Sensitivity, Informativeness, and Misspecification in GMM Estimation

Seojeong Lee[*] Fangzhou Yu[†]

April 16, 2026

Abstract

This paper develops a sensitivity and informativeness framework for GMM estimators that remains valid under general misspecification of moment conditions. Sensitivity is defined through the conditional expectation of the estimator given the moments and is characterized using the influence functions of the estimator and the moments. Under misspecification, sensitivity must be evaluated at the estimator's pseudo-true probability limit, and additional sources of variation arising from the Jacobian and from estimated weight matrices become first-order relevant. We derive misspecification-robust sensitivity measures for one-step, two-step and iterated GMM estimators, and introduce an informativeness measure that quantifies the fraction of an estimator's asymptotic variance explained by sampling variation in the moments themselves. This measure provides a notion of structural efficiency under misspecification that complements standard specification tests. Applications to the automobile demand model of Berry et al. (1995) and the consumption insurance model of Blundell et al. (2008) illustrate that accounting for misspecification can substantially alter sensitivity rankings and reveal large losses in informativeness.

[*]Department of Economics, Seoul National University. s.jay.lee@snu.ac.kr

[†]School of Economics, University of Sydney. yfz.1017@gmail.com

1 Introduction

Empirical researchers increasingly rely on sensitivity diagnostics to assess how GMM estimates depend on particular moment conditions. In many applications, however, overidentifying restrictions are strongly rejected, and the maintained moment conditions fail to hold exactly at any parameter value. In such settings, the GMM estimator does not converge to a structurally "true" parameter satisfying the moment restrictions. Instead, it converges to a pseudo-true value defined as the minimizer of a misspecified population objective function.

Under misspecification in this sense, local sensitivity can no longer be interpreted as a perturbation around a correctly specified benchmark. The relevant object is instead the mapping from the moment conditions to the estimator's misspecified probability limit. This paper develops sensitivity measures defined relative to that pseudo-true population solution and shows that doing so yields diagnostics with different economic content and interpretation, even when the perturbations themselves are local.

Assessing sensitivity to model specification is central to evaluating the robustness of empirical conclusions. In reduced-form settings, researchers have access to a rich and well-developed toolkit. For example, the omitted variable bias formula in linear regression provides a transparent benchmark for assessing the impact of confounding variables, and a large literature has developed formal sensitivity diagnostics for a wide range of reduced-form estimators (Rosenbaum and Rubin, 1983; Imbens, 2003; Oster, 2019; Cinelli and Hazlett, 2020). By contrast, comparable tools for structural estimators are far less standardized, despite the central role such estimators play in empirical work. This reflects the implicit and often nonlinear relationship between parameters and identifying moment conditions in structural models, as well as the prevalence of misspecification in applied work.

A growing literature has begun to formalize sensitivity analysis for structural estimators using large-sample approximations. Gentzkow and Shapiro (2014) introduce the concepts of asymptotic sensitivity and asymptotic sufficiency, which characterize how parameter estimates covary with auxiliary sample statistics. Building on these ideas, Andrews et al. (2017) formalize sensitivity for GMM estimators through local perturbations of the moment conditions, showing that such perturbations are mapped linearly into asymptotic bias through a closed-form sensitivity matrix. These contributions provide powerful and computationally simple diagnostics and have become influential in applied structural analysis.

All of these approaches, however, implicitly evaluate sensitivity relative to a benchmark at which the moment conditions are correctly specified, meaning that there exists a parameter value at which the population moments are exactly satisfied. When this condition fails, the model is misspecified, and the GMM estimator converges instead to the population minimizer of the GMM criterion, commonly referred to as the pseudo-true parameter. Because pseudo-true values are defined through an approximation criterion rather than through the moment conditions themselves, they do not generally have a direct economic interpretation beyond representing the best-fitting parameter within the maintained model. Regardless of their economic interpretation, pseudo-true parameters are the natural reference point for sensitivity analysis, as they characterize the probability limit of the estimator under the maintained estimation procedure. Furthermore, there are empirically important settings in which pseudo-true parameters remain economically meaningful objects. A canonical example arises in overidentified instrumental variables models with heterogeneous treatment effects. When different instruments identify different local average treatment effects, no single parameter value can satisfy all moment conditions simultaneously, and the two-stage least squares estimator converges to a weighted average of the underlying local average treatment effects (Lee, 2018). Inference and interpretation in these models therefore require tools that remain valid under misspecification and that describe the behavior of the estimator around its pseudo-true limit.

This paper extends local sensitivity analysis to explicitly allow for misspecified moment conditions. We propose a misspecification-robust sensitivity (MRS) measure that characterizes how local deviations in the moment conditions affect the first-order behavior of GMM estimators around their pseudo-true probability limits when the model is globally misspecified. To construct MRS, we derive influence function representations for a broad class of estimators, including one-step, two-step and iterated GMM, explicitly accounting for misspecification and for variation arising from estimated weight matrices. This analysis reveals that, relative to existing sensitivity measures, MRS contains additional terms that capture the effects of misspecification operating through the Jacobian of the moment conditions and the estimated weight matrix. These terms are closely related to components of the misspecification-robust asymptotic variance (e.g., Imbens, 1997; Hall and Inoue, 2003; Windmeijer, 2005; Hansen and Lee, 2021; Hwang et al., 2022).

This connection has an important practical implication. Recent work argues that applied researchers should default to misspecification-robust standard errors when conducting inference for GMM estimators (Andrews et al., 2025). Our results show that, once these robust variance components are computed, misspecification-robust sensitivity measures can be obtained at essentially no additional computational cost. When the moment conditions are correctly specified, the additional terms vanish, and MRS reduces to existing sensitivity measures.

Our results also shed light on the informativeness of moment conditions under misspecification.

In the context of an estimator's sensitivity to its own moments, Gentzkow and Shapiro (2014) show that sufficiency, referred to as informativeness in Andrews et al. (2020), is equal to one under correct specification. This equivalence breaks down under misspecification. By analyzing the variance of the influence function, we reinterpret informativeness as a measure of *structural efficiency*, defined as the proportion of total asymptotic variance attributable to sampling variation in the moment conditions themselves. Unlike classical GMM efficiency, which concerns the choice of optimal weights to minimize variance, structural efficiency captures additional variance arising from stochastic variation in the Jacobian and the estimated weight matrix that is not attributable to sampling variation in the moment conditions. We show that this structural efficiency loss provides a diagnostic of misspecification that complements traditional specification tests such as Hansen's J-test (Hansen, 1982). While the J-test assesses whether moment conditions are jointly violated, structural efficiency quantifies the practical implications of such violations for the precision and stability of the estimator.

Finally, our paper is related to several recent developments in local sensitivity analysis. Jørgensen (2023) adapts derivative-based sensitivity measures to study the dependence of GMM estimators on calibrated parameters. Bonhomme and Weidner (2022) develop estimators designed to minimize the impact of model misspecification on key quantities of interest. Christensen and Connault (2023) analyze sensitivity of counterfactual conclusions to assumptions on unobserved heterogeneity, while Armstrong and Kolesár (2021) study sensitivity to moment selection and propose optimal weighting schemes. Our contribution is complementary to this literature in focusing on sensitivity to moment conditions evaluated at the estimator's misspecified probability limit.

The remainder of the paper is organized as follows. Section 2 defines sensitivity in terms of the conditional expectation of the estimator given the moments and establishes its relationship with the influence function of GMM estimators. Section 3.1 derives influence function representations for one step, two step, and iterated GMM, explicitly allowing for misspecification. Section 4 discusses the proposed misspecification robust sensitivity measure and informativeness through illustrative examples. We consider applications based on the Berry et al. (1995) model of automobile demand and the Blundell et al. (2008) model of household consumption insurance to illustrate how sensitivity analysis can be used to assess the implications of violations of moment conditions in potentially misspecified GMM models.

2 Framework

2.1 Setup and Definitions

Consider i.i.d. random vectors $X_1, \ldots, X_n$ with unknown distribution P_0. Let $g(X_i, \theta)$ be a $q \times 1$ vector of moment functions, and let $\theta \in \Theta \subset \mathbb{R}^p$ be a $p \times 1$ parameter vector with $q > p$. A GMM estimator is defined as

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \widehat{g}(\theta)' W \widehat{g}(\theta),$$

where $\widehat{g}(\theta) = n^{-1} \sum_{i=1}^{n} g(X_i, \theta)$ is the sample moment vector and W is a positive semidefinite weight matrix. For expositional clarity, the weight matrix W is treated as deterministic in this section; stochastic weight matrices are considered in Section 3.1. The GMM model is said to be correctly specified if there exists $\theta \in \Theta$ such that $\mathbb{E}[g(X_i, \theta)] = 0$; otherwise the model is misspecified.

We denote by θ_0 the probability limit of the GMM estimator $\widehat{\theta}$. Under standard regularity conditions, θ_0 is the unique minimizer of the population GMM criterion

$$Q(\theta) = \mathbb{E}[g(X_i, \theta)]' W \mathbb{E}[g(X_i, \theta)].$$

When the model is correctly specified, this minimizer satisfies $\mathbb{E}[g(X_i, \theta_0)] = 0$ and therefore coincides with the true parameter. When the model is misspecified, θ_0 represents the pseudo-true parameter that best fits the moment conditions in the GMM objective. Let $g(\theta) = E[g(X_i, \theta)]$ and $g_0 = g(\theta_0)$.

Suppose the standard regularity conditions hold, e.g., Theorem 1 of Hall and Inoue (2003). Then,

$$\sqrt{n} \begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{g}(\theta_0) - g_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \widetilde{\theta} \\ \widetilde{g} \end{pmatrix} \sim N(0, \Sigma), \text{ with } \Sigma = \begin{pmatrix} \sigma_{\theta\theta} & \sigma_{\theta g} \\ \sigma_{\theta g} & \sigma_{gg} \end{pmatrix}. \tag{1}$$

By properties of the multivariate normal distribution, the conditional expectation of $\widetilde{\theta}$ given $\widetilde{g}$ is linear and given by

$$\mathbb{E}[\widetilde{\theta} \mid \widetilde{g}] = \sigma_{\theta g} \sigma_{gg}^{-1} \widetilde{g}. \tag{2}$$

We interpret the corresponding population regression coefficient as a measure of the sensitivity of the estimator to the moments.

Definition 1 (Misspecification-Robust Sensitivity)**.** *The misspecification-robust sensitivity (MRS)* *of the GMM estimator $\widehat{\theta}$ with respect to the moment vector $\widehat{g}(\theta_0)$ is defined as*

$$\Lambda = \sigma_{\theta g}\sigma_{gg}^{-1}.$$

The sensitivity matrix Λ is $p \times q$. Its (k, l) element measures how a local deviation in the lth moment affects the first-order behavior of the kth component of the estimator around its probability limit, holding the other moments fixed. The linear relationship in equation (2) also appears in Andrews et al. (2017), who study sensitivity under correct specification and analyze the response of $\widehat{\theta}$ to local perturbations of the data distribution. Our framework extends their sensitivity measure by allowing for misspecified moment conditions. In Example 1, we show that MRS nests the sensitivity measure of Andrews et al. (2017) as a special case when the model is correctly specified.

Sensitivity captures how the estimator responds locally to deviations in the realized moments. A complementary concept concerns the extent to which the moments are informative for the estimator. We adapt the notion of informativeness from Andrews et al. (2020) to the GMM setting under misspecification.

Definition 2 (Informativeness of Moments)**.** *The informativeness of the moment vector for the kth component of the GMM estimator, $\widehat{\theta}_k$, is defined as*

$$\Delta_k = \frac{\sigma_{\theta_k g}\sigma_{gg}^{-1}\sigma_{g\theta_k}}{\sigma_{\theta_k\theta_k}},$$

where $\sigma_{\theta_k\theta_k}$ denotes the kth diagonal element of $\sigma_{\theta\theta}$ and $\sigma_{\theta_k g}$ is the kth row of $\sigma_{\theta g}$.

The kth row of Λ and the scalar $\Delta_k \in [0, 1]$ correspond to the population regression coefficients and the R^2 from the linear projection of $\widetilde{\theta}_k$ on $\widetilde{g}$, respectively. Under correct specification, the asymptotic variance of the GMM estimator is fully explained by sampling variation in its estimation moments, implying $\Delta_k = 1$. Under misspecification, Δ_k is generally less than one. As shown in Section 3.2, this occurs because the estimator is influenced by additional sources of asymptotic variation arising from the estimation of the Jacobian, the curvature of the GMM criterion, and the weight matrix. We therefore interpret Δ_k as a measure of *structural efficiency*. Lower values of Δ_k indicate that a larger share of the estimator's asymptotic variance is driven by features of the GMM structure rather than by sampling variation in the moment conditions themselves.

2.2 Influence Function Representations of Sensitivity and Informativeness

To implement the proposed sensitivity and informativeness measures, we require estimates of the asymptotic covariance matrix Σ. We use influence function representations for this purpose. For an asymptotically linear estimator such as the GMM estimator $\widehat{\theta}$, an influence function $\psi(x)$ satisfies

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1), \tag{3}$$
$$\mathbb{E}[\psi(X_i)] = 0, \qquad \mathbb{E}[\psi(X_i)\psi(X_i)'] < \infty.$$

Let ψ and ν denote influence functions associated with the GMM estimator $\widehat{\theta}$ and the sample moment vector $\widehat{g}(\theta_0)$, respectively. Then

$$\sqrt{n} \begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{g}(\theta_0) - g_0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \psi(X_i) \\ \nu(X_i) \end{pmatrix} + o_p(1). \tag{4}$$

As a consequence, the misspecification-robust sensitivity defined in Definition 1 can be written as

$$\Lambda = \mathbb{E}[\psi\nu']\mathbb{E}[\nu\nu']^{-1}. \tag{5}$$

Similarly, the informativeness of the moments for the kth parameter, Δ_k in Definition 2, is given by

$$\Delta_k = \frac{\mathbb{E}[\psi_k\nu']\mathbb{E}[\nu\nu']^{-1}\mathbb{E}[\nu\psi_k]}{\mathbb{E}[\psi_k^2]}, \tag{6}$$

where ψ_k is the kth component of ψ.

A closely related expression to equation (5) appears in Iskrev (2019), who studies the sensitivity of structural parameters to calibrated parameters. In that setting, the sensitivity measure can be interpreted as arising from treating the calibrated parameter analogously to an estimated moment. Jørgensen (2023) discusses limitations of this approach and emphasizes that the key conceptual distinction lies in the treatment of the calibrated parameter. While Iskrev (2019) effectively treats the calibrated parameter as a stochastic object estimated from the data, Jørgensen (2023) treats it as fixed and non-stochastic.

Equations (5) and (6) are expressed in terms of population moments and influence functions. In practice, the sensitivity matrix Λ and the informativeness measure Δ_k can be consistently estimated using a plug-in approach based on estimated influence functions. Specifically, we construct estimated influence functions, denoted $\widehat{\psi}_i$ and $\widehat{\nu}_i$, for each observation $i = 1, \ldots, n$, by replacing unknown population quantities with their sample counterparts evaluated at the GMM estimate $\widehat{\theta}$.

First, the estimated influence function for the moments, $\widehat{\nu}_i$, is obtained by centering the moment function evaluated at the GMM estimate,

$$\widehat{\nu}_i = g(X_i, \widehat{\theta}) - \widehat{g}(\widehat{\theta}).$$

Second, the estimated influence function for the GMM estimator, $\widehat{\psi}_i$, is obtained by substituting sample estimators into the analytic expressions for the influence function $\psi(X_i)$ derived in Proposition 1 below. The population influence function depends on the Jacobian of the moment conditions, the weight matrix, and additional terms involving derivatives of the GMM criterion. We replace these population quantities with their sample counterparts evaluated at $\widehat{\theta}$ and compute

$$\widehat{\psi}_i = \psi\left(X_i; \widehat{\theta}, \widehat{G}, \widehat{W}\right).$$

With these estimated influence functions in hand, the sensitivity matrix $\widehat{\Lambda}$ can be computed using a sample analogue of equation (5). Specifically, $\widehat{\Lambda}$ is obtained by regressing $\widehat{\psi}_i$ on $\widehat{\nu}_i$, which yields

$$\widehat{\Lambda} = \left(\sum_{i=1}^{n} \widehat{\psi}_i \widehat{\nu}_i'\right) \left(\sum_{i=1}^{n} \widehat{\nu}_i \widehat{\nu}_i'\right)^{-1}.$$

Similarly, the informativeness measure $\widehat{\Delta}_k$ can be estimated using a sample analogue of equation (6), based on the sample variances and covariances of $\widehat{\psi}_{ik}$ and $\widehat{\nu}_i$. This implementation requires only the analytic form of the influence function ψ, which is derived in the next section for a range of GMM estimators.

3 Sensitivity Measures for GMM Estimators

3.1 Influence Functions for GMM Estimators

Equation (5) establishes that the misspecification-robust sensitivity, Λ, and informativeness, Δ, can be computed directly from the influence functions of the estimator and the moments. The influence function for the moment vector is immediate:

$$\nu(X_i) = g(X_i, \theta_0) - g_0.$$

Deriving influence functions for GMM estimators requires more care. As shown in Hall and Inoue (2003) and subsequent work, the asymptotic behavior of GMM under misspecification depends critically on both the choice of the weight matrix and the estimation procedure.

We consider a class of GMM estimators defined as minimizers of

$$\widehat{Q}_n(\theta) = \widehat{g}(\theta)'\widehat{W}\widehat{g}(\theta),$$

with different choices of the weight matrix.

In one-step GMM, the weight matrix is fixed across iterations. We consider both the case of a deterministic weight matrix, where $\widehat{W} = W$ is non-stochastic, and the case of an estimated weight matrix satisfying $\widehat{W} \xrightarrow{p} W$, where $\widehat{W}$ is asymptotically linear and does not depend on θ.

In two-step efficient GMM, the weight matrix is evaluated at a preliminary estimator $\widehat{\phi}$. Specifically, the second-step estimator minimizes the criterion using the efficient weight matrix

$$\widehat{W}(\widehat{\phi}) = \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i, \widehat{\phi})g(X_i, \widehat{\phi})'\right)^{-1},$$

where $\widehat{\phi}$ is a consistent first-step estimator.

Iterated GMM updates the weight matrix repeatedly by re-evaluating it at the current parameter estimate until convergence. At convergence, the estimator minimizes the criterion with weight matrix $\widehat{W}(\widehat{\theta})$.

We now introduce notation used in the influence function derivations. Let $G(X_i, \theta) = \frac{\partial g(X_i,\theta)}{\partial \theta'}$. Define $g = \mathbb{E}[g(X_i, \theta_0)]$ and $G = \mathbb{E}[G(X_i, \theta_0)]$. Let $\widehat{W}(\theta)$ be a parameter-dependent weight matrix estimator, and its probability limit be $W(\theta)$. Let $W = W(\theta_0)$.

The influence function derivations rely on the population curvature of the GMM objective. Let

$$R = \mathbb{E}\left[\frac{\partial}{\partial \theta'}\mathrm{vec}(G(X_i, \theta_0)')\right], \qquad H = (g'W \otimes I_p)R.$$

H appears in the influence functions of one-step GMM. Let

$$S = \frac{\partial}{\partial \theta'}\mathrm{vec}(W(\theta))\bigg|_{\theta_0}, \qquad J = (g' \otimes G')S.$$

In addition to H, J appears in the influence function of two-step and iterated GMM.

The matrices H and J depend on the misspecification vector g and vanish under correct specification. The population curvature matrices are defined as

$$A = G'WG + H. \tag{7}$$

We assume the following regularity conditions.

Assumption 1. *(i) The observations $\{X_i\}_{i=1}^n$ are i.i.d.*

(ii) The parameter space $\Theta \subset \mathbb{R}^p$ is compact. For one-step, two-step, and iterated GMM, the population objective $Q(\theta)$ has a unique minimizer θ_0 in the interior of Θ.

(iii) The moment function $g(X, \theta)$ is three times continuously differentiable in $\theta \in \Theta$ almost surely and satisfies the uniform integrability conditions

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \|g(X,\theta)\|^2\right] < \infty, \qquad \mathbb{E}\left[\sup_{\theta\in\Theta} \|G(X,\theta)\|^2\right] < \infty,$$

and for all $j_1, j_2 \in \{1, \ldots, p\}$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left\|\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}} g(X,\theta)\right\|^2\right] < \infty.$$

(iv) The population curvature matrices A, $A + J$ are nonsingular.

(v) For iterated GMM, the population updating map is a contraction at θ_0.

(vi) Weight matrix conditions:

(a) (Estimated weight, one-step GMM) The weight matrix $\widehat{W}$ does not depend on θ, $\widehat{W} \xrightarrow{p} W$ where W is symmetric positive definite, and admits the asymptotically linear representation:

$$\sqrt{n}\,\mathrm{vec}(\widehat{W} - W) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi^W(X_i) + o_p(1), \quad \textit{with } \mathbb{E}[\|\psi^W(X_i)\|^2] < \infty.$$

(b) (Parameter-dependent weight) Let $\widehat{\phi}$ be a preliminary estimator converging in probability to ϕ_0 that admits the asymptotically linear representation as in Equation (3). There exists a neighborhood $\mathcal{N}$ of ϕ_0 such that $W(\phi)$ is continuously differentiable and $W(\phi_0)$

is positive definite. Furthermore, with probability approaching one, the sample weight matrix $\widehat{W}(\phi)$ is continuously differentiable on $\mathcal{N}$ and satisfies

$$\sup_{\phi \in \mathcal{N}} \|\widehat{W}(\phi) - W(\phi)\| \xrightarrow{p} 0, \qquad \sup_{\phi \in \mathcal{N}} \left\| \frac{\partial \operatorname{vec}(\widehat{W}(\phi))}{\partial \phi'} - \frac{\partial \operatorname{vec}(W(\phi))}{\partial \phi'} \right\| \xrightarrow{p} 0.$$

Finally, the sample weight matrix admits the asymptotically linear representation:

$$\sqrt{n} \operatorname{vec}(\widehat{W}(\phi_0) - W(\phi_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi^W(X_i) + o_p(1), \quad \text{with } \mathbb{E}[\|\psi^W(X_i)\|^2] < \infty.$$

Before deriving the influence functions, we formally establish the consistency of the GMM estimators. The consistency of one-step, two-step, and iterated GMM under misspecification follows established results in the literature (e.g., Hall and Inoue, 2003; Hansen, 2022; Hwang et al., 2022).

Proposition 1 (Influence Functions for GMM under Misspecification)**.** *Suppose Assumption 1 holds. The influence functions of GMM estimators under misspecification are given as follows. Define the components of GMM first-order conditions as*

$$\begin{aligned} f_1(X_i) &= G'Wg(X_i, \theta_0) + G(X_i, \theta_0)'Wg, \\ f_2(X_i) &= f_1(X_i) + (g' \otimes G')\psi^W(X_i). \end{aligned}$$

(i) One-step GMM with deterministic weight W:

$$\psi^{1s}(X_i) = -A^{-1} f_1(X_i). \tag{8}$$

(ii) One-step GMM with estimated weight $\widehat{W}$: If $\widehat{W}$ is asymptotically linear and does not depend on θ,

$$\psi^{1s,W}(X_i) = -A^{-1} f_2(X_i). \tag{9}$$

(iii) Two-step efficient GMM: Let $\psi^{\phi}(X_i)$ denote the influence function of the first-step estimator $\widehat{\phi}$. Then

$$\psi^{2s}(X_i) = -A^{-1} \left[f_2(X_i) + J\psi^{\phi}(X_i) \right] \tag{10}$$

where A, J, and ψ^W in f_2 are evaluated using the preliminary weight $W(\phi_0)$.

(iv) Iterated GMM:

$$\psi^{it}(X_i) = -(A + J)^{-1} f_2(X_i) \tag{11}$$

where A, J, and ψ^W in f_2 are evaluated using the weight $W(\theta_0)$.

A common choice of weight matrix is the uncentered efficient weight $\widehat{W} = (\frac{1}{n}\sum_{i=1}^{n} g_i g_i')^{-1}$. Let $\Omega = \mathbb{E}[g(X_i,\theta_0)g(X_i,\theta_0)']$. In this case, the delta method gives $\psi^W(X_i) = \text{vec}(-W(g_i g_i' - \Omega)W)$. Substituting this yields

$$(g' \otimes G')\psi^W(X_i) = -G'Wg_i g_i'Wg + G'W\Omega Wg.$$

Since $W = \Omega^{-1}$ and the population FOC guarantees $G'Wg = 0$, the second term drops out, yielding $f_2(X_i) = f_1(X_i) - G'Wg_i g_i'Wg$.

The influence functions in Proposition 1 have appeared, in various forms, in the existing literature; see, for example, Hall and Inoue (2003), Hwang et al. (2022), and Hansen and Lee (2021). For Continuously Updating GMM (CUGMM), which jointly updates the parameter and the weight matrix, it presents unique theoretical challenges under misspecification, including non-convexity and the risk of variance collapse (Kleibergen and Zhan, 2025). We provide analysis of CUGMM, including its consistency and novel misspecification-robust influence function, in Appendix B.

The influence function for two-step GMM derived in Proposition 1(iii) has an intuitive decomposition. It consists of a direct estimation effect and an indirect effect arising from the first-step estimator $\widehat{\phi}$. This structure allows us to formally characterize the sensitivity of the second-step estimator to the first-step estimator.

Proposition 2 (Sensitivity to the First-Step Estimator)**.** *Suppose Assumption 1 holds. Let $\widehat{\phi}$ be a first-step estimator with probability limit ϕ_0, and let $\widehat{\theta}(\widehat{\phi})$ denote the corresponding second-step GMM estimator defined as the solution to the sample first-order condition given $\widehat{\phi}$. Then the sensitivity of the second-step estimator to the first-step estimator is given by*

$$\Lambda_\phi = plim \frac{\partial \widehat{\theta}(\widehat{\phi})}{\partial \widehat{\phi}'} = -A^{-1}J,$$

where A and J are evaluated at (θ_0, ϕ_0).

Proposition 2 shows that the indirect component of the two-step GMM influence function arises because first-step estimation uncertainty enters through the weight updating map. The matrix $A^{-1}J$ represents the linearization of a weight update, as in the two-step estimator. Iterated GMM repeatedly applies the same updating map, so the local behavior of the iterated procedure is governed by successive applications of this operator. The contraction condition in Assumption 1(vi) is imposed on the population updating map underlying the two-step estimator. Linearizing this map around θ_0 yields the Jacobian $-A^{-1}J$, so the contraction condition is equivalent to requiring $\rho(-A^{-1}J) < 1$, where $\rho(\cdot)$ denotes the spectral radius.

Under this condition, the influence functions generated by an s-step GMM procedure have a recursive representation driven by repeated application of $A^{-1}J$. As the number of updating steps increases, this recursion converges to the influence function of the iterated GMM estimator, yielding a fixed-point interpretation of iterated GMM under misspecification.

Proposition 3 (Limit of s-step GMM influence functions)**.** *Suppose Assumption 1 holds. Let $\psi^{ss}(X_i)$ denote the influence function of the s-step GMM estimator obtained by iteratively updating the efficient weight matrix using the previous-step estimate. Then, under the contraction condition in Assumption 1(v),*

$$\lim_{s\to\infty} \psi^{ss}(X_i) = \psi^{it}(X_i) = -(A+J)^{-1} f_2(X_i).$$

Remark 1 (Variance reduction of iterated GMM)**.** *The recursion in Proposition 3 implies the exact decomposition*

$$\psi^{ss}(X_i) - \psi^{it}(X_i) = (-A^{-1}J)^{s-1} d(X_i), \qquad d(X_i) = \psi^{1s}(X_i) + (A+J)^{-1} f_2(X_i).$$

Under the contraction condition, the step-specific component $(-A^{-1}J)^{s-1}d(X_i)$ shrinks geometrically in s (in mean square under mild moment conditions). This provides a formal sense in which iterated GMM eliminates step-specific variation induced by incomplete weight updating under misspecification.

3.2 Discussions on GMM Sensitivity

3.2.1 One-Step GMM with Weight Independent of Parameter

When the weight matrix W is deterministic, the influence function is given by Proposition 1(i):

$$\psi(X_i) = -A^{-1}\left[G'Wg(X_i,\theta_0) + G(X_i,\theta_0)'Wg\right].$$

This structure reveals how misspecification alters the sensitivity relative to the correctly specified case. The population curvature matrix $A = G'WG + H$ now includes the term $H = (g'W \otimes I_p)R$, which arises from the curvature of the moment function evaluated at a parameter value where the population moments are nonzero. Example 1 shows that AGS is obtained as a special case of our MRS under correct specification.

Example 1 (AGS under Correct Specification)**.** *If $\mathbb{E}[g(X_i,\theta_0)] = 0$, then $g = 0$, implying $H = 0$. Consequently, A reduces to the standard Hessian $G'WG$. The influence function simplifies to*

$$\psi(X_i) = -(G'WG)^{-1}G'Wg(X_i,\theta_0).$$

Since $g(X_i, \theta_0) = \nu(X_i)$ under correct specification, the sensitivity coefficient calculated via Theorem 1 becomes

$$\Lambda = \mathbb{E}[\psi\nu']\mathbb{E}[\nu\nu']^{-1} = -(G'WG)^{-1}G'W = \Lambda_{AGS}.$$

In this scenario, the estimator is fully determined by the linear projection of the moments, yielding an informativeness measure of $\Delta_k = 1$ for all k.

Under misspecification, the estimator's behavior also depends on stochastic variation in the Jacobian $G(X_i, \theta_0)$. When this variation is not fully spanned by the variation in the moment vector, the estimator is influenced by additional components that are not explained by the moments themselves, causing the informativeness measure Δ to fall below one.

When the weight matrix is estimated from the data but does not depend on θ, the influence function includes an additional term reflecting the sampling variation of $\widehat{W}$. From Proposition 1(ii),

$$\psi^{1s,W}(X_i) = -A^{-1}\left(f_1(X_i) + (g' \otimes G')\psi^W(X_i)\right).$$

The term $(g' \otimes G')\psi^W(X_i)$ captures the effect of estimating the weight matrix. A leading example is the 2SLS estimator, where the estimated weight matrix $\widehat{W}$ is given by $(n^{-1}Z'Z)^{-1}$. In this case,

$$\psi^W(X_i) = -\operatorname{vec}\left(\mathbb{E}[Z'Z]^{-1}(Z'Z - \mathbb{E}[Z'Z])\mathbb{E}[Z'Z]^{-1}\right).$$

Lee (2018) uses $\psi^{1s,W}(X_i)$ to calculate the asymptotic variance of the 2SLS estimator under treatment effect heterogeneity, which implies misspecified moment condition.

Note that $(g' \otimes G')\psi_W(X_i)$ is proportional to g. If the model is correctly specified with $g = 0$, the uncertainty in the weight matrix does not affect the asymptotic distribution of the estimator to the first order. Under misspecification, however, this term contributes to the estimator's variance and generally reduces informativeness.

Example 2 (2SLS)**.** *Consider a linear IV model $Y_i = D_i\theta + \epsilon_i$ with instruments Z_i. The moment function is $g(X_i, \theta) = Z_i(Y_i - D_i\theta)$, which is linear in θ, implying $R = 0$ and thus $H = 0$. Consider the 2SLS estimator with weight $\widehat{W} = (n^{-1}Z'Z)^{-1}$. In Appendix C.1, we apply Proposition 1(ii) to decompose the influence function of θ into three components,*

$$\psi^{1s,W}(X_i) = \Lambda_{AGS}\, g(X_i, \theta_0) + \Lambda_{Bias}\operatorname{vec}(G(X_i)') + \Lambda_{Weight}\operatorname{vec}(Z_iZ_i')$$

where

$$\Lambda_{AGS} = -(G'WG)^{-1}G'W$$
$$\Lambda_{Bias} = -(G'WG)^{-1}(g'W \otimes I_p)$$
$$\Lambda_{Weight} = (G'WG)^{-1}\left[(g'W) \otimes (G'W)\right].$$

The MRS is

$$\Lambda = \Lambda_{AGS} + \Lambda_{Bias}\sigma_{G\nu}\sigma_{\nu\nu}^{-1} + \Lambda_{Weight}\xi \tag{12}$$

where $\sigma_{G\nu}$ is the covariance between $G(X_i, \theta_0)$ and $\nu(X_i)$, ξ is the projection coefficient of $\mathrm{vec}(Z_i Z_i')$ *onto $\nu(X_i)$. The second component in Equation (12) captures the sensitivity to sampling variation in the Jacobian of the moment conditions. Under misspecification, sampling variation in $G(X_i, \theta_0)$ interacts with the nonzero population moment g, generating additional variation in the estimator that is not explained by the moments themselves. The third component reflects the contribution of estimating the weight matrix $\widehat{W}$. These two components are zero under correct specification as $g = 0$. The loss of informativeness $1 - \Delta$ can therefore be attributed to variation in the Jacobian and the estimated weight matrix that is not explained by the moment variation, scaled by the degree of misspecification through Λ_{Bias} and Λ_{Weight}, respectively.*

3.2.2 Efficient GMM Estimators

Efficient GMM estimators use a weight matrix that depends on the parameter, $W(\theta) = \Omega(\theta)^{-1}$. In the two-step GMM estimator, the weight matrix is evaluated at a preliminary estimator $\widehat{\phi}$, so the second-step estimator inherits uncertainty from the first step. Proposition 1(iii) shows that the influence function takes the form

$$\psi^{2s}(X_i) = -A^{-1}f_2(X_i) - A^{-1}J\,\psi(X_i),$$

where $\psi(X_i)$ is the influence function of the first-step estimator. The matrix J captures the sensitivity of the optimal weight matrix to the parameter, and the term $-A^{-1}J\,\psi(X_i)$ quantifies how first-step estimation uncertainty enters the second-step estimator. Under correct specification, $g = 0$ and hence $J = 0$, recovering the standard result that the choice of the first-step estimator does not affect asymptotic efficiency.

Iterated GMM repeatedly updates the weight matrix until the parameter used in the weight matrix coincides with the estimated parameter. As shown in Proposition 1(iv), this alters the effective curvature of the problem from A to $A + J$. While the estimator no longer depends on an external preliminary estimator, sensitivity to misspecification remains embedded in the curvature through J.

The following example illustrates a central implication of these results: efficient GMM estimators can share the same local sensitivity while exhibiting substantially different informativeness under misspecification.

Example 3 (Nonlinear normal mean model of Schennach (2007))**.** *Consider i.i.d. data* $X_i \sim N(\mu, \sigma^2)$ *and the overidentified moment vector*

$$g(X_i, \theta) = \begin{pmatrix} X_i - \theta \\ (X_i - \theta)^2 - 1 \end{pmatrix}.$$

The parameter of interest is the mean θ*, while the variance restriction may be misspecified when* $\sigma^2 \neq 1$*.*

For all efficient GMM estimators considered below, the pseudo-true value satisfies $\theta_0 = \mu$*, regardless of the degree of misspecification. Moreover, the sensitivity measure is identical across estimators,* $\Lambda = (1, 0)$*, reflecting that* θ *locally tracks the sample mean.*

Despite identical sensitivity, informativeness differs sharply across estimators. When the weight matrix depends on the parameter, the influence function contains additional components arising from estimation of the weight matrix under misspecification. These components inflate the variance of the estimator without affecting local sensitivity, leading to informativeness $\Delta < 1$*. Table 1 summarizes the sensitivity and informativeness measures for one-step, two-step, and iterated estimators.*

Table 1: Sensitivity and Informativeness in Schennach (2007) Model

Estimator	**Condition for $\theta_0 = \mu$**	**Sensitivity Λ**	**Informativeness Δ**
One-Step ($W = I$)	$\sigma^2 \geq 1/2$	$(1, 0)$	1
Two-Step GMM	$\sigma^2 > 0$	$(1, 0)$	< 1 (if $a \neq 0$)
Iterated GMM	$\sigma^2 > 0$	$(1, 0)$	$\frac{2\sigma^4}{5\sigma^4 - 6\sigma^2 + 3}$

3.2.3 Informativeness as Structural Efficiency

A key implication of our framework is that informativeness Δ_k has a natural interpretation as a measure of structural efficiency. To see this, consider the population linear projection of the influence function of the kth parameter on the influence function of the moments

$$\psi_k = \Lambda_k \nu + \varepsilon_k,$$

where Λ_k denotes the kth row of the sensitivity matrix Λ, and the residual ε_k satisfies $\mathbb{E}[\varepsilon_k \nu] = 0$ by construction. This decomposition implies

$$\mathrm{Var}(\psi_k) = \mathrm{Var}(\Lambda_k \nu) + \mathrm{Var}(\varepsilon_k).$$

Recall that informativeness is defined as

$$\Delta_k = \frac{\mathrm{Var}(\Lambda_k \nu)}{\mathrm{Var}(\psi_k)}.$$

Thus, Δ_k measures the share of the estimator's asymptotic variance that is linearly explained by sampling variation in the moment conditions themselves. The complement $1 - \Delta_k$ captures variance generated by other components of the estimation procedure.

This interpretation clarifies the role of misspecification. Under correct specification, the Jacobian behaves asymptotically as a constant linear map, and the estimator is asymptotically a linear function of the moments. In this case, $\varepsilon_k = 0$ and $\Delta_k = 1$. Under misspecification, however, the influence functions derived in Proposition 1 generally contain stochastic components arising from variation in the Jacobian and, when applicable, the estimated weight matrix. When this variation is not fully spanned by the moments, it contributes to $\mathrm{Var}(\varepsilon_k)$ and reduces informativeness.

From this perspective, informativeness provides a diagnostic measure of structural efficiency: it quantifies how much of the estimator's variability is driven by the moment conditions versus auxiliary features of the GMM procedure. This interpretation complements classical efficiency notions based on optimal weighting, which focus on minimizing variance under correct specification, but do not distinguish variance components arising from misspecification.

The role of moment functional form is illustrated in Appendix C.4, which compares two misspecified GMM estimators that identify the same pseudo-true parameter using different moment functions. Although both estimators have identical sensitivity, their informativeness differs substantially, reflecting differences in how Jacobian variation enters the influence function. This example highlights that, under misspecification, informativeness depends not only on which moments are used, but also on how they enter the model.

4 Applications

4.1 BLP Automobile Market

We apply our misspecification-robust sensitivity (MRS) measure to the BLP automobile demand and supply model (Berry et al., 1995), following the implementation in Andrews et al. (2017). The purpose of this application is to illustrate how allowing for misspecified moments alters sensitivity diagnostics relative to AGS, holding the model, data, and estimation procedure fixed.

The model and instruments are identical to those used in AGS. Let X_j denote the observables for market-product j, Z_j be the vector of instruments, and $u_j(\theta) = (\xi_j(\theta), \omega_j(\theta))'$ be the vector of unobserved product quality and marginal cost shocks. The identifying moments are based on the orthogonality of these unobservables to the instruments. Stacking the demand and supply moments yields a 31-dimensional sample moment vector $\widehat{g}(\theta) = \frac{1}{n}\sum_{j=1}^{n} g(X_j, \theta)$, where $g(X_j, \theta) = Z_j \otimes u_j(\theta)$.

Estimation is conducted using efficient two-step GMM. The second-step estimator $\widehat{\theta}$ minimizes the objective function

$$\widehat{Q}(\theta) = \widehat{g}(\theta)'\widehat{W}(\widehat{\phi})\widehat{g}(\theta)$$

where $\widehat{W}(\widehat{\phi}) = \left(\frac{1}{n}\sum j = 1^n g(X_j, \widehat{\phi}) g(X_j, \widehat{\phi})'\right)^{-1}$ is the estimated efficient weight matrix, and $\widehat{\phi}$ is the first-step GMM estimator with identity weight matrix. Under global misspecification, the influence function of $\widehat{\theta}$ is derived in Proposition 1(iii).

The parameter of interest is the average markup,

$$c(\theta) = \frac{1}{n}\sum_{j} \frac{p_j - mc_j(\theta)}{p_j},$$

whose influence function is given by the gradient $\partial c(\theta)/\partial \theta'$ multiplied by the influence function of the GMM estimator.

AGS study sensitivity under local violations of instrument validity, modeled as direct effects of instruments entering demand or cost equations at rate $1/\sqrt{n}$. We compute the corresponding normalized sensitivity coefficients using both AGS and our MRS measure for the same excluded instruments. The AGS sensitivities are taken directly from their replication files, while MRS is computed using the misspecification-robust influence function derived in Proposition 1(iii).

Figure 1 compares the two sensitivity measures. For both demand and supply side moments, MRS is generally smaller than AGS in magnitude. Specifically, on the demand-side, AGS suggests that the estimated markup is sensitive to instruments related to own-firm product characteristics, whereas MRS indicates that demand-side sensitivities are uniformly small. Allowing for global misspecification therefore substantially attenuates the role of demand-side moments and changes both the magnitude and the ranking of influential moments. Table 2 revisits the asymptotic bias calculations in AGS using MRS. It shows that violations related to economies of scope and to cross-firm demand spillovers have significantly smaller effects under MRS relative to AGS.

To assess the model specification, we calculate the J-statistic for overidentifying restrictions. The test yields a J-statistic of 947 with 25 degrees of freedom, corresponding to a p-value < 0.001. The strong rejection of the overidentifying restrictions suggests significant model misspecification.

Finally, we compute the informativeness of the average markup estimator. The informativeness is 0.58, indicating that 58% of the estimator's asymptotic variance is linearly explained by the moments themselves. The remaining 42% reflects structural noise arising from the interaction of misspecification with the Jacobian and estimated weight matrix. This highlights the practical relevance of sensitivity analysis under misspecification in structural models.

Figure 1: MRS and Replication of AGS for BLP Model

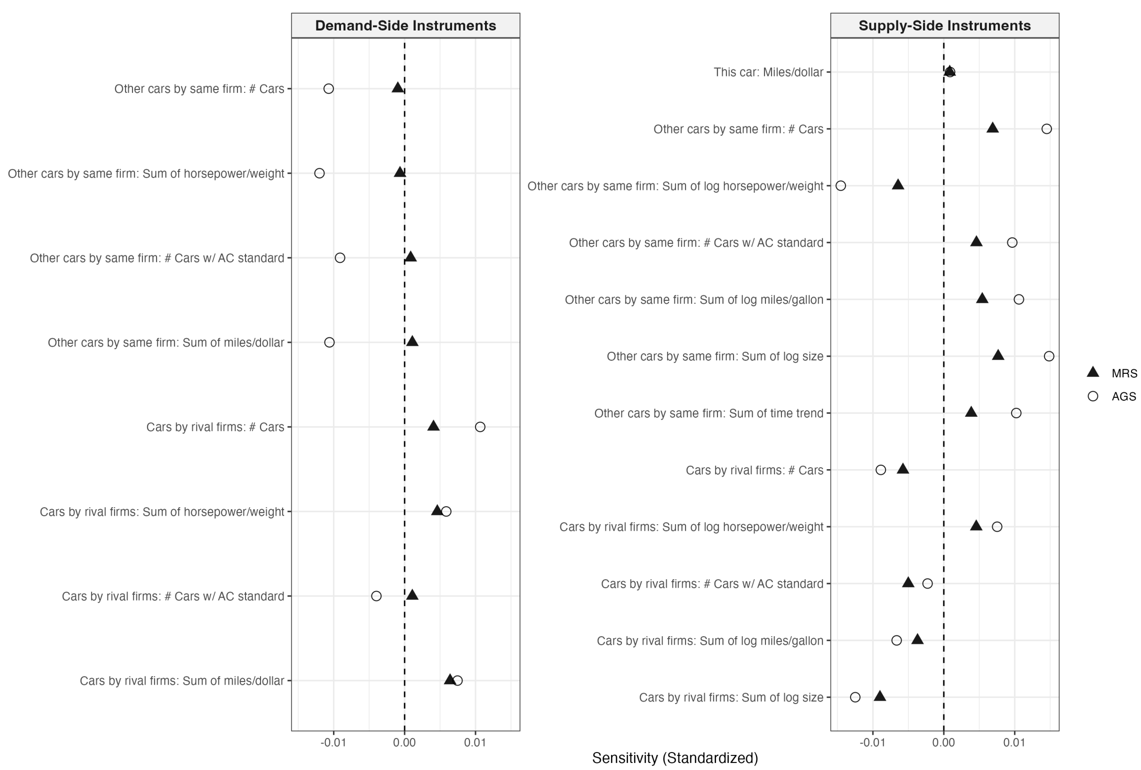


Table 2: Asymptotic Bias in Average Markup Based on MRS and AGS

	Bias in Average Markup	
	AGS	**MRS**
Supply-Side Violations (1% increase in marginal cost)		
Removing own car	−0.1731	−0.0827
Removing rival's car	0.2095	0.1546
Demand-Side Violations (1% decrease in willingness to pay)		
Removing own car	−0.1277	−0.0137
Removing rival's car	0.2515	0.0881
Baseline Estimate	0.3272	0.3272

4.2 BPP Income Dynamics and Consumption Insurance

Our second application revisits the BPP model of household income dynamics and consumption insurance (Blundell et al., 2008). The BPP framework is a canonical setting for assessing how income shocks are transmitted into consumption, and has been extensively studied in subsequent work (Kaplan and Violante, 2010; Chatterjee et al., 2021). This application is particularly well suited to our framework because estimation relies on a large set of second-moment conditions and an estimated optimal weight matrix.

In the BPP model, unexplained log income is decomposed into a permanent and a transitory component. The permanent component follows a random walk, $y^P_{i,t} = y^P_{i,t-1} + \zeta_{i,t}$, while the transitory component follows an MA(1) process. Income growth is therefore

$$\Delta y_{i,t} = \zeta_{i,t} + \Delta\epsilon_{i,t} + \theta\Delta\epsilon_{i,t-1}.$$

Consumption growth is modeled as

$$\Delta c_{i,t} = \phi\zeta_{i,t} + \psi\epsilon_{i,t} + \xi_{i,t} + \Delta u_{i,t},$$

where ϕ and ψ measure partial insurance against permanent and transitory shocks.

Estimation proceeds by Optimal Minimum Distance (OMD), matching empirical covariances of income and consumption growth to their theoretical counterparts. We focus on the simplified stationary model discussed in Appendix C of Blundell et al. (2008), with parameter vector $\beta = (\phi, \psi, \sigma^2_\zeta, \sigma^2_\epsilon, \sigma^2_\xi)$. Let m_i denote the vector of sample cross-products of income and consumption growth for household i, and let $m(\beta)$ denote the corresponding theoretical covariances.

The sample moment conditions are defined as the distance between the data and the theory: $\widehat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \beta)$, where $g(X_i, \beta) = m_i - m(\beta)$. The OMD estimator minimizes the criterion

$$\widehat{Q}(\beta) = \widehat{g}(\beta)'\widehat{W}\widehat{g}(\beta)$$

where the optimal weight matrix

$$\widehat{W} = \left[\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})(m_i - \bar{m})'\right]^{-1}$$

is the inverse of the sample covariance matrix of the data moments.

This OMD estimator of BPP model is the one-step GMM with an estimated weight matrix. The Jacobian is $G(X_i, \beta) = -(\partial/\partial\beta')m(\beta)$, which is deterministic across households. Because there is zero sampling variation in the Jacobian, the Minimum Distance estimator effectively isolates the impact of estimating the optimal weight matrix. Accordingly, differences between AGS and MRS arise entirely from weight-matrix uncertainty, captured by the term $(g' \otimes G')\psi^W(X_i)$ from $f_2(X_i)$ in Equation (9).

To avoid near-singularity of $\widehat{W}$, we restrict attention to a set of 11 unique moments. The construction of these moments is detailed in Appendix D.3. Parameter estimates are reported in Table 3. We also perform a J-test for overidentifying restrictions. For the simplified BPP model, the J-statistic is 306.77 with 6 degrees of freedom, resulting in a p-value < 0.001.[1] This result strongly reject the null hypothesis of correct model specification, motivating the need for misspecification-robust inference.

Figures 2 and 3 compare AGS and MRS for the insurance coefficients ϕ and ψ. Because the Jacobian of the moments depends only on θ, this setting isolates the effect of estimating the optimal weight matrix. Accordingly, differences between AGS and MRS arise entirely from weight-matrix uncertainty.

The estimated informativeness is 0.57 for ϕ and 0.26 for ψ, indicating substantial structural efficiency loss, especially for the response to transitory shocks. For ϕ, both AGS and MRS show that identification is primarily driven by contemporaneous covariances between income and consumption. For ψ, however, MRS assigns substantially lower sensitivity to several dynamic covariance moments than AGS, reflecting the fact that their influence operates largely through variation in the estimated weight matrix.

[1] For the full model, the J-statistic is approximately 9.5×10^5 with $df = 290$ and $p < 0.001$.

This application illustrates that even when the Jacobian is deterministic, accounting for weight-matrix estimation uncertainty is crucial for valid sensitivity analysis under potential misspecification. Ignoring this channel, as in AGS, can substantially overstate the effective information content of the moments.

5 Conclusion

This paper develops a sensitivity and informativeness framework for GMM estimators that remains valid under general moment misspecification. By expressing sensitivity measures through influence functions, we provide a unified characterization of how deviations in moment conditions affect GMM estimators across one-step, two-step, iterated, and continuously updating procedures.

Our analysis shows that allowing for misspecification fundamentally alters both the magnitude and the interpretation of sensitivity measures. In particular, estimators may exhibit substantial sensitivity even when conventional measures suggest robustness, and the choice of weight matrix and updating scheme plays a central role in shaping this behavior. The informativeness measure introduced in this paper quantifies the extent to which sampling variation in an estimator is driven by the moment conditions themselves, as opposed to variation arising from the Jacobian or the estimated weight matrix.

Applications to the BLP automobile market model and the BPP consumption insurance model illustrate that these distinctions are empirically important. In both settings, accounting for misspecification leads to materially different sensitivity rankings and reveals substantial losses in structural efficiency that are not captured by existing approaches.

Overall, the proposed framework provides a diagnostic tool for assessing the robustness of GMM-based structural estimates when moment validity is uncertain. Rather than correcting for misspecification, the approach clarifies how and through which channels misspecification affects estimation, thereby improving the transparency of empirical conclusions drawn from GMM procedures.

Table 3: BPP Parameter Estimates

Parameter	Toy Model	Full Model
θ		0.13 (0.02) [0.03]
ϕ	0.27 (0.04) [0.05]	0.33 (0.03) [0.06]
ψ	0.01 (0.05) [0.05]	0.07 (0.03) [0.05]
σ_u		0.09 (0.01) [0.01]
$\sigma_{\zeta,1}$	0.17 (0.01) [0.01]	0.12 (0.01) [0.01]
$\sigma_{\zeta,2}$		0.14 (0.01) [0.01]
$\sigma_{\zeta,3}$		0.13 (0.01) [0.02]
$\sigma_{\zeta,4}$		0.08 (0.01) [0.02]
$\sigma_{\zeta,5}$		0.14 (0.02) [0.03]
$\sigma_{\zeta,6}$		0.12 (0.01) [0.02]
$\sigma_{\zeta,7}$		0.12 (0.02) [0.02]
$\sigma_{\zeta,8}$		0.14 (0.02) [0.04]
$\sigma_{\zeta,9}$		0.12 (0.01) [0.02]
$\sigma_{\zeta,10}$		0.10 (0.01) [0.02]
$\sigma_{\epsilon,1}$	0.17 (0.00) [0.00]	0.14 (0.01) [0.02]
$\sigma_{\epsilon,2}$		0.12 (0.01) [0.01]
$\sigma_{\epsilon,3}$		0.14 (0.01) [0.01]
$\sigma_{\epsilon,4}$		0.16 (0.01) [0.01]
$\sigma_{\epsilon,5}$		0.14 (0.01) [0.01]
$\sigma_{\epsilon,6}$		0.16 (0.01) [0.01]
$\sigma_{\epsilon,7}$		0.17 (0.01) [0.01]
$\sigma_{\epsilon,8}$		0.18 (0.01) [0.02]
$\sigma_{\epsilon,9}$		0.18 (0.01) [0.01]
$\sigma_{\epsilon,10}$		0.16 (0.01) [0.01]
$\sigma_{\epsilon,11}$		0.17 (0.01) [0.01]
$\sigma_{\epsilon,12}$		0.16 (0.01) [0.01]
$\sigma_{\xi,1}$	0.24 (0.01) [0.01]	0.23 (0.01) [0.02]
$\sigma_{\xi,2}$		0.20 (0.01) [0.02]
$\sigma_{\xi,3}$		0.20 (0.01) [0.02]
$\sigma_{\xi,4}$		0.21 (0.01) [0.02]
$\sigma_{\xi,5}$		0.20 (0.01) [0.02]
$\sigma_{\xi,6}$		0.25 (0.02) [0.03]
$\sigma_{\xi,7}$		0.25 (0.02) [0.02]
$\sigma_{\xi,8}$		0.23 (0.01) [0.02]
$\sigma_{\xi,9}$		0.22 (0.01) [0.01]

Note: Point estimates with conventional standard errors in parentheses and misspecification-robust standard errors in brackets, calculated by the variance of the influence function in Proposition 1(ii). Column 1 is the Minimum Distance estimator of β with optimal weight matrix. Column 2 is the replication of Table B1 in Chatterjee et al. (2021), which replicates Table 6 of BPP but with a correction of code error in BPP.

Figure 2: ϕ Sensitivity w.r.t 1SD Change in Moment Conditions

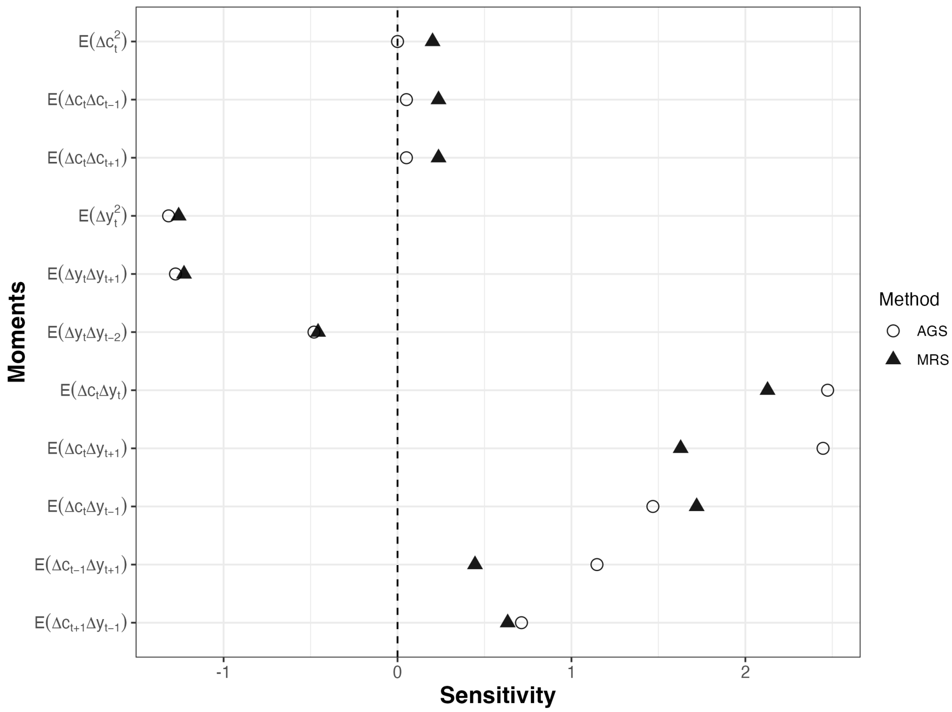
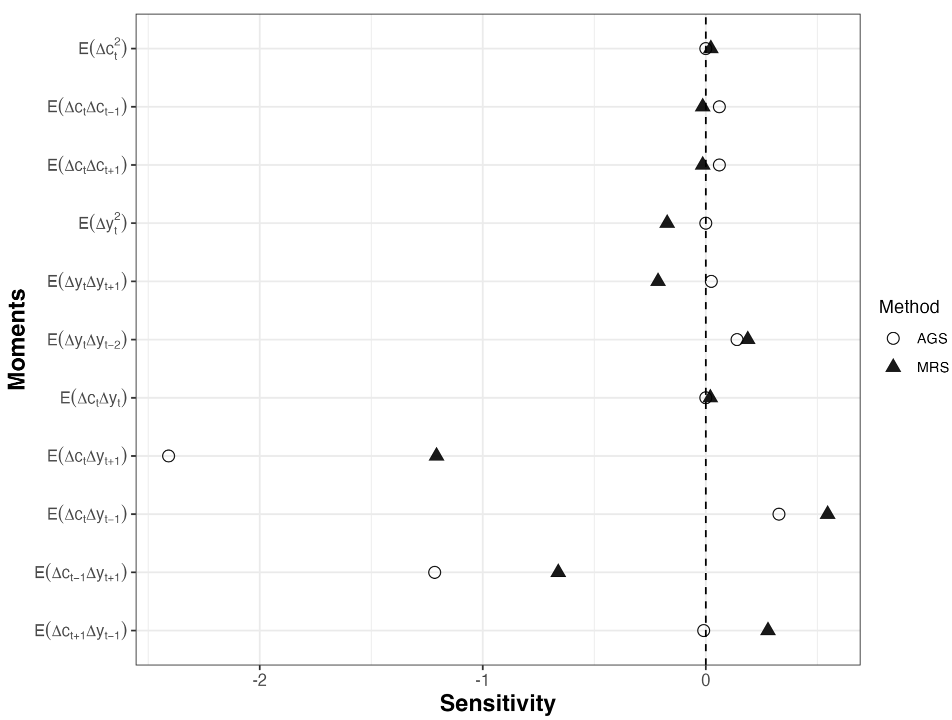


Figure 3: ψ Sensitivity w.r.t 1SD Change in Moment Conditions



References

Andrews, I., J. Chen, and O. Tecchio (2025): "The purpose of an estimator is what it does: Misspecification, estimands, and over-identification," *Econometric Society 2025 World Congress Volume*.

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.

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

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98, 1887–1921.

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

Chatterjee, A., J. Morley, and A. Singh (2021): "Estimating household consumption insurance," *Journal of Applied Econometrics*, 36, 628–635.

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

Cinelli, C. and C. Hazlett (2020): "Making sense of sensitivity: Extending omitted variable bias," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82, 39–67.

Gentzkow, M. and J. M. Shapiro (2014): *Measuring the sensitivity of parameter estimates to sample statistics*, National Bureau of Economic Research.

Hall, A. R. and A. Inoue (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114, 361–394.

Hansen, B. (2022): *Econometrics*, Princeton University Press.

Hansen, B. E. and S. Lee (2021): "Inference for iterated GMM under misspecification," *Econometrica*, 89, 1419–1447.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the econometric society*, 1029–1054.

Hwang, J., B. Kang, and S. Lee (2022): "A doubly corrected robust variance estimator for linear GMM," *Journal of Econometrics*, 229, 276–298.

Imbens, G. W. (1997): "One-step estimators for over-identified generalized method of moments models," *The Review of Economic Studies*, 64, 359–383.

——— (2003): "Sensitivity to exogeneity assumptions in program evaluation," *American Economic Review*, 93, 126–132.

Iskrev, N. (2019): "What to expect when you're calibrating: Measuring the effect of calibration on the estimation of macroeconomic models," *Journal of Economic Dynamics and Control*, 99, 54–81.

Jørgensen, T. H. (2023): "Sensitivity to calibrated parameters," *Review of Economics and Statistics*, 105, 474–481.

Kaplan, G. and G. L. Violante (2010): "How much consumption insurance beyond self-insurance?" *American Economic Journal: Macroeconomics*, 2, 53–87.

Kleibergen, F. and Z. Zhan (2025): "Double robust inference for continuous updating GMM," *Quantitative Economics*, 16, 295–327.

Lee, S. (2018): "A consistent variance estimator for 2sls when instruments identify different lates," *Journal of Business & Economic Statistics*, 36, 400–410.

Oster, E. (2019): "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Rosenbaum, P. R. and D. B. Rubin (1983): "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *Journal of the Royal Statistical Society: Series B (Methodological)*, 45, 212–218.

Schennach, S. M. (2007): "Point estimation with exponentially tilted empirical likelihood," *The Annals of Statistics*, 35, 634–672.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of econometrics*, 126, 25–51.

A Appendix: Proofs

Proof of Proposition 1. We begin by establishing preliminary uniform convergence results implied by Assumption 1.

Assumptions 1(i)–(iii) imply a Uniform Law of Large Numbers (ULLN) over the parameter space Θ for the sample moments and their derivatives that appear in the sample first-order conditions. In particular,

$$\sup_{\theta\in\Theta} \|\widehat{g}(\theta) - \mathbb{E}[g(X_i,\theta)]\| \xrightarrow{p} 0,$$
$$\sup_{\theta\in\Theta} \|\widehat{G}(\theta) - \mathbb{E}[G(X_i,\theta)]\| \xrightarrow{p} 0,$$
$$\sup_{\theta\in\Theta} \|\widehat{R}(\theta) - \mathbb{E}[R(X_i,\theta)]\| \xrightarrow{p} 0,$$

where $R(X_i,\theta) = \frac{\partial}{\partial\theta'}\operatorname{vec}(G(X_i,\theta)')$. Since $H(\theta)$ is an affine function of $R(\theta)$ with coefficients depending on $g(\theta)$ and the weight matrix W, it follows that

$$\sup_{\theta\in\Theta} \|\widehat{H}(\theta) - \mathbb{E}[H(X_i,\theta)]\| \xrightarrow{p} 0.$$

For estimators with parameter-dependent weight matrices, including two-step and iterated GMM, uniform convergence of the sample objective function and its derivatives follows from Assumption 1(i)–(iii) and (vi)(b). In particular, the sample criterion $Q_n(\theta)$ converges uniformly to its population counterpart $Q(\theta)$. Standard extremum estimator arguments (e.g., Newey and McFadden, 1994, Theorem 2.1) therefore imply consistency, $\widehat{\theta} \xrightarrow{p} \theta_0$ (and $\widehat{\phi} \xrightarrow{p} \phi_0$ for the first step).

For iterated GMM, Assumption 1(vi) imposes a contraction condition on the population updating map at θ_0. Together with the uniform law of large numbers over Θ, this ensures that the sample updating map converges uniformly to its population counterpart and that the iterated GMM sequence converges with probability approaching one to the unique fixed point θ_0. These conditions are directly analogous to those in Hansen and Lee (2021), and guarantee existence and consistency of the iterated GMM estimator under misspecification.

Since θ_0 is the unique interior minimizer of $Q(\theta)$, the population first-order condition holds

$$G'Wg = 0,$$

where W is the corresponding limit weight matrix.

With probability approaching one, the estimator satisfies the sample first-order condition $F_n(\widehat{\theta}) = 0$. Because $F_n(\theta)$ is a $p \times 1$ vector, we apply the mean value theorem row-by-row around θ_0 yielding

$$0 = F_n(\theta_0) + F_{n,\theta}(\widetilde{\theta})(\widehat{\theta} - \theta_0),$$

where $F_{n,\theta}(\widetilde{\theta})$ is a matrix whose k-th row is evaluated at an intermediate point $\widetilde{\theta}^{(k)}$ on the line segment between $\widehat{\theta}$ and θ_0. Because $\widehat{\theta} \xrightarrow{p} \theta_0$, every $\widetilde{\theta}^{(k)} \xrightarrow{p} \theta_0$, and by the ULLN, the sample Jacobian $F_{n,\theta}(\widetilde{\theta})$ converges uniformly in probability to the appropriate population curvature matrix.

Rearranging,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -\left[F_{n,\theta}(\widetilde{\theta})\right]^{-1}\sqrt{n}F_n(\theta_0).$$

Thus, for each estimator, the derivation proceeds by (i) establishing the probability limit of $F_{n,\theta}(\widetilde{\theta})$, and (ii) obtaining an asymptotic linear expansion

$$\sqrt{n}F_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_i) + o_p(1).$$

The influence function then follows as $\psi(X_i) = -A^{-1}f(X_i)$.

For two-step GMM, additional moment conditions are required to derive asymptotic linear representations. In particular, Assumption 1(vi)(b) guarantees the finite fourth moments needed to apply a central limit theorem to $g(X_i,\theta_0)g(X_i,\theta_0)'$ and hence to justify the influence function expansions for efficient GMM estimators.

To formalize the $o_p(1)$ remainder terms in the following multivariable Taylor expansions, let $\Delta G = \widehat{G}(\theta_0) - G$, $\Delta g = \widehat{g}(\theta_0) - g$, and $\Delta W = \widehat{W} - W$ (evaluated at the relevant probability limit). By the Central Limit Theorem and Assumption 1, ΔG, Δg, and ΔW are all $O_p(n^{-1/2})$. Consequently, any pairwise product (e.g., $\Delta G' W \Delta g$) is $O_p(n^{-1})$, and any three-way product (e.g., $\Delta G' \Delta W \Delta g$) is $O_p(n^{-3/2})$. When scaled by $\sqrt{n}$, these higher-order cross-products become at most $O_p(n^{-1/2})$, which vanish into $o_p(1)$.

One-Step GMM with Deterministic Weight W

The sample FOC is $F_n(\theta) = \widehat{G}(\theta)'W\widehat{g}(\theta)$ with derivative

$$F_{n,\theta}(\theta) = \widehat{G}(\theta)'W\widehat{G}(\theta) + \big(\widehat{g}(\theta)'W \otimes I_p\big)\widehat{R}(\theta). \tag{A1}$$

By the ULLN and Continuous Mapping Theorem (CMT),

$$F_{n,\theta}(\widetilde{\theta}) \xrightarrow{p} G'WG + H = A.$$

We expand $F_n(\theta_0)$ by decomposing $\widehat{G}(\theta_0)$ and $\widehat{g}(\theta_0)$ around their population limits G and g.

$$\begin{aligned} F_n(\theta_0) &= (G + \Delta G)'W(g + \Delta g) \\ &= G'Wg + G'W\Delta g + \Delta G'Wg + \Delta G'W\Delta g. \end{aligned}$$

We know $G'Wg = 0$. As established, the cross-term multiplied by $\sqrt{n}$ is

$$\sqrt{n}\Delta G'W\Delta g = O_p(n^{-1/2}) = o_p(1).$$

Thus, the expansion is

$$\sqrt{n}F_n(\theta_0) = G'W\sqrt{n}(\widehat{g} - g) + \sqrt{n}(\widehat{G} - G)'Wg + o_p(1).$$

This provides the asymptotic linear representation. The influence function of the FOC is

$$f_1(X_i) = G'W(g(X_i, \theta_0) - g) + (G(X_i, \theta_0) - G)'Wg,$$

which simplifies to the expression defined in the main text

$$f_1(X_i) = G'Wg(X_i, \theta_0) + G(X_i, \theta_0)'Wg.$$

We recover $\psi^{1s}(X_i) = -A^{-1}f_1(X_i)$.

One-Step GMM with $\widehat{W}$ estimated independent of θ

The FOC is $F_n(\theta) = \widehat{G}(\theta)'\widehat{W}\widehat{g}(\theta)$. Since $\widehat{W}$ does not depend on θ and $\widehat{W} \xrightarrow{p} W$ by Assumption 1(vi)(a), the derivative $F_{n,\theta}(\theta)$ has the same form as in Equation (A1), replacing W with $\widehat{W}$. By the ULLN, CMT, and consistency of $\widetilde{\theta}$, the limit of the Hessian remains A.

Expand $\widehat{G}, \widehat{g}$, and $\widehat{W}$ around their limits

$$\sqrt{n}F_n(\theta_0) = G'W\sqrt{n}(\widehat{g} - g) + \sqrt{n}(\widehat{G} - G)'Wg + G'\sqrt{n}(\widehat{W} - W)g + \mathcal{R}_n,$$

where the remainder strictly collects all higher-order cross-products $\mathcal{R}_n = \sqrt{n}(\Delta G'W\Delta g + G'\Delta W\Delta g + \Delta G'\Delta Wg + \Delta G'\Delta W\Delta g)$. As established, every individual term in $\mathcal{R}_n$ is bounded by at most

$O_p(n^{-1/2})$, hence $\mathcal{R}_n = o_p(1)$.

We analyze the third term using the vectorization identity $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}(B)$

$$G'\sqrt{n}(\widehat{W} - W)g = (g' \otimes G')\sqrt{n}\,\mathrm{vec}(\widehat{W} - W).$$

By the assumed asymptotic linearity of $\widehat{W}$, the influence function of this term is $(g' \otimes G')\psi^W(X_i)$. The influence function of the FOC is $f_1(X_i) + (g' \otimes G')\psi^W(X_i)$. This equals $f_2(X_i)$, and thus, the total influence function is

$$\psi^{1s,W}(X_i) = -A^{-1}f_2(X_i).$$

Two-Step GMM

Let $\widehat{\phi}$ be the first-step estimator. It converges in probability to the pseudo-true value ϕ_0. Following our derivation for One-Step GMM, we assume $\widehat{\phi}$ is asymptotically linear with influence function $\psi^\phi(X_i)$.

The second-step estimator $\widehat{\theta}$ utilizes the estimated weight matrix $\widehat{W}(\widehat{\phi})$. $\widehat{\theta}$ converges in probability to the pseudo-true value θ_0. The population weight matrix used in the second step is evaluated at ϕ_0, yielding $W(\phi_0)$. The pseudo-true value θ_0 minimizes the population objective function $Q(\theta) = g(\theta)'W(\phi_0)g(\theta)$. Under misspecification, generally $\phi_0 \neq \theta_0$.

The second-step estimator $\widehat{\theta}$ satisfies the sample FOC: $F_n(\widehat{\theta}, \widehat{\phi}) = \widehat{G}(\widehat{\theta})'\widehat{W}(\widehat{\phi})\widehat{g}(\widehat{\theta}) = 0$. We apply the mean value theorem row-by-row to expand $F_n(\widehat{\theta}, \widehat{\phi})$ around the pseudo-true values (θ_0, ϕ_0)

$$0 = F_n(\theta_0, \phi_0) + F_{n,\theta}(\widetilde{\theta}, \widetilde{\phi})(\widehat{\theta} - \theta_0) + F_{n,\phi}(\widetilde{\theta}, \widetilde{\phi})(\widehat{\phi} - \phi_0),$$

where $(\widetilde{\theta}, \widetilde{\phi})$ are intermediate values such that $(\widetilde{\theta}, \widetilde{\phi}) \xrightarrow{p} (\theta_0, \phi_0)$. Rearranging yields the expansion for the estimator:

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -\left[F_{n,\theta}(\widetilde{\theta}, \widetilde{\phi})\right]^{-1}\left(\sqrt{n}F_n(\theta_0, \phi_0) + F_{n,\phi}(\widetilde{\theta}, \widetilde{\phi})\sqrt{n}(\widehat{\phi} - \phi_0)\right).$$

We analyze the probability limits of the derivatives $F_{n,\theta}$ and $F_{n,\phi}$. The derivative with respect to θ, $F_{n,\theta}(\theta, \phi)$, is the Hessian of the second-step objective function. By the ULLN, consistency of $(\widetilde{\theta}, \widetilde{\phi})$, and the Continuous Mapping Theorem (CMT), $F_{n,\theta}(\widetilde{\theta}, \widetilde{\phi})$ converges in probability to the curvature matrix evaluated explicitly using the preliminary weight $W(\phi_0)$. Hence, its limit is exactly the matrix $A = G'W(\phi_0)G + H$, where $H = (g'W(\phi_0) \otimes I_p)R$.

The derivative with respect to ϕ, $F_{n,\phi}(\theta,\phi)$, captures the effect of the weight matrix estimation: $F_{n,\phi}(\theta,\phi) = (\widehat{g}(\theta)' \otimes \widehat{G}(\theta)')\frac{\partial \operatorname{vec}(\widehat{W}(\phi))}{\partial \phi'}$. Let $\widehat{S}(\phi) = \frac{\partial \operatorname{vec}(\widehat{W}(\phi))}{\partial \phi'}$. Assumption 1(vi)(b) ensures that $\widehat{S}(\widetilde{\phi}) \xrightarrow{p} S(\phi_0) = \left.\frac{\partial \operatorname{vec}(W(\phi))}{\partial \phi'}\right|_{\phi_0}$. By the ULLN and CMT, $F_{n,\phi}(\widetilde{\theta},\widetilde{\phi})$ converges in probability to the matrix $J = (g' \otimes G')S(\phi_0)$.

We analyze the asymptotic behavior of $\sqrt{n}F_n(\theta_0,\phi_0) = \sqrt{n}\widehat{G}(\theta_0)'\widehat{W}(\phi_0)\widehat{g}(\theta_0)$. We expand the sample quantities around their population limits $G, W(\phi_0), g$. Utilizing the identical bounding logic for the cross-products as in the One-Step estimated weight case, all higher-order interactions vanish into $o_p(1)$ because $\widehat{W}(\phi_0) - W(\phi_0) = O_p(n^{-1/2})$. Expanding, utilizing the population FOC $G'W(\phi_0)g = 0$, and collecting terms yields

$$\begin{aligned}\sqrt{n}F_n(\theta_0,\phi_0) = &\underbrace{G'W(\phi_0)\sqrt{n}(\widehat{g} - g) + \sqrt{n}(\widehat{G} - G)'W(\phi_0)g}_{\text{Term 1}} \\ &+ \underbrace{G'\sqrt{n}(\widehat{W}(\phi_0) - W(\phi_0))g}_{\text{Term 2}} + o_p(1).\end{aligned}$$

We identify the influence function for each term. Term 1 is the One-Step component evaluated with weight $W(\phi_0)$:

$$f^{1S}(X_i) = G'W(\phi_0)(g(X_i,\theta_0) - g) + (G(X_i,\theta_0) - G)'W(\phi_0)g.$$

Since $G'W(\phi_0)g = 0$, this simplifies precisely to $f_1(X_i)$ evaluated at $W(\phi_0)$:

$$f^{1S}(X_i) = G'W(\phi_0)g(X_i,\theta_0) + G(X_i,\theta_0)'W(\phi_0)g.$$

Term 2 is the Weight Matrix component. By the vectorization identity, this is $(g'\otimes G')\sqrt{n}\operatorname{vec}(\widehat{W}(\phi_0) - W(\phi_0))$. By assumed asymptotic linearity, its influence function is $(g' \otimes G')\psi^W(X_i)$. Therefore, Term 1 + Term 2 recovers $f_2(X_i) = f_1(X_i) + (g' \otimes G')\psi^W(X_i)$ evaluated at $W(\phi_0)$.

We substitute the evaluated limits A and J. By Slutsky's theorem,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -A^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_2(X_i) + J\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{\phi}(X_i)\right) + o_p(1).$$

The influence function for the Two-Step GMM estimator under general misspecification is

$$\psi^{2s}(X_i) = -A^{-1}\left(f_2(X_i) + J\psi^{\phi}(X_i)\right),$$

where it is mathematically understood, matching the main text, that A, J, and $f_2(X_i)$ are evaluated using the preliminary weight $W(\phi_0)$.

Iterated GMM

Note that $A^{-1}J$ in the proof above is the local updating function. At convergence, the Iterated GMM estimator $\widehat{\theta}$ is defined by the fixed point condition satisfying

$$F_n(\widehat{\theta}) = \widehat{G}(\widehat{\theta})'\widehat{W}(\widehat{\theta})\widehat{g}(\widehat{\theta}) = 0.$$

We apply the mean value theorem row-by-row around θ_0 yielding

$$0 = F_n(\theta_0) + F_{n,\theta}(\widetilde{\theta})(\widehat{\theta} - \theta_0).$$

We obtain the total derivative of $F_n(\theta)$ with respect to θ via the product rule

$$F_{n,\theta}(\theta) = \Big[\widehat{G}(\theta)'\widehat{W}(\theta)\widehat{G}(\theta) + \big(\widehat{g}(\theta)'\widehat{W}(\theta) \otimes I_p\big)\widehat{R}(\theta)\Big] + \left[(\widehat{g}(\theta)' \otimes \widehat{G}(\theta)')\frac{\partial\,\mathrm{vec}(\widehat{W}(\theta))}{\partial\theta'}\right].$$

Because $\phi_0 = \theta_0$ identically at the fixed point, the weight matrix limit is precisely $W = W(\theta_0)$. By the ULLN, consistency of $\widetilde{\theta}$, and Assumption 1(vi)(b), the limit of the first bracketed part evaluates at $W(\theta_0)$, yielding A. The limit of the second bracketed part evaluates at $W(\theta_0)$, yielding J. Thus,

$$F_{n,\theta}(\widetilde{\theta}) \xrightarrow{p} A + J.$$

The expansion of $\sqrt{n}F_n(\theta_0)$ follows the exact logic of the Two-Step expansion evaluated precisely at $W = W(\theta_0)$

$$\sqrt{n}F_n(\theta_0) = G'W\sqrt{n}(\widehat{g} - g) + \sqrt{n}(\widehat{G} - G)'Wg + G'\sqrt{n}(\widehat{W}(\theta_0) - W)g + \mathcal{R}_n,$$

where the remainder $\mathcal{R}_n = o_p(1)$. The influence function of this expansion exactly produces $f_2(X_i)$ evaluated at $W(\theta_0)$. Therefore, the total influence function is

$$\psi^{it}(X_i) = -(A + J)^{-1}f_2(X_i).$$

Q.E.D.

Proof of Proposition 2. Let $F_n(\theta, \phi) = \widehat{G}(\theta)'\widehat{W}(\phi)\widehat{g}(\theta)$ denote the sample FOC function for the second-step GMM estimator. By definition, the second-step estimator $\widehat{\theta}(\widehat{\phi})$ exactly satisfies the sample FOC evaluated at the first-step estimate $\widehat{\phi}$

$$F_n(\widehat{\theta}(\widehat{\phi}), \widehat{\phi}) = 0.$$

Under Assumption 1, the sample moments and their derivatives are continuously differentiable in a neighborhood of the estimates. Applying the Implicit Function Theorem to the FOC with respect to $\widehat{\phi}$ yields the finite-sample derivative

$$\frac{\partial\widehat{\theta}(\widehat{\phi})}{\partial\widehat{\phi}'} = -[F_{n,\theta}(\widehat{\theta},\widehat{\phi})]^{-1}F_{n,\phi}(\widehat{\theta},\widehat{\phi}),$$

where $F_{n,\theta}(\theta,\phi) = \frac{\partial F_n(\theta,\phi)}{\partial\theta'}$, $F_{n,\phi}(\theta,\phi) = \frac{\partial F_n(\theta,\phi)}{\partial\phi'}$, and $\widehat{\theta}$ is the second-step GMM estimator.

As established in the proof of Proposition 1, the uniform law of large numbers over the compact parameter space combined with the consistency of the preliminary and second-step estimators, $(\widehat{\theta},\widehat{\phi}) \xrightarrow{p} (\theta_0,\phi_0)$, ensures that the sample derivatives evaluated at the estimates converge in probability to their population counterparts evaluated at the pseudo-true values

$$F_{n,\theta}(\widehat{\theta},\widehat{\phi}) \xrightarrow{p} A \quad \text{and} \quad F_{n,\phi}(\widehat{\theta},\widehat{\phi}) \xrightarrow{p} J,$$

where $A = G'W(\phi_0)G + H$ is the population curvature matrix of the second-step objective, and $J = (g'\otimes G')S(\phi_0)$ captures the sensitivity of the optimal weight matrix to the first-step parameter.

By Assumption 1(iv), the population curvature matrix A is nonsingular. Applying the Continuous Mapping Theorem to the matrix inverse, we obtain the probability limit of the finite-sample sensitivity

$$\Lambda_\phi = \text{plim}\frac{\partial\widehat{\theta}(\widehat{\phi})}{\partial\widehat{\phi}'} = -A^{-1}J.$$

Q.E.D.

Proof of Proposition 3. Under misspecification, the effect of updating the weight matrix enters the influence function linearly through the matrix J. For the s-step procedure, the influence functions satisfy the recursion

$$\psi^{(s)}(X_i) = -A^{-1}f_2(X_i) - A^{-1}J\psi^{(s-1)}(X_i), \qquad s \geq 2, \tag{A2}$$

where $\psi^{(1)}(X_i) = \psi^{1s}(X_i)$. Iterating (A2) yields

$$\psi^{(s)}(X_i) = \sum_{k=0}^{s-2}(-A^{-1}J)^k(-A^{-1}f_2(X_i)) + (-A^{-1}J)^{s-1}\psi^{(1)}(X_i).$$

By Assumption 1(vi), the spectral radius of $A^{-1}J$ is less than one, so $(-A^{-1}J)^{s-1}\psi^{(1)}(X_i) \to 0$

and the finite sum converges to the Neumann series

$$\lim_{s\to\infty}\psi^{(s)}(X_i) = \left(I + A^{-1}J\right)^{-1}(-A^{-1}f_2(X_i)).$$

Finally,

$$\left(I + A^{-1}J\right)^{-1}(-A^{-1}) = -(A+J)^{-1},$$

which implies $\lim_{s\to\infty}\psi^{(s)}(X_i) = -(A+J)^{-1}f_2(X_i) = \psi^{it}(X_i)$. *Q.E.D.*

B Appendix: CUGMM

The Continuously Updating GMM (CUGMM) estimator treats the weight matrix as a function of the parameter and updates it jointly with θ. This parameter-dependent objective function introduces unique theoretical challenges under global misspecification. This appendix provides a formal treatment of the CUGMM estimator. We establish its consistency, derive its misspecification-robust influence function, and provide the explicit derivations for the Schennach (2007) example.

The CUGMM estimator is defined as the minimizer of the sample objective function

$$\widehat{Q}_{cue}(\theta) = \widehat{g}(\theta)'\widehat{\Omega}(\theta)^{-1}\widehat{g}(\theta),$$

where $\widehat{\Omega}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta)g(X_i,\theta)'$ is the uncentered sample second-moment matrix.

In practice, efficient GMM weight matrices are constructed using either the uncentered sample second-moment matrix $\widehat{\Omega}(\theta)$ or the centered sample covariance matrix $\widehat{V}(\theta) = \widehat{\Omega}(\theta) - \widehat{g}(\theta)\widehat{g}(\theta)'$. Under correct specification, this choice is asymptotically irrelevant. Under misspecification, however, they converge to different limits ($\Omega(\theta)$ vs. $V(\theta)$). As discussed in Section 3.1, this generally yields different pseudo-true parameters and local sensitivities for two-step or s-step GMM. However, the iterated and CUGMM estimator are invariant to this choice of centering. Let $\widehat{Q}_c(\theta) = \widehat{g}(\theta)'\widehat{V}(\theta)^{-1}\widehat{g}(\theta)$ denote the centered objective function. By applying the Sherman-Morrison formula to $\widehat{\Omega}(\theta) = \widehat{V}(\theta) + \widehat{g}(\theta)\widehat{g}(\theta)'$, we obtain

$$\widehat{Q}_{cue}(\theta) = \widehat{g}(\theta)'\left(\widehat{V}(\theta) + \widehat{g}(\theta)\widehat{g}(\theta)'\right)^{-1}\widehat{g}(\theta) = \frac{\widehat{g}(\theta)'\widehat{V}(\theta)^{-1}\widehat{g}(\theta)}{1+\widehat{g}(\theta)'\widehat{V}(\theta)^{-1}\widehat{g}(\theta)} = \frac{\widehat{Q}_c(\theta)}{1+\widehat{Q}_c(\theta)}.$$

Because the function $f(x) = x/(1+x)$ is increasing for $x \geq 0$, minimizing the uncentered objective is identical to minimizing the centered objective. An identical monotonic relationship holds for their population counterparts. Thus, the sample estimator $\widehat{\theta}_{cue}$, its unique population pseudo-true value θ_0, and its local sensitivity are invariant to the choice of centering. This allows us to evaluate the local curvature of CUGMM using the uncentered matrix $\Omega(\theta)$ without loss of generality.

To characterize the local behavior of CUGMM, we define the following higher-order population curvature matrices evaluated at θ_0. Let $W = \Omega(\theta_0)^{-1}$ and $v = Wg$. Furthermore, let $S =$

$\frac{\partial}{\partial \theta'}\text{vec}(\Omega(\theta))\big|_{\theta_0}$. We define

$$\begin{aligned} T &= \frac{\partial}{\partial \theta'}\text{vec}(S(\theta)')\Big|_{\theta_0}, \\ K &= ((v \otimes v)' \otimes I_p)\, T + 2S'(v \otimes I_q)\,(WG - (g'W \otimes W)S)\,. \end{aligned}$$

The misspecification-robust population curvature matrix for CUGMM is defined as $A_{cu} = 2(A + J) - K$, where A and J are the population curvature matrices of iterated GMM evaluated at W, as defined in Section 3.1.

We impose the following regularity conditions

Assumption 2. *(i) There exists a unique θ_0 in the interior of Θ such that for any $\epsilon > 0$, $\inf_{\theta\in\Theta:\|\theta-\theta_0\|\geq\epsilon} Q_{cue}(\theta) > Q_{cue}(\theta_0)$, where $Q_{cue}(\theta) = \mathbb{E}[g(X_i,\theta)]'\Omega(\theta)^{-1}\mathbb{E}[g(X_i,\theta)]$.*

(ii) The population curvature matrix A_{cu} is nonsingular.

(iii) For the population centered covariance matrix $V(\theta) = \Omega(\theta) - \mathbb{E}[g(X_i,\theta)]\mathbb{E}[g(X_i,\theta)]'$, there exists a constant $c > 0$ such that $\inf_{\theta\in\Theta} \lambda_{\min}(V(\theta)) \geq c$.

Assumption 2(iii) precludes the variance collapse highlighted by Kleibergen and Zhan (2025). By ensuring uniform non-singularity, we guarantee that the estimator does not spuriously minimize the objective by pushing the parameter to the boundary and inflating the variance determinant, while Assumption 2(i) ensures identification strength dominates the degree of misspecification.

Proposition 4 (Consistency of CUGMM under Misspecification)**.** *Suppose Assumption 1 and Assumption 2 hold. Then, the CUGMM estimator converges in probability to the unique pseudo-true value $\widehat{\theta}_{cue} \xrightarrow{p} \theta_0$.*

Proof of Proposition 4. First, we show that Assumption 2(i) implies θ_0 is the unique minimizer of the population objective $Q_{cue}(\theta)$. Suppose, for a contradiction, there exists another minimizer $\theta_1 \neq \theta_0$. Then $Q_{cue}(\theta_1) \leq Q_{cue}(\theta_0)$. Let $\epsilon = \|\theta_1 - \theta_0\| > 0$. By strong pseudo-identification, $\inf_{\|\theta-\theta_0\|\geq\epsilon} Q_{cue}(\theta) > Q_{cue}(\theta_0)$. Since $\|\theta_1 - \theta_0\| \geq \epsilon$, it follows that $Q_{cue}(\theta_1) \geq \inf_{\|\theta-\theta_0\|\geq\epsilon} Q_{cue}(\theta) > Q_{cue}(\theta_0)$, which contradicts $Q_{cue}(\theta_1) \leq Q_{cue}(\theta_0)$. Thus, θ_0 is unique.

By the uniform integrability conditions in Assumption 1, the Uniform Law of Large Numbers (ULLN) applies to the sample moments and the uncentered sample covariance matrix over the compact set Θ:

$$\sup_{\theta\in\Theta} \|\widehat{g}(\theta) - \mu(\theta)\| \xrightarrow{p} 0, \quad \text{and} \quad \sup_{\theta\in\Theta} \|\widehat{\Omega}(\theta) - \Omega(\theta)\| \xrightarrow{p} 0,$$

where $\mu(\theta) = \mathbb{E}[g(X_i, \theta)]$.

Because $\Omega(\theta) = V(\theta) + \mu(\theta)\mu(\theta)'$, we have $\Omega(\theta) \geq V(\theta)$ in the positive semi-definite sense. Therefore, Assumption 2(iii) implies $\inf_{\theta\in\Theta} \lambda_{\min}(\Omega(\theta)) \geq c > 0$ holds as well.

Let $\mathcal{M}_c$ be the space of symmetric matrices with eigenvalues bounded below by $c/2$. With probability approaching one, $\widehat{\Omega}(\theta) \in \mathcal{M}_c$ for all $\theta \in \Theta$. Because the matrix inversion mapping is uniformly continuous on the compact space $\mathcal{M}_c$, the continuous mapping theorem yields

$$\sup_{\theta\in\Theta} \|\widehat{\Omega}(\theta)^{-1} - \Omega(\theta)^{-1}\| \xrightarrow{p} 0.$$

We bound the difference between the sample and population objective functions using the triangle inequality and sub-multiplicativity of norms

$$\begin{aligned}|\widehat{Q}_{cue}(\theta) - Q_{cue}(\theta)| \leq\ & \|\widehat{g}(\theta) - \mu(\theta)\|^2 \|\widehat{\Omega}(\theta)^{-1}\| \\ & + 2\|\mu(\theta)\| \|\widehat{\Omega}(\theta)^{-1}\| \|\widehat{g}(\theta) - \mu(\theta)\| \\ & + \|\mu(\theta)\|^2 \|\widehat{\Omega}(\theta)^{-1} - \Omega(\theta)^{-1}\|.\end{aligned}$$

Since $\mu(\theta)$ and $\Omega(\theta)^{-1}$ are uniformly bounded over Θ, each term converges to zero uniformly in probability. Thus, $\sup_{\theta\in\Theta} |\widehat{Q}_{cue}(\theta) - Q_{cue}(\theta)| \xrightarrow{p} 0$.

By definition, the estimator satisfies $\widehat{Q}_{cue}(\widehat{\theta}_{cue}) \leq \widehat{Q}_{cue}(\theta_0)$. Using the uniform convergence established above, it follows that $Q_{cue}(\widehat{\theta}_{cue}) \leq Q_{cue}(\theta_0) + o_p(1)$. Finally, for any $\epsilon > 0$, there exists an $\eta > 0$ such that $\inf_{\|\theta-\theta_0\|\geq\epsilon} Q_{cue}(\theta) > Q_{cue}(\theta_0) + \eta$. The event $\|\widehat{\theta}_{cue} - \theta_0\| \geq \epsilon$ implies $\{Q_{cue}(\widehat{\theta}_{cue}) > Q_{cue}(\theta_0) + \eta\}$. However, since $Q_{cue}(\widehat{\theta}_{cue}) \leq Q_{cue}(\theta_0) + o_p(1)$, the probability of the latter converges to zero. Therefore, $\lim_{n\to\infty} \mathbb{P}(\|\widehat{\theta}_{cue} - \theta_0\| \geq \epsilon) = 0$, establishing that $\widehat{\theta}_{cue} \xrightarrow{p} \theta_0$. *Q.E.D.*

Proposition 5 (Influence Function for CUGMM)**.** *Suppose the conditions of Proposition 4 hold. The influence function for the CUGMM estimator under misspecification is given by*

$$\psi^{cu}(X_i) = -A_{cu}^{-1} f_{cu}(X_i),$$

where $f_{cu}(X_i) = f_1(X_i) - f_2(X_i)$*. Letting* $v = Wg$*,* $\psi_\Omega(X_i) = g(X_i, \theta_0)g(X_i, \theta_0)' - \Omega(\theta_0)$*, and* $\widetilde{\psi}_v(X_i) = W\nu(X_i) - W\psi_\Omega(X_i)v$*, the components are defined as*

$$\begin{aligned}f_1(X_i) &= 2(G(X_i, \theta_0) - G)'v + 2G'\widetilde{\psi}_v(X_i), \\ f_2(X_i) &= \widetilde{\psi}_{Sv}(X_i) + S'\left(v \otimes \widetilde{\psi}_v(X_i) + \widetilde{\psi}_v(X_i) \otimes v\right),\end{aligned}$$

with $\widetilde{\psi}_{Sv}(X_i) = 2(g(X_i,\theta_0)'v)G(X_i,\theta_0)'v - S'(v \otimes v)$.

Proof of Proposition 5. The CUGMM sample FOC is obtained by taking the total derivative of the objective function. By the chain rule and the identity $\frac{\partial \text{vec}(\widehat{W}(\theta))}{\partial \theta'} = -(\widehat{W}(\theta) \otimes \widehat{W}(\theta))\widehat{S}(\theta)$, the FOC is

$$\begin{aligned} F_n(\theta) &= 2\widehat{G}(\theta)'\widehat{W}(\theta)\widehat{g}(\theta) - \widehat{S}(\theta)'(\widehat{W}(\theta) \otimes \widehat{W}(\theta))'(\widehat{g}(\theta) \otimes \widehat{g}(\theta)) \\ &= 2\widehat{G}(\theta)'\widehat{W}(\theta)\widehat{g}(\theta) - \widehat{S}(\theta)'(\widehat{W}(\theta)\widehat{g}(\theta) \otimes \widehat{W}(\theta)\widehat{g}(\theta)). \end{aligned}$$

Letting $\widehat{v}(\theta) = \widehat{W}(\theta)\widehat{g}(\theta)$, the estimator satisfies $F_n(\widehat{\theta}_{cue}) = 0$. At θ_0, the population FOC is $F(\theta_0) = 2G'v - S'(v \otimes v) = 0$.

Applying the Mean Value Theorem around θ_0 gives $0 = F_n(\theta_0) + F_{n,\theta}(\widetilde{\theta})(\widehat{\theta}_{cue} - \theta_0)$. By the continuous mapping theorem and the consistency of $\widetilde{\theta}$, the Hessian converges to the CUGMM curvature matrix defined earlier: $F_{n,\theta}(\widetilde{\theta}) \xrightarrow{p} A_{cu}$.

Next, we expand $\sqrt{n}F_n(\theta_0)$ around population limits G, v, and S. Let $\widehat{G} = G + (\widehat{G} - G)$, $\widehat{v} = v + (\widehat{v} - v)$, and $\widehat{S} = S + (\widehat{S} - S)$. Substituting these into the sample FOC and collecting first-order terms yields

$$\begin{aligned} F_n(\theta_0) = 2&\left[(\widehat{G} - G)'v + G'(\widehat{v} - v)\right] \\ &- \left[(\widehat{S} - S)'(v \otimes v) + S'\left(v \otimes (\widehat{v} - v) + (\widehat{v} - v) \otimes v\right)\right] + O_p(n^{-1}). \end{aligned}$$

We seek the influence function $f_{cu}(X_i)$ such that $\sqrt{n}F_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_i f_{cu}(X_i) + o_p(1)$. We define $f_{cu}(X_i) = f_1(X_i) - f_2(X_i)$ corresponding to the two bracketed terms.

To find $f_{cu}(X_i)$, we determine the influence function for the composite vector $\widehat{v}$. Since $\sqrt{n}(\widehat{v} - v) = W\sqrt{n}(\widehat{g} - g) + \sqrt{n}(\widehat{W} - W)g + o_p(1)$, and by the delta method $\sqrt{n}(\widehat{W} - W)$ has influence function $-W\psi_\Omega(X_i)W$, the influence function for $\sqrt{n}(\widehat{v} - v)$ is $\widetilde{\psi}_v(X_i) = W\nu(X_i) - W\psi_\Omega(X_i)v$. Substituting $\widetilde{\psi}_v(X_i)$ into the first bracket yields $f_1(X_i) = 2(G(X_i,\theta_0) - G)'v + 2G'\widetilde{\psi}_v(X_i)$.

For the second bracket, the term $\sqrt{n}(\widehat{S} - S)'(v \otimes v)$ requires the influence function of $\widehat{S}$, evaluated in projection with $v \otimes v$. Let $M_i(\theta_0) = \frac{\partial \text{vec}(g(X_i,\theta)g(X_i,\theta)')}{\partial \theta'}\Big|_{\theta_0}$, such that $\widehat{S} = \frac{1}{n}\sum_i M_i(\theta_0)$. Using the product rule, $M_i = [(g_i \otimes I_q) + (I_q \otimes g_i)]G_i$, where $g_i = g(X_i,\theta_0)$ and $G_i = G(X_i,\theta_0)$. Post-multiplying by $(v \otimes v)$ gives

$$M_i'(v \otimes v) = 2(g_i'v)G_i'v.$$

Demeaning this gives the influence function $\widetilde{\psi}_{Sv}(X_i) = 2(g_i'v)G_i'v - S'(v \otimes v)$. Combining these components yields $f_2(X_i) = \widetilde{\psi}_{Sv}(X_i) + S' \left(v \otimes \widetilde{\psi}_v(X_i) + \widetilde{\psi}_v(X_i) \otimes v \right)$.

Collecting all terms, the influence function of the CUGMM estimator is $\psi^{cu}(X_i) = -A_{cu}^{-1}(f_1(X_i) - f_2(X_i))$. *Q.E.D.*

C Appendix: Derivation of Examples

C.1 Derivation of Example 2

We derive the influence function for the linear IV model with estimated weight matrix $\widehat{W} = (n^{-1}Z'Z)^{-1}$. The moment is $g(X_i, \theta) = Z_i(Y_i - D_i\theta)$, with Jacobian $G(X_i) = -Z_iD_i$ and curvature $A = G'WG$. From Proposition 1, we have

$$f_{1s}(X_i) = G'Wg(X_i, \theta_0) + G(X_i)'Wg.$$

We decompose the sample quantities into their expectations and influence functions: $g(X_i, \theta_0) = g + \nu(X_i)$ and $G(X_i) = G + \widetilde{G}(X_i)$. Substituting these into the expression,

$$\begin{aligned} f_{1s}(X_i) &= G'W(g + \nu(X_i)) + (G + \widetilde{G}(X_i))'Wg \\ &= G'W\nu(X_i) + \widetilde{G}(X_i)'Wg. \end{aligned}$$

The terms $G'Wg$ vanish because the pseudo-true value θ_0 satisfies the population FOC. To separate the sensitivity coefficient from the data in the term $\widetilde{G}(X_i)'Wg$, we use the vectorization identity to get

$$\widetilde{G}(X_i)'Wg = \text{vec}(\widetilde{G}(X_i)'Wg) = (g'W \otimes I_p)\,\text{vec}(\widetilde{G}(X_i)').$$

Proposition 1(ii) defines the weight estimation effect as

$$f^W(X_i) = (g' \otimes G')\,\text{vec}(\psi_W(X_i))$$

We first determine $\psi_W(X_i)$. Let $\Omega = \mathbb{E}[Z_iZ_i']$ so that $W = \Omega^{-1}$. The influence function for Ω is $Z_iZ_i' - \Omega$. By the delta method for the inverse matrix, the influence function for the weight matrix is $\psi_W(X_i) = -W(Z_iZ_i' - \Omega)W$.

Apply the vectorization identity again to get

$$\begin{aligned} f^W(X_i) &= \text{vec}\left(G'\psi_W(X_i)g\right) \\ &= G'\left[-W(Z_iZ_i' - \Omega)W\right]g \\ &= -G'WZ_iZ_i'Wg + G'W\Omega Wg \end{aligned}$$

Since $\Omega = W^{-1}$, the second term becomes $G'Wg$, which is zero by the population FOC. Thus,

$$f^W(X_i) = -G'W(Z_iZ_i')Wg$$

We combine the components into the full influence function

$$\psi(X_i) = -A^{-1}\left[G'W\nu(X_i) + (g'W \otimes I_p)\operatorname{vec}(\widetilde{G}(X_i)') - G'W(Z_iZ_i')Wg\right]$$

We simplify this notation by separating the deterministic parameters from the stochastic data vectors

$$\psi(X_i) = \Lambda_{AGS}\nu(X_i) + \Lambda_{Bias}\operatorname{vec}(\widetilde{G}(X_i)') + \Lambda_{Weight}\operatorname{vec}(Z_iZ_i')$$

where

$$\begin{aligned}
\Lambda_{AGS} &= -(G'WG)^{-1}G'W \\
\Lambda_{Bias} &= -(G'WG)^{-1}(g'W \otimes I_p) \\
\Lambda_{Weight} &= (G'WG)^{-1}\left[(g'W) \otimes (G'W)\right].
\end{aligned}$$

C.2 Derivation of Example 3

This appendix details the derivations for Example 3. We consider the moment function $g(X_i, \theta) = (X_i - \theta, (X_i - \theta)^2 - 1)'$, assuming the data X_i are i.i.d. draws from a normal distribution $N(\mu, \sigma^2)$. The parameter of interest is the mean θ, while the variance is constrained to one. Let $U_i = X_i - \mu$ denote the demeaned data, $\delta = \mu - \theta$ the parameter bias, and $a = \sigma^2 - 1$ the degree of misspecification. The model is correctly specified if $\sigma^2 = 1$ (i.e., $a = 0$).

We first establish the population quantities required by Proposition 1. The expected moment vector is $g = [0, a]'$, and the Jacobian is $G = [-1, 0]'$. The population Hessian of the moments is $R = \mathbb{E}[\partial \operatorname{vec}(G(X_i, \theta)')/\partial\theta']|_\mu = [0, 2]'$. The covariance matrix of the moments is $\Omega(\mu) = \operatorname{diag}(\sigma^2, \tau)$, where $\tau = \mathbb{E}[(U_i^2 - 1)^2] = 3\sigma^4 - 2\sigma^2 + 1$. The efficient weight matrix is $W = \operatorname{diag}(1/\sigma^2, 1/\tau)$. We also define $S = \partial \operatorname{vec}(\Omega(\theta))/\partial\theta'|_\mu$. Noting that $\partial/\partial\theta = -\partial/\partial\delta$, we calculate the derivatives of the covariance terms at μ, yielding $S = [0, s_{12}, s_{12}, 0]'$ with $s_{12} = 1 - 3\sigma^2$.

One-Step GMM (W=I)

The One-Step GMM estimator minimizes the objective function $Q_I(\theta) = g(\theta)'g(\theta) = \delta^2 + (a + \delta^2)^2$.

The first-order condition (FOC) with respect to δ is:

$$\frac{\partial Q_I}{\partial \delta} = 2\delta + 4\delta(a + \delta^2) = 2\delta(1 + 2a + 2\delta^2) = 0.$$

Substituting $a = \sigma^2 - 1$, the condition becomes $2\delta(2\sigma^2 - 1 + 2\delta^2) = 0$. Therefore,

$$\theta_0 = \begin{cases} \mu, & \text{if } \sigma^2 \geq \frac{1}{2}, \\ \mu \pm \sqrt{\frac{1}{2} - \sigma^2}, & \text{if } \sigma^2 < \frac{1}{2}. \end{cases}$$

We analyze the sensitivity and informativeness at $\theta_0 = \mu$ under the condition $\sigma^2 \geq 1/2$. The curvature components at $\theta_0 = \mu$ are $H = 2a$ and $A = G'IG + H = 2\sigma^2 - 1$. Applying Proposition 1(i), the influence function is $\psi^{1s}(X_i) = -A^{-1}(G'Ig(X_i, \mu) + G(X_i, \mu)'Ig) = U_i$. It follows that $\Lambda^{1s} = [1, 0]$ and $\Delta^{1s} = 1$.

Two-Step GMM

We assume the first-step estimator is the One-Step GMM estimator, which requires $\sigma^2 \geq 1/2$ to converge to μ. The second-step estimator minimizes $Q_{2S}(\theta) = g(\theta)'W(\mu)g(\theta) = \delta^2/\sigma^2 + (a+\delta^2)^2/\tau$. The FOC is

$$\frac{\partial Q_{2S}}{\partial \delta} = 2\delta \left(\frac{1}{\sigma^2} + \frac{2(a + \delta^2)}{\tau} \right) = 0.$$

Evaluating the term in the parenthesis at $\delta = 0$ yields $A = 1/\sigma^2 + 2a/\tau$. Under normality, $A = (5\sigma^4 - 4\sigma^2 + 1)/(\sigma^2\tau)$, which is positive for all real σ^2. Since $A > 0$ and $\delta^2/\tau \geq 0$, the term in the parenthesis is positive, ensuring $\delta = 0$ is the unique solution. Thus, the pseudo-true value is $\theta_0 = \mu$.

We derive the influence function using the common curvature matrices for efficient GMM: $H = 2a/\tau$, $J = as_{12}/(\sigma^2\tau)$, and $A_{it} = 2\sigma^2/\tau$. Using Proposition 1(iii) with first-step influence function $\psi^{\phi}(X_i) = U_i$, we obtain

$$\psi^{2s}(X_i) = U_i - A^{-1} \frac{a}{\sigma^2 \tau} Z_i,$$

where $Z_i = U_i^3 - 3\sigma^2 U_i$. Since $\mathbb{E}[U_i Z_i] = 0$, the sensitivity remains $\Lambda^{2s} = [1, 0]$. However, the variance increases, leading to $\Delta^{2s} < 1$ whenever $a \neq 0$.

Iterated GMM

The Iterated GMM estimator is defined by the fixed point condition $G(\theta)'W(\theta)g(\theta) = 0$. At $\theta = \mu$, we have $[-1, 0] \cdot \text{diag}(1/\sigma^2, 1/\tau) \cdot [0, a]' = 0$, so μ satisfies the condition. For convergence and uniqueness locally, we verify the contraction mapping condition $|A^{-1}J| < 1$. Under normality, this inequality takes the form:

$$\left|\frac{a(1-3\sigma^2)}{\sigma^2\tau}\right| < \frac{1}{\sigma^2} + \frac{2a}{\tau} \quad \Longleftrightarrow \quad |-3\sigma^4 + 4\sigma^2 - 1| < 5\sigma^4 - 4\sigma^2 + 1.$$

This holds for all $\sigma^2 > 0$. Thus, $\theta_0 = \mu$ is the unique fixed point.

Using Proposition 1(iv), the influence function is $\psi^{it}(X_i) = -A_{it}^{-1} f_{2s}(X_i)$, which simplifies to:

$$\psi^{it}(X_i) = \frac{U_i}{2\sigma^4}\left[5\sigma^4 - 3\sigma^2 - (\sigma^2 - 1)U_i^2\right].$$

We verify $\mathbb{E}[\psi^{it}U_i] = \sigma^2$, yielding $\Lambda^{it} = [1, 0]$. The informativeness is $\Delta^{it} = 2\sigma^4/(5\sigma^4 - 6\sigma^2 + 3)$, which is less than 1 under misspecification.

Continuously Updating GMM

The CUE estimator minimizes $Q_{cue}(\theta) = g(\theta)'\Omega(\theta)^{-1}g(\theta)$. Note that $\Omega(\theta) = V(\theta) + g(\theta)g(\theta)'$, where $V(\theta) = \text{Var}(g(X_i, \theta))$. Substituting in the objective function to get $Q_{cue}(\theta) = g(\theta)'(V(\theta) + g(\theta)g(\theta)')^{-1}g(\theta)$.

Using the Sherman-Morrison formula,

$$\begin{aligned}
Q_{cue} &= g'\left(V^{-1} - \frac{V^{-1}gg'V^{-1}}{1 + g'V^{-1}g}\right)g \\
&= g'V^{-1}g - \frac{g'V^{-1}gg'V^{-1}g}{1 + g'V^{-1}g} \\
&= Q_{Var} - \frac{Q_{Var}^2}{1 + Q_{Var}} \\
&= \frac{Q_{Var}}{1 + Q_{Var}}.
\end{aligned}$$

Minimizing $Q_{cue}(\theta)$ is equivalent to minimizing $Q_{Var}(\theta) = g(\theta)'V(\theta)^{-1}g(\theta)$, as $Q_{cue} = Q_{Var}/(1 + Q_{Var})$ is a monotonically increasing function of Q_{Var}.

We calculate $V(\theta)$ explicitly as a function of δ. Utilizing the properties of the normal distribution, we derived:

$$V(\theta) = \begin{pmatrix} \sigma^2 & 2\delta\sigma^2 \\ 2\delta\sigma^2 & 2\sigma^4 + 4\delta^2\sigma^2 \end{pmatrix}.$$

The determinant is $|V(\theta)| = 2\sigma^6$, which is constant and independent of θ. Substituting $g(\theta) = [\delta, a + \delta^2]'$ and computing the quadratic form yields:

$$Q_{Var}(\delta) = \frac{1}{2\sigma^6} \left[\sigma^2\delta^4 + 2\sigma^2\delta^2(1 - a + \sigma^2) + \sigma^2 a^2 \right].$$

Using $a = \sigma^2 - 1$, the expression simplifies to $Q_{Var}(\delta) = \frac{1}{2\sigma^4}(\delta^4 + 2\delta^2 + a^2)$. This function is convex in δ^2 and achieves a unique global minimum at $\delta = 0$ for any $\sigma^2 > 0$. Thus, the pseudo-true value is uniquely $\theta_0 = \mu$, provided $\sigma^2 > 0$.

The curvature matrix is $A_{cue} = 2A_{it} - (H_S + K_V)$. Calculating the derivatives of the weight matrix terms under normality yields $A_{cue} = 8\sigma^4/\tau^2$, which is positive, confirming a local minimum. The influence function derived via Proposition 1(v) is

$$\psi^{cue}(X_i) = U_i - \frac{a}{2\sigma^2} Z_i.$$

Therefore, $\Lambda^{cue} = [1, 0]$. The informativeness is $\Delta^{cue} = (1 + \frac{3}{2}(\sigma^2 - 1)^2)^{-1}$, which is less than 1 when $a \neq 0$.

C.3 Example 4

Example 4 (Schennach Cont'd)**.** *In Example 3, we have shown that different choices of weight matrix can lead to different informativeness under misspecification. In this example, we consider the same weight matrix but illustrate how the functional form of moments can affect informativeness.*

For simplicity, we assume that the data X_i is i.i.d. from $N(0, \sigma^2)$. Two researchers consider One-Step GMM with identity weight matrix $W = I$ to estimate the mean θ. Two researchers incorrectly assume $\sigma^2 = 1$ and both models are misspecified whenever $\sigma^2 \neq 1$. Due to the symmetry of the Normal distribution, the pseudo-true value is $\theta_0 = 0$ in both cases.

The first researcher uses the moment function in Example 3

$$g_1(X_i, \theta) = \begin{pmatrix} X_i - \theta \\ (X_i - \theta)^2 - 1 \end{pmatrix}.$$

The influence function is

$$\psi_1(X_i) = X_i,$$

which is perfectly correlated with the first moment. Consequently, $\Delta_1 = 1$.

The second researcher uses the moment function utilizing the kurtosis of the Normal distribution

$$g_2(X_i, \theta) = \begin{pmatrix} X_i - \theta \\ (X_i - \theta)^4 - 3 \end{pmatrix}.$$

As derived in Appendix C.4, the influence function is

$$\psi_2(X_i) = \frac{X_i + 4bX_i^3}{1 + 12b\sigma^2}$$

where $b = 3(\sigma^4 - 1)$ is the degree of misspecification. While X_i^3 is correlated with the moment X_i, it is not perfectly spanned by the moment, which reduces the informativeness of the estimator. Specifically,

$$\Delta_2 = \frac{(1 + 12b\sigma^2)^2}{1 + 24b\sigma^2 + 240b^2\sigma^4} < 1,$$

where $b = \sigma^2 - 1$ is the degree of misspecification. For example, if $\sigma^2 = 2$, $\Delta_2 \approx 0.6$. Moreover, let $V_1 = \mathrm{Var}(\widetilde{\psi}_1(X_i))$ and $V_2 = \mathrm{Var}(\widetilde{\psi}_2(X_i))$ be the asymptotic variances of the estimators for the *first and second researchers, respectively. Then*

$$\Delta_2 = \frac{V_1}{V_2}$$

This shows that informativeness Δ_2 is exactly the ratio of the efficient variance V_1 to the inflated variance V_2.

C.4 Derivation of Example 4

We consider the estimation of the mean from i.i.d. data $X_i \sim N(0, \sigma^2)$. Two researchers utilize One-Step GMM with identity weight matrix and misspecified moments due to the incorrect assumption that $X_i \sim N(\theta, 1)$. The first researcher uses the moment function $g_1(X_i, \theta) = (X_i - \theta, (X_i - \theta)^2 - 1)'$ and the second researcher uses the moment function $g_2(X_i, \theta) = (X_i - \theta, (X_i - \theta)^4 - 3)'$. In both cases, the pseudo-true value is $\theta_0 = 0$.

For the first researcher, at $\theta_0 = 0$, the population moment is:

$$g_1 = \mathbb{E}[g_1(X_i, 0)] = \begin{pmatrix} 0 \\ \sigma^2 - 1 \end{pmatrix}.$$

Let $a = \sigma^2 - 1$. The Jacobian is $G_1(X_i, \theta) = (-1, -2(X_i - \theta))'$. The expected Jacobian at $\theta_0 = 0$ is:

$$G_1 = \mathbb{E}[G_1(X_i, 0)] = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

The curvature matrix is $A_1 = G_1'G_1 + (g_1' \otimes I_1)R_1$. The Hessian of the second moment is $\mathbb{E}[\nabla_\theta(-2(X_i - \theta))] = 2$. Thus $H_1 = a(2) = 2a$.

$$A_1 = (-1)^2 + 2a = 1 + 2a.$$

From Proposition 1(i), the influence function is

$$\psi_1(X_i) = -A_1^{-1} f_{1s,1}(X_i),$$

where $f_{1s,1}(X_i) = G_1' g_1(X_i, 0) + G_1(X_i, 0)' g_1$, and

$$G_1' g_1(X_i, 0) = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} X_i \\ X_i^2 - 1 \end{pmatrix} = (-1)(X_i) + (0)(X_i^2 - 1) = -X_i,$$

$$G_1(X_i,0)'g_1 = \begin{pmatrix} -1 & -2X_i \end{pmatrix} \begin{pmatrix} 0 \\ a \end{pmatrix} = (-1)(0) + (-2X_i)(a) = -2aX_i.$$

Summing these terms yields $f_{1s,1}(X_i) = -X_i - 2aX_i = -(1+2a)X_i$. Substituting into the influence function formula

$$\psi_1(X_i) = -\frac{1}{1+2a}\left(-(1+2a)X_i\right) = X_i.$$

The asymptotic variance is $V_1 = \mathbb{E}[\psi_1(X_i)^2] = \mathbb{E}[X_i^2] = \sigma^2$. Since $\psi_1(X_i)$ is perfectly correlated with the first element of $\nu(X_i)$, $\Delta_1 = 1$.

For the second researcher, at $\theta_0 = 0$,

$$g_2 = \mathbb{E}[g_2(X_i,0)] = \begin{pmatrix} 0 \\ 3\sigma^4 - 3 \end{pmatrix}.$$

Let $b = 3(\sigma^4 - 1)$. The Jacobian is $G_2(X_i,\theta) = (-1, -4(X_i-\theta)^3)'$. The expected Jacobian is:

$$G_2 = \mathbb{E}[G_2(X_i,0)] = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

The curvature matrix A_2 involves the expected Hessian of the second moment, $\mathbb{E}[12X_i^2] = 12\sigma^2$. Thus, $A_2 = 1 + b(12\sigma^2) = 1 + 12b\sigma^2$.

$$G_2'g_2(X_i,0) = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} X_i \\ X_i^4 - 3 \end{pmatrix} = -X_i.$$

$$G_2(X_i,0)'g_2 = \begin{pmatrix} -1 & -4X_i^3 \end{pmatrix} \begin{pmatrix} 0 \\ b \end{pmatrix} = -4bX_i^3.$$

The influence function is

$$\psi_2(X_i) = -A_2^{-1}(-X_i - 4bX_i^3) = \frac{1}{A_2}(X_i + 4bX_i^3).$$

To calculate the variance $V_2 = \mathbb{E}[\psi_2^2]$, we utilize the higher moments of the normal distribution ($\mathbb{E}[X^4] = 3\sigma^4, \mathbb{E}[X^6] = 15\sigma^6$),

$$V_2 = \frac{1}{A_2^2}\left(\sigma^2 + 24b\sigma^4 + 240b^2\sigma^6\right).$$

Under symmetry, ψ_2, is uncorrelated with the residuals of the second moment condition $X_i^4 - 3$.

Therefore, the informativeness Δ_2 depends only on the correlation with the first moment X_i. Using Definition 2

$$\Delta_2 = \frac{(\mathbb{E}[\psi_2 X_i])^2}{\mathbb{E}[\psi_2^2]\mathbb{E}[X_i^2]} = \frac{(\mathbb{E}[\psi_2 X_i])^2}{V_2 \sigma^2}.$$

The covariance is calculated as

$$\mathbb{E}[\psi_2 X_i] = \frac{1}{A_2}(\mathbb{E}[X_i^2] + 4b\mathbb{E}[X_i^4]) = \frac{\sigma^2 + 12b\sigma^4}{1 + 12b\sigma^2} = \sigma^2.$$

Substituting this back into the expression for Δ_2

$$\Delta_2 = \frac{(\sigma^2)^2}{V_2 \sigma^2} = \frac{\sigma^2}{V_2} = \frac{V_1}{V_2}.$$

This shows that informativeness Δ_2 is exactly the ratio of the efficient variance V_1 to the inflated variance V_2.

D Appendix: Extra Results

D.1 A Geometric Framework

This appendix provides a rigorous geometric foundation for the sensitivity and informativeness measures presented in the main text. By treating the statistical model as a Riemannian manifold, we show that the concepts of sensitivity and informativeness arise naturally from the geometric structure of the tangent space. This framework unifies the asymptotic distribution perspective and the local perturbation perspective through the language of information geometry.

Let $\mathcal{P}$ denote a statistical model, defined as a family of probability measures P on a sample space $\mathcal{X}$. We assume $\mathcal{P}$ possesses the structure of a smooth Banach manifold. The local behavior of functionals on this manifold is characterized by the tangent space. At any point $P \in \mathcal{P}$, the tangent space $T_P\mathcal{P}$ is the vector space representing all possible infinitesimal deformations of the model P.

An element $v \in T_P\mathcal{P}$, referred to as a score vector, describes a local direction of deviation. Formally, consider a smooth one-dimensional submodel $\{P_t : t \in (-\epsilon, \epsilon)\} \subset \mathcal{P}$ passing through P at $t = 0$. The tangent vector v associated with this path is the derivative of the log-likelihood:

$$v(x) = \frac{d}{dt} \log(dP_t(x))\Big|_{t=0}.$$

The tangent space $T_P\mathcal{P}$ is a subspace of $L_0^2(P)$, the Hilbert space of square-integrable functions with mean zero with respect to P.

We consider differentiable statistical functionals defined on $\mathcal{P}$. Let $\psi : \mathcal{P} \to \mathbb{R}$ be a scalar target parameter of interest, and let $\nu : \mathcal{P} \to \mathbb{R}$ represent a control functional, such as a moment condition. The local response of a functional to perturbations in the model is captured by its differential. The differential of ψ at P, denoted $d\psi_P$, is a continuous linear functional on the tangent space $T_P\mathcal{P}$ defined by the Fréchet derivative:

$$d\psi_P(v) = \lim_{t \to 0} \frac{\psi(P_t) - \psi(P)}{t}.$$

To define geometric properties such as lengths and angles, we endow the manifold $\mathcal{P}$ with a Riemannian metric g. The metric $g_P(\cdot, \cdot)$ is an inner product on the tangent space $T_P\mathcal{P}$ that varies smoothly with P. This metric defines the squared norm of a tangent vector v as $\|v\|_g^2 = g_P(v, v)$.

The geometric structure allows us to represent the differential of a functional as a vector in the tangent space. By the Riesz Representation Theorem, for a given metric g, there exists a unique vector $\nabla^g \psi_P \in T_P\mathcal{P}$, called the gradient of ψ, such that for all $v \in T_P\mathcal{P}$:

$$d\psi_P(v) = g_P(\nabla^g \psi_P, v).$$

The gradient $\nabla^g \psi_P$ points in the direction of the steepest ascent of the functional ψ with respect to the distance defined by the metric g.

Here, we define sensitivity and informativeness purely in terms of the geometric relationship between the gradients of the target parameter and the control functional. The sensitivity of ψ to ν is defined as the directional derivative of ψ along the direction of the gradient of ν. Using the definition of the gradient, this is equivalent to the inner product of the two gradients:

$$\partial_\nu \psi(P) := d\psi_P(\nabla^g \nu_P) = g_P(\nabla^g \psi_P, \nabla^g \nu_P).$$

We define the geometric sensitivity coefficient, denoted $S(\psi, \nu)_P$, as the projection of the target gradient onto the control gradient, normalized by the magnitude of the control gradient. This is analogous to a regression coefficient in the tangent space:

$$S(\psi, \nu)_P := \frac{g_P(\nabla^g \psi_P, \nabla^g \nu_P)}{\|\nabla^g \nu_P\|_g^2}.$$

Complementarily, we define the geometric informativeness, denoted $R(\psi, \nu)_P$, as the squared correlation between the gradients. This measure captures the extent to which the local variation in ψ is explained by the local variation in ν:

$$R(\psi, \nu)_P := \frac{[g_P(\nabla^g \psi_P, \nabla^g \nu_P)]^2}{\|\nabla^g \psi_P\|_g^2 \cdot \|\nabla^g \nu_P\|_g^2}.$$

Geometrically, $R(\psi, \nu)_P = \cos^2(\theta)$, where θ is the angle between the vectors $\nabla^g \psi_P$ and $\nabla^g \nu_P$ in the tangent space. A value of 1 implies that the gradients are collinear, meaning the parameter ψ is locally fully determined by ν.

The abstract geometric definitions connect to standard asymptotic theory through the Information Metric. The Information Metric, denoted I, is defined by the covariance of the score vectors, which corresponds to the inner product in $L^2(P)$:

$$I_P(v, w) = \mathbb{E}_P[v(X)w(X)].$$

A fundamental result in semiparametric theory links the information geometry to the influence function. Under the Information Metric, the gradient of a regular functional ψ is precisely its influence function, denoted ψ. That is, $\nabla^I \psi_P = \psi_P$.

By substituting the Information Metric and the influence functions into the geometric definitions, we recover the statistical measures used in the main text. The geometric sensitivity coefficient becomes the ratio of the asymptotic covariance to the asymptotic variance of the moments:

$$S(\psi, \nu)_P \Big|_{g=I} = \frac{\mathbb{E}_P[\psi\nu]}{\mathbb{E}_P[\nu^2]} = \sigma_{\psi\nu}\sigma_{\nu\nu}^{-1}.$$

This corresponds exactly to the Misspecification-Robust Sensitivity (MRS) Λ defined in Definition 1. Similarly, the geometric informativeness becomes the squared correlation of the influence functions:

$$R(\psi, \nu)_P \Big|_{g=I} = \frac{(\mathbb{E}_P[\psi\nu])^2}{\mathbb{E}_P[\psi^2]\mathbb{E}_P[\nu^2]} = \frac{\sigma_{\psi\nu}^2}{\sigma_{\psi\psi}\sigma_{\nu\nu}}.$$

This corresponds to the informativeness measure Δ defined in Definition 2. Thus, the sensitivity and informativeness measures proposed in this paper are intrinsic geometric properties of the statistical manifold under the Fisher information metric.

The geometric framework provides an intuitive interpretation of the relationship between informativeness and bias reduction under misspecification. Consider a local misspecification represented by a perturbation vector $z \in T_P\mathcal{P}$. The magnitude of this misspecification is measured by its information norm $\|z\|_I$. The induced asymptotic bias in the estimator of ψ is given by the inner product $I_P(\psi, z)$.

Following Andrews et al. (2020), worst-case bias is defined as b^N, since it is the maximum possible bias resulting from a perturbation with bounded norm $\|z\|_I \leq \mu$. By the Cauchy-Schwarz inequality, this maximum occurs when z is collinear with ψ, yielding $b^N = \mu\|\psi\|_I$.

Consider now the restricted worst-case bias, b^{RN}, where the misspecification is constrained to be orthogonal to the moments used for estimation. This constraint requires $I_P(\nu, z) = 0$. Geometrically, we seek to maximize the projection of z onto ψ while z remains in the subspace orthogonal to ν. The maximum bias is determined by the component of ψ orthogonal to ν. The magnitude of this orthogonal component is $\|\psi\|_I \sin(\theta)$, where θ is the angle between ψ and ν.

Since the informativeness measure is given by $\Delta = \cos^2(\theta)$, it follows that $\sin(\theta) = \sqrt{1-\Delta}$. Therefore, the ratio of the restricted worst-case bias to the unrestricted worst-case bias is:

$$\frac{b^{RN}}{b^N} = \sqrt{1-\Delta}.$$

This result demonstrates that Δ measures the proportion of the potential bias that is eliminated by the moment conditions. A low value of Δ implies that a significant portion of the gradient ψ lies orthogonal to ν, leaving the estimator vulnerable to misspecification in directions that the moment conditions cannot detect.

D.2 Influence Function for Clustered Data

Let $X_i = (X_{i1}, X_{i2}, ..., X_{iN_i})$. Assume $(X_i, N_i) \sim P$ with $i = 1, ..., n$ are iid, and $N_i \perp X_{ij}$. Let $N = \sum_{i=1}^{n} N_i$. For a GMM estimator $\widehat{\theta}$,

$$\sqrt{N}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \psi(X_{ij}) + o_p(1)$$

and then it follows

$$\sqrt{N}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\psi}(X_i) + o_p(1)$$

with

$$\bar{\psi}^*(X_i) = \frac{n}{N} \sum_{j=1}^{N_i} \psi(X_{ij})$$

which are iid and therefore admits the usual CLT to derive the asymptotic variance of $\widehat{\theta}$.

D.3 Unique Moments in the BPP Model

In the BPP model, the structural equations are

$$\begin{aligned}\Delta y_t &= \zeta_t + \epsilon_t + (\theta - 1)\epsilon_{t-1} - \theta\epsilon_{t-2} \\ \Delta c_t &= \phi\zeta_t + \psi\epsilon_t + \xi_t + \Delta u_t\end{aligned}$$

Below is the list of the 11 unique non-collinear moments.

1. $\mathbb{E}[(\Delta c_t)^2] = \phi^2\sigma_\zeta^2 + \psi^2\sigma_\epsilon^2 + \sigma_\xi^2 + 2\sigma_u^2 = C_0$
2. $\mathbb{E}[\Delta c_t \Delta c_{t-1}] = -\sigma_u^2 = C_1$
3. $\mathbb{E}[\Delta c_{t-1} \Delta c_{t+1}] = 0$
4. $\mathbb{E}[(\Delta y_t)^2] = \sigma_\zeta^2 + 2(1 - \theta + \theta^2)\sigma_\epsilon^2 = I_0$
5. $\mathbb{E}[\Delta y_t \Delta y_{t-1}] = -(1-\theta)^2\sigma_\epsilon^2 = I_1$
6. $\mathbb{E}[\Delta y_{t-1} \Delta y_{t+1}] = -\theta\sigma_\epsilon^2 = I_2$
7. $\mathbb{E}[\Delta c_t \Delta y_t] = \phi\sigma_\zeta^2 + \psi\sigma_\epsilon^2 = R_0$
8. $\mathbb{E}[\Delta c_t \Delta y_{t+1}] = \psi(\theta - 1)\sigma_\epsilon^2 = R_1$
9. $\mathbb{E}[\Delta c_{t-1} \Delta y_{t+1}] = -\psi\theta\sigma_\epsilon^2 = R_2$
10. $\mathbb{E}[\Delta c_t \Delta y_{t-1}] = 0$
11. $\mathbb{E}[\Delta c_{t+1} \Delta y_{t-1}] = 0$

Let $\Delta c = (\Delta c_{t-1}, \Delta c_t, \Delta c_{t+1})'$ and $\Delta y = (\Delta y_{t-1}, \Delta y_t, \Delta y_{t+1})'$. Then the 3-period covariance matrix for each t is

$$\Omega = \begin{pmatrix} C_0 & C_1 & 0 & R_0 & R_1 & R_2 \\ C_1 & C_0 & C_1 & 0 & R_0 & R_1 \\ 0 & C_1 & C_0 & 0 & 0 & R_0 \\ R_0 & 0 & 0 & I_0 & I_1 & I_2 \\ R_1 & R_0 & 0 & I_1 & I_0 & I_1 \\ R_2 & R_1 & R_0 & I_2 & I_1 & I_0 \end{pmatrix}$$

The full BPP model allows for time-varying variances of the permanent and temporary shocks, σ_ζ^2 and σ_ϵ^2, respectively, which results in 26 parameters with 325 moments. To illustrate our method, we consider the simplified BPP model presented in Appendix C of Blundell et al. (2008)

The income process consists of a random walk permanent component ζ_t and a serially uncorrelated transitory component ϵ_t. The growth equation is:

$$\Delta y_t = \zeta_t + \Delta v_t = \zeta_t + \epsilon_t - \epsilon_{t-1}$$

The consumption process follows a random walk with sensitivity to permanent shocks (ϕ) and transitory shocks (ψ), plus an independent innovation ξ_t. Measurement error is assumed to be zero in this specific derivation:

$$\Delta c_t = \phi\zeta_t + \psi\epsilon_t + \xi_t$$

Assume that $\zeta_t, \epsilon_t, \xi_t$ are mutually uncorrelated with means zero and variances $\sigma_\zeta^2, \sigma_\epsilon^2, \sigma_\xi^2$, and that all parameters are stationary. Let $\Delta c = (\Delta c_{t-1}, \Delta c_t, \Delta c_{t+1})'$ and $\Delta y = (\Delta y_{t-1}, \Delta y_t, \Delta y_{t+1})'$. We obtain the following elements in the 3-period covariance matrix.

$$\begin{aligned}
\mathbb{E}[(\Delta c_t)^2] &= \phi^2\sigma_\zeta^2 + \psi^2\sigma_\epsilon^2 + \sigma_\xi^2 \equiv C \\
\mathbb{E}[\Delta c_t \Delta c_{t-1}] &= 0 \\
\mathbb{E}[\Delta c_t \Delta c_{t+1}] &= 0 \\
\mathbb{E}[(\Delta y_t)^2] &= \sigma_\zeta^2 + 2\sigma_\epsilon^2 \equiv I \\
\mathbb{E}[\Delta y_t \Delta y_{t+1}] &= -\sigma_\epsilon^2 \\
\mathbb{E}[\Delta y_{t-1} \Delta y_{t+1}] &= 0 \\
\mathbb{E}[\Delta c_t \Delta y_t] &= \phi\sigma_\zeta^2 + \psi\sigma_\epsilon^2 \equiv R \\
\mathbb{E}[\Delta c_t \Delta y_{t+1}] &= -\psi\sigma_\epsilon^2 \\
\mathbb{E}[\Delta c_t \Delta y_{t-1}] &= 0 \\
\mathbb{E}[\Delta c_{t-1} \Delta y_{t+1}] &= 0 \\
\mathbb{E}[\Delta c_{t+1} \Delta y_{t-1}] &= 0
\end{aligned}$$

And the covariance matrix becomes

$$\Omega = \begin{pmatrix} C & 0 & 0 & R & -\psi\sigma_\epsilon^2 & 0 \\ 0 & C & 0 & 0 & R & -\psi\sigma_\epsilon^2 \\ 0 & 0 & C & 0 & 0 & R \\ R & 0 & 0 & I & -\sigma_\epsilon^2 & 0 \\ -\psi\sigma_\epsilon^2 & R & 0 & -\sigma_\epsilon^2 & I & -\sigma_\epsilon^2 \\ 0 & -\psi\sigma_\epsilon^2 & R & 0 & -\sigma_\epsilon^2 & I \end{pmatrix}$$