ in the sense that there does not exist $\theta \neq \bar{\theta}$ with $P_\theta = P_{\bar{\theta}}$.

Proposition 1 (Uniqueness of estimand)**.** *Under Assumption 1, there exists* $D < \infty$ *and* $g \in L^2(P_{\bar{\theta}}; \mathbb{R}^D)$ *such that* $\bar{\theta}$ *is the unique solution to the population moment-matching problem,* $\Theta_g = \{\bar{\theta}\}$.

Lemma 1 and Proposition 1 establish an "anything-goes" feature of moment matching under misspecification: under mild conditions, one can construct moments so that any pre-specified parameter value $\bar{\theta} \in \Theta$ is a unique solution to the population moment conditions, and thus the unique pseudo-true value.

Remark 1 (Absolute continuity)*.* Our characterization assumes that P is known and absolutely continuous with respect $P_{\bar{\theta}}$. If absolute continuity fails, manipulation is even easier. Specifically, if there exists a set A with $P(A) > 0$ and $P_{\bar{\theta}}(A) = 0$, an adversary can add functions such as $c \cdot \mathbb{I}\{X \in A\}$ to the moments, which shift $\mathrm{E}_P[g(X)]$ while leaving both $\mathrm{E}_{\bar{\theta}}[g(X)]$ and $\mathrm{Var}_{\bar{\theta}}(g(X))$ unchanged, and thus create "free" moment shifts absent under

absolute continuity. Similarly, in finite samples an adversary can often engineer any estimate they wish by choosing a sample-dependent g such that $\mathrm{E}_{\bar{\theta}}\left[g\left(X\right)\right] = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right)$. In particular, if the model implies the data are continuous then an adversarial researcher can add indicators for the observed data points to the moments without changing $\mathrm{E}_{\bar{\theta}}[g(X)]$ at all. Our focus on the population mapping $g \mapsto \Theta_g$ thus makes the manipulation problem harder than if we allowed moment selection based on the observed data.

4 Minimizing the Asymptotic Variance

The "anything-goes" result in Lemma 1 gives a stark warning about what an adversarial researcher can achieve through moment engineering. One might hope, however, that this result will be more troubling in theory than in practice. In particular, this could have limited practical bite because the moment functions needed to engineer a particular pseudo-true value are either socially unacceptable (e.g. too obviously contrived to be taken seriously), or alternatively deliver estimates with such extreme imprecision that they cannot hope to be persuasive. This section addresses the second possibility by characterizing the moments that minimize asymptotic variance subject to targeting a given pseudo-true value. In so doing, we find that, unfortunately, an adversarial researcher can often engineer not only their desired estimand but also model-implied standard errors close to those of maximum likelihood.

4.1 Asymptotic Variance for a Fixed Moment Function

Proposition 1 shows that for any $\bar{\theta}$, there exists a finite set of moments from $\mathcal{C}\left(\bar{\theta}\right)$ so that $\bar{\theta}$ is the unique population solution. Indeed, there are many such choices of moments, and different moments in this set imply different standard errors for the resulting estimates.

Specifically, under standard regularity conditions as in Newey and McFadden (1994), if we consider estimation based on moments g and weight matrix $W_n \to_p W$, under the model

$$\sqrt{n}\left(\hat{\theta}_{g,W} - \theta_g\right) \overset{P_\theta}{\rightsquigarrow} \mathcal{N}\left(0, \mathrm{AVar}\left(\hat{\theta}_{g,W}\right)\right), \tag{3}$$

where $\overset{P_\theta}{\rightsquigarrow}$ denotes convergence in distribution under P_θ, and

$$\mathrm{AVar}\left(\hat{\theta}_{g,W}\right) := (G^\top W G)^{-1} G^\top W \Omega W G (G^\top W G)^{-1} \tag{4}$$

for

$$G = \frac{\partial}{\partial \theta} E_\theta[g(X)] = E_\theta\left[g(X)s_\theta(X)^\top\right], \quad \Omega = \mathrm{Var}_\theta\left(g(X)\right).$$

Model-implied standard errors are then based on plug-in estimates for Ω and G. We focus on such implied standard errors for two reasons: first, it yields a tractable expression that depends only on the chosen moments g and the parametric model. Second, such standard errors are used in practice, and are implicit, for example, in moment matching papers that quantify uncertainty using the parametric bootstrap (e.g. Su and Judd 2012; Bourreau et al. 2021; Lagakos et al. 2023; Simonovska and Waugh 2025).

4.2 Efficient Moment Function for Specific Pseudo-True Value

We next ask how precise an estimator can be when attention is restricted to moments that target a given pseudo-true value. As a first step, we note that it suffices to consider just-identifying sets of moments, i.e. taking $D = \dim(g) = \dim(\theta) = K$. In particular, observe that the variance given in equation (4) is the same as that we would obtain from moments $\tilde{g}(X) := (G^\top W G)^{-1} G^\top W g(X)$, where if $\bar{\theta}$ solves the moments based on g, it also solves those based on $\tilde{g}$. Thus, to lower-bound the variance it suffices to consider the case with $D = K$. Since the weight matrix drops out in the just-identified case, we write the resulting estimate as $\hat{\theta}_g$.

With this simplification, consider the problem of minimizing the asymptotic variance:

$$g^*_{\bar{\theta}} \in \underset{g \in \mathcal{C}(\bar{\theta})}{\arg\min}\ \mathrm{AVar}\left(\hat{\theta}_g\right).$$

Thanks to our focus on the model-implied variance this problem admits a simple solution.

Proposition 2 (Efficient moment function within $\mathcal{C}\left(\bar{\theta}\right)$)**.** *If $\chi^2\left(P||P_{\bar{\theta}}\right) \in (0, \infty)$, then any asymptotically efficient, just-identifying moment function subject to $g \in \mathcal{C}\left(\bar{\theta}\right)$ takes the form*

$$g^*_{\bar{\theta}} := A s_{\bar{\theta},\perp} + c,$$

for some invertible matrix $A \in \mathbb{R}^{K \times K}$ and vector $c \in \mathbb{R}^K$, where

$$s_{\bar{\theta},\perp} := s_{\bar{\theta}} - \frac{\mathrm{E}_P\left[s_{\bar{\theta}}(X)\right]}{\chi^2\left(P||P_{\bar{\theta}}\right)}\left(\frac{dP}{dP_{\bar{\theta}}} - 1\right).$$

The corresponding minimum asymptotic variance is

$$\begin{aligned}\text{AVar}\left(\hat{\theta}_{g^*_{\bar{\theta}}}\right) &= \mathrm{E}_{\bar{\theta}}\left[s_{\bar{\theta},\perp}(X)\, s_{\bar{\theta},\perp}(X)^\top\right]^{-1} \\ &= \left\{\mathrm{E}_{\bar{\theta}}\left[s_{\bar{\theta}}(X)\, s_{\bar{\theta}}(X)^\top\right] - \frac{\mathrm{E}_P\left[s_{\bar{\theta}}(X)\right]\mathrm{E}_P\left[s_{\bar{\theta}}(X)\right]^\top}{\chi^2\left(P||P_{\bar{\theta}}\right)}\right\}^{-1}.\end{aligned}$$

The efficient moment is essentially the score $s_{\bar{\theta}}$ "purged" of the component aligned with the likelihood-ratio residual r. Specifically, the correction term subtracts the projection of $s_{\bar{\theta}}$ onto span $\{r\}$, which is exactly what is needed to enforce the constraint $g \in \mathcal{C}\left(\bar{\theta}\right)$ while remaining as close as possible, in an efficiency sense, to the score.

Proposition 2 characterizes the just-identified moments which minimize the model-implied asymptotic variance at a given $\bar{\theta}$. While these moments need not suffice to uniquely pin down $\bar{\theta}$, we next show that provided one uses the efficient weighting matrix, augmenting with additional moments does not increase the asymptotic variance.

Lemma 2 (Augmentation does not affect asymptotic variance)**.** *Fix $\bar{\theta} \in \Theta$ with $\chi^2\left(P||P_{\bar{\theta}}\right) \in (0,\infty)$ and let $g^*_{\bar{\theta}}$ be an efficient moment function from Proposition 2. Let $q : \mathcal{X} \to \mathbb{R}^J$ be a finite collection of moments in $\mathcal{C}\left(\bar{\theta}\right)$, and form the augmented moment vector $\tilde{g} := (g^{*\top}_{\bar{\theta}}, q^\top)^\top$. Under efficient weighting $W = \mathrm{Var}_{\bar{\theta}}(\tilde{g})^{-1}$, where we assume for simplicity $Var_{\bar{\theta}}(\tilde{g})$ has full rank,*

$$\text{AVar}\left(\hat{\theta}_{\tilde{g},W}\right) = \text{AVar}\left(\hat{\theta}_{g^*_{\bar{\theta}}}\right).$$

Lemma 2 shows that Proposition 2's restriction to just-identified moments is purely for analytic convenience: given an efficient set of moments, one can always augment it with moments as in Proposition 1 and weight efficiently to construct an estimator which both delivers unique estimand $\bar{\theta}$ and minimizes asymptotic variance.

4.3 Variance Cost of Estimand Targeting

We began this section by asking whether statistical considerations could mitigate the impact of moment engineering. In particular, is it the case that the variance inflation due to moment engineering substantially constrains the scope for an adversarial researcher to drive a given conclusion? We close by showing that this is, unfortunately, not the case.

Under $P_{\bar{\theta}}$, canonical results in statistics tell us that maximum likelihood estimation is asymptotically efficient relative to the class of asymptotically normal and unbiased estimators. Correspondingly, under $P_{\bar{\theta}}$, the score $s_{\bar{\theta}}$ delivers the smallest asymptotic variance

among all moments,

$$\text{AVar}\left(\hat{\theta}_g\right) \geq \text{AVar}\left(\hat{\theta}_{s_{\bar{\theta}}}\right)$$

for all moments g, where for square matrices A and B we say $A \geq B$ if and only if $A - B$ is positive semidefinite. We can thus use the model-implied variance of the MLE as a benchmark: if the engineered moments in Proposition 2 deliver a variance which is close to the MLE we know that there is little cost of moment engineering, in the sense that a researcher "organically" reaching the same estimate could not have meaningfully smaller model-implied standard errors.

To quantify the efficiency loss relative to the MLE, define the *variance inflation factor* $L\left(\bar{\theta}\right) \geq 1$ as the worst-case ratio of scalar asymptotic variances over all one-dimensional linear combinations of parameters,

$$L\left(\bar{\theta}\right) := \sup_{\alpha \neq 0} \frac{\alpha^\top \text{AVar}\left(\hat{\theta}_{g^*_{\bar{\theta}}}\right) \alpha}{\alpha^\top \text{AVar}\left(\hat{\theta}_{s_{\bar{\theta}}}\right) \alpha}.$$

The square root of this quantity, $\sqrt{L\left(\bar{\theta}\right)}$, is the largest factor by which any one-dimensional confidence interval (i.e. confidence interval for a scalar linear transformation $\alpha^\top \theta$) based on the engineered moments must widen relative to the MLE benchmark.

The inflation factor $L\left(\bar{\theta}\right)$ turns out to have a close connection to measures of statistical distinguishability. To state this connection, we use the variational characterization of the chi-squared divergence, which states that we can also express the root chi-squared divergence $\sqrt{\chi^2\left(P||P_{\bar{\theta}}\right)}$ as a "maximum population t-statistic" for distinguishing P and $P_{\bar{\theta}}$,

$$\sqrt{\chi^2\left(P||P_{\bar{\theta}}\right)} = \sup_{h \in L^2(P_{\bar{\theta}};\mathbb{R}), \text{Var}_{\bar{\theta}}(h(X))>0} \frac{\left|\text{E}_P\left[h\left(X\right)\right] - \text{E}_{\bar{\theta}}\left[h\left(X\right)\right]\right|}{\sqrt{\text{Var}_{\bar{\theta}}\left(h\left(X\right)\right)}}. \tag{5}$$

Intuitively, this measures the largest shift, measured in standard deviation units, in the mean of any scalar function of the data when the distribution changes from $P_{\bar{\theta}}$ to P. Hence, the chi-squared divergence will be small if and only if the two distributions are close in a variance-normalized sense.

This variational representation also suggests a class of generalized chi-squared divergences, where rather than searching over $L^2\left(P_{\bar{\theta}};\mathbb{R}\right)$ we consider more restricted function

classes. Formally, for any linear class $\mathcal{H} \subseteq L^2\left(P_{\bar{\theta}}; \mathbb{R}\right)$, define

$$\sqrt{\chi^2_{\mathcal{H}}\left(P||P_{\bar{\theta}}\right)} := \sup_{h \in \mathcal{H}, \mathrm{Var}_{\bar{\theta}}(h(X))>0} \frac{\left|\mathrm{E}_P\left[h\left(X\right)\right] - \mathrm{E}_{\bar{\theta}}\left[h\left(X\right)\right]\right|}{\sqrt{\mathrm{Var}_{\bar{\theta}}\left(h\left(X\right)\right)}}. \tag{6}$$

Like (5), (6) asks how large a mean shift, measured in standard deviation units, is induced by the change from $P_{\bar{\theta}}$ to P, now taking the upper bound over the restricted class of functions $\mathcal{H}$. Since we take the supremum over a smaller set, $\sqrt{\chi^2_{\mathcal{H}}\left(P||P_{\bar{\theta}}\right)} \leq \sqrt{\chi^2\left(P||P_{\bar{\theta}}\right)}$ by construction. The inflation factor $L(\bar{\theta})$ turns out to be precisely determined by the ratio of these two divergences for $\mathcal{H}$ the linear span of the score $s_{\bar{\theta}}$.

Proposition 3 (Discrepancy ratio)**.** *Suppose that* $0 < \chi^2(P\|P_{\bar{\theta}}) < \infty$*, and that*

$$I_{\bar{\theta}} := \mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta}}(X)s_{\bar{\theta}}(X)^\top], \quad Q_{\bar{\theta}} := \mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta},\perp}(X)s_{\bar{\theta},\perp}(X)^\top]$$

are positive definite. If $\mathcal{S}_{\bar{\theta}} := \left\{v^\top s_{\bar{\theta}} : v \in \mathbb{R}^K\right\}$*, then*

$$L(\bar{\theta}) = \left\{1 - \kappa(\bar{\theta})\right\}^{-1} \text{ for } \kappa(\bar{\theta}) := \frac{\chi^2_{\mathcal{S}_{\bar{\theta}}}(P\|P_{\bar{\theta}})}{\chi^2(P\|P_{\bar{\theta}})}.$$

The quantity $\kappa\left(\bar{\theta}\right) \in [0,1]$ measures the fraction of the total chi-squared separation between P and $P_{\bar{\theta}}$ that is captured by linear combinations of the score. This reflects the ease with which $P_{\bar{\theta}}$ and P can be distinguished using linear combinations of the score, relative to using any square integrable function. Proposition 3 thus implies that the variance penalty is small (so $\kappa\left(\bar{\theta}\right) \approx 0$ and hence $L\left(\bar{\theta}\right) \approx 1$) if and only if the score has little power to distinguish P from $P_{\bar{\theta}}$ (informally, the two distributions are close in the sense that the model "cares about" local to $\bar{\theta}$), even though some other functions clearly distinguish the two distributions.

5 Adversarial Construction in an Application

Our theoretical results in hand, we close by returning to the motivating example discussed in Section 2. In particular, we discuss how we apply our adversarial construction to this example, and provide further details on model-implied standard errors.

Since our theoretical results consider an adversary who knows the data distribution, our first step is to construct an estimate for P. This is complicated by the fact that the model-implied distribution of X has parameter-dependent support $[0, \sqrt{\bar{y}}]$. If we estimate

P with support extending beyond this range, the chi-squared divergence will be infinite and as discussed in Remark 1 our adversarial construction will mechanically imply very poor performance for moment matching.

To avoid this possibility, we consider an adversary who knows only the distribution of the censored absolute centered log price change

$$X_c = \begin{cases} X, & X \le c, \\ c^+, & X > c, \end{cases}$$

where we use the censoring value $c = 0.22$, which is somewhat below the smallest upper bound on the support, $\sqrt{0.06} \approx 0.245$, obtained over our grid of target parameters.[12] Under P_θ, the continuous part on $[0, c]$ is the original model density $p_\theta(x)$, while the point c^+ has mass $\pi_\theta(c) = P_\theta(X > c)$. We then construct an estimate $\hat{P}$ for the unknown law of X_c. Specifically, for the continuous component of $\hat{P}$ on $[0, c]$ we use a logspline density estimator and combine it with the plug-in estimate of the mass at c^+.[13]

We then directly implement our adversarial construction using the estimated distribution $\hat{P}$, constructing the target-specific adversarial moment $g_{\bar{\theta},\hat{P}}$ and computing estimates $\hat{\theta}$ which solve

$$\frac{1}{n}\sum_{i=1}^{n} g_{\bar{\theta},\hat{P}}(X_{c,i}) - \mathrm{E}_\theta[g_{\bar{\theta},\hat{P}}(X_c)] = 0.$$

Since we use the estimated distribution $\hat{P}$, the resulting estimates (plotted e.g. in Figures 3 and 4) use the data twice, first to form $\hat{P}$ and then to construct $\hat{\theta}$.

To reduce this re-use of the data, we also report cross-fit estimates. For this approach, reported in Figure 5, we split the sample into 10 disjoint folds, say $\{\mathcal{I}_1, \ldots, \mathcal{I}_{10}\}$ with $\cup_j \mathcal{I}_j = \{1, \ldots, n\}$, where for each fold we estimate $\hat{P}_{-j}$ on the data excluding fold j, evaluate $g_{\bar{\theta},\hat{P}_{-j}}$ on fold j, and solve the average moment equation

$$\frac{1}{10}\sum_{j=1}^{10}\left(\frac{1}{|\mathcal{I}_j|}\sum_{i\in\mathcal{I}_j} g_{\bar{\theta},\hat{P}_{-j}}(X_{c,i}) - \mathrm{E}_\theta[g_{\bar{\theta},\hat{P}_{-j}}(X_c)]\right) = 0.$$

[12] Censoring in this way preserves more information than truncating to price changes below the threshold, since it also preserves the tail probability $P_\theta(X > c)$.

[13] The logspline estimator models the log conditional density on $[0, c]$ as a cubic B-spline, as in Kooperberg and Stone (1991, 1992), using three evenly-space knots. For $B(x)'\gamma$ the log-density under coefficients γ, we choose γ to minimize $-\sum_{i:X_i\le c} B(X_i)'\gamma + n_c \log \int_0^c \exp\{B(u)'\gamma\}\, du$ for n_c the number of uncensored observations. The fitted conditional density is multiplied by the empirical mass $\frac{n_c}{n}$ below c, and paired with the empirical $\frac{n-n_c}{n}$ mass at c^+.

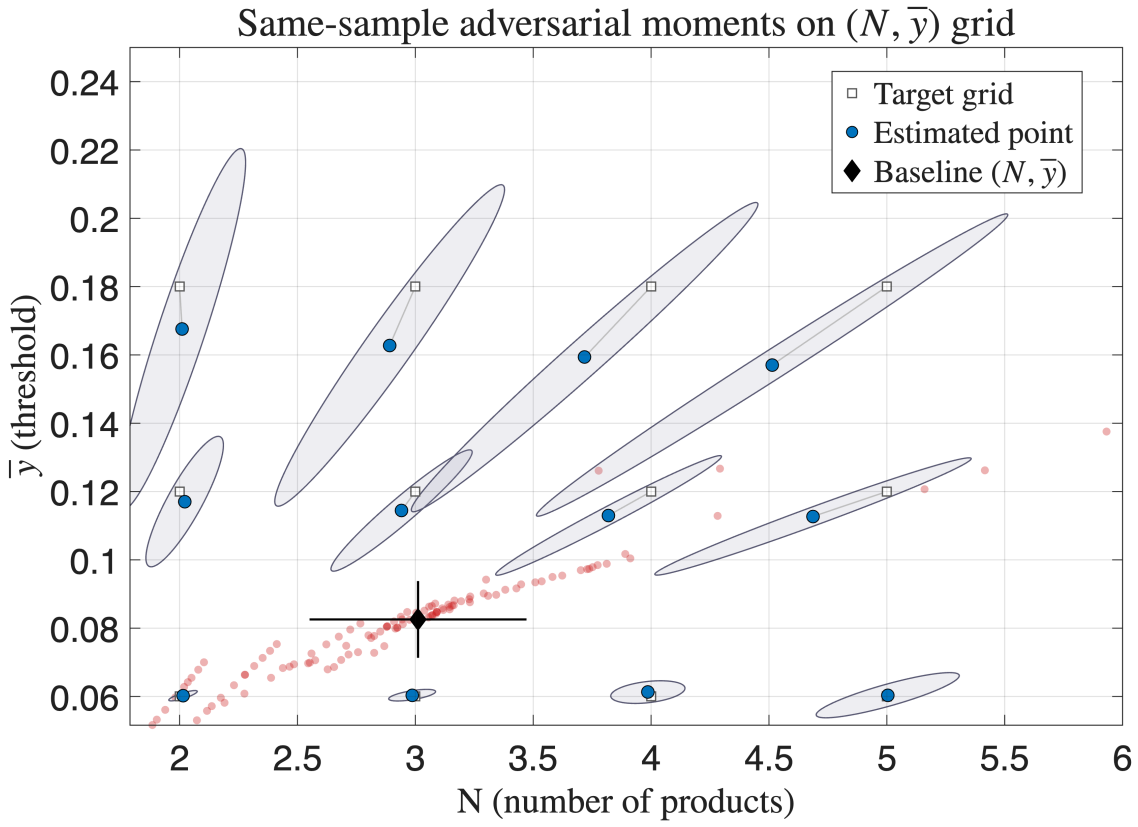

Figure 4: Same-sample adversarial estimates from Proposition 2 with model-implied uncertainty. Shaded regions are model-implied Wald confidence ellipses around each $\hat{\theta}$, using χ^2_1 critical values to ensure correct coverage for one-dimensional projections. The red scatter shows, for reference, the varying-moment estimates from Figure 2; the black diamond and bars are the baseline $(N, \bar{y})$ estimate and confidence intervals.

Cross-fitting reduces the dependence between the estimated data distribution and the sample moments, but comes at the cost of higher variability in the estimated data distribution due to the smaller sample size (though with 10-fold cross-fitting, this cost is limited).

In both Figures 4 and 5, we include ellipses around each estimate which reflect the confidence intervals for one-dimensional linear combinations $\alpha^\top \theta$, using model-implied standard errors. To read off the 95% confidence interval for a given parameter, e.g. N, for a given choice of moments, simply take the projection of the corresponding ellipse on the corresponding axis.[14] As these ellipses highlight, the deviation of the engineered estimates from their targets is on the same order as sampling variability.

[14]Formally, we plot Wald confidence ellipses constructed from the sandwich variance $\text{AVar}(\hat{\theta}_{g_{\tilde{\theta},\hat{P}}})$ and χ^2_1 critical values, where the use of χ^2_1 critical values ensures the projections are non-conservative but does not imply joint coverage of both parameters.

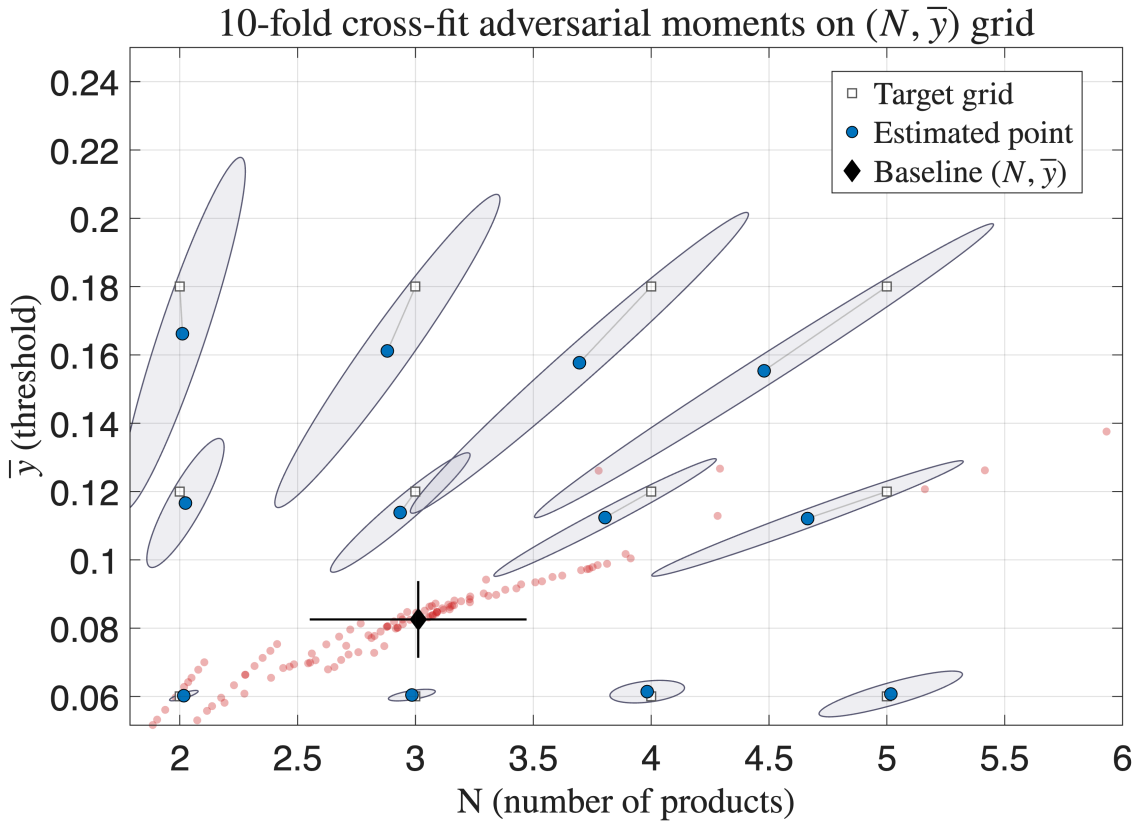

Figure 5: 10-fold cross-fit adversarial estimates from Proposition 2 with model-implied uncertainty. Shaded regions are model-implied Wald confidence ellipses around each $\hat{\theta}$, using χ^2_1 critical values to ensure correct coverage for one-dimensional projections. The red scatter shows, for reference, the varying-moment estimates from Figure 2; the black diamond and bars are the baseline $(N, \bar{y})$ estimate and confidence intervals.

6 What to Do with This?

Given concerns about model misspecification, it is natural for researchers to want discretion over which moments to use for estimation. As we highlight above, however, this flexibility also introduces substantial researcher degrees of freedom, which could be abused by an adversarial researcher. Even with the best of intentions, these researcher degrees of freedom may lead researchers working with the same data and model to reach very different conclusions. In this section we discuss some potential responses to this issue.

These concerns are likely to grow more pressing as AI-assisted coding tools make it easier to iterate over a wide range of specifications. Where searching over moment choices once required substantial time and expertise, much of this cost has already fallen, and is likely to continue falling. The economics profession will therefore need to grapple with researcher discretion in a setting where the practical friction against exploiting that discretion, whether

intentionally or not, is lower than it used to be. Thus, while some moment matching papers already abide by informal practices which reflect some of the points we discuss below, we think it is a good time to revisit these practices and ask if there is room for improvement.

Restrict the choice of estimators One response to researcher discretion is to try and remove it. This could take the form of explicit policies (e.g., journal requirements) or professional norms which limit the class of "acceptable" estimators researchers may report. For instance, if researchers estimating parametric models were required to report only the maximum likelihood estimator (MLE), the issues we explore above would not arise. However, researchers have good reasons to avoid relying exclusively on the MLE. In particular, the efficiency of the MLE under correct specification provides little comfort when the model is clearly misspecified, and the MLE estimand, while minimizing Kullback–Leibler divergence as shown by White (1982), is often not particularly interpretable in economic applications.

A more modest step in a similar direction would be to restrict the class of moments that are considered acceptable for estimation, while leaving some degree of discretion. However, this approach faces the same trade-off: if the restriction is very tight, it may inherit the same drawbacks as requiring the MLE or another specific estimator; if it is loose, it simply reintroduces researcher discretion in the choice of moments.

Explicitly motivating the choice of estimator A different approach would be to maintain the current practice of allowing researchers to choose their estimators on a case-by-case basis, but to require that they explicitly motivate why their choices, e.g. the specific set of moments and weighting matrix, are appropriate for their setting. This would amount to a formalization of some current practices: researchers often provide informal discussion of why at least some of their choices are well-suited to their estimation objectives, but the motivation is rarely spelled out formally.

If researchers instead articulated the specific forms of misspecification they are concerned about and why, this would in principle imply bounds on the performance of different estimators under consideration. One could then select the estimator, confidence interval, or other procedure with the "best" performance according to some criterion. Versions of this approach have been developed for particular misspecification settings by, for example, Armstrong and Kolesár (2021), Bonhomme and Weidner (2022), Christensen and Connault (2023), Adusumilli (2026), and Andrews et al. (2026), who derive optimal procedures under some types of misspecification concern. If widely adopted, this approach would shift the

focus from the choice of estimator to the misspecification concerns researchers wish to guard against, which may provide a more productive basis for discussion.

Committing to moments in advance In settings where there is a meaningful pre-data stage, researchers could commit to the moments used for estimation before accessing the data, analogous to pre-analysis plans in randomized controlled trials. Such commitments could reduce the scope for researchers to search over different moment sets in pursuit of a preferred result. A limitation of this approach, however, is that in many economic applications (for instance analyses using widely studied datasets like the PSID or US aggregate macro time-series) it may be difficult for researchers to credibly demonstrate that their choices were not informed by the data. In part, this reflects the fact that researchers' choices may reasonably be informed by previous work using the same dataset.

This idea is related to the common practice of comparing untargeted moments with their model-implied counterparts to assess overall model fit. One possibility would be for researchers to commit to the set of untargeted moments they will report before conducting the estimation. To the extent researchers can credibly guarantee that the "held out" moments are, in fact, held out, this could help ensure that model fit is evaluated using a predetermined set of empirical patterns rather than moments selected ex post. What econometric gains can be ensured by such an approach, given that all moments are computed based on the same underlying observations, is an interesting topic for future work.

All of the above The approaches discussed above are complementary in important respects. For example, if researchers were to commit to their moment choices in advance, it could be appealing to select those moments based on a formal analysis of misspecification concerns. Similarly, requiring researchers to use estimators that are optimal under an explicitly specified model of misspecification would effectively restrict the class of permissible estimators. In this way, combining these approaches may help reduce researcher degrees of freedom while still allowing moment selection to reflect the economic features that researchers consider most relevant.

References

Adusumilli, K. (2026): "You've Got to be Efficient: Ambiguity, Misspecification and Variational Preferences," *arXiv preprint arXiv:2604.05327*.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Alvarez, F., and F. Lippi (2014): "Price setting with menu cost for multiproduct firms," *Econometrica*, 82, 89–135.

Andrews, I., J. Chen, and O. Tecchio (2025): "The purpose of an estimator is what it does: Misspecification, estimands, and over-identification," *arXiv preprint arXiv:2508.13076*.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

——— (2020): "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 88, 2231–2258.

Andrews, I., R. Li, and Y. Shang (2026): "Misspecification-Averse Estimation."

Armstrong, T. B., and M. Kolesár (2021): "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 12, 77–108.

Benhabib, J., A. Bisin, and M. Luo (2019): "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, 109, 1623–1647.

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

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in macroeconomic expectations," *American Economic Review*, 110, 2748–2782.

Bourreau, M., Y. Sun, and F. Verboven (2021): "Market entry, fighting brands, and tacit collusion: Evidence from the French mobile telecommunications market," *American Economic Review*, 111, 3459–3499.

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

Cocci, M. D., and M. Plagborg-Møller (2024): "Standard errors for calibrated parameters," *Review of Economic Studies*, rdae099.

Cooley (1997): "Calibrated models," *Oxford Review of Economic Policy*, 13, 55–69.

Gallant, A. R., and G. Tauchen (1996): "Which moments to match?" *Econometric theory*, 12, 657–681.

Hansen, L. P., and J. J. Heckman (1996): "The empirical foundations of calibration," *Journal of economic perspectives*, 10, 87–104.

Kaji, T., E. Manresa, and G. Pouliot (2023): "An adversarial approach to structural estimation," *Econometrica*, 91, 2041–2063.

Kooperberg, C., and C. J. Stone (1991): "A study of logspline density estimation," *Computational Statistics & Data Analysis*, 12, 327–347.

——— (1992): "Logspline density estimation for censored data," *Journal of Computational and Graphical Statistics*, 1, 301–328.

Lagakos, D., A. M. Mobarak, and M. E. Waugh (2023): "The welfare effects of encouraging rural–urban migration," *Econometrica*, 91, 803–837.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Prescott, E. C. (1986): "Theory ahead of business-cycle measurement," in *Carnegie-Rochester conference series on public policy* Volume 25, 11–44, Elsevier.

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011): "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 22, 1359–1366, http://dx.doi.org/10.1177/0956797611417632rm 10.1177/0956797611417632.

Simonovska, I., and M. E. Waugh (2025): "Trade models, trade elasticities, and the gains from trade,"Technical report, National Bureau of Economic Research.

Su, C.-L., and K. L. Judd (2012): "Constrained optimization approaches to estimation of structural models," *Econometrica*, 80, 2213–2230.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the econometric society*, 1–25.

WOOLDRIDGE, J. M. (2001): "Applications of generalized method of moments estimation," *Journal of Economic perspectives*, 15, 87–100.

Online Appendix to

"The Incredible Flexibility of Moment Matching"

Isaiah Andrews and Bas Sanders

This supplemental appendix collects proofs of all results stated in the main paper. Throughout, notation, definitions, and numbered results refer to the main paper.

S1 Proof of Lemma 1

Define the likelihood ratios

$$L := \frac{dP}{dP_{\bar{\theta}}},\ r := L - 1.$$

Take any $g \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right)$. Since P is absolutely continuous with respect to $P_{\bar{\theta}}$, we have

$$\mathrm{E}_P\left[g\left(X\right)\right] = \mathrm{E}_{\bar{\theta}}\left[g\left(X\right)L\left(X\right)\right].$$

Therefore the equality

$$\mathrm{E}_{\bar{\theta}}\left[g\left(X\right)\right] = \mathrm{E}_P\left[g\left(X\right)\right]$$

holds if and only if

$$\mathrm{E}_{\bar{\theta}}\left[g\left(X\right)r\left(X\right)\right] = 0.$$

So we can rewrite the constraint set as

$$\mathcal{C}\left(\bar{\theta}\right) = \left\{g \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right) : \mathrm{E}_{\bar{\theta}}\left[g\left(X\right)r\left(X\right)\right] = 0\right\}.$$

Now, to show the displayed set is included in $\mathcal{C}\left(\bar{\theta}\right)$, fix any $f \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right)$ and define

$$g := f - \frac{\mathrm{E}_{\bar{\theta}}\left[f\left(X\right)r\left(X\right)\right]}{\mathrm{E}_{\bar{\theta}}\left[r\left(X\right)^2\right]}r.$$

Because $\mathrm{E}_{\bar{\theta}}\left[r\left(X\right)^2\right] = \chi^2\left(P||P_{\bar{\theta}}\right) \in (0, \infty)$, the denominator is finite and nonzero, so g is well-defined. Moreover, since $f \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right)$ and $r \in L^2\left(P_{\bar{\theta}}; \mathbb{R}\right)$, we have $g \in L^2\left(P_{\bar{\theta}}; \mathbb{R}^D\right)$. Then since $\mathrm{E}_{\bar{\theta}}\left[g\left(X\right)r\left(X\right)\right] = 0$, we have $g \in \mathcal{C}\left(\bar{\theta}\right)$.

To show the reverse inclusion, take any $g \in \mathcal{C}\left(\bar{\theta}\right)$. Then $\mathrm{E}_{\bar{\theta}}\left[g\left(X\right)r\left(X\right)\right] = 0$ by definition of $\mathcal{C}\left(\bar{\theta}\right)$. Set $f = g$. Then

$$g = f - \frac{\mathrm{E}_{\bar{\theta}}\left[f\left(X\right)r\left(X\right)\right]}{\mathrm{E}_{\bar{\theta}}\left[r\left(X\right)^2\right]}r.$$

So g belongs to the displayed set. □

S2 Proof of Proposition 1

We make use of the following lemma.

Lemma 3 (Collinearity and mixtures)**.** *Fix $\bar{\theta}$ and suppose $0 < \chi^2(P\|P_{\bar{\theta}}) < \infty$. Then*

$$\inf_{a\in\mathbb{R}} \mathrm{E}_{\bar{\theta}}\left[\left(r_\theta(X) - ar(X)\right)^2\right] = 0$$

if and only if one of $P_{\bar{\theta}}$, P_θ, and P is a possibly degenerate mixture of the other two.

Proof. Because $\mathrm{span}\{r\}$ is a finite-dimensional closed subspace of $L^2(P_{\bar{\theta}})$, the displayed infimum equals zero if and only if there exists $a \in \mathbb{R}$ such that $r_\theta = ar$ $P_{\bar{\theta}}$-almost surely. Equivalently,

$$\frac{dP_\theta}{dP_{\bar{\theta}}} = 1 + a\left(\frac{dP}{dP_{\bar{\theta}}} - 1\right) = (1-a) + a\frac{dP}{dP_{\bar{\theta}}}.$$

Integrating both sides over an arbitrary measurable set A gives

$$P_\theta(A) = (1-a)P_{\bar{\theta}}(A) + aP(A),$$

so $P_\theta = (1-a)P_{\bar{\theta}} + aP$. If $0 < a < 1$, then P_θ is a nontrivial mixture of $P_{\bar{\theta}}$ and P. If $a > 1$, then $P = \frac{1}{a}P_\theta + \frac{a-1}{a}P_{\bar{\theta}}$. If $a < 0$, then $P_{\bar{\theta}} = \frac{1}{1-a}P_\theta + \frac{-a}{1-a}P$. Finally, $a = 0$ gives $P_\theta = P_{\bar{\theta}}$, and $a = 1$ gives $P_\theta = P$.

Conversely, if one of the three distributions is a possibly degenerate mixture of the other two, elementary rearrangement yields $P_\theta = (1-a)P_{\bar{\theta}} + aP$ for some $a \in \mathbb{R}$. Taking Radon–Nikodym derivatives with respect to $P_{\bar{\theta}}$ gives $r_\theta = ar$, and hence the infimum is zero. □

That lemma in hand, note that, for $C_d(\bar{\theta})$ the analog of $\mathcal{C}(\bar{\theta})$ for $\mathbb{R}^d$-valued functions, $s_{\bar{\theta},\perp} = s_{\bar{\theta}} - \frac{E_{\bar{\theta}}[s_{\bar{\theta}}r]}{E_{\bar{\theta}}[r^2]}r \in \mathcal{C}_K(\bar{\theta})$. Indeed, $\mathrm{E}_{\bar{\theta}}[s_{\bar{\theta},\perp}(X)] = 0$ because $\mathrm{E}_{\bar{\theta}}[s_{\bar{\theta}}(X)] = 0$ and $\mathrm{E}_{\bar{\theta}}[r(X)] = 0$, while

$$\mathrm{E}_{\bar{\theta}}[s_{\bar{\theta},\perp}(X)r(X)] = 0$$

by construction. Hence $\mathrm{E}_P[s_{\bar{\theta},\perp}(X)] = \mathrm{E}_{\bar{\theta}}[s_{\bar{\theta},\perp}(X)L(X)] = \mathrm{E}_{\bar{\theta}}[s_{\bar{\theta},\perp}(X)]$.

Next, define $\Psi(\theta) := \mathrm{E}_{P_\theta}[s_{\bar\theta,\perp}(X)]$, and observe that $\Psi(\bar\theta) = 0$. By differentiating under the integral sign, $\dot\Psi(\bar\theta) = \mathrm{E}_{\bar\theta}\left[s_{\bar\theta,\perp}(X)s_{\bar\theta}(X)^\top\right]$. Since $s_{\bar\theta,\perp}$ is the residual from projecting $s_{\bar\theta}$ on span$\{r\}$, the last display equals $\mathrm{E}_{\bar\theta}\left[s_{\bar\theta,\perp}(X)s_{\bar\theta,\perp}(X)^\top\right] := Q_{\bar\theta}$. By Assumption 1(iv), $Q_{\bar\theta}$ is nonsingular. The inverse function theorem therefore implies that there exists $\varepsilon > 0$ such that

$$\Psi(\theta) = \Psi(\bar\theta) \quad \text{and} \quad \|\theta - \bar\theta\| < \varepsilon \quad \Longrightarrow \quad \theta = \bar\theta.$$

Moment matching on $s_{\bar\theta,\perp}$ admits no local solutions other than $\bar\theta$.

Now consider the nonlocal solution set based on moments $s_{\bar\theta,\perp}$

$$S := \left\{\theta \in \Theta : \|\theta - \bar\theta\| \geq \varepsilon,\ \mathrm{E}_\theta[s_{\bar\theta,\perp}(X)] = \mathrm{E}_{\bar\theta}[s_{\bar\theta,\perp}(X)]\right\}.$$

By the $L^2(P_{\bar\theta})$-continuity of $\theta \mapsto r_\theta$, the moment maps are continuous for every fixed square-integrable moment. Hence S is compact. Fix $\theta \in S$. Define $v_\theta := r_\theta - \frac{\mathrm{E}_{\bar\theta}[r_\theta(X)r(X)]}{\mathrm{E}_{\bar\theta}[r(X)^2]}\, r$. Then $v_\theta \in L^2(P_{\bar\theta})$ and $\mathrm{E}_{\bar\theta}[v_\theta(X)r(X)] = 0$, so $v_\theta \in \mathcal{C}_1(\bar\theta)$. Moreover,

$$\mathrm{E}_\theta[v_\theta(X)] - \mathrm{E}_{\bar\theta}[v_\theta(X)] = \mathrm{E}_{\bar\theta}[v_\theta(X)r_\theta(X)] = \mathrm{E}_{\bar\theta}[v_\theta(X)^2].$$

The last quantity is strictly positive by Assumption 1(v) and Lemma 3, since v_θ is the residual from projecting r_θ on span$\{r\}$. Therefore v_θ separates this θ from $\bar\theta$.

By continuity of the moment map associated with v_θ, this separation persists on an open neighborhood U_θ of θ. The collection $\{U_\theta : \theta \in S\}$ covers the compact set S, so there exist $\theta_1, \ldots, \theta_J \in S$ such that

$$S \subseteq \bigcup_{j=1}^{J} U_{\theta_j}.$$

Let $v_j := v_{\theta_j}$. Each $v_j \in \mathcal{C}_1(\bar\theta)$, so $\bar\theta$ satisfies all augmented moment conditions. Local alternatives are excluded by $s_{\bar\theta,\perp}$, and nonlocal alternatives are excluded by at least one of $v_1, \ldots, v_J$. Hence for g that collects these moment functions, $\{\theta \in \Theta : \mathrm{E}_\theta[g(X)] = \mathrm{E}_P[g(X)]\} = \{\bar\theta\}$, as desired. □

S3 Proof of Proposition 2

Again let $C_d(\bar\theta)$ denote the analog of $\mathcal{C}(\bar\theta)$ for $\mathbb{R}^d$-valued functions. Adding a constant to g does not affect the asymptotic variance, so we work with centered moments $h(X) :=$

$g(X) - \mathrm{E}_{P_{\bar{\theta}}}[g(X)]$. The constraint $g \in \mathcal{C}_K(\bar{\theta})$ is equivalent to

$$\mathrm{E}_{P_{\bar{\theta}}}[h(X)] = 0, \qquad \mathrm{E}_{P_{\bar{\theta}}}[h(X)r(X)] = 0.$$

Let $\mathcal{H}$ denote the closed linear subspace of such centered moments.

For $\mathbb{R}^K$-valued h, define

$$G_h := \mathrm{E}_{P_{\bar{\theta}}}\left[h(X)s_{\bar{\theta}}(X)^\top\right], \qquad \Omega_h := \mathrm{E}_{P_{\bar{\theta}}}\left[h(X)h(X)^\top\right].$$

The asymptotic variance is $G_h^{-1}\Omega_h(G_h^{-1})^\top$. Because this expression is invariant to nonsingular linear transformations of h, and the variance is infinite when G has reduced-rank, we may impose the normalization $\mathrm{E}_{P_{\bar{\theta}}}[h(X)s_{\bar{\theta}}(X)^\top] = I_K$. Since every $h \in \mathcal{H}$ is orthogonal to r, and $s_{\bar{\theta},\perp}$ is the projection of $s_{\bar{\theta}}$ onto $\mathcal{H}$,

$$\mathrm{E}_{P_{\bar{\theta}}}\left[h(X)s_{\bar{\theta}}(X)^\top\right] = \mathrm{E}_{P_{\bar{\theta}}}\left[h(X)s_{\bar{\theta},\perp}(X)^\top\right].$$

Let $Q_{\bar{\theta}} := \mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta},\perp}(X)s_{\bar{\theta},\perp}(X)^\top]$, and consider the normalized candidate $h^*(X) := Q_{\bar{\theta}}^{-1}s_{\bar{\theta},\perp}(X)$. This candidate is feasible and satisfies the normalization because

$$\mathrm{E}_{P_{\bar{\theta}}}[h^*(X)s_{\bar{\theta}}(X)^\top] = \mathrm{E}_{P_{\bar{\theta}}}[h^*(X)s_{\bar{\theta},\perp}(X)^\top] = Q_{\bar{\theta}}^{-1}Q_{\bar{\theta}} = I_K.$$

Now take any other normalized feasible h. Since $h \in \mathcal{H}$, $\mathrm{E}_{P_{\bar{\theta}}}[h(X)r(X)] = 0$. Using $s_{\bar{\theta},\perp} = s_{\bar{\theta}} - \frac{\mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta}}(X)r(X)]}{\mathrm{E}_{P_{\bar{\theta}}}[r(X)^2]}r$, we have

$$\mathrm{E}_{P_{\bar{\theta}}}[h(X)s_{\bar{\theta},\perp}(X)^\top] = \mathrm{E}_{P_{\bar{\theta}}}[h(X)s_{\bar{\theta}}(X)^\top] - \mathrm{E}_{P_{\bar{\theta}}}[h(X)r(X)]\frac{\mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta}}(X)r(X)]^\top}{\mathrm{E}_{P_{\bar{\theta}}}[r(X)^2]} = I_K.$$

Therefore, defining $u(X) := h(X) - Q_{\bar{\theta}}^{-1}s_{\bar{\theta},\perp}(X)$, we obtain

$$\begin{aligned}\mathrm{E}_{P_{\bar{\theta}}}[u(X)s_{\bar{\theta},\perp}(X)^\top] &= \mathrm{E}_{P_{\bar{\theta}}}[h(X)s_{\bar{\theta},\perp}(X)^\top] - \mathrm{E}_{P_{\bar{\theta}}}[Q_{\bar{\theta}}^{-1}s_{\bar{\theta},\perp}(X)s_{\bar{\theta},\perp}(X)^\top] \\ &= I_K - Q_{\bar{\theta}}^{-1}Q_{\bar{\theta}} \\ &= 0.\end{aligned}$$

Taking transposes also gives $\mathrm{E}_{P_{\bar{\theta}}}[s_{\bar{\theta},\perp}(X)u(X)^\top] = 0$.

Hence

$$\begin{aligned}\mathrm{E}_{P_{\bar\theta}}[h(X)h(X)^\top] &= \mathrm{E}_{P_{\bar\theta}}[\{Q_{\bar\theta}^{-1}s_{\bar\theta,\perp}(X)+u(X)\}\{Q_{\bar\theta}^{-1}s_{\bar\theta,\perp}(X)+u(X)\}^\top]\\ &= Q_{\bar\theta}^{-1}+\mathrm{E}_{P_{\bar\theta}}[u(X)u(X)^\top],\end{aligned}$$

because the cross terms vanish. Since $\mathrm{E}_{P_{\bar\theta}}[u(X)u(X)^\top] \succeq 0$, we have $\mathrm{E}_{P_{\bar\theta}}[h(X)h(X)^\top] \succeq Q_{\bar\theta}^{-1}$, where equality holds if and only if $u(X) = 0$ $P_{\bar\theta}$-almost surely. Thus $Q_{\bar\theta}^{-1}s_{\bar\theta,\perp}$ is the unique normalized minimizer. □

S4 Proof of Lemma 2

By the definition of the efficient weighting, we know that the asymptotic variance obtained by using these weights is weakly smaller (in the positive semidefinite sense) than the asymptotic variance obtained under any other weighting. However, the variance obtained in Proposition 2 corresponds to a weight matrix W which puts weight only on the first K elements of $\tilde{g}$. Thus, the variance from the efficient weighting must be weakly smaller than that obtained in Proposition 2 which, by efficiency of the Proposition 2 moments, proves the result. □

S5 Proof of Proposition 3

Fix $\bar\theta \in \Theta$, and denote $I\left(\bar\theta\right) := \mathrm{E}_{\bar\theta}\left[s_{\bar\theta}\left(X\right)s_{\bar\theta}\left(X\right)^\top\right]$ and $m\left(\bar\theta\right) := \mathrm{E}_P\left[s_{\bar\theta}\left(X\right)\right]$. From Proposition 2, we know

$$\mathrm{AVar}\left(\hat\theta_{g^*_{\bar\theta}}\right) = \left(I\left(\bar\theta\right) - \frac{m\left(\bar\theta\right)m\left(\bar\theta\right)^\top}{\chi^2\left(P||P_{\bar\theta}\right)}\right)^{-1},$$

while the MLE benchmark has $\mathrm{AVar}\left(\hat\theta_{s_{\bar\theta}}\right) = I\left(\bar\theta\right)^{-1}$. For any non-zero $\alpha \in \mathbb{R}^K$, define

$$L_\alpha\left(\bar\theta\right) := \frac{\alpha^\top \mathrm{AVar}\left(\hat\theta_{g^*_{\bar\theta}}\right)\alpha}{\alpha^\top \mathrm{AVar}\left(\hat\theta_{s_{\bar\theta}}\right)\alpha}.$$

Writing $y = I\left(\bar\theta\right)^{-1/2}\alpha$ and $w = I\left(\bar\theta\right)^{-1/2}\frac{m}{\sqrt{\chi^2(P||P_{\bar\theta})}}$, we have

$$L_\alpha\left(\bar\theta\right) = \frac{\alpha^\top\left\{I\left(\bar\theta\right) - \frac{m\left(\bar\theta\right)m\left(\bar\theta\right)^\top}{\chi^2(P||P_{\bar\theta})}\right\}^{-1}\alpha}{\alpha^\top I\left(\bar\theta\right)^{-1}\alpha} = \frac{y^\top\left(I_K - ww^\top\right)^{-1}y}{y^\top y}.$$

Since $I\left(\bar{\theta}\right) - \frac{m\left(\bar{\theta}\right)m\left(\bar{\theta}\right)^{\top}}{\chi^2(P||P_{\bar{\theta}})}$ is positive definite, $I_K - ww^{\top}$ is positive definite. So from the Sherman-Morrison formula we know

$$\left(I_K - ww^{\top}\right)^{-1} = I_K + \frac{ww^{\top}}{1 - w^{\top}w},$$

so we obtain

$$L_{\alpha}\left(\bar{\theta}\right) = 1 + \frac{\left(y^{\top}w\right)^2}{\left(1 - w^{\top}w\right)y^{\top}y} = 1 + \frac{\left(\alpha^{\top}I\left(\bar{\theta}\right)^{-1}m\left(\bar{\theta}\right)\right)^2}{\left(\chi^2\left(P||P_{\bar{\theta}}\right) - m\left(\bar{\theta}\right)^{\top}I\left(\bar{\theta}\right)^{-1}m\left(\bar{\theta}\right)\right)\alpha^{\top}I\left(\bar{\theta}\right)^{-1}\alpha}.$$

To find $L\left(\bar{\theta}\right)$, note that

$$L\left(\bar{\theta}\right) = \sup_{\alpha \neq 0} L_{\alpha}\left(\bar{\theta}\right) = \max_{y \neq 0} \frac{y^{\top}\left(I_K - ww^{\top}\right)^{-1}y}{y^{\top}y}.$$

But the last expression is the maximum of a Rayleigh quotient and it follows that

$$L\left(\bar{\theta}\right) = \lambda_{\max}\left\{\left(I_K - ww^{\top}\right)^{-1}\right\} = \frac{1}{1 - w^{\top}w} = \frac{1}{1 - \frac{m\left(\bar{\theta}\right)^{\top}I\left(\bar{\theta}\right)^{-1}m\left(\bar{\theta}\right)}{\chi^2(P||P_{\bar{\theta}})}}.$$

Since $\chi^2\left(P||P_{\bar{\theta}}\right) = \chi^2_{L^2}\left(P||P_{\bar{\theta}}\right)$, it now remains to show that

$$\chi^2_{\mathcal{S}_{\bar{\theta}}}\left(P||P_{\bar{\theta}}\right) = m\left(\bar{\theta}\right)^{\top}I\left(\bar{\theta}\right)^{-1}m\left(\bar{\theta}\right). \tag{S1}$$

Towards this, note that any $h \in \mathcal{S}_{\bar{\theta}}$ will look like $h\left(Z\right) = \beta^{\top}s_{\bar{\theta}}\left(Z\right)$ and that $\mathrm{E}_{\bar{\theta}}\left[s_{\bar{\theta}}\left(X\right)\right] = 0$. The left-hand side becomes

$$\sup_{\beta \neq 0} \frac{\left(\mathrm{E}_P\left[\beta^{\top}s_{\bar{\theta}}\left(X\right)\right]\right)^2}{\mathrm{Var}_{\bar{\theta}}\left(\beta^{\top}s_{\bar{\theta}}\left(X\right)\right)} = \sup_{\beta \neq 0} \frac{\left(\beta^{\top}m\left(\bar{\theta}\right)\right)^2}{\beta^{\top}I\left(\bar{\theta}\right)\beta}.$$

The maximum is achieved at $\beta \propto I\left(\bar{\theta}\right)^{-1}m\left(\bar{\theta}\right)$ and the conclusion follows. □

S6 Further Results for Alvarez and Lippi (2014) Application

Table 1 reports the finite-sample counterpart of the asymptotic-variance bound in Proposition 3 for the same-sample censored logspline implementation of Section 5. For each target $\bar{\theta} = (N, \bar{y})$ we report the proportional increase in the model-implied standard errors of the

adversarial estimator, relative to the MLE benchmark, for N and $\bar{y}$ separately,

$$\sqrt{L_N(\theta)} - 1 \;=\; \sqrt{\frac{\mathrm{AVar}_{\mathrm{adv}}(\theta)_{N,N}}{\mathrm{AVar}_{\mathrm{MLE}}(\theta)_{N,N}}} - 1, \qquad \sqrt{L_{\bar{y}}(\theta)} - 1 \;=\; \sqrt{\frac{\mathrm{AVar}_{\mathrm{adv}}(\theta)_{\bar{y},\bar{y}}}{\mathrm{AVar}_{\mathrm{MLE}}(\theta)_{\bar{y},\bar{y}}}} - 1,$$

with both asymptotic-variance matrices evaluated under P_θ. The left panel of the table evaluates these at the target $\theta = \bar{\theta}$, and the right panel at the same-sample adversarial estimate $\theta = \hat{\theta}_{g_{\bar{\theta},\hat{P}}}$. Each entry is at least zero by Proposition 3; a value close to zero indicates that the adversarial moment achieves close to the MLE standard error at that point.

		At target $\bar{\theta}$		At estimate $\hat{\theta}_{g_{\bar{\theta},\hat{P}}}$	
N	$\bar{y}$	$\sqrt{L_N} - 1$	$\sqrt{L_{\bar{y}}} - 1$	$\sqrt{L_N} - 1$	$\sqrt{L_{\bar{y}}} - 1$
2	0.06	24.8%	5.3%	23.1%	4.8%
2	0.12	2.6%	3.3%	2.5%	3.4%
2	0.18	0.1%	5.0%	0.2%	4.9%
3	0.06	4.9%	12.8%	4.8%	10.5%
3	0.12	5.7%	15.4%	4.8%	14.4%
3	0.18	5.7%	15.2%	4.5%	14.4%
4	0.06	0.5%	100.6%	1.9%	75.5%
4	0.12	20.0%	22.8%	18.0%	20.8%
4	0.18	17.0%	24.7%	14.5%	22.9%
5	0.06	41.6%	218.7%	39.0%	214.2%
5	0.12	27.9%	21.8%	26.0%	20.0%
5	0.18	26.8%	31.0%	23.8%	28.4%
Median		11.3%	18.6%	9.7%	17.2%

Table 1: Proportional increase $\sqrt{L_N} - 1$ and $\sqrt{L_{\bar{y}}} - 1$ in the model-implied standard error of the adversarial estimator relative to the MLE, evaluated at the target $\theta = \bar{\theta}$ (left panel) or at the same-sample adversarial estimate $\theta = \hat{\theta}_{g_{\bar{\theta},\hat{P}}}$ (right panel), with both asymptotic-variance matrices evaluated under P_θ.