

Choosing What to Calibrate and What to Estimate in Structural Models

Joan Alegre

Department of Economics
Universidad Carlos III de Madrid

joan.alegre.canton@gmail.com

[CLICK HERE FOR THE LATEST VERSION OF THE PAPER](#)

Abstract

Structural models often fix (calibrate) some parameters and estimate the rest, but this calibration–estimation partition is usually chosen by convention. This paper treats that choice as an econometric partition-selection problem. For each admissible partition, we construct a scalar sensitivity statistic measuring the local response of a target object—such as a policy effect, welfare measure, impulse response, or treatment effect—to perturbations of the calibrated parameters. One rule selects the partition that minimizes this statistic and therefore minimizes worst-case local bias from calibration errors. A second rule minimizes worst-case mean squared error, balancing calibration bias and sampling variance at a specified miscalibration size. We first illustrate the decision problem in two canonical examples. We then apply it to the New Keynesian model of Nakamura and Steinsson (2018), where the partition choice has large implications for credibility: some partitions remain reliable under sizeable miscalibrations, whereas others generate large bias from small calibration errors. The sensitivity rule requires only local derivatives; the MSE rule also uses the joint moment covariance. Both avoid repeated re-estimation and apply to a broad class of structural models.

Keywords: Model selection; Structural parameters; Sensitivity analysis; Partial identification.

JEL classification: C10; C12; C16.

I am grateful to Juan Carlos Escanciano, Andrey Ramos, Juan José Dolado, Telmo Pérez, Facundo Argañaraz, Jesús Carro and Anna Mikusheva for their advice and helpful comments. Research funded by Ministerio de Ciencia e Innovación, grants PGC2018-096732-B-I00 and PID2021-127794NB-I00, and Comunidad de Madrid, grants EPUC3M11 (VPRICIT) and H2019/HUM-589.

1 Introduction

A common practice in structural model estimation is to calibrate some structural parameters—that is, to hold them fixed at prespecified values—and estimate the rest.¹ This approach is often justified when full estimation is computationally infeasible (Jørgensen, 2023), when the model is not fully identified (Iskrev, 2019), when fixing parameters can improve estimation precision, or when population values are known. In practice, however, the calibration–estimation partition is rarely chosen by a formal econometric criterion. It is often inherited from previous studies, justified informally, or dictated by computational convenience.

This paper develops two econometric criteria for choosing what to calibrate and what to estimate in a general class of structural models. We treat the calibration–estimation split as a partition-selection problem. For each admissible partition, we construct a scalar sensitivity statistic and use it to approximate the worst-case bias induced by small calibration errors. The first criterion selects the partition that minimizes this sensitivity, and hence worst-case calibration. The second is based on worst-case mean squared error (MSE), combining the squared worst-case bias approximation with sampling variance. It therefore accounts for the possibility that accepting some calibration bias can improve precision enough to reduce MSE. This perspective connects the paper to robust estimation under local misspecification, especially Bonhomme and Weidner (2022).

A key feature of the implementation is that the locally robust partition is computed relative to two inputs. The first is a reference point for the structural parameters, typically obtained from a first-stage estimation. This reference point is where local sensitivities are evaluated. It also allows the researcher to compare candidate partitions without re-estimating the model separately for each one. The second input is a normalization of the calibrated parameters. This normalization specifies the scale on which calibration errors are measured and makes perturbations comparable across parameters expressed in different units. Hence, the locally robust partition should be interpreted as conditional on both the reference point and the chosen normalization of calibration errors.

The method is developed in a minimum-distance framework (Newey and McFadden, 1994), broad enough to cover several estimation approaches commonly used in structural work, including the Generalized Method of Moments (GMM), Classical Minimum Distance (CMD), Simulated Method of Moments (SMM), or Indirect Inference (II).

We illustrate the importance of partition choice in three settings. The first two are New Keynesian models, based on An and Schorfheide (2007) and Nakamura and Steinsson (2018). The third is the dynamic discrete-choice model in Kalouptzidi et al. (2021). Across these examples, we show that the calibration–estimation partition has large implications for robustness. In the Nakamura–Steinsson application (Section 7), where the object of interest is the monetary policy’s information-

¹Throughout the paper, a parameter is *calibrated* if it is held fixed during the estimation problem. This usage is narrower than in some applied and econometric work, where *calibration* refers to choosing parameters to match empirical moments, i.e. to minimum-distance estimation; see, for example, Cocci and Plagborg-Møller (2021). In our terminology, the latter is estimation.

effect parameter, the sensitivity statistic varies substantially across admissible partitions. For some partitions, miscalibrations of only a few percent are sufficient to overturn the qualitative conclusion about the object of interest. By contrast, miscalibrations of comparable magnitude generate only a small worst-case bias under the locally robust partition. These results show that fragility can be avoided purely by choosing more carefully which parameters to fix and which to estimate.

The paper is related to several strands of literature. It builds most directly on work on sensitivity analysis for structural models. [Iskrev \(2019\)](#) and [Jørgensen \(2023\)](#) study how estimates and model outputs respond to perturbations in calibrated parameters, while [Lau \(2024\)](#) develops local and global sensitivity measures for dynamic discrete choice models with fixed parameters. Relative to [Jørgensen \(2023\)](#), our proposed statistic is scalar and independent of calibrated parameter units. Related work studies sensitivity to moments or identifying assumptions, including [Andrews et al. \(2017\)](#), [Honore et al. \(2020\)](#), and [Escanciano and Zhu \(2013\)](#). Our contribution differs in that sensitivity is not the final object of interest. Instead, it is the input into a decision rule that selects which parameters should be fixed and which should be estimated.

The paper is also related to the literature on identification and partial identification of counterfactuals in dynamic discrete choice models. Building on [Aguirregabiria \(2010\)](#), [Aguirregabiria and Suzuki \(2014\)](#), [Norets and Tang \(2014\)](#), and [Arcidiacono and Miller \(2020\)](#), [Kalouptside et al. \(2021\)](#) characterize which counterfactuals are point identified despite payoff nonidentification, while [Kalouptside et al. \(2026\)](#) characterize the corresponding identified set for the counterfactual. We reverse the direction of this exercise. In that literature, the calibration or identifying strategy is taken as given, and the question is which counterfactuals are point or set identified. Here, the counterfactual exercise is taken as given, and the question is which normalization—or calibration—estimation partition—makes that same object of interest robust to local calibration error.

The minimax interpretation connects the paper to the literature on estimation and inference under local misspecification (e.g., [Newey, 1985](#); [Conley et al., 2012](#); [Kitamura et al., 2013](#)). The closest paper is [Bonhomme and Weidner \(2022\)](#), who choose an estimator, through its influence function, to minimize worst-case mean squared error under local misspecification. Our decision variable is the calibration–estimation partition rather than the influence function. Also closely related is [Armstrong and Kolesár \(2021\)](#), who show that GMM weights should account for both sampling variance and misspecification bias when moments may be only approximately valid. We apply a similar local-robustness logic to a different margin: the choice of which parameters to calibrate and which to estimate.

The remainder of the paper is organized as follows. Section 2 presents two motivational examples. Section 3 defines the framework and the locally robust partition, while Section 4 gives local approximations for the bias and MSE criteria and shows how to obtain their locally robust partitions. Section A provides a step-by-step implementation guide. Section 7 applies the method to [Nakamura and Steinsson \(2018\)](#). Section 8 presents Monte Carlo evidence, and Section 9 concludes.

2 Motivational Examples

This section illustrates the importance of choosing the calibration–estimation partition. In both examples, the model requires fixing at least one structural parameter before the remaining parameters can be estimated. The central point is that different choices of the fixed parameter can generate very different sensitivity of the object of interest to calibration error.

2.1 Motivational Example I: A dynamic discrete choice model

Consider the entry–exit model in [Kalouptsi et al. \(2021\)](#). In each period, a firm chooses whether to be inactive, $a = 0$, or active, $a = 1$. The state is $x = (k, w)$, where $k \in \{0, 1\}$ is the firm’s lagged activity status and $w \in \{w_L, w_M, w_H\}$ is an exogenous demand state. Next period’s lagged activity is $k' = a$. Up to the idiosyncratic choice shock, flow payoffs are

$$\pi_a(k, w; \eta) = \begin{cases} k \, sv, & a = 0, \\ k\{\bar{\pi}(w) - fc\} - (1 - k)ec, & a = 1, \end{cases} \quad \eta = (ec, fc, sv).$$

Here ec is the entry cost, fc is the fixed cost, and sv is the scrap value. Given η , the choice-specific value and CCP are

$$\begin{aligned} v_a(x; \eta) &= \pi_a(x; \eta) + \rho \mathbb{E}[V(x'; \eta) \mid a, x], \\ p_a(x; \eta) &= \frac{\exp\{v_a(x; \eta)\}}{\exp\{v_0(x; \eta)\} + \exp\{v_1(x; \eta)\}}, \end{aligned}$$

where the second expression uses the type-I extreme-value normalization.²

The counterfactual exercise multiplies the deterministic active-choice payoff by $\tau = 1.10$, leaving inactive payoffs and the distribution of choice shocks unchanged. The counterfactual flow payoff is

$$\pi_0^\tau(k, w; \eta) = \pi_0(k, w; \eta), \quad \pi_1^\tau(k, w; \eta) = \tau[k\{\bar{\pi}(w) - fc\} - (1 - k)ec].$$

The object of interest is the counterfactual inactive CCP vector,

$$p_0^\tau(x; \eta).$$

The model is not point identified without one additional normalization: one of ec , fc , or sv must be fixed ([Aguirregabiria and Suzuki, 2014](#)). This is not merely a computational detail. Observationally equivalent parameterizations of the model can imply different values of $p_0^\tau(x; \eta)$. In applications,

²The numerical primitives are those in [Kalouptsi et al. \(2021\)](#). Writing ξ for the inverse-demand slope, residual demand is $P_t = w_t - \xi Q_t$, marginal cost is $c = 11$, $\xi = 1.5$, $w = (20, 17, 12)$, and $\bar{\pi}(w) = (w - c)^2 / (4\xi)$. The remaining primitives are $\rho = 0.95$, $ec = 9$, $fc = 5.5$, $sv = 10$, and

$$F_w = \begin{pmatrix} 0.40 & 0.35 & 0.25 \\ 0.30 & 0.40 & 0.30 \\ 0.20 & 0.20 & 0.60 \end{pmatrix}.$$

The idiosyncratic shocks are type-I extreme value.

researchers most often fix either the scrap value or the fixed cost. Table 1 summarizes this choice in related studies.

TABLE 1 Calibration Strategy in Recent Studies

<i>Calibrate the scrap value</i>	<i>Calibrate the fixed cost</i>
Das et al. (2007)	Ryan (2012)
Kryukov (2010)	Sweeting (2013)
Collard-Wexler (2011)	Santos (2017)
Collard-Wexler (2013)	Beresteanu et al. (2024)
Aguirregabiria and Ho (2012)	
Suzuki (2013)	
Barwick and Pathak (2015)	
Lin (2015)	
Igami (2017)	
Varela (2018)	

Note: No study in this list calibrated the entry cost.

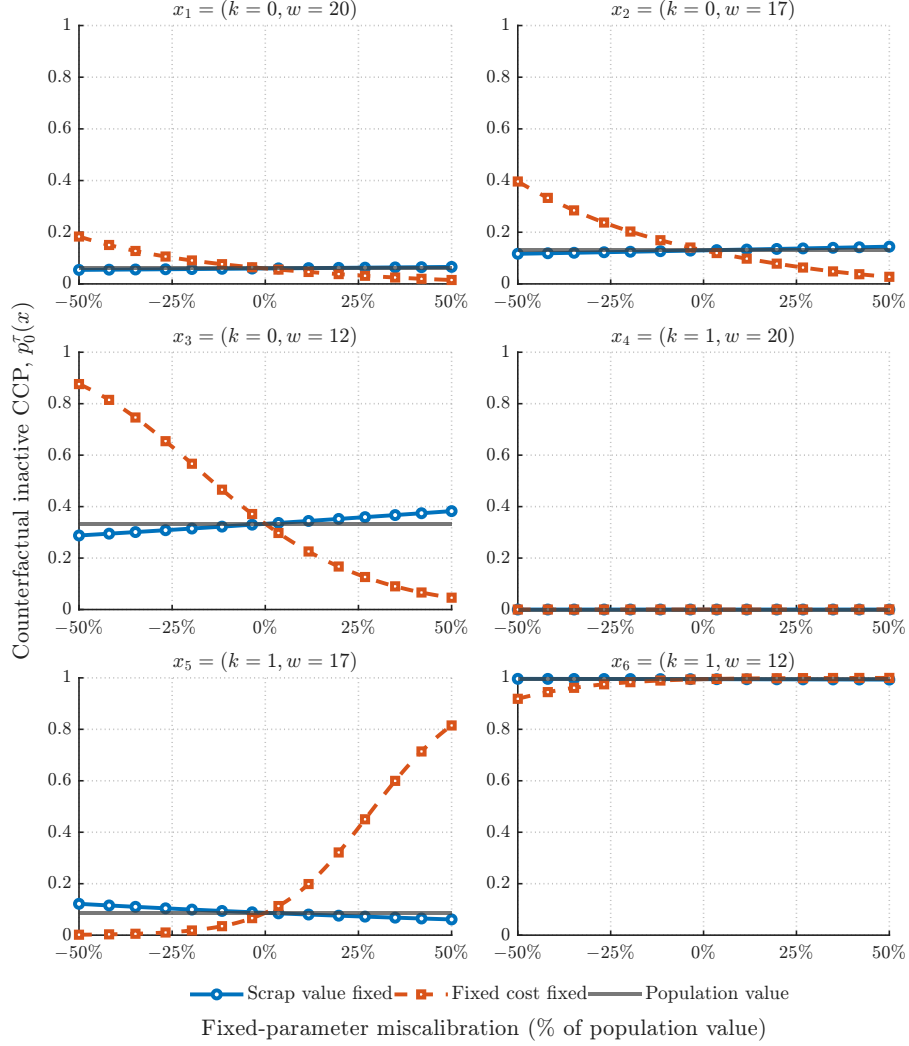
Figure 1 plots the object of interest as a function of the fixed parameter. For each curve, either sv or fc is fixed at a value around its population value, the remaining cost parameters are then estimated, and the counterfactual CCP is recomputed. The horizontal axis measures percentage deviations from the population value of the fixed parameter, ranging from -50% to 50% , so zero corresponds to correct calibration. The horizontal line is the population counterfactual CCP; the vertical distance from this line is the bias induced by miscalibration.

The figure shows that fixing the scrap value produces a relatively flat curve, while fixing the fixed cost produces a much steeper one. Thus, two common ways of identifying the same entry–exit model imply very different robustness properties for the same counterfactual object.

This motivational example is related to Kalouptside et al. (2021) and Kalouptside et al. (2026). Those papers take the calibration strategy as given and ask which counterfactuals are point identified, or what set of counterfactual values is implied when point identification fails. We ask the dual question. We take the counterfactual object, $p_0^T(x; \eta)$, as given and compare the admissible calibration strategies themselves. In this example, fixing sv , fc , or ec generates a different local range for the same counterfactual CCP under calibration errors; our method selects the strategy that makes this local range smallest.

2.2 Motivation Example II: New Keynesian Model

This section illustrates the importance of choosing the calibration–estimation partition in a canonical New Keynesian Model used in An and Schorfheide (2007). In this example, the model requires fixing at least one structural parameter before the remaining parameters can be estimated. The central point is that different admissible choices of the fixed parameter can generate very different

FIGURE 1 Counterfactual CCP as a function of the fixed parameter

sensitivity of the object of interest to calibration error.

Consider the three-equation forward-looking New Keynesian model:

Euler Equation: $y_t = \mathbb{E}_t y_{t+1} + g_t - \mathbb{E}_t g_{t+1} - \frac{1}{\tau}(r_t - \mathbb{E}_t \pi_{t+1} - \mathbb{E}_t z_{t+1}),$

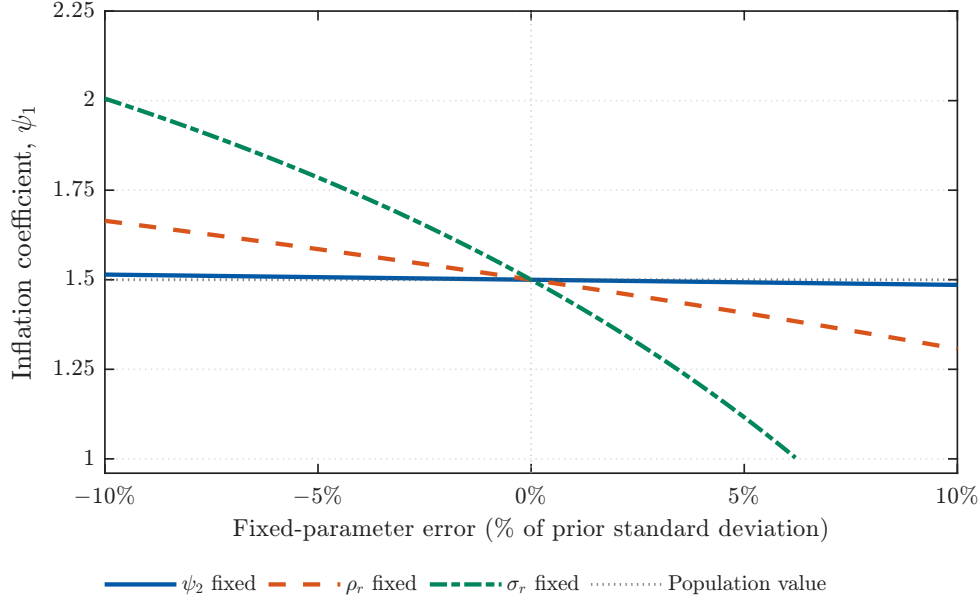
Phillips curve: $\pi_t = \rho \mathbb{E}_t \pi_{t+1} + \kappa(y_t - g_t),$

$$c_t = y_t - g_t,$$

Monetary Policy: $r_t = \rho_r r_{t-1} + (1 - \rho_r)\psi_1 \pi_t + (1 - \rho_r)\psi_2(y_t - g_t) + \varepsilon_{r,t},$

$$g_t = \rho_g g_{t-1} + \varepsilon_{g,t},$$

$$z_t = \rho_z z_{t-1} + \varepsilon_{z,t}.$$

FIGURE 2 Value of ψ_1 as a function of miscalibration in the fixed parameter

The shocks $\varepsilon_{r,t}$, $\varepsilon_{g,t}$, and $\varepsilon_{z,t}$ are mutually uncorrelated, with variances σ_r^2 , σ_g^2 , and σ_z^2 ³.

We estimate the model by maximum likelihood (MLE), using time series for the interest rate r_t , inflation π_t , and output y_t .

Let the object of interest be ψ_1 , the coefficient on inflation in the Taylor rule. In this model, ψ_1 cannot be jointly identified with all elements of $\{\psi_2, \rho_r, \sigma_r^2\}$; one of these three parameters must be fixed in order to estimate the others (e.g. Iskrev, 2010; Komunjer and Ng, 2011; Qu and Tkachenko, 2012, 2017; Kocięcki and Kolasa, 2023). The researcher must therefore decide whether to fix ψ_2 , ρ_r , or σ_r , where fixing the positive shock standard deviation is equivalent to fixing its variance.

Because these parameters have different units, Figure 2 reports miscalibrations in normalized percent-of-prior-standard-deviation units. Thus, $x = 0$ means correct calibration, and $x = 0.1$ means that the fixed parameter is set 10% of its prior standard deviation above its baseline value⁴. For each curve, we fix one parameter at nearby values, re-estimate the remaining parameters by MLE, and record the implied value of ψ_1 . The true value is $\psi_1 = 1.5$. Hence, the vertical distance from 1.5 reflects estimation error, including the effect of miscalibration. Fixing ψ_2 is relatively ro-

³We use the calibration in An and Schorfheide (2007) as the baseline:

$$(\rho, \tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r^2, \sigma_g^2, \sigma_z^2) = (0.9975, 2, 0.33, 1.5, 0.125, 0.75, 0.95, 0.9, 0.04, 0.36, 0.09).$$

⁴We normalize by the marginal prior standard deviations from An and Schorfheide (2007, Table 3), before joint determinacy truncation. In the application's percentage units, these are

$$s_{\psi_2} = 0.25, \quad s_{\rho_r} = 0.20, \quad s_{\sigma_r} = 0.30\sqrt{2 - \pi/2} \simeq 0.19654.$$

For a fixed parameter b , the normalized deviation is $x = (b - b_0)/s_b$. We normalize σ_r rather than σ_r^2 , since the latter has infinite prior variance under the paper's inverse-gamma prior with four degrees of freedom.

bust: the curve is nearly flat around zero. Fixing ρ_r or σ_r , in contrast, can generate large changes in ψ_1 even for small normalized perturbations.

This example motivates the criterion developed below. Full re-estimation curves are informative, but they are costly to compute, difficult to compare when the target or the calibrated block is multidimensional, and sensitive to the units of the fixed parameters. The statistic K_S replaces these curves with a local, scalar measure of worst-case sensitivity of the object of interest to calibration error.

3 Framework and locally robust partition

We first present the framework, then define worst-case bias and worst-case mean squared error. Section 3.4 uses these two criteria to select locally robust partitions. Section 4 gives local approximations that make both rules easy to implement.

Let $\eta \in \mathcal{N} \subset \mathbb{R}^{d_\eta}$ denote the structural parameter vector and let $\gamma = \Gamma(\eta) \in \mathbb{R}^{d_\gamma}$ denote the object of interest. A partition is an index set $S \subseteq \{1, \dots, d_\eta\}$ selecting which components of η are estimated. Define

$$\alpha := (\eta_j)_{j \in S}, \quad \beta := (\eta_j)_{j \in S^c},$$

so that $\eta = (\alpha, \beta)$, where α denotes the estimated parameters and β denotes the calibrated (fixed) parameters⁵.

Consider the Minimum Distance estimation framework implied by a system of d_g equations $g(\eta) : \mathcal{N} \rightarrow \mathcal{G} \subseteq \mathbb{R}^{d_g}$. For a fixed β , a local branch for the minimum-distance solution of the estimated parameters α is given by

$$\alpha(\beta; S) \in \arg \min_{\alpha \in \mathcal{A}} g(\alpha, \beta)' W g(\alpha, \beta), \quad (1)$$

where W is a positive-definite weighting matrix. The implicit function $\alpha(\beta; S)$ induces the partition-specific map

$$\gamma(\beta; S) = \Gamma(\alpha(\beta; S), \beta).$$

Let

$$\mathcal{N}_0 := \{\eta \in \mathcal{N} : g(\eta) = 0\}$$

denote the identified set for the structural parameter vector, and define the identified set for the object of interest as

$$\mathcal{G}_0 := \{\Gamma(\eta) : \eta \in \mathcal{N}_0\} \subseteq \mathbb{R}^{d_\gamma}.$$

Throughout the paper, we assume correct specification of the model and smoothness of the relevant functions:

⁵Equivalently, S induces $\mathcal{N} = \mathcal{A} \times \mathcal{B}$ with $\mathcal{A} := \prod_{j \in S} \mathcal{N}_j$ and $\mathcal{B} := \prod_{j \in S^c} \mathcal{N}_j$, where $\mathcal{N}_j$ is the parameter space associated with the j th element of η .

ASSUMPTION 1 *i)* There exists $\eta_0 \in \mathcal{N}_0$; *ii)* $g(\alpha, \beta; S)$ is twice continuously differentiable in (α, β) ; *iii)* $\Gamma(\alpha, \beta)$ is continuously differentiable.

As noted in the introduction, researchers may fix parameters for three main reasons. First, calibration can improve computational feasibility: the minimum-distance optimization problem may be difficult to solve, and fixing parameters reduces its dimension. Second, calibration can improve precision by imposing restrictions that reduce sampling variance. Third, calibration may supply restrictions needed to identify the structural parameters or the object of interest.

REMARK 1 (THREE IDENTIFICATION REGIMES) These considerations motivate a distinction between three cases:

1. Case 1: The model and the object of interest are point identified. Both $\mathcal{N}_0$ and $\mathcal{G}_0$ are singletons.
2. Case 2: The model is not point identified, but the object of interest is point identified. The set $\mathcal{N}_0$ contains multiple parameter values, while $\mathcal{G}_0$ is a singleton. Consequently,

$$\Gamma(\eta^*) = \Gamma(\eta_0) \quad \text{for every } \eta^* \in \mathcal{N}_0,$$

where η_0 denotes the population structural parameter.

3. Case 3: Neither the model nor the object of interest is point identified.

Our Nakamura and Steinsson (2018) application is a Case 2 situation: the information effect of monetary policy is identified even though the full structural parameter vector is not. The two previous motivational examples illustrate Case 3. Their respective objects of interest—the counterfactual probability and the inflation coefficient in the Taylor rule—are not point identified without additional restrictions.

Partition selection is relevant in all three cases. In Cases 1 and 2, calibration may improve computational feasibility or reduce sampling variance, even though it is unnecessary for identifying the target. In Case 3, calibration additionally supplies restrictions needed to identify the object of interest. In all three cases, incorrect calibration can introduce bias. The MSE criterion evaluates the tradeoff between this calibration bias and sampling variance.

3.1 Worst-Case Bias

To define what the worst-case bias is, we first construct the set of possible miscalibrations. To begin with, we consider a reference point $\eta^* \in \mathcal{N}$ as the baseline value around which local calibration errors are evaluated. The whole methodology will be conditional on this. For a partition S , write

$$\alpha^* := (\eta_j^*)_{j \in S}, \quad \beta^* := (\eta_j^*)_{j \in S^c}.$$

The locally robust partition defined below is conditional on η^* . For a discussion of how to obtain the reference point and its regularity conditions, see Section 5.

For a scalar $\epsilon > 0$, define the local miscalibration neighborhood

$$\mathcal{B}_\epsilon(\beta^*; S) := \left\{ \beta \in \mathcal{B} : \left\| \Sigma_S^{-1}(\beta - \beta^*) \right\|_2 \leq \epsilon \sqrt{|S^c|} \right\}, \quad (2)$$

where Σ_S is a calibration-unit-normalization matrix and the factor $\sqrt{|S^c|}$ adjusts the neighborhood so that it is comparable across partitions with different numbers of calibrated parameters. The term $1/\sqrt{|S^c|} \cdot \left\| \Sigma_S^{-1}(\beta - \beta^*) \right\|_2$ can be interpreted as the distance defining the ball. Alternative metrics are possible, but they would change the interpretation of what constitutes a plausible miscalibration. The neighborhood includes calibrated values for which no choice of α satisfies $g(\alpha, \beta) = 0$; the minimum-distance solution $\alpha(\beta; S)$ remains defined locally at such values.

When the object of interest is vector-valued, its components may also have different scales. We therefore normalize them using a diagonal normalization matrix Σ_γ .

DEFINITION 1 (WORST-CASE BIAS) The worst-case bias for partition S is defined by

$$\mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) := \sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \left\| \Sigma_\gamma (\gamma(\beta; S) - \gamma(\beta^*; S)) \right\|_2. \quad (3)$$

For $S^c = \emptyset$, the worst-case bias is zero by convention.

The goal is to compare worst-case bias over different partitions. To make such a comparison, the point $\gamma(\beta^*; S)$ must be common across partitions. The following assumption ensures this.

ASSUMPTION 2 *i)* The reference point satisfies $\eta^* \in \mathcal{N}_0$; *ii)* $\eta^* \in \text{int}(\mathcal{N})$.

For each admissible partition, we use the local solution branch satisfying $\alpha(\beta^*; S) = \alpha^*$, so that $\gamma(\beta^*; S) = \Gamma(\eta^*)$ is the common benchmark. See Section 5 for a full discussion of primitive conditions for the reference point to be in the identified set.

3.2 Worst-Case Mean Squared Error

Let $\hat{\gamma}(\beta; S)$ denote the estimator of the object of interest obtained by estimating α from the sample MD problem while fixing β . Its MSE accounts for both calibration bias and sampling uncertainty. Using the same reference point and calibration neighborhoods, define

$$\text{WMSE}_n(S, \epsilon) := \sup_{\Sigma_S \in \mathcal{F}_S} \sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \mathbb{E} \left[\left\| \Sigma_\gamma (\hat{\gamma}(\beta; S) - \gamma(\beta^*; S)) \right\|_2^2 \right]. \quad (4)$$

The expectation uses the reference sampling distribution. The family $\mathcal{F}_S$ collects the available normalization matrices and is described in the next subsection. When $S^c = \emptyset$, there is no calibration bias, but the estimator can still have sampling variance.

ASSUMPTION 3 (SAMPLING MOMENTS) For each admissible partition S , in a fixed neighborhood of β^* ,

$$\mathbb{E}[\hat{\gamma}(\beta; S)] - \gamma(\beta; S) = o(n^{-1/2}), \quad \text{Var}(\hat{\gamma}(\beta; S)) = \frac{V_S(\beta)}{n} + o(n^{-1}).$$

The remainders are uniform in β in that neighborhood, and $V_S(\beta)$ is continuous at β^* . Write $V_S := V_S(\beta^*)$.

For each admissible partition, under the smoothness and reference-point conditions above and Assumption 3,

$$\begin{aligned} \text{WMSE}_n(S, \epsilon) &= \left[\max_{\Sigma_S \in \mathcal{F}_S} \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) \right]^2 \\ &\quad + \frac{\text{tr}(\Sigma_\gamma V_S \Sigma_\gamma')}{n} + o(\epsilon^2 + n^{-1}), \end{aligned} \tag{5}$$

as $\epsilon \downarrow 0$ and $n \rightarrow \infty$. The first term is squared worst-case calibration bias; the second is sampling variance. The proof is in Appendix D, and Appendix A.2.3 explains how to estimate V_S . To keep squared calibration bias and sampling variance comparable as the sample grows, we use $\epsilon_n = c/\sqrt{n}$ for a fixed $c > 0$ in the local MSE comparison below.

3.3 The matrix Σ_S

The normalization matrix determines the scale at which calibration errors are compared across parameters and therefore constitutes one of the key elements of the methodology. This section briefly introduces Σ_S to provide context for the definition of worst-case bias. For a detailed discussion of the normalization matrix and its construction, see Section 6. We consider normalization matrices of the form

$$\Sigma_S = \text{diag}(s_j : j \in S^c), \quad s_j > 0, \tag{6}$$

where s_j determines the scale of the j th calibrated parameter. We consider two different types of normalization; unit-normalization and informative-normalization.

As discussed in Section 6, we allow for the possibility that multiple normalization matrices Σ_S are available for a given partition S . We denote by $\mathcal{F}_S$ the set of normalization matrices available for partition S , and define $\mathcal{F} = \{\mathcal{F}_S\}_{S \in \mathcal{S}_{AD}}$ as the corresponding family of sets. We assume each $\mathcal{F}_S$ is nonempty and compact.

3.4 Locally robust partition and Admissible Set

We consider two alternatives over the same admissible set, $\mathcal{S}_{AD}$, defined below: minimize worst-case bias, or minimize worst-case MSE.

DEFINITION 2 (LOCALLY ROBUST PARTITIONS) Conditional on the reference point η^* , the normalization families $\mathcal{F}$, the target scale Σ_γ and the radius ϵ , a partition that minimizes worst-case

bias is any solution to

$$S^*(\eta^*, \mathcal{F}, \Sigma_\gamma, \epsilon) \in \arg \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \max_{\Sigma_S \in \mathcal{F}_S} \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon). \quad (7)$$

For a sample size n , a partition that minimizes worst-case MSE is any solution to

$$S_{\text{MSE},n}^*(\eta^*, \mathcal{F}, \Sigma_\gamma, \epsilon) \in \arg \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \text{WMSE}_n(S, \epsilon). \quad (8)$$

The bias rule focuses on calibration robustness. Its local implementation below does not require choosing a particular ϵ . The MSE rule also accounts for sampling uncertainty, so its choice can depend on ϵ and n .

To define the set of admissible partitions, let $\mathcal{H} \subseteq 2^{\{1, \dots, d_\eta\}}$ denote any hard restrictions imposed by the researcher, such as parameters that must always be estimated or always be fixed. If there are no such restrictions, take $\mathcal{H} = 2^{\{1, \dots, d_\eta\}}$.

ASSUMPTION 4 (ADMISSIBILITY) A partition $S \in \mathcal{H}$ is admissible at the reference point η^* if:

1. *Non-triviality.* The object of interest depends on the estimated parameters around α^* .
2. *Conditional local identification.* With β^* fixed, the estimated block α is locally identified: $\text{Rank}(G_\alpha) = |S|$, where $G_\alpha := \partial g(\alpha^*, \beta^*) / \partial \alpha$.

The admissible set is

$$\mathcal{S}_{\text{AD}}(\eta^*) := \{S \in \mathcal{H} : S \text{ satisfies Assumption 4}\}.$$

4 Approximating Worst-Case Bias and MSE

Directly computing worst-case bias and MSE can be costly. This section uses the sensitivity statistic to approximate calibration bias. Adding the sampling variance gives a local MSE criterion for comparing the same admissible partitions.

DEFINITION 3 (SENSITIVITY STATISTIC) The sensitivity statistic is defined by

$$K_S := \sqrt{|S^c|} \left\| \Sigma_\gamma \frac{\partial \gamma(\beta^*)}{\partial \beta} \Sigma_S \right\|_2. \quad (9)$$

For $S^c = \emptyset$, set $K_S = 0$.

A closed form for $\partial \gamma(\beta^*) / \partial \beta$ is discussed below. Let $\Gamma_\alpha, \Gamma_\beta, G_\alpha, G_\beta$ be the respective Jacobians of $\Gamma(\cdot)$ and $g(\cdot)$ with respect to α and β , evaluated at the reference point (α^*, β^*) .

REMARK 2 The sensitivity $\partial \gamma(\beta^*) / \partial \beta$ might be obtained by applying the implicit function

theorem to the first-order conditions of the MD problem and using the chain rule:⁶

$$\frac{\partial \gamma(\beta^*)}{\partial \beta} = \Gamma_\alpha \frac{\partial \alpha(\beta^*)}{\partial \beta} + \Gamma_\beta, \quad \frac{\partial \alpha(\beta^*)}{\partial \beta} = -(G'_\alpha W G_\alpha)^{-1} G'_\alpha W G_\beta, \quad (10)$$

where $\Gamma_\alpha, \Gamma_\beta, G_\alpha, G_\beta$ are the respective Jacobians of $\Gamma(\cdot)$ and $g(\cdot)$ with respect to α and β , evaluated at the reference point (α^*, β^*) .

Expression (10) can also be found in Escanciano and Zhu (2013) and Jørgensen (2023). Computing $\partial \gamma(\beta^*)/\partial \beta$ requires only the relevant Jacobians of $g(\cdot)$ and $\Gamma(\cdot)$; no re-estimation is necessary.

For each admissible partition S , Assumptions 1 and 2 imply that the minimum-distance first-order conditions have a continuously differentiable local solution $\alpha(\beta; S)$ around β^* . A first-order expansion of the induced map $\gamma(\beta; S)$ yields the following approximation.

THEOREM 1 (WORST-CASE BIAS AND MSE APPROXIMATIONS) *Fix $S \in \mathcal{S}_{\text{AD}}(\eta^*)$ and suppose Assumptions 1 and 2 hold.*

(i) *The worst-case bias satisfies*

$$\mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) = \epsilon K_S + o(\epsilon), \quad \epsilon \downarrow 0. \quad (11)$$

(ii) *If Assumption 3 also holds, then*

$$\text{WMSE}_n(S, \epsilon) = \epsilon^2 \left(\max_{\Sigma_S \in \mathcal{F}_S} K_S \right)^2 + \frac{\text{tr}(\Sigma_\gamma V_S \Sigma'_\gamma)}{n} + o(\epsilon^2 + n^{-1}), \quad (12)$$

as $\epsilon \downarrow 0$ and $n \rightarrow \infty$.

PROOF: See Appendix D. ■

These approximations avoid the maximization over calibration values. Minimizing sensitivity gives the first-order bias rule for small miscalibrations. Adding sampling variance to the squared bias approximation gives the corresponding local MSE rule.

ASSUMPTION 5 (BEST SENSITIVITY PARTITIONS) *The admissible set $\mathcal{S}_{\text{AD}}(\eta^*)$ is nonempty. Define*

$$\mathcal{S}^* := \arg \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \max_{\Sigma_S \in \mathcal{F}_S} K_S. \quad (13)$$

The set $\mathcal{S}^*$ may contain more than one partition.

THEOREM 2 (LOCALLY ROBUST PARTITIONS) *Suppose Assumptions 1, 2, and 5 hold.*

⁶We omit the Hessian terms in the implicit function theorem derivation of the first-order conditions after assuming $\eta^* \in \mathcal{N}_0$ in Assumption 2. For the full discussion on this see Appendix C.4.

(i) There exists $\mathcal{E} > 0$ such that, for all $0 < \epsilon < \mathcal{E}$, every partition that minimizes worst-case bias satisfies

$$S^*(\eta^*, \mathcal{F}, \Sigma_\gamma, \epsilon) \in \mathcal{S}^*. \quad (14)$$

(ii) Suppose Assumption 3 holds. For a fixed $c > 0$, define

$$\mathcal{S}_{\text{MSE}}^*(c) := \arg \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \left\{ c^2 \left(\max_{\Sigma_S \in \mathcal{F}_S} K_S \right)^2 + \text{tr}(\Sigma_\gamma V_S \Sigma_\gamma') \right\}. \quad (15)$$

Then, for all sufficiently large n , every partition that minimizes worst-case MSE satisfies

$$S_{\text{MSE},n}^*(\eta^*, \mathcal{F}, \Sigma_\gamma, c/\sqrt{n}) \in \mathcal{S}_{\text{MSE}}^*(c). \quad (16)$$

If either minimizing set contains only one partition, the corresponding result identifies that partition.

PROOF: See Appendix D. ■

REMARK 3 (RELATION TO BONHOMME AND WEIDNER (2022)) Bonhomme and Weidner (2022) obtain similar bias–variance expressions under local misspecification. They minimize worst-case MSE over influence functions, whereas we minimize it over calibration–estimation partitions.

5 Reference point from a first-stage estimation

Consider the case where the reference point is obtained from a first-stage estimation. Define the first-stage calibration strategy as (β_1, S_1) , where S_1 is the partition used and β_1 is the calibrated value. In population terms, the first-stage estimation is defined by

$$\alpha(\beta_1; S_1) \in \arg \min_{\alpha \in \mathcal{A}} g(\alpha, \beta_1)' W g(\alpha, \beta_1), \quad (17)$$

where we assume standard conditional local identification of α . The reference point is $\eta^* := (\alpha(\beta_1; S_1), \beta_1)$.

A key assumption is that the reference point belongs to the identified set, $\eta^* \in \mathcal{N}_0$. A sufficient primitive condition is that the restricted model implied by the first-stage calibration is correctly specified. Formally, there exists $\alpha \in \mathcal{A}$ such that

$$g(\alpha, \beta_1) = 0. \quad (18)$$

Since the minimum-distance objective is nonnegative, this condition ensures that its population minimum is zero and hence that $\eta^* \in \mathcal{N}_0$.

REMARK 4 In Case 1, any reference point being in the identified set is equivalent to $\eta^* = \eta_0$. In

Case 2, for any $\eta^* \in \mathcal{N}_0$, $\gamma(\beta^*) = \gamma(\beta_0)$. Thus, in both cases, the object of interest evaluated at the reference point $\gamma(\beta^*)$ equals the population object of interest. Minimizing worst-case calibration bias can therefore be interpreted as selecting the partition that provides the greatest local protection of the population target against calibration errors.

In overidentified GMM or CMD settings, the condition in (18) is empirically testable. One can use identification-robust calibration-validity tests, as in [Alegre and Escanciano \(2023\)](#), or related weak-identification robust methods such as [Forneron \(2024\)](#).

When (18) is too strong an assumption to maintain, but we are willing to assume correct specification of the full model, there exist α, β such that $g(\alpha, \beta) = 0$. We may then estimate the full model using ridge regularization to select a reference point among multiple solutions. For this construction to satisfy our framework, the regularization must be chosen so that the resulting estimator converges to a point $\eta^* \in \mathcal{N}_0$.

REMARK 5 Any paper that estimates a structural model provides a potential reference point η^* for applying our methodology. In Case 3, interpreting the local analysis relative to the population target implicitly requires η^* to be sufficiently close to η_0 for $\Gamma(\eta^*)$ to provide a credible local approximation to $\Gamma(\eta_0)$. This proximity is a substantive assumption, since membership in the identified set alone does not guarantee equality of these target values. If η^* is too far from η_0 for the local analysis to be informative, this would also raise concerns about the reliability of the original paper’s estimates.

6 The normalization matrix

The normalization matrix determines the scale at which calibration errors are compared across parameters. This section develops two possible methods to obtain Σ_S . The first method, unit-normalization, uses the model and the reference parameter values to define a common scale for calibration errors. The second method, informative normalization, uses external prior standard deviations to determine these scales. In both cases, we use a diagonal scale matrix

$$\Sigma_S = \text{diag}(s_j : j \in S^c), \quad s_j > 0, \quad (19)$$

where the same coordinate scale s_j applies whenever a candidate partition calibrates parameter j . These scales determine the metric in the ϵ -neighborhood.

6.1 Unit-normalization

Unit-normalization is designed for environments where priors about structural parameters are scarce. A good example is our previous DDC motivational example. Information about the firm’s scrap value or fixed cost is scarce, as [Kalouptside et al. \(2021\)](#) pointed out: “There is little guidance to justify these assumptions [the calibrated values] because cost or scrap values data are extremely

rare.” In such cases, we construct a common scale for calibration errors using the model and the reference parameter values.

We consider two unit-normalization strategies:

$$s_j^{\text{mom}} = (G_j' W G_j)^{-1/2}, \quad \text{or} \quad s_j^{\text{rel}} = |\eta_j^*|, \quad (20)$$

where G_j is column j of the population moment Jacobian at η^* . The first strategy measures a coordinate perturbation by its first-order effect on the weighted moment norm, holding the remaining coordinates fixed. The second measures it relative to the parameter’s reference value, so equal proportional errors receive equal weight in the neighborhood. They require, respectively, $G_j' W G_j > 0$ and $\eta_j^* \neq 0$ for every coordinate that may be calibrated.

These normalizations provide comparable units without specifying how likely different calibration errors are. After choosing one unit-normalization strategy, each admissible partition has a unique normalization, so $\mathcal{F}_S = \{\Sigma_S\}$. In the following subsection, external information may provide several plausible scales for the same parameter. Then $\mathcal{F}_S$ need not be a singleton, and the locally robust partition uses the worst-case Σ_S in $\mathcal{F}_S$.

6.2 Informative normalization

In environments such as DSGE models, the widespread use of Bayesian estimation provides rich prior information about structural parameters. We use the standard deviations of these priors to construct normalization scales for calibration errors.

Suppose we have a prior for the parameter η_j with finite, positive variance τ_j^2 . We use the prior standard deviation as the normalization scale,

$$s_j := \tau_j. \quad (21)$$

Calibration errors are therefore measured in units of external prior dispersion. The prior supplies the coordinate scale, while the reference point and ϵ determine the center and radius of the neighborhood.

REMARK 6 The Informative normalization Σ_S makes sensitivity comparable across parameter units and encodes calibration uncertainty. To illustrate this consider a scalar object of interest and a single calibrated parameter, $S^c = \{j\}$, and $\Sigma_\gamma = 1$

$$K_S = \underbrace{\left| \frac{\partial \gamma(\beta^*; S)}{\partial \beta} \right|}_{\text{Model-implied sensitivity}} \cdot \underbrace{(\tau_j)}_{\text{Uncertainty about fixed parameter}}. \quad (22)$$

Thus, K_S combines model sensitivity with an uncertainty measure of miscalibration.

Since more than one prior distribution may be available in the same literature, we allow for a pool of $M_j > 0$ priors, with finite, positive variances $\tau_{j,1}^2, \dots, \tau_{j,M_j}^2$. We group these alternatives by parameter and retain their standard deviations as possible scales, allowing every combination across the parameters fixed under S . The available scales, normalization matrices, and their family are

$$\mathcal{D}_j := \{\tau_{j,m} : m = 1, \dots, M_j\}, \quad (23)$$

$$\mathcal{F}_S := \left\{ \text{diag}(s) : s \in \prod_{j \in S^c} \mathcal{D}_j \right\}, \quad (24)$$

$$\mathcal{F} := \{\mathcal{F}_S\}_{S \in \mathcal{S}_{AD}(\eta^*)}. \quad (25)$$

All scales are assumed positive. The least-sensitive partition evaluates each partition using the largest sensitivity over $\mathcal{F}_S$.

To add a prior to the pool, the parameter must have the same structural interpretation as in our model, possibly after a transparent transformation of units, timing, or parameterization. Priors need not come from identical models, but must describe comparable economic objects. The empirical application illustrates this procedure parameter by parameter. The family $\mathcal{F}$ collects the prior standard deviations as alternative normalization scales, without combining the distributions into a new prior. The locally robust partition is conditional on the information in $\mathcal{F}$. Different sets of priors may therefore result in different locally robust partitions.

7 Empirical Application

[Nakamura and Steinsson \(2018\)](#) provide evidence of monetary non-neutrality using high-frequency responses of real interest rates, expected inflation, and output growth to Federal Reserve announcements. A central implication of their analysis is that monetary policy affects the economy not only through conventional channels, but also by shifting market beliefs about broader economic fundamentals via *information effects*. They study this mechanism by extending a New Keynesian model to incorporate an information effect and estimating the model by classical minimum distance (CMD), matching model-implied impulse response functions (IRFs) to reduced-form IRFs obtained via local projections.

The model extends the standard New Keynesian framework with internal consumption habits, price indexation, and information effects of monetary announcements. Let $x_t = y_t - y_t^n$ denote the output gap, $\lambda_{x,t} = \lambda_t - \lambda_t^n$ the marginal utility gap, and i_t , π_t , and r_t^n the nominal interest rate, inflation, and the natural real interest rate, respectively. Superscript n denotes flexible-price outcomes, and all variables are expressed as deviations from steady state. The log-linearized equilibrium conditions

are

$$\begin{aligned}
\text{Euler equation:} \quad & \lambda_{x,t} = E_t \lambda_{x,t+1} + i_t - E_t \pi_{t+1} - r_t^n, \\
\text{Internal habits:} \quad & \lambda_{x,t} = -\sigma_c [(1 + \rho b^2)x_t - b x_{t-1} - \rho b E_t x_{t+1}], \\
\text{Phillips curve:} \quad & \Delta \pi_t = \rho E_t \Delta \pi_{t+1} + \kappa \zeta (\omega x_t - \lambda_{x,t}), \\
\text{Monetary policy:} \quad & i_t - E_t \pi_{t+1} = \bar{r}_t + \phi_\pi \pi_t, \\
& \bar{r}_t = (\rho_1 + \rho_2) \bar{r}_{t-1} - \rho_1 \rho_2 \bar{r}_{t-2} + \varepsilon_t,
\end{aligned}$$

where $\Delta \pi_t = \pi_t - \pi_{t-1}$, $\sigma_c = [\sigma(1-b)(1-\rho b)]^{-1}$, and ε_t is the innovation to the exogenous policy component $\bar{r}_t$. The parameter b controls habit formation, while full indexation to lagged inflation gives the Phillips curve its backward-looking component. The benchmark sets $\phi_\pi = 0.01$.

For the impulse response to an announcement at time 0, the information effect is specified by

$$E_0 r_h^n = \gamma E_0 \bar{r}_h, \quad h \geq 0, \quad E_0 y_0^n = \gamma E_0 \bar{r}_0.$$

The subsequent path of natural output follows the flexible-price household equations. Thus, γ measures how strongly policy announcements revise private-sector beliefs about the natural rate of interest; $\gamma = 0$ corresponds to no information effect.

We use their environment as a laboratory for our partition-selection criterion. Our object of interest is the information-effect parameter γ . Throughout, we hold fixed the model, the CMD objective, and the reduced-form moments; what varies across implementations is only the partition of structural parameters into estimated and calibrated (fixed) components. In this setting, the mapping $g(\cdot; S)$ depends on S only through the relabeling of estimated and fixed parameters, while its functional form remains constant across partitions. The number of reduced-form parameters matched in CMD is $d_m = 24$, and the number of structural parameters is $d_\eta = 10$.

This application illustrates Case 2: the information-effect parameter γ is locally identified even though the full structural parameter vector is not. There are therefore some partitions for which the information effect does not depend on the calibrated values to first order. Nevertheless, additional calibration restrictions may improve the precision with which γ is estimated, at the cost of potential calibration bias when those restrictions are incorrectly calibrated. Our partition-selection rule weighs this reduction in variance against the potential calibration bias to improve approximate worst-case MSE.

Table 2 reports the structural parameters and the values reported in Nakamura and Steinsson (2018). For rank estimation, we obtain the empirical reference point from the corrected joint CMD problem in Appendix A.1, and denote the corresponding information-effect estimate by $\hat{\gamma}$. The table retains the values reported in the original study. Alegre and Escanciano (2023) test the Nakamura–Steinsson benchmark calibration on a feasible subset of moments and do not reject it.

TABLE 2 Estimated and Fixed Parameters in Nakamura and Steinsson (2018)

<i>Notation</i>	<i>Definition</i>	<i>Value</i>
<i>Fixed parameters</i>		
ρ	<i>Subjective discount factor</i>	0.99
δ	<i>Nominal rigidity</i>	0.75
ω	<i>Elasticity of marginal cost to output</i>	2
θ	<i>Elasticity of substitution across varieties</i>	10
σ	<i>Intertemporal elasticity of substitution</i>	0.50
b	<i>Consumption habit</i>	0.90
<i>Estimated parameters</i>		
ρ_1	<i>Autoregressive first root of monetary shock</i>	0.873
ρ_2	<i>Autoregressive second root of monetary shock</i>	0.872
χ	<i>Additional real rigidity multiplier</i>	0.035
γ	<i>Information effect</i>	0.754

Notes: In the original paper, three additional fixed structural parameters (the inflation target shock, the endogenous feedback of the Taylor rule, and the coefficient of lagged inflation) are omitted here to streamline the analysis.

7.1 Informative normalization

We construct the normalization matrix from priors in the surveyed literature on Bayesian New Keynesian models, focusing on papers published between 2000 and 2026 in top-five and leading field journals. We retain a prior standard deviation only when the parameter has a comparable structural interpretation.

For each parameter j , let $\mathcal{D}_j$ collect the retained standard deviations. Since sensitivity is nondecreasing in each diagonal scale, we only need the largest standard deviation for each parameter to evaluate the worst case. For each partition, $\mathcal{F}_S$ allows all combinations of the retained scales, so the maximum sensitivity is attained at

$$\bar{\Sigma}_S = \text{diag}(\bar{s}_j : j \in S^c), \quad \bar{s}_j = \max \mathcal{D}_j.$$

Table 3 reports these maximum standard deviations. Here, M_j counts the retained prior or reference specifications for parameter j , including specifications with the same standard deviation.

The following subsections describe the prior pool for each parameter and explain how we obtain the standard deviations in the table. The largest value in each $\mathcal{D}_j$ is shown in bold.

7.1.1 Quarterly discount factor, ρ

Milani (2007) and Bianchi and Melosi (2018) provide direct Beta priors with standard deviations 0.01 and 0.005, respectively. Transforming the interest-rate prior in Lubik and Schorfheide (2004) into a quarterly discount factor gives 0.002425. Smets and Wouters (2007), Del Negro and Schorfheide (2009), and Justiniano et al. (2010, 2011, 2013) provide five further specifications, each

TABLE 3 Maximum Standard Deviations for Informative Normalization

<i>Notation</i>	<i>Definition</i>	M_j	<i>Maximum SD</i>
ρ	<i>Subjective discount factor</i>	8	0.01
δ	<i>Nominal rigidity</i>	17	0.20
ω	<i>Elasticity of marginal cost to output</i>	3	0.40
θ	<i>Elasticity of substitution across varieties</i>	5	24
σ	<i>Intertemporal elasticity of substitution</i>	7	0.75
b	<i>Consumption habit</i>	12	0.29
ρ_1	<i>Autoregressive first root of monetary shock</i>	1	0.58
ρ_2	<i>Autoregressive second root of monetary shock</i>	1	0.58
χ	<i>Additional real rigidity multiplier</i>	1	12.23

Notes: M_j counts the number of used priors for parameter j . The maximum standard deviation is $\bar{s}_j = \max \mathcal{D}_j$.

giving 0.000994 after the corresponding discount-rate transformation. Thus,

$$M_\rho = 8, \quad \mathcal{D}_\rho \simeq \{0.000994, 0.002425, 0.005, \mathbf{0.01}\}.$$

7.1.2 Nominal price rigidity, δ

For the quarterly Calvo non-reset probability, [Smets and Wouters \(2003\)](#) provide a standard deviation of 0.05. The value 0.10 appears in [Smets and Wouters \(2007\)](#), [Justiniano and Primiceri \(2008\)](#), [Justiniano et al. \(2010, 2011, 2013\)](#), [Christiano et al. \(2014\)](#), [Christiano et al. \(2016\)](#), [Milani \(2017\)](#), and [Christiano et al. \(2021\)](#). [Del Negro and Schorfheide \(2009\)](#) and [Christiano et al. \(2015\)](#) provide 0.15. [Rabanal and Rubio-Ramírez \(2005\)](#) provide a prior on expected price duration T ; transforming it through $\delta = 1 - 1/T$ gives 0.172509. Finally, [Del Negro et al. \(2007\)](#), [Rabanal \(2007\)](#), [Leeper et al. \(2017\)](#), and [Bianchi and Melosi \(2018\)](#) provide 0.20. Hence,

$$M_\delta = 17, \quad \mathcal{D}_\delta \simeq \{0.05, 0.10, 0.15, 0.172509, \mathbf{0.20}\}.$$

7.1.3 Marginal-cost elasticity, ω

We use three priors on the output-curvature parameter entering the marginal-cost expression: [Milani \(2007\)](#), [Cúrdia et al. \(2015\)](#), and [Cole and Milani \(2019\)](#), with standard deviations 0.40, 0.20, and 0.25, respectively. We do not include inverse-Frisch priors that require a separate production-technology mapping. The pool is

$$M_\omega = 3, \quad \mathcal{D}_\omega = \{0.20, 0.25, \mathbf{0.40}\}.$$

7.1.4 Substitution across goods, θ

[Bianchi and Melosi \(2018\)](#) provide a direct prior on the CES elasticity with standard deviation 3. We also transform markup priors using $\theta = 1 + 1/m$ for a net markup m , or $\theta = \Lambda/(\Lambda - 1)$ for

a gross markup Λ . The net-markup prior in [Del Negro et al. \(2007\)](#) gives 24. The gross-markup priors in [Christiano et al. \(2015, 2016\)](#) give 2.139010, while [Christiano et al. \(2021\)](#) give 22.935105. These gross-markup transformations include the documented lower bound $\Lambda \geq 1.001$. Thus,

$$M_\theta = 5, \quad \mathcal{D}_\theta \simeq \{2.139010, 3, 22.935105, \mathbf{24}\}.$$

The large transformed standard deviations reflect the nonlinear relationship between markups and substitution elasticities.

7.1.5 Intertemporal elasticity of substitution, σ

We retain seven specifications from six papers. [Milani \(2008, 2011\)](#) provide direct IES prior standard deviations of 0.70 and 0.75. Transforming Gamma priors on inverse IES gives 0.142539 in [Lubik and Schorfheide \(2004\)](#), [Del Negro and Schorfheide \(2004\)](#), [Milani \(2017\)](#), and the Japanese specification of [Aruoba et al. \(2018\)](#). The U.S. specification in the latter paper gives 0.064508. All transformations concern the primitive IES, not a habit-adjusted Euler-equation coefficient. The pool is therefore

$$M_\sigma = 7, \quad \mathcal{D}_\sigma \simeq \{0.064508, 0.142539, 0.70, \mathbf{0.75}\}.$$

7.1.6 Internal consumption habit, b

We retain only priors on internal habits. [Del Negro et al. \(2007\)](#), [Del Negro and Schorfheide \(2009\)](#), and [Christiano et al. \(2021\)](#) provide standard deviations of 0.05. [Justiniano and Primiceri \(2008\)](#), [Justiniano et al. \(2010, 2011, 2013\)](#), [Christiano et al. \(2014\)](#), and [Cole and Milani \(2019\)](#) provide 0.10. [Christiano et al. \(2016\)](#) provide 0.15, and [Cúrdia et al. \(2015\)](#) provide 0.20. [Milani \(2007\)](#) uses a uniform prior on $[0, 1]$, with standard deviation $1/\sqrt{12}$. Hence,

$$M_b = 12, \quad \mathcal{D}_b \simeq \{0.05, 0.10, 0.15, 0.20, \mathbf{0.288675}\}.$$

7.1.7 Monetary-shock roots, ρ_1 and ρ_2

We found no sufficiently comparable literature prior for the two factorized roots of the monetary disturbance. We instead use a uniform reference distribution on $(-1, 1)$ for each root, allowing stationary roots of either sign. This choice follows the support of the roots in this parameterization; it does not imply noninformativeness under other parameterizations. It gives

$$M_{\rho_1} = M_{\rho_2} = 1, \quad \mathcal{D}_{\rho_1} = \mathcal{D}_{\rho_2} = \{1/\sqrt{3}\} \simeq \{\mathbf{0.577350}\}.$$

7.1.8 Additional real rigidity, χ

[Cúrdia et al. \(2015\)](#) assign a Gamma prior with mean 0.10 and standard deviation 0.05 to the coefficient multiplying the marginal-cost expression in their Phillips curve. Their model includes

internal habits and price indexation. We treat this coefficient as comparable to our composite $\kappa\zeta$ and use its prior standard deviation, $s_{\kappa\zeta} = 0.05$. Since the independently varied parameter is χ , we use

$$\kappa\zeta = \chi C(\rho, \delta, \omega, \theta), \quad C(\rho, \delta, \omega, \theta) = \frac{(1 - \delta)(1 - \rho\delta)}{\delta(1 + \omega\theta)}.$$

Holding the other inputs at their reference values for this conversion gives

$$C^* = C(0.99, 0.75, 2, 10) = 0.0040873016,$$

and therefore

$$M_\chi = 1, \quad \mathcal{D}_\chi = \{0.05/C^*\} \simeq \{\mathbf{12.233010}\}.$$

This scale comes from converting the prior standard deviation of the Phillips-curve coefficient while holding the other inputs fixed. It is not a prior on χ reported by [Cúrdia et al. \(2015\)](#).

7.2 Main Results

We estimate the admissible set using the normalized SVD rule in [Appendix A.1](#). At the corrected reference point, the ten-column Jacobian has rank eight. Among the 512 candidate partitions containing the information-effect parameter γ , 256 satisfy conditional local identification. The rank deficiency arises because δ , θ , and χ enter through a single Phillips-curve coefficient. An admissible partition can therefore estimate at most one of these three parameters. This also explains the groups of three in [Table 4](#): within each group, the partitions differ only in which of these parameters is estimated. The ties in these local criteria are allowed by [Theorems 2 and 5](#): either approximate rule can select any one of its tied minimizers.

The three partitions $S_{0,1}$, $S_{0,2}$, and $S_{0,3}$ identify the information effect locally without relying on the calibrated values: γ does not change to first order when those values change. Their sensitivity statistic is therefore zero, and any of them minimizes approximate worst-case bias. This robustness comes at a substantial cost in precision. Their estimator standard deviation is 2.018, consistent with weak identification: the information effect is locally identified, but estimated very imprecisely. In the approximation reported in [Table 5](#), their MSE is entirely driven by sampling variance.

Allowing some calibration sensitivity can greatly reduce this variance. The partitions $S_{1,1}$, $S_{1,2}$, and $S_{1,3}$ retain σ and b as calibrated parameters. Although this introduces some potential calibration bias, the gain in precision more than compensates for it. These partitions achieve the smallest approximate worst-case MSE at every miscalibration size reported in [Table 5](#).

The comparison with the original partition is particularly clear for $S_{1,3}$, which additionally estimates ρ and ω . Its sensitivity falls from 0.714 to 0.256, while its estimator standard deviation falls from 0.121 to 0.105. Thus, the original partition is almost three times as sensitive to calibration errors as each of these MSE-minimizing partitions. In this case, changing the partition improves both

TABLE 4 Selected admissible partitions, sensitivity and estimator standard errors

<i>Name</i>	<i>Estimated</i>	<i>Calibrated</i>	$\widehat{K}_S$ (Sensitivity)	$\sqrt{\widehat{V}_S/n}$ (Std. error)
<i>original</i>	$(\rho_1, \rho_2, \chi, \gamma)$	$(\rho, \delta, \omega, \theta, \sigma, b)$	0.714	0.121
Least sensitive				
$S_{0,1}$	$(\rho, \delta, \omega, \sigma, \rho_1, \rho_2, \gamma, b)$	(θ, χ)	0.000	2.018
$S_{0,2}$	$(\rho, \omega, \theta, \sigma, \rho_1, \rho_2, \gamma, b)$	(δ, χ)	0.000	2.018
$S_{0,3}$	$(\rho, \omega, \sigma, \rho_1, \rho_2, \chi, \gamma, b)$	(δ, θ)	0.000	2.018
Least worst-case MSE for sizes $\varepsilon \in [0.5\%, 84.5\%]$				
$S_{1,1}$	$(\rho, \delta, \omega, \rho_1, \rho_2, \gamma)$	$(\theta, \sigma, \chi, b)$	0.256	0.105
$S_{1,2}$	$(\rho, \omega, \theta, \rho_1, \rho_2, \gamma)$	$(\delta, \sigma, \chi, b)$	0.256	0.105
$S_{1,3}$	$(\rho, \omega, \rho_1, \rho_2, \chi, \gamma)$	$(\delta, \theta, \sigma, b)$	0.256	0.105
Second least worst-case MSE				
$S_{2,1}$	$(\rho, \delta, \omega, \rho_1, \rho_2, \gamma, b)$	(θ, σ, χ)	0.241	0.129
$S_{2,2}$	$(\rho, \omega, \theta, \rho_1, \rho_2, \gamma, b)$	(δ, σ, χ)	0.241	0.129
$S_{2,3}$	$(\rho, \omega, \rho_1, \rho_2, \chi, \gamma, b)$	(δ, θ, σ)	0.241	0.129

robustness and precision.

The alternative group $S_{2,1}$, $S_{2,2}$, and $S_{2,3}$ additionally estimates the habit parameter b . This reduces sensitivity slightly further but increases sampling variance. Relative to the original partition, $S_{2,3}$ additionally estimates ρ , ω , and b , yielding lower sensitivity but slightly less precision. This group becomes optimal for the larger miscalibration sizes reported in Table 4. The MSE criterion therefore balances protection against calibration bias with the precision needed to estimate the information effect.

TABLE 5 Approximate worst-case bias, sampling variance and MSE at miscalibration sizes from 3% to 14%

Panel A. Miscalibration sizes: 3%, 5%, 8%									
Partition	$\varepsilon = 3\%$			$\varepsilon = 5\%$			$\varepsilon = 8\%$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
<i>original</i>	0.021	0.015	0.015	0.036	0.015	0.016	0.057	0.015	0.018
$S_{0,1}, S_{0,2}, S_{0,3}$	0.000	4.072	4.072	0.000	4.072	4.072	0.000	4.072	4.072
$S_{1,1}, S_{1,2}, S_{1,3}$	0.008	0.011	0.011	0.013	0.011	0.011	0.021	0.011	0.011
$S_{2,1}, S_{2,2}, S_{2,3}$	0.007	0.017	0.017	0.012	0.017	0.017	0.019	0.017	0.017
Panel B. Miscalibration sizes: 10%, 12%, 14%									
Partition	$\varepsilon = 10\%$			$\varepsilon = 12\%$			$\varepsilon = 14\%$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
<i>original</i>	0.071	0.015	0.020	0.086	0.015	0.022	0.100	0.015	0.025
$S_{0,1}, S_{0,2}, S_{0,3}$	0.000	4.072	4.072	0.000	4.072	4.072	0.000	4.072	4.072
$S_{1,1}, S_{1,2}, S_{1,3}$	0.026	0.011	0.012	0.031	0.011	0.012	0.036	0.011	0.012
$S_{2,1}, S_{2,2}, S_{2,3}$	0.024	0.017	0.017	0.029	0.017	0.018	0.034	0.017	0.018

8 Monte Carlo Simulation

This section uses Monte Carlo simulations to study the finite-sample implications of the partition-selection criterion in the [Nakamura and Steinsson \(2018\)](#) application. We compare the original partition with $S_{1,3}$, one of the partitions selected by the approximate worst-case MSE criterion, and focus on estimation of the information effect γ . The exercise considers four sample sizes and miscalibration along the positive and negative local worst-case directions.

Table 6 shows that $S_{1,3}$ has smaller absolute bias and lower MSE in every case. In the positive direction, the reduction in bias often comes with higher sampling variance, but the overall MSE is still lower. In the negative direction, both bias and variance are lower. Larger samples generally reduce sampling variance, but substantial calibration bias remains under the original partition. Increasing the sample size alone therefore does not resolve the effects of miscalibration.

TABLE 6 Monte Carlo bias, sampling variance and MSE under directional miscalibration

Panel A. Positive direction: 15%, 20%, 30%									
Partition	$\varepsilon = 15\%$			$\varepsilon = 20\%$			$\varepsilon = 30\%$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
Sample size: $N = 150$									
<i>original</i>	0.077	0.015	0.021	0.107	0.015	0.027	0.142	0.022	0.042
$S_{1,3}$	0.010	0.022	0.022	0.010	0.021	0.021	0.002	0.021	0.021
Sample size: $N = 250$									
<i>original</i>	0.091	0.008	0.016	0.124	0.006	0.022	0.166	0.007	0.034
$S_{1,3}$	0.031	0.013	0.014	0.032	0.013	0.014	0.024	0.014	0.015
Sample size: $N = 750$									
<i>original</i>	0.113	0.001	0.014	0.143	0.001	0.021	0.180	0.000	0.033
$S_{1,3}$	0.052	0.004	0.006	0.052	0.003	0.006	0.044	0.004	0.006
Sample size: $N = 1000$									
<i>original</i>	0.114	0.001	0.014	0.144	0.000	0.021	0.181	0.000	0.033
$S_{1,3}$	0.054	0.003	0.006	0.054	0.003	0.006	0.046	0.003	0.005

Table 6 (continued)

Panel B. Negative direction: 5%, 10%, 14%									
<i>Partition</i>	$\varepsilon = 5\%$			$\varepsilon = 10\%$			$\varepsilon = 14\%$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
Sample size: $N = 150$									
<i>original</i>	-0.088	0.025	0.033	-0.134	0.036	0.054	-0.234	0.046	0.100
$S_{1,3}$	-0.051	0.019	0.022	-0.074	0.020	0.025	-0.093	0.021	0.030
Sample size: $N = 250$									
<i>original</i>	-0.069	0.016	0.021	-0.113	0.026	0.039	-0.206	0.033	0.075
$S_{1,3}$	-0.028	0.011	0.012	-0.049	0.012	0.014	-0.067	0.014	0.018
Sample size: $N = 750$									
<i>original</i>	-0.049	0.004	0.006	-0.103	0.008	0.019	-0.182	0.010	0.043
$S_{1,3}$	-0.007	0.004	0.004	-0.024	0.004	0.004	-0.040	0.004	0.006
Sample size: $N = 1000$									
<i>original</i>	-0.046	0.003	0.005	-0.099	0.006	0.016	-0.178	0.006	0.038
$S_{1,3}$	-0.007	0.003	0.003	-0.022	0.003	0.003	-0.037	0.003	0.004

FIGURE 3 Population information effect γ along the positive and negative local worst-case miscalibration directions for the original partition and $S_{1,3}$

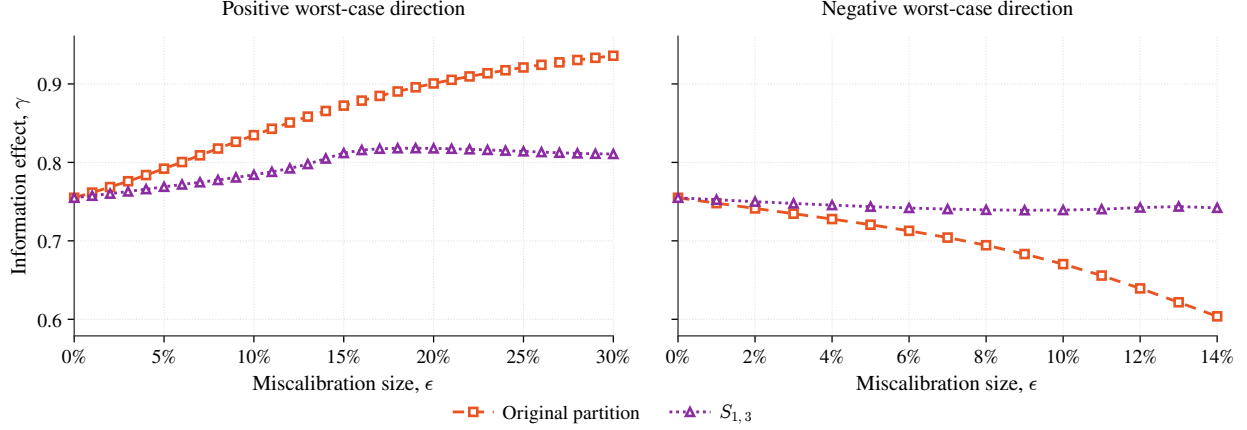


Figure 3 shows the same pattern at the population level. As miscalibration increases in either direction, the information effect moves much further from its reference value under the original partition than under $S_{1,3}$. The lower MSE in the simulations is therefore consistent with the greater protection against calibration errors offered by the selected partition.

9 Conclusion

This paper provides a criterion for choosing which parameters to estimate and which to calibrate in structural models. For each admissible partition, a scalar sensitivity statistic measures first-order worst-case calibration bias in the object of interest. The least-sensitive rule minimizes this statistic, while the MSE rule balances calibration bias against sampling variance. Both compare partitions using local derivatives, without repeated re-estimation; the MSE rule additionally requires the joint moment covariance.

The role of calibration depends on identification of the model and the target. When the target is point identified, calibration may still improve computational feasibility or estimation precision, even if some structural parameters are not identified. Partition selection then evaluates whether these gains justify the potential calibration bias. When the target is not point identified, calibration also supplies identifying restrictions for the object of interest.

The Nakamura–Steinsson application illustrates this distinction. The information effect is locally identified, but the partitions with zero first-order calibration sensitivity estimate it very imprecisely. Additional calibration restrictions substantially improve precision, and the MSE criterion evaluates their cost in potential bias.

The analysis remains local and conditional on the reference point and the scales used to measure calibration errors. In Cases 1 and 2, a reference point in the population identified set provides

the correct target benchmark. In Case 3, interpreting the results relative to the population target additionally requires a credible reference point sufficiently close to the population structural parameters. Different calibration-error scales may also lead to different selected partitions. By making these inputs explicit, the proposed criteria provide a practical way to choose calibration restrictions according to their consequences for the object of interest.

Appendices

A Estimation and Implementation

This section describes how to estimate the admissible set, the sensitivity statistic and the MSE criterion at a common reference point. We first estimate the relevant ranks, then compute the local derivatives for every partition using the same estimated Jacobian. Throughout, the dimensions and the candidate family $\mathcal{H}$ are fixed, and $d_\eta \leq d_g$.

A.1 Rank estimation

Sequential tests, bootstrap procedures and singular-value rules provide alternative ways to estimate rank; see, for example, [Robin and Smith \(2000\)](#) and [Chen and Fang \(2019\)](#). Their finite-sample performance can differ across models and sample sizes. We therefore compare candidate procedures in an application-specific Monte Carlo exercise. For the present application, we use the normalized singular-value decomposition (SVD) rule below. Appendix ?? reports its performance and tuning checks.

A.1.1 The reference point and numerical derivatives

For the first-stage calibration (β_1, S_1) , estimate

$$\hat{\alpha}_1 \in \arg \min_{\alpha} \hat{g}_n(\alpha, \beta_1)' \widehat{W} \hat{g}_n(\alpha, \beta_1), \quad \hat{\eta} = (\hat{\alpha}_1, \beta_1). \quad (26)$$

The weight is held fixed during minimization. In the application, both components of the original minimum-distance criterion enter this single optimization problem. Here $\alpha_1 = \alpha(\beta_1; S_1)$ and the maintained population reference point $\eta^* = (\alpha_1, \beta_1)$ satisfies $g(\eta^*) = 0$.

Write $G_\eta(\eta) = D_\eta g(\eta)$. Distinguish the population derivative, the exact sample derivative, and its numerical approximation:

$$G_\eta = G_\eta(\eta^*), \quad G_n^\circ = D_\eta \hat{g}_n(\hat{\eta}), \quad \hat{G}_\eta \simeq G_n^\circ. \quad (27)$$

The data enter through $\hat{g}_n$ and $\hat{\eta}$. In CMD with $\hat{g}_n(\eta) = \hat{\mu}_n - h(\eta)$, the exact sample derivative is $-D_\eta h(\hat{\eta})$. Any additional estimated inputs to h are also held at their sample values.

We obtain entrywise numerical-error diagnostics $\hat{E}_n^{\text{num}}$ from successive Richardson derivative estimates, with a roundoff floor, and propagate them through analytic chain rules. These diagnostics guide derivative refinement; they do not measure sampling variation or determine the rank cutoff. Known algebraic relations between derivative columns are preserved when computing the Jacobian. Benchmark ranks used to score a simulation are never inputs to the rank procedure.

A.1.2 Normalizing the Jacobian

Choose a nonsingular row normalization $\hat{L}$ and normalize each nonzero column to unit Euclidean length. Define

$$\hat{B}_n = \hat{L}\hat{G}_\eta, \quad \hat{d}_j = \|\hat{B}_{n,j}\|_2^{-1}, \quad \hat{D} = \text{diag}(\hat{d}_j), \quad \hat{Y}_n = \hat{B}_n\hat{D}. \quad (28)$$

Compute these scales once from the full Jacobian. For a nonempty block J , let E_J select its columns, write $p_J = |J|$ and $\hat{D}_J = E_J'\hat{D}E_J$, and extract $\hat{Y}_{n,J} = \hat{Y}_n E_J$. Known identically zero columns remain zero and receive a fixed positive scale. An unresolved near-zero column requires derivative refinement before normalization.

In the application, use

$$L = \text{diag}(s_i^{-1} : i = 1, \dots, d_g), \quad s_i > 0, \quad (29)$$

where s_i are fixed reference standard errors of the empirical targets. This row normalization does not change the MD weight. A row whose derivative is identically zero may be omitted from the rank calculation.

The normalization removes changes in individual parameter units; changes in moment units also cancel when the reference standard errors are transformed accordingly. A common positive rescaling of L cancels in (28). Thus, if estimated target standard errors have rate $n^{-1/2}$, use $\hat{L}/\sqrt{n}$ in the consistency argument. The row and column normalizations must have nonsingular finite limits. These numerical scales are separate from the calibration-error normalizations Σ_S .

A.1.3 The SVD rank rule

Let $\hat{\sigma}_{1,J} \geq \dots \geq \hat{\sigma}_{p_J,J}$ be the computed singular values of $\hat{Y}_{n,J}$. For an arithmetic multiplier $\gamma_{\text{num}} > 0$, define the common arithmetic cutoff

$$h_n(\gamma_{\text{num}}) = \gamma_{\text{num}} \max(d_g, d_\eta) \epsilon_n \|\hat{Y}_n\|_2, \quad \tau_{n,J} = h_n(\gamma_{\text{num}}) + b_{n,J}, \quad (30)$$

where ϵ_n is the working machine precision and $b_{n,J} \geq 0$ is an optional sampling-error floor. Select

$$\hat{r}_J = \sum_{k=1}^{p_J} \mathbf{1}\{\hat{\sigma}_{k,J} > \tau_{n,J}\}. \quad (31)$$

In the application, $b_{n,J} = 0$: every block uses the same full-matrix cutoff. This makes estimated rank monotone when columns are added. The rule requires neither bootstrap critical values nor a covariance estimate for the Jacobian.

An arithmetic cutoff is appropriate when the derivative construction preserves rank locally and positive directions remain numerically separated from zero, as specified in Appendix B.1. For a generic sample Jacobian, a vanishing floor must dominate sampling and derivative approximation

error. With a root- n rate, for example, $b_{n,J} = \sqrt{\log n/n}$ is one possible choice. Machine precision alone does not provide a general statistical rank estimator.

A.1.4 Application defaults and tuning

The application defaults are

Quantity	Default
Arithmetic multiplier	$\gamma_{\text{num}} = 1$
Sampling-error floor	$b_{n,J} = 0$
Arithmetic	Double precision; 40-digit evaluation if needed
Initial derivative step	$2 \times 10^{-4} \max\{1, \hat{\eta}_j \}$
Step refinement	Halve the step; at most 10 steps
Derivative tolerances	Absolute 2×10^{-7} ; relative 2×10^{-5}

We accept a derivative column when its entrywise diagnostics satisfy the absolute-plus-relative tolerance. The routine uses analytic chain rules where available and increases function-evaluation precision for difficult columns. An unresolved derivative failure is reported rather than assigned a rank.

The multiplier $\gamma_{\text{num}} = 1$ is a standard arithmetic convention, not a certified bound on all numerical errors. The Monte Carlo comparison supports SVD for this application. Multipliers 10 and 100, together with an alternative row normalization, provide robustness checks; all give the same rank decisions in the saved-sample comparison. The sufficient conditions below state separately the numerical separation needed for consistency.

A.1.5 The estimated admissible set

Let E_S select the coordinates indexed by a candidate partition S , and write $p_S = |S|$. Extract

$$\hat{G}_{\alpha,S} = \hat{G}_\eta E_S. \quad (32)$$

Apply (31) to each nonempty candidate block, using the full-matrix normalization and common application cutoff, and denote its estimated rank by $\hat{r}_S$. Let ι_S indicate the other admissibility requirements, including non-triviality, and let $\hat{\iota}_S$ estimate it. Then

$$\hat{\mathcal{S}}_{\text{AD}}(\hat{\eta}) = \{S \in \mathcal{H} : 1 \leq p_S \leq d_g, \hat{\iota}_S = 1, \hat{r}_S = p_S\}. \quad (33)$$

For a coordinate target $\Gamma(\eta) = \eta_j$, non-triviality is the known condition $j \in S$. For a general target, consistency of its derivative alone does not justify classifying an exact zero by $\hat{\Gamma}_{\alpha,S} \neq 0$; this condition must be checked separately.

Apply the same rule to the full Jacobian and denote its estimated rank by $\hat{r}$. This rank is reported as a diagnostic; it does not replace the block rank decisions. A correct exact rank restriction implied

by the model may instead be imposed explicitly.

A.2 Sensitivity estimation

A.2.1 Local derivatives and the QR calculation

Compute $\hat{\Gamma}_\eta = D_\eta \Gamma(\hat{\eta})$. For each retained partition, let $q_S = |S^c|$ and extract

$$\hat{G}_{\beta,S} = \hat{G}_\eta E_{S^c}, \quad \hat{\Gamma}_{\alpha,S} = \hat{\Gamma}_\eta E_S, \quad \hat{\Gamma}_{\beta,S} = \hat{\Gamma}_\eta E_{S^c}.$$

With $\widehat{W}^{1/2}$ denoting the symmetric positive-definite square root of $\widehat{W}$, compute the thin QR factorization

$$\widehat{W}^{1/2} \hat{G}_{\alpha,S} = \hat{Q}_S \hat{R}_S, \quad \hat{Q}_S' \hat{Q}_S = I_{p_S}. \quad (34)$$

The matrix $\hat{R}_S$ is square and nonsingular when $\hat{G}_{\alpha,S}$ has full column rank. Solve

$$\hat{R}_S \hat{D}_{\alpha\beta,S} = -\hat{Q}_S' \widehat{W}^{1/2} \hat{G}_{\beta,S}, \quad \hat{D}_{\gamma\beta,S} = \hat{\Gamma}_{\alpha,S} \hat{D}_{\alpha\beta,S} + \hat{\Gamma}_{\beta,S}.$$

For each $\Sigma_S \in \mathcal{F}_S$, compute

$$\widehat{K}_S(\Sigma_S) = \sqrt{q_S} \|\Sigma_\gamma \hat{D}_{\gamma\beta,S} \Sigma_S\|_2. \quad (35)$$

The norm is the spectral norm, including for a vector-valued target. There is no need to form an inverse. Set the sensitivity to zero when $q_S = 0$. A nonpositive-definite weight or a singular $\hat{R}_S$ must be reported as a failure of the sensitivity calculation.

To see the population counterpart, for an admissible partition define

$$D_{\alpha\beta,S} = -(G_{\alpha,S}' W G_{\alpha,S})^{-1} G_{\alpha,S}' W G_{\beta,S}, \quad D_{\gamma\beta,S} = \Gamma_{\alpha,S} D_{\alpha\beta,S} + \Gamma_{\beta,S}.$$

The QR calculation is an algebraically equivalent way to compute the plug-in estimates of these derivatives. Hence the population version of (35) equals the statistic K_S in the main text. The sample calculation estimates this statistic without re-estimating the model for each partition.

A.2.2 Selecting the least-sensitive partition

For each retained partition, take the largest sensitivity over the allowed normalizations, then minimize over partitions:

$$\hat{\kappa}_S = \max_{\Sigma_S \in \mathcal{F}_S} \widehat{K}_S(\Sigma_S), \quad \hat{S}^* \in \arg \min_{S \in \hat{\mathcal{S}}_{AD}(\hat{\eta})} \hat{\kappa}_S. \quad (36)$$

For a single normalization, $\hat{\kappa}_S = \widehat{K}_S$. When a normalization depends on unknown quantities, use its plug-in estimate throughout the sensitivity calculation and the comparison. This applies, for example, to the relative and moment-based unit normalizations in the main text. The consistency result below also covers finite families of such estimated normalizations. Use a fixed rule to resolve sample ties. With population ties, consistency means selecting a member of the population min-

imizing set; the selected label may vary across samples. If the estimated admissible set is empty, report that no partition was selected.

A.2.3 Sampling variance and the MSE rule

Let Ω be the limiting covariance matrix of $\sqrt{n}\hat{g}_n(\eta^*)$. For an interior, regular MD estimator at correct calibration, suppose its usual linearization holds in mean square:

$$\begin{aligned} \sqrt{n}[\hat{\gamma}(\beta^*; S) - \gamma(\beta^*; S)] &= -\Gamma_{\alpha,S}(G'_{\alpha,S}WG_{\alpha,S})^{-1}G'_{\alpha,S}W\sqrt{n}\hat{g}_n(\eta^*) \\ &\quad + o_{L^2}(1). \end{aligned}$$

Here $o_{L^2}(1)$ means that the expected squared norm of the remainder tends to zero. If the first two moments of the scaled sample moments converge, the limiting covariance is

$$\begin{aligned} V_S &= \Gamma_{\alpha,S}(G'_{\alpha,S}WG_{\alpha,S})^{-1}G'_{\alpha,S}W\Omega WG_{\alpha,S} \\ &\quad \times (G'_{\alpha,S}WG_{\alpha,S})^{-1}\Gamma'_{\alpha,S}. \end{aligned} \tag{37}$$

This follows by taking covariances in the linearization; the remainder and its cross terms vanish by Cauchy–Schwarz.

Given a consistent positive-semidefinite estimate $\hat{\Omega}$, use the same QR factors as in (34) to compute

$$\hat{V}_S = \hat{\Gamma}_{\alpha,S}\hat{R}_S^{-1}\hat{Q}'_S\hat{W}^{1/2}\hat{\Omega}\hat{W}^{1/2}\hat{Q}_S\hat{R}_S^{-'}\hat{\Gamma}'_{\alpha,S}. \tag{38}$$

Evaluate the inverse notation with triangular solves. The covariance estimate must describe the joint sampling uncertainty of the moments used in the MD problem, including their cross-covariances.

For a chosen radius ϵ and sample size n , select

$$\hat{S}_{\text{MSE},n}^*(\epsilon) \in \arg \min_{S \in \hat{\mathcal{S}}_{\text{AD}}(\hat{\eta})} \left\{ \epsilon^2 \hat{\kappa}_S^2 + \frac{\text{tr}(\Sigma_\gamma \hat{V}_S \Sigma'_\gamma)}{n} \right\}. \tag{39}$$

For $\epsilon = c/\sqrt{n}$, this is equivalent to minimizing $c^2 \hat{\kappa}_S^2 + \text{tr}(\Sigma_\gamma \hat{V}_S \Sigma'_\gamma)$. Use the same fixed rule for sample ties. No covariance estimate is needed for the least-sensitive rule.

A.3 Implementation algorithm

Inputs. Sample moments $\hat{g}_n$, MD weight $\hat{W}$, first-stage calibration (β_1, S_1) , candidate family $\mathcal{H}$, normalization families $\{\mathcal{F}_S\}$, target scale Σ_γ , row scales, arithmetic multiplier γ_{num} , sampling-error floors, and derivative settings. Supply plug-in calibration scales when the normalizations are estimated. For the MSE rule, also supply n , ϵ and $\hat{\Omega}$.

Step 1. *Estimate the reference point.* Solve (26) and set $\hat{\eta} = (\hat{\alpha}_1, \beta_1)$.

- Step 2.** *Compute the derivatives and their numerical diagnostics.* Evaluate $\hat{G}_\eta$, $\hat{E}_n^{\text{num}}$ and $\hat{\Gamma}_\eta$ at $\hat{\eta}$. Report unresolved derivative failures.
- Step 3.** *Normalize and estimate ranks.* Use (28) and (29). Apply (30)–(31) to the full Jacobian and each nonempty candidate block, using (32). Compute the column scales and the application cutoff once from the full matrix. Report unavailable or nonpositive scales.
- Step 4.** *Construct the admissible set.* Combine the conditional rank estimates with the other requirements as in (33). Stop if this set is empty.
- Step 5.** *Compute local derivatives.* For each retained partition, use (34) and solve the triangular system for $\hat{D}_{\alpha\beta,S}$, then compute $\hat{D}_{\gamma\beta,S}$. Report a nonpositive-definite weight or a singular triangular factor; do not use a pseudoinverse to assign a sensitivity in these cases.
- Step 6.** *Compute sensitivities.* For each retained partition and each allowed normalization, use (35). Set sensitivity to zero when $q_S = 0$. For the MSE rule, also compute $\hat{V}_S$ using (38).
- Step 7.** *Select the partition.* If all retained partitions have valid sensitivity calculations and, for the MSE rule, valid variance estimates, apply (36) for the bias rule or (39) for the MSE rule. Report the selected partition, criterion values, admissible set, ranks, cutoffs and numerical diagnostics.

A small singular value relative to its cutoff should prompt a refinement check. Numerical rank recovery does not by itself establish strong identification or that the population reference point lies in the identified set.

B Consistency results

B.1 Rank estimation and the admissible set of partitions

B.1.1 Sufficient conditions

Let $\mathcal{J}$ contain the full coordinate set and all nonempty candidate blocks to which the rank procedure is applied. For $J \in \mathcal{J}$, write $p_J = |J|$, $G_J = G_\eta E_J$, and $r_J = \text{rank}(G_J)$. Let $\bar{G}_n$ denote the derivative construction before final floating-point assembly, normalization and singular-value computation. It may contain approximation error in the independent derivative columns, but preserves any exact column identities used by the routine. With exact derivatives, $\bar{G}_n = G_n^\circ$. Define

$$Y_{n,J}^\circ = \hat{L} \bar{G}_n E_J \hat{D}_J, \quad \delta_{n,J} = \max_{1 \leq k \leq p_J} \left| \hat{\sigma}_{k,J} - \sigma_k(Y_{n,J}^\circ) \right|.$$

Thus $\delta_{n,J}$ measures the remaining numerical error. Approximation error in $\bar{G}_n$ is controlled separately below.

ASSUMPTION 6 (RANK ESTIMATION) The dimensions and $\mathcal{J}$ are fixed. The following conditions hold.

- (i) $\hat{L} \xrightarrow{p} L$ and $\hat{D} \xrightarrow{p} D$, where L, D are nonsingular.
- (ii) The arithmetic multiplier $\gamma_{\text{num},n}$ belongs to a fixed finite set $\mathcal{C} \subset (0, \infty)$. Write $\gamma_- = \min \mathcal{C}$ and $\gamma_+ = \max \mathcal{C}$. Then

$$\Pr\{\delta_{n,J} \leq h_n(\gamma_-) \text{ for every } J \in \mathcal{J}\} \longrightarrow 1. \quad (40)$$

- (iii) For each block J , one of the following holds:

- (a) $b_{n,J} = 0$ and

$$\Pr\left\{ \begin{array}{l} \text{rank}(\bar{G}_n E_J) = r_J, \\ \sigma_{r_J}(Y_{n,J}^\circ) > h_n(\gamma_+) + \delta_{n,J} \quad \text{if } r_J > 0 \end{array} \right\} \longrightarrow 1;$$

- (b) for a deterministic $a_n \rightarrow \infty$,

$$\|\bar{G}_n E_J - G_J\|_2 = O_p(a_n^{-1}), \quad h_n(\gamma_+) \xrightarrow{p} 0, \quad b_{n,J} \rightarrow 0, \quad a_n b_{n,J} \rightarrow \infty.$$

Condition (iii)(a) applies when the derivative construction preserves rank. Its positive-singular-value margin follows from $\bar{G}_n \xrightarrow{p} G_\eta$ and a vanishing arithmetic cutoff, or can hold directly at fixed precision. Constant rank of the full Jacobian alone does not ensure constant rank of every column block.

In the application, the three rigidity columns associated with δ , θ and χ are nonzero multiples of one column. The remaining seven columns, together with that column, are locally independent at the reference point. Preserving these identities while differentiating the independent columns therefore preserves the rank of every candidate block. Approximation errors in the independent columns need to preserve this independence; they need not be smaller than machine precision.

Condition (iii)(b) covers derivatives whose sample variation need not preserve rank. With a root- n rate, $b_{n,J} = \sqrt{\log n/n}$ is one possible floor. Neither route requires a Gaussian limiting distribution.

Condition (ii) and the margin in (iii)(a) are sufficient numerical requirements. The default multiplier, derivative diagnostics and Monte Carlo success do not establish them by themselves. Precision checks and numerical refinement assess their plausibility in the application. The results are point-wise at the maintained reference point.

B.1.2 Rank and admissible-set consistency

THEOREM 3 (CONSISTENCY OF THE ESTIMATED RANKS) *Under Assumption 6,*

$$\Pr\{\hat{r}_J = r_J \text{ for every } J \in \mathcal{J}\} \longrightarrow 1.$$

The conclusion also holds if some ranks are imposed directly from correct exact model restrictions.

PROOF: See Appendix D.1. ■

COROLLARY 1 (CONSISTENCY OF THE ADMISSIBLE SET) Suppose Theorem 3 applies to the relevant blocks and $\Pr\{\hat{\iota}_S = \iota_S\} \rightarrow 1$ for every candidate. Then

$$\Pr\{\hat{\mathcal{S}}_{\text{AD}}(\hat{\eta}) = \mathcal{S}_{\text{AD}}(\eta^*)\} \rightarrow 1.$$

PROOF: The sets coincide whenever all rank decisions and other admissibility indicators are correct. A finite union bound proves the result. No independence across partitions is required. ■

B.1.3 Fixed numerical settings

For rank-preserving CMD, condition (iii)(a) permits fixed arithmetic settings when exact zero directions remain below the cutoff and all positive directions remain above it. This is a numerical separation condition, not a consequence of the default multiplier. Subsequent sensitivity consistency still requires convergence of the computed derivatives.

B.2 The sensitivity statistic and the selected partitions

ASSUMPTION 7 (SENSITIVITY ESTIMATION) The reference point satisfies $g(\eta^*) = 0$ and is interior to the smooth structural parameter domain; g is twice continuously differentiable and Γ is continuously differentiable near η^* . In addition:

- (i) $\hat{\eta} \xrightarrow{p} \eta^*$, $\hat{G}_\eta \xrightarrow{p} G_\eta$, $\hat{\Gamma}_\eta \xrightarrow{p} \Gamma_\eta$, $\hat{W} \xrightarrow{p} W \succ 0$.
- (ii) The admissible set is consistently estimated, and $\mathcal{S}_{\text{AD}}(\eta^*)$ is nonempty.
- (iii) Σ_γ is fixed. For each admissible S , the family $\mathcal{F}_S$ is a fixed nonempty compact set of positive diagonal matrices, with scales bounded away from zero. Alternatively, it is a fixed finite indexed family $\{\Sigma_{S,k} : k = 1, \dots, m_S\}$ of positive diagonal matrices, estimated by $\hat{\Sigma}_{S,k} \xrightarrow{p} \Sigma_{S,k}$.

The rank and admissible-set results above provide sufficient conditions for (ii). Reference-point consistency, uniform convergence of exact sample derivatives, and vanishing computational error give derivative convergence in (i). A persistent numerical approximation error does not automatically satisfy that condition. Positive limiting scales exclude zero denominators in the normalizations.

LEMMA 4 (CONSISTENCY OF SENSITIVITY AND VARIANCE ESTIMATES) *Under Assumption 7, the QR calculations are well defined for all population-admissible partitions with probability tending*

to one. For fixed families,

$$\max_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \sup_{\Sigma_S \in \mathcal{F}_S} |\widehat{K}_S(\Sigma_S) - K_S(\Sigma_S)| \xrightarrow{p} 0.$$

For estimated finite families, the corresponding conclusion is

$$\max_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \max_{1 \leq k \leq m_S} |\widehat{K}_S(\widehat{\Sigma}_{S,k}) - K_S(\Sigma_{S,k})| \xrightarrow{p} 0.$$

Consequently, for $\kappa_S = \max_{\Sigma_S \in \mathcal{F}_S} K_S(\Sigma_S)$,

$$\max_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} |\widehat{\kappa}_S - \kappa_S| \xrightarrow{p} 0.$$

For estimated families, use $\widehat{\kappa}_S = \max_k \widehat{K}_S(\widehat{\Sigma}_{S,k})$. If the regular MD variance formula (37) applies and $\widehat{\Omega} \xrightarrow{p} \Omega$, then, for every admissible S ,

$$\widehat{V}_S \xrightarrow{p} V_S.$$

PROOF: See Appendix D.1. ■

THEOREM 5 (CONSISTENCY OF THE SELECTED PARTITIONS) *Suppose Assumption 7 holds.*

(i) *For the set $\mathcal{S}^*$ of best sensitivity partitions,*

$$\Pr\{\widehat{\mathcal{S}}^* \in \mathcal{S}^*\} \rightarrow 1.$$

(ii) *Suppose the variance-consistency conditions in Lemma 4 also hold. For a fixed $c > 0$, let $\mathcal{S}_{\text{MSE}}^*(c)$ be the set of minimizers in (15). Then*

$$\Pr\{\widehat{\mathcal{S}}_{\text{MSE},n}^*(c/\sqrt{n}) \in \mathcal{S}_{\text{MSE}}^*(c)\} \rightarrow 1.$$

If either minimizing set contains only one partition, the corresponding conclusion gives consistency for that partition.

PROOF: See Appendix D.1. ■

Under the conditions in the main text, every exact worst-case bias minimizer belongs to $\mathcal{S}^*$ for sufficiently small calibration neighborhoods. For $\epsilon_n = c/\sqrt{n}$ and sufficiently large n , every exact worst-case MSE minimizer belongs to $\mathcal{S}_{\text{MSE}}^*(c)$.

C Further Discussions

C.1 Framework and notation for the appendix

This appendix uses the notation of the main text. Unless otherwise stated, all local objects are evaluated at the reference point η^* . For a partition S , write $\eta = (\alpha_S, \beta_S)$, where α_S collects the estimated coordinates and β_S collects the fixed coordinates. When no confusion can arise, the subscript S is suppressed and we write (α, β) . The Jacobian blocks are

$$G_{\alpha,S} := \frac{\partial g(\eta^*; S)}{\partial \alpha'}, \quad G_{\beta,S} := \frac{\partial g(\eta^*; S)}{\partial \beta'},$$

and

$$\Gamma_{\alpha,S} := \frac{\partial \Gamma(\eta^*)}{\partial \alpha'}, \quad \Gamma_{\beta,S} := \frac{\partial \Gamma(\eta^*)}{\partial \beta'}.$$

For any partition satisfying the rank condition in the main text, define

$$D_{\alpha\beta,S} := - \left(G'_{\alpha,S} W G_{\alpha,S} \right)^{-1} G'_{\alpha,S} W G_{\beta,S}, \quad D_{\gamma\beta,S} := \Gamma_{\alpha,S} D_{\alpha\beta,S} + \Gamma_{\beta,S}. \quad (41)$$

We omit the S subscript when the context is clear.

C.2 Local derivative and first-order expansion

The next lemma records the only differentiability fact needed repeatedly in the appendix and is used in the proof of Theorem 1.

LEMMA 6 (LOCAL DERIVATIVE AND UNIFORM FIRST-ORDER EXPANSION) *Fix $S \in \mathcal{S}_{\text{AD}}(\eta^*)$. Under Assumptions 1 and 2, the following statements hold.*

- (i) $\alpha^* = \alpha(\beta^*; S)$.
- (ii) $\partial\alpha(\beta^*; S)/\partial\beta' = D_{\alpha\beta,S}$ and $\partial\gamma(\beta^*; S)/\partial\beta' = D_{\gamma\beta,S}$, where the matrices are defined in (41).
- (iii) If

$$R_S(\beta) := \gamma(\beta; S) - \gamma(\beta^*; S) - D_{\gamma\beta,S}(\beta - \beta^*),$$

then

$$\sup_{0 < \|\beta - \beta^*\|_2 \leq \delta} \frac{\|R_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \longrightarrow 0 \quad \text{as } \delta \downarrow 0.$$

In particular, for

$$\mathcal{B}_\epsilon(\beta^*; S) := \left\{ \beta \in \mathcal{B} : \left\| \Sigma_S^{-1}(\beta - \beta^*) \right\|_2 \leq \epsilon \sqrt{|S^c|} \right\},$$

one has

$$\sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \|R_S(\beta)\|_2 = o(\epsilon).$$

PROOF: See Appendix D. ■

C.3 Derivative computation without re-estimation

The statistic $\widehat{D}_{\alpha\beta,S}$, proposed in the algorithm, should be interpreted as a direct plug-in computation of the local derivative, not as the exact finite-sample derivative of every partition-specific re-estimation map. To see the distinction, fix a candidate partition S and write the reference point under the S -specific ordering as

$$\widehat{\eta} = (\widehat{\alpha}, \widehat{\beta}).$$

A literal finite-sample computation of the slope of the re-estimation map would first solve

$$\widehat{\alpha}(\widehat{\beta}; S) \in \arg \min_{\alpha \in A_S} \widehat{g}(\alpha, \widehat{\beta}; S)' \widehat{W} \widehat{g}(\alpha, \widehat{\beta}; S),$$

and then differentiate the corresponding first-order conditions with respect to the fixed parameters. Doing this separately for each partition would require re-estimating the model many times. Importantly, $\widehat{\alpha}$, the S -coordinates of the reference point, need not coincide with $\widehat{\alpha}(\widehat{\beta}; S)$, the MD solution obtained after fixing $\widehat{\beta}$.

The exact finite-sample derivative would also include additional Hessian terms weighted by the residuals $\widehat{g}(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})$; under Assumption 2, which places the reference point in the identified set, these terms vanish asymptotically.

The algorithm avoids this re-estimation step. Instead, for each partition S , it uses the derivative bundle evaluated once at the common reference point $\widehat{\eta}$ and computes

$$\widehat{D}_{\alpha\beta}(S) = - \left(\widehat{G}_{\alpha}(\widehat{\eta}; S)' \widehat{W} \widehat{G}_{\alpha}(\widehat{\eta}; S) \right)^{-1} \widehat{G}_{\alpha}(\widehat{\eta}; S)' \widehat{W} \widehat{G}_{\beta}(\widehat{\eta}; S).$$

The justification is asymptotic. Under the regularity conditions in Appendix C.4, and using the reference point in the identified set, $\widehat{D}_{\alpha\beta}(S)$ provides a consistent asymptotic approximation to the population slope,

$$\left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta = \widehat{\beta}} = \widehat{D}_{\alpha\beta}(S) + o_p(1), \quad \widehat{D}_{\alpha\beta}(S) = \frac{\partial \alpha(\beta^*; S)}{\partial \beta'} + o_p(1).$$

Therefore, re-estimating $\widehat{\alpha}(\widehat{\beta}; S)$ for every candidate partition would only change the derivative by an $o_p(1)$ term. The same argument applies to the derivative of the target object.

C.4 Direct and re-estimated derivatives

This subsection formalizes why Algorithm in Section A can compute the local derivative at the common reference point rather than re-estimating the model separately for every candidate partition.

Fix a partition S . For a generic fixed value β , define

$$\hat{\alpha}(\beta; S) \in \arg \min_{\alpha \in A_S} \hat{g}(\alpha, \beta; S)' \widehat{W} \hat{g}(\alpha, \beta; S).$$

Let

$$\widehat{G}_\alpha(\eta; S) := \frac{\partial \hat{g}(\eta; S)}{\partial \alpha'}, \quad \widehat{G}_\beta(\eta; S) := \frac{\partial \hat{g}(\eta; S)}{\partial \beta'}.$$

For the S -specific ordering of the common reference point, write $\hat{\eta} = (\hat{\alpha}_S, \hat{\beta}_S)$. The direct derivative used by the algorithm is

$$\widehat{D}_{\alpha\beta}(S) := - \left[\widehat{G}_\alpha(\hat{\eta}; S)' \widehat{W} \widehat{G}_\alpha(\hat{\eta}; S) \right]^{-1} \widehat{G}_\alpha(\hat{\eta}; S)' \widehat{W} \widehat{G}_\beta(\hat{\eta}; S). \quad (42)$$

ASSUMPTION 8 (LOCAL REGULARITY FOR DIRECT DERIVATIVE COMPUTATION) For the fixed partition S , let $\eta^* = (\alpha^*, \beta^*)$ denote the population reference point under the S -specific ordering. The following conditions hold.

- (i) $\eta^* \in \text{int}(N)$, $g(\eta^*; S) = 0$, and $G_{\alpha,S}$ has full column rank.
- (ii) $\hat{\eta} \xrightarrow{P} \eta^*$ and $\hat{\beta}_S \xrightarrow{P} \beta^*$.
- (iii) $\widehat{W} \xrightarrow{P} W$, where W is symmetric positive definite.
- (iv) The local sample solution $\hat{\alpha}(\hat{\beta}_S; S)$ exists with probability approaching one, satisfies the sample first-order conditions, and

$$(\hat{\alpha}(\hat{\beta}_S; S), \hat{\beta}_S) \xrightarrow{P} (\alpha^*, \beta^*).$$

- (v) In a neighborhood of η^* , $\hat{g}$, its first derivatives, and its second derivatives converge uniformly in probability to their population counterparts.

THEOREM 7 (EQUIVALENCE OF DIRECT AND RE-ESTIMATED DERIVATIVES) *Suppose Assumption 8 holds. Then, with probability approaching one, the derivative of the sample re-estimation map exists at $\hat{\beta}_S$, and*

$$\left. \frac{\partial \hat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta = \hat{\beta}_S} = \widehat{D}_{\alpha\beta}(S) + o_p(1) = D_{\alpha\beta,S} + o_p(1),$$

where $D_{\alpha\beta,S}$ is defined in (41).

PROOF: See Appendix D. ■

COROLLARY 2 (DERIVATIVE OF THE OBJECT OF INTEREST) Suppose the conditions of Theo-

rem 7 hold and Γ is continuously differentiable in a neighborhood of η^* . Define

$$\widehat{D}_{\gamma\beta}(S) := \Gamma_\alpha(\widehat{\eta}; S) \widehat{D}_{\alpha\beta}(S) + \Gamma_\beta(\widehat{\eta}; S).$$

Then

$$\left. \frac{\partial \widehat{\gamma}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} = \widehat{D}_{\gamma\beta}(S) + o_p(1) = D_{\gamma\beta, S} + o_p(1).$$

PROOF: See Appendix D. ■

REMARK 7 (UNIFORMITY OVER PARTITIONS) Let $\mathcal{S}_0$ be a finite collection of candidate partitions. If Assumption 8 holds uniformly over $S \in \mathcal{S}_0$ and

$$\inf_{S \in \mathcal{S}_0} \lambda_{\min} \left(G'_{\alpha, S} W G_{\alpha, S} \right) > 0,$$

then

$$\max_{S \in \mathcal{S}_0} \left\| \left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} - \widehat{D}_{\alpha\beta}(S) \right\| = o_p(1),$$

and the analogous statement holds for $\widehat{D}_{\gamma\beta}(S)$. Hence the direct algorithmic derivative is uniformly asymptotically equivalent to the derivative obtained after partition-specific re-estimation.

D Main Proofs

PROOF OF LEMMA 6: Fix a partition $S \in \mathcal{S}_{AD}(\eta^*)$. Under the S -specific ordering, write $\eta^* = (\alpha^*, \beta^*)$. By Assumption 2, the reference point satisfies $g(\alpha^*, \beta^*; S) = 0$. The minimum-distance criterion at the fixed value β^* is

$$Q(\alpha, \beta^*; S) := g(\alpha, \beta^*; S)' W g(\alpha, \beta^*; S).$$

Since W is positive definite, $Q(\alpha, \beta^*; S) \geq 0$ for every α . At $\alpha = \alpha^*$, the criterion equals zero. Hence α^* solves the conditional MD problem at β^* , that is,

$$\alpha^* = \alpha(\beta^*; S).$$

This proves part (i).

To obtain the derivative, define the population first-order condition map

$$F(\alpha, \beta; S) := \frac{\partial Q(\alpha, \beta; S)}{\partial \alpha} = 2G_\alpha(\alpha, \beta; S)'Wg(\alpha, \beta; S) \in \mathbb{R}^{|S|},$$

where

$$G_\alpha(\alpha, \beta; S) := \frac{\partial g(\alpha, \beta; S)}{\partial \alpha'}.$$

By Assumption 1, g is twice continuously differentiable, so F is continuously differentiable in a neighborhood of (α^*, β^*) . Since $g(\alpha^*, \beta^*; S) = 0$,

$$F(\alpha^*, \beta^*; S) = 0.$$

The derivative of F with respect to α , evaluated at the reference point, is

$$\frac{\partial F(\alpha^*, \beta^*; S)}{\partial \alpha'} = 2G'_{\alpha, S}WG_{\alpha, S},$$

because all terms involving second derivatives of g are multiplied by $g(\alpha^*, \beta^*; S) = 0$. Similarly,

$$\frac{\partial F(\alpha^*, \beta^*; S)}{\partial \beta'} = 2G'_{\alpha, S}WG_{\beta, S}.$$

The rank condition in $S \in \mathcal{S}_{AD}(\eta^*)$ gives $\text{rank}(G_{\alpha, S}) = |S|$. Since W is positive definite, $G'_{\alpha, S}WG_{\alpha, S}$ is nonsingular. Therefore the implicit function theorem applied to $F(\alpha, \beta; S) = 0$ yields a neighborhood of β^* and a unique differentiable map $\beta \mapsto \alpha(\beta; S)$ satisfying

$$F(\alpha(\beta; S), \beta; S) = 0, \quad \alpha(\beta^*; S) = \alpha^*.$$

Differentiating the identity $F(\alpha(\beta; S), \beta; S) = 0$ at $\beta = \beta^*$ gives

$$0 = \frac{\partial F}{\partial \alpha'} \frac{\partial \alpha(\beta^*; S)}{\partial \beta'} + \frac{\partial F}{\partial \beta'}.$$

Substituting the two derivative blocks above and cancelling the common factor 2,

$$\frac{\partial \alpha(\beta^*; S)}{\partial \beta'} = -\left(G'_{\alpha, S}WG_{\alpha, S}\right)^{-1} G'_{\alpha, S}WG_{\beta, S} = D_{\alpha\beta, S}.$$

Since Γ is continuously differentiable by Assumption 1, the chain rule applied to

$$\gamma(\beta; S) = \Gamma(\alpha(\beta; S), \beta)$$

gives

$$\frac{\partial \gamma(\beta^*; S)}{\partial \beta'} = \Gamma_{\alpha, S}D_{\alpha\beta, S} + \Gamma_{\beta, S} = D_{\gamma\beta, S}.$$

This proves part (ii).

It remains to prove the first-order expansion. Since $\gamma(\cdot; S)$ is differentiable at β^* , it is Fréchet differentiable in the finite-dimensional Euclidean space, and therefore

$$\frac{\|\gamma(\beta; S) - \gamma(\beta^*; S) - D_{\gamma\beta, S}(\beta - \beta^*)\|_2}{\|\beta - \beta^*\|_2} \rightarrow 0 \quad \text{as } \beta \rightarrow \beta^*, \beta \neq \beta^*.$$

With the notation

$$R_S(\beta) := \gamma(\beta; S) - \gamma(\beta^*; S) - D_{\gamma\beta, S}(\beta - \beta^*),$$

this is exactly the pointwise statement

$$\frac{\|R_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \rightarrow 0 \quad \text{as } \beta \rightarrow \beta^*.$$

We now turn this pointwise statement into the displayed uniform statement. Suppose, toward a contradiction, that

$$\sup_{0 < \|\beta - \beta^*\|_2 \leq \delta} \frac{\|R_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \not\rightarrow 0 \quad \text{as } \delta \downarrow 0.$$

Then there exist $c > 0$, a sequence $\delta_m \downarrow 0$, and points β_m with $0 < \|\beta_m - \beta^*\|_2 \leq \delta_m$ such that

$$\frac{\|R_S(\beta_m)\|_2}{\|\beta_m - \beta^*\|_2} \geq c.$$

But $\|\beta_m - \beta^*\|_2 \leq \delta_m \rightarrow 0$, so $\beta_m \rightarrow \beta^*$, contradicting the pointwise Fréchet differentiability limit. Hence

$$\sup_{0 < \|\beta - \beta^*\|_2 \leq \delta} \frac{\|R_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \rightarrow 0 \quad \text{as } \delta \downarrow 0.$$

Finally, consider the normalized neighborhood

$$\mathcal{B}_\epsilon(\beta^*; S) := \left\{ \beta \in \mathcal{B} : \left\| \Sigma_S^{-1}(\beta - \beta^*) \right\|_2 \leq \epsilon \sqrt{|S^c|} \right\}.$$

For any $\beta \in \mathcal{B}_\epsilon(\beta^*; S)$,

$$\|\beta - \beta^*\|_2 = \left\| \Sigma_S \Sigma_S^{-1}(\beta - \beta^*) \right\|_2 \leq \|\Sigma_S\|_2 \left\| \Sigma_S^{-1}(\beta - \beta^*) \right\|_2 \leq \|\Sigma_S\|_2 \epsilon \sqrt{|S^c|}.$$

Thus $\mathcal{B}_\epsilon(\beta^*; S)$ is contained in a Euclidean ball around β^* of radius $C_S \epsilon$, where $C_S := \|\Sigma_S\|_2 \sqrt{|S^c|}$. Applying the uniform differentiability bound on this ball gives

$$\sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \|R_S(\beta)\|_2 \leq \left(\sup_{0 < \|\beta - \beta^*\|_2 \leq C_S \epsilon} \frac{\|R_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \right) C_S \epsilon = o(\epsilon).$$

This proves part (iii) and completes the proof of the lemma. ■

PROOF OF THEOREM 1: (i) Fix $S \in \mathcal{S}_{\text{AD}}(\eta^*)$ and let $p := |S^c|$. If $p = 0$, both the worst-case

bias and K_S are zero by convention, so the result is immediate. Suppose henceforth that $p > 0$ and write

$$D := D_{\gamma\beta, S}, \quad B := \Sigma_\gamma D \Sigma_S.$$

By Lemma 6,

$$\Sigma_\gamma [\gamma(\beta; S) - \gamma(\beta^*; S)] = \Sigma_\gamma D(\beta - \beta^*) + e_S(\beta),$$

where $e_S(\beta^*) = 0$ and

$$\sup_{0 < \|\beta - \beta^*\|_2 \leq \delta} \frac{\|e_S(\beta)\|_2}{\|\beta - \beta^*\|_2} \longrightarrow 0 \quad \text{as } \delta \downarrow 0.$$

Every $\beta \in \mathcal{B}_\epsilon(\beta^*; S)$ satisfies

$$\|\beta - \beta^*\|_2 \leq \epsilon \sqrt{p} \|\Sigma_S\|_2.$$

Consequently,

$$\sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \|e_S(\beta)\|_2 = o(\epsilon).$$

Since the reference point is interior, for all sufficiently small ϵ the entire normalized ellipsoid lies in the calibration domain and in the neighborhood where the selected MD branch is defined. We may therefore write

$$\mathcal{B}_\epsilon(\beta^*; S) = \{\beta^* + \epsilon \sqrt{p} \Sigma_S v : \|v\|_2 \leq 1\}.$$

The reverse triangle inequality and the uniform remainder bound give

$$\left| \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) - \epsilon \sqrt{p} \sup_{\|v\|_2 \leq 1} \|Bv\|_2 \right| \leq \sup_{\beta \in \mathcal{B}_\epsilon(\beta^*; S)} \|e_S(\beta)\|_2 = o(\epsilon).$$

By the definition of the spectral norm, $\sqrt{p} \sup_{\|v\|_2 \leq 1} \|Bv\|_2 = \sqrt{p} \|B\|_2 = K_S$. Hence

$$\mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) = \epsilon K_S + o(\epsilon),$$

as claimed.

(ii) Compactness of $\mathcal{F}_S$ places all its calibration neighborhoods in a Euclidean ball of radius $C_S \epsilon$. The remainder bound in Lemma 6 therefore also holds over all those normalizations. Taking their maximum in part (i) gives

$$\max_{\Sigma_S \in \mathcal{F}_S} \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) = \epsilon \max_{\Sigma_S \in \mathcal{F}_S} K_S + o(\epsilon).$$

Square this expression and substitute it into (5) to obtain (12). ■

PROOF OF THEOREM 2: (i) Let

$$\mathcal{S}_0 := \mathcal{S}_{\text{AD}}(\eta^*), \quad \bar{R}_S(\epsilon) := \max_{\Sigma_S \in \mathcal{F}_S} \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon), \quad \kappa_S := \max_{\Sigma_S \in \mathcal{F}_S} K_S(\Sigma_S),$$

where β^* denotes the calibrated reference block associated with partition S . Since $d_\eta < \infty$, the set $\mathcal{S}_0$ is finite.

The compactness of $\mathcal{F}_S$ implies $\sup_{\Sigma_S \in \mathcal{F}_S} \|\Sigma_S\|_2 < \infty$. Thus every calibration neighborhood for partition S is contained in a Euclidean ball of radius $C_S \epsilon$, with C_S independent of $\Sigma_S \in \mathcal{F}_S$. The uniform differentiability bound in Lemma 6 and the proof of Theorem 1 therefore give

$$\sup_{\Sigma_S \in \mathcal{F}_S} |\mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) - \epsilon K_S(\Sigma_S)| = o(\epsilon).$$

Taking maxima over $\mathcal{F}_S$ yields

$$\bar{R}_S(\epsilon) = \epsilon \{\kappa_S + r_S(\epsilon)\}, \quad r_S(\epsilon) \rightarrow 0.$$

Finiteness of $\mathcal{S}_0$ implies

$$\rho(\epsilon) := \max_{S \in \mathcal{S}_0} |r_S(\epsilon)| \rightarrow 0.$$

If $\mathcal{S}^* = \mathcal{S}_0$, the result is immediate. Otherwise, choose any $T \in \mathcal{S}^*$. Finiteness gives a strictly positive gap

$$\Delta := \min_{S \in \mathcal{S}_0 \setminus \mathcal{S}^*} (\kappa_S - \kappa_T) > 0.$$

Choose $\mathcal{E} > 0$ sufficiently small that

$$\rho(\epsilon) < \frac{\Delta}{3} \quad \text{for every } 0 < \epsilon < \mathcal{E}.$$

Then, for every $S \in \mathcal{S}_0 \setminus \mathcal{S}^*$ and every $0 < \epsilon < \mathcal{E}$,

$$\begin{aligned} \bar{R}_S(\epsilon) - \bar{R}_T(\epsilon) &= \epsilon(\kappa_S - \kappa_T) + \epsilon[r_S(\epsilon) - r_T(\epsilon)] \\ &\geq \epsilon\Delta - 2\epsilon\rho(\epsilon) \\ &> \frac{\epsilon\Delta}{3} > 0. \end{aligned}$$

Thus no partition outside $\mathcal{S}^*$ can minimize worst-case bias for $0 < \epsilon < \mathcal{E}$. By the definition of a

locally robust partition,

$$S^*(\eta^*, \mathcal{F}, \Sigma_\gamma, \epsilon) \in \mathcal{S}^* \quad \text{for every } 0 < \epsilon < \mathcal{E}.$$

(ii) With $\epsilon_n = c/\sqrt{n}$, Theorem 1(ii) gives, for every $S \in \mathcal{S}_0$,

$$n \text{WMSE}_n(S, c/\sqrt{n}) = c^2 \kappa_S^2 + \text{tr}(\Sigma_\gamma V_S \Sigma'_\gamma) + o(1).$$

If $\mathcal{S}_{\text{MSE}}^*(c) = \mathcal{S}_0$, the result is immediate. Otherwise, choose any $T \in \mathcal{S}_{\text{MSE}}^*(c)$. Finiteness gives

$$\Delta := \min_{S \in \mathcal{S}_0 \setminus \mathcal{S}_{\text{MSE}}^*(c)} \left\{ c^2 (\kappa_S^2 - \kappa_T^2) + \text{tr}(\Sigma_\gamma (V_S - V_T) \Sigma'_\gamma) \right\} > 0.$$

For sufficiently large n , the remainder for each of the finitely many partitions is smaller than $\Delta/3$ in absolute value. The scaled risk of every $S \notin \mathcal{S}_{\text{MSE}}^*(c)$ then exceeds that of T by at least $\Delta/3$. Thus every exact worst-case MSE minimizer belongs to $\mathcal{S}_{\text{MSE}}^*(c)$, as claimed. ■

PROOF OF THE EXPANSION IN (5): Fix an admissible partition S and write

$$b_S(\beta) := \Sigma_\gamma [\hat{\gamma}(\beta; S) - \gamma(\beta^*; S)], \quad m_{n,S}(\beta) := \mathbb{E}[\hat{\gamma}(\beta; S)] - \gamma(\beta; S).$$

The mean–variance decomposition gives

$$\begin{aligned} \mathbb{E} \left[\left\| \Sigma_\gamma [\hat{\gamma}(\beta; S) - \gamma(\beta^*; S)] \right\|_2^2 \right] &= \|b_S(\beta)\|_2^2 + 2b_S(\beta)' \Sigma_\gamma m_{n,S}(\beta) \\ &\quad + \|\Sigma_\gamma m_{n,S}(\beta)\|_2^2 \\ &\quad + \text{tr} \left\{ \Sigma_\gamma \text{Var}(\hat{\gamma}(\beta; S)) \Sigma'_\gamma \right\}. \end{aligned}$$

The normalization families are compact, so all their calibration neighborhoods lie in a Euclidean ball of radius $C_S \epsilon$. Differentiability of the target map implies $\|b_S(\beta)\|_2 = O(\epsilon)$ uniformly on these neighborhoods. Assumption 3 bounds the cross term by $\epsilon o(n^{-1/2}) = o(\epsilon^2 + n^{-1})$ and the squared sampling bias by $o(n^{-1})$. The uniform variance assumption and continuity of $V_S(\beta)$ at β^* imply

$$\text{tr} \left\{ \Sigma_\gamma \text{Var}(\hat{\gamma}(\beta; S)) \Sigma'_\gamma \right\} = \frac{\text{tr}(\Sigma_\gamma V_S \Sigma'_\gamma)}{n} + o(n^{-1}),$$

uniformly over these neighborhoods as $\epsilon \downarrow 0$ and $n \rightarrow \infty$. Taking the suprema over calibrations and normalizations gives

$$\text{WMSE}_n(S, \epsilon) = \left[\max_{\Sigma_S \in \mathcal{F}_S} \mathcal{R}(S, \beta^*, \Sigma_S, \Sigma_\gamma, \epsilon) \right]^2 + \frac{\text{tr}(\Sigma_\gamma V_S \Sigma'_\gamma)}{n} + o(\epsilon^2 + n^{-1}).$$

This proves (5). ■

PROOF OF THEOREM 7: Define the sample first-order condition map

$$\widehat{F}(\alpha, \beta; S) := \widehat{G}_\alpha(\alpha, \beta; S)' \widehat{W} \widehat{g}(\alpha, \beta; S).$$

Let

$$\widetilde{\eta}_S := (\widehat{\alpha}(\widehat{\beta}_S; S), \widehat{\beta}_S).$$

By Assumption 8,

$$\widetilde{\eta}_S \xrightarrow{p} \eta^*.$$

The sample first-order condition is

$$\widehat{F}(\widetilde{\eta}_S; S) = 0.$$

Let

$$\widehat{F}_\alpha(\eta; S) := \frac{\partial \widehat{F}(\eta; S)}{\partial \alpha'}, \quad \widehat{F}_\beta(\eta; S) := \frac{\partial \widehat{F}(\eta; S)}{\partial \beta'}.$$

By the product rule, evaluated at $\widetilde{\eta}_S$,

$$\widehat{F}_\alpha = \widehat{G}'_\alpha \widehat{W} \widehat{G}_\alpha + \widehat{R}_{\alpha,n},$$

where $\widehat{R}_\alpha$ collects the terms involving second derivatives of $\widehat{g}$ multiplied by $\widehat{W} \widehat{g}$. Similarly,

$$\widehat{F}_\beta = \widehat{G}'_\alpha \widehat{W} \widehat{G}_\beta + \widehat{R}_\beta,$$

where $\widehat{R}_\beta$ collects the mixed second-derivative terms multiplied by $\widehat{W} \widehat{g}$. All matrices in these two displays are evaluated at $\widetilde{\eta}_S$.

Because second derivatives are locally bounded with probability approaching one, $\widehat{W} = O_p(1)$, and

$$\widehat{g}(\tilde{\eta}_S; S) \xrightarrow{p} g(\eta^*; S) = 0,$$

we have

$$\widehat{R}_\alpha = o_p(1), \quad \widehat{R}_\beta = o_p(1).$$

Therefore,

$$\widehat{F}_\alpha(\tilde{\eta}_S; S) = \widehat{G}_\alpha(\tilde{\eta}_S; S)' \widehat{W} \widehat{G}_\alpha(\tilde{\eta}_S; S) + o_p(1),$$

and

$$\widehat{F}_\beta(\tilde{\eta}_S; S) = \widehat{G}_\alpha(\tilde{\eta}_S; S)' \widehat{W} \widehat{G}_\beta(\tilde{\eta}_S; S) + o_p(1).$$

Uniform convergence of the Jacobians, consistency of $\tilde{\eta}_S$, and $\widehat{W} \xrightarrow{p} W$ imply

$$\widehat{F}_\alpha(\tilde{\eta}_S; S) \xrightarrow{p} G'_{\alpha,S} W G_{\alpha,S},$$

and

$$\widehat{F}_\beta(\tilde{\eta}_S; S) \xrightarrow{p} G'_{\alpha,S} W G_{\beta,S}.$$

Since $G_{\alpha,S}$ has full column rank and W is positive definite, $G'_{\alpha,S} W G_{\alpha,S}$ is nonsingular. Hence $\widehat{F}_\alpha(\tilde{\eta}_S; S)$ is nonsingular with probability approaching one. The sample implicit function theorem gives

$$\left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta = \widehat{\beta}_S} = - \left[\widehat{F}_\alpha(\tilde{\eta}_S; S) \right]^{-1} \widehat{F}_\beta(\tilde{\eta}_S; S).$$

By the continuous mapping theorem,

$$\left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta = \widehat{\beta}_S} \xrightarrow{p} -(G'_{\alpha,S} W G_{\alpha,S})^{-1} G'_{\alpha,S} W G_{\beta,S} = D_{\alpha\beta,S}.$$

Now consider the direct derivative used in the algorithm. Since $\widehat{\eta}_n \xrightarrow{p} \eta^*$, the same uniform

convergence and continuous mapping argument implies

$$\widehat{D}_{\alpha\beta}(S) \xrightarrow{p} D_{\alpha\beta,S}.$$

Consequently,

$$\left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} - \widehat{D}_{\alpha\beta}(S) = o_p(1).$$

Finally, applying the population implicit function theorem to

$$F(\alpha, \beta; S) := G_\alpha(\alpha, \beta; S)' W g(\alpha, \beta; S)$$

at (α^*, β^*) gives

$$\frac{\partial \alpha(\beta^*; S)}{\partial \beta'} = -(G'_{\alpha,S} W G_{\alpha,S})^{-1} G'_{\alpha,S} W G_{\beta,S} = D_{\alpha\beta,S}.$$

The second-derivative terms in the derivative of the population first-order condition vanish because $g(\eta^*; S) = 0$. This completes the proof. $\blacksquare$

PROOF OF COROLLARY 2: By the chain rule,

$$\left. \frac{\partial \widehat{\gamma}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} = \Gamma_\alpha(\widetilde{\eta}_S; S) \left. \frac{\partial \widehat{\alpha}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} + \Gamma_\beta(\widetilde{\eta}_S; S).$$

Theorem 7 gives the convergence of the derivative of the re-estimation map. Continuity of Γ_α and Γ_β , together with

$$\widetilde{\eta}_S \xrightarrow{p} \eta^*, \quad \widehat{\eta} \xrightarrow{p} \eta^*,$$

gives

$$\left. \frac{\partial \widehat{\gamma}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} = D_{\gamma\beta,S} + o_p(1).$$

The same continuity and continuous-mapping argument applied to the direct derivative gives

$$\widehat{D}_{\gamma\beta}(S) = D_{\gamma\beta,S} + o_p(1).$$

Therefore

$$\left. \frac{\partial \widehat{\gamma}(\beta; S)}{\partial \beta'} \right|_{\beta=\widehat{\beta}_S} = \widehat{D}_{\gamma\beta}(S) + o_p(1),$$

and the proof is complete. $\blacksquare$

D.1 Consistency proofs

PROOF OF THEOREM 3: Fix $J \in \mathcal{J}$. Under condition (iii)(a), $Y_{n,J}^\circ$ has rank r_J with probability tending to one. By (40), every computed singular value corresponding to an exact zero is at most

$$\delta_{n,J} \leq h_n(\gamma_-) \leq h_n(\gamma_{\text{num},n}) = \tau_{n,J}$$

with probability tending to one. If $r_J > 0$, the margin in (iii)(a) gives

$$\hat{\sigma}_{r_J,J} \geq \sigma_{r_J}(Y_{n,J}^\circ) - \delta_{n,J} > h_n(\gamma_+) \geq \tau_{n,J}.$$

Hence exactly r_J singular values exceed the cutoff.

Under condition (iii)(b), compare $Y_{n,J}^\circ$ with $\hat{L}G_J\hat{D}_J$, whose rank is r_J . By condition (i),

$$\|\hat{L}(\bar{G}_n E_J - G_J)\hat{D}_J\|_2 = O_p(a_n^{-1}) = o_p(b_{n,J}).$$

The singular-value perturbation inequality and (40) therefore bound every computed zero direction by $h_n(\gamma_-) + o_p(b_{n,J})$. With probability tending to one, this is at most

$$h_n(\gamma_{\text{num},n}) + b_{n,J} = \tau_{n,J}.$$

The positive singular values of $\hat{L}G_J\hat{D}_J$ have positive limits, while $\tau_{n,J} \xrightarrow{p} 0$ and both perturbations vanish. The computed positive singular values therefore exceed the cutoff with probability tending to one. These arguments cover $r_J = 0$ and $r_J = p_J$. A finite union bound over $\mathcal{J}$ completes the proof. $\blacksquare$

PROOF OF LEMMA 4: Fix an admissible partition S and write $q_S = |S^c|$. Conditional full column rank and positive definiteness of W imply that $G'_{\alpha,S}WG_{\alpha,S}$ is positive definite. By Assumption 7,

$$\hat{G}'_{\alpha,S}\hat{W}\hat{G}_{\alpha,S} \xrightarrow{p} G'_{\alpha,S}WG_{\alpha,S}.$$

Hence the estimated weighted Jacobian has full column rank, and its thin QR factorization and the corresponding triangular solve are well defined, with probability tending to one. The QR

computation equals

$$\widehat{D}_{\alpha\beta,S} = -(\widehat{G}'_{\alpha,S}\widehat{W}\widehat{G}_{\alpha,S})^{-1}\widehat{G}'_{\alpha,S}\widehat{W}\widehat{G}_{\beta,S}.$$

Continuity of multiplication and inversion gives

$$\widehat{D}_{\alpha\beta,S} \xrightarrow{p} D_{\alpha\beta,S}, \quad \widehat{D}_{\gamma\beta,S} = \widehat{\Gamma}_{\alpha,S}\widehat{D}_{\alpha\beta,S} + \widehat{\Gamma}_{\beta,S} \xrightarrow{p} D_{\gamma\beta,S}.$$

For $q_S > 0$, the reverse triangle inequality and submultiplicativity of the spectral norm yield

$$\begin{aligned} \sup_{\Sigma_S \in \mathcal{F}_S} |\widehat{K}_S(\Sigma_S) - K_S(\Sigma_S)| &\leq \sqrt{q_S} \|\Sigma_\gamma\|_2 \|\widehat{D}_{\gamma\beta,S} - D_{\gamma\beta,S}\|_2 \sup_{\Sigma_S \in \mathcal{F}_S} \|\Sigma_S\|_2 \\ &= o_p(1), \end{aligned}$$

because $\mathcal{F}_S$ is compact. For finite estimated families, consistency of each $\widehat{\Sigma}_{S,k}$ and continuity of matrix multiplication and the spectral norm give

$$\max_{1 \leq k \leq m_S} |\widehat{K}_S(\widehat{\Sigma}_{S,k}) - K_S(\Sigma_{S,k})| = o_p(1).$$

The case $q_S = 0$ follows from the zero convention. There are finitely many admissible partitions, so these conclusions hold uniformly over the population admissible set.

For fixed families, the conclusion for the maximized sensitivities follows from

$$|\widehat{\kappa}_S - \kappa_S| \leq \sup_{\Sigma_S \in \mathcal{F}_S} |\widehat{K}_S(\Sigma_S) - K_S(\Sigma_S)|.$$

For estimated finite families, use the analogous bound with the maximum over the matched normalization indices. A finite maximum over partitions completes the proof.

For the variance claim, the QR expression (38) equals the plug-in version of (37). All estimated matrices converge, including $\widehat{\Omega}$, and $G'_{\alpha,S}WG_{\alpha,S}$ is nonsingular. Continuity of multiplication and inversion gives $\widehat{V}_S \xrightarrow{p} V_S$ for each admissible partition. ■

PROOF OF THEOREM 5: (i) If $\mathcal{S}^* = \mathcal{S}_{\text{AD}}(\eta^*)$, admissible-set consistency proves the result. Otherwise, write $\kappa_{\min} = \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*)} \kappa_S$. Finiteness gives a positive gap

$$\Delta = \min_{S \in \mathcal{S}_{\text{AD}}(\eta^*) \setminus \mathcal{S}^*} (\kappa_S - \kappa_{\min}) > 0.$$

With probability tending to one, the estimated admissible set is correct and $\max_S |\widehat{\kappa}_S - \kappa_S| < \Delta/3$. On this event, every partition outside $\mathcal{S}^*$ has a larger estimated criterion than every partition in $\mathcal{S}^*$. Hence $\widehat{\mathcal{S}}^* \in \mathcal{S}^*$, regardless of the sample tie rule.

(ii) Lemma 4 and continuity of the square and trace give, for every admissible S ,

$$c^2 \hat{\kappa}_S^2 + \text{tr}(\Sigma_\gamma \hat{V}_S \Sigma'_\gamma) \xrightarrow{p} c^2 \kappa_S^2 + \text{tr}(\Sigma_\gamma V_S \Sigma'_\gamma).$$

There are finitely many admissible partitions, so convergence is uniform over them. If every admissible partition belongs to $\mathcal{S}_{\text{MSE}}^*(c)$, admissible-set consistency proves the result. Otherwise, the minimum is separated from the criterion values outside $\mathcal{S}_{\text{MSE}}^*(c)$ by a positive gap. The same argument as in (i) shows that the estimated MSE minimizer belongs to $\mathcal{S}_{\text{MSE}}^*(c)$ with probability tending to one.

■

References

- Aguirregabiria, Victor (2010), “Another look at the identification of dynamic discrete decision processes: An application to retirement behavior.” *Journal of Business & Economic Statistics*, 28 (2), 201–218.
- Aguirregabiria, Victor and Chun-Yu Ho (2012), “A dynamic oligopoly game of the us airline industry: Estimation and policy experiments.” *Journal of Econometrics*, 168 (1), 156–173.
- Aguirregabiria, Victor and Junichi Suzuki (2014), “Identification and counterfactuals in dynamic models of market entry and exit.” *Quantitative Marketing and Economics*, 12, 267–304.
- Alegre, Joan and Juan Carlos Escanciano (2023), “Robust minimum distance inference in structural models.” *arXiv preprint arXiv:2310.05761*.
- An, Sungbae and Frank Schorfheide (2007), “Bayesian analysis of dsge models.” *Econometric reviews*, 26 (2-4), 113–172.
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro (2017), “Measuring the sensitivity of parameter estimates to estimation moments.” *The Quarterly Journal of Economics*, 132 (4), 1553–1592.
- Antoine, Bertille and Eric Renault (2024), “Gmm with nearly-weak identification.” *Econometrics and Statistics*, 30, 36–59.
- Arcidiacono, Peter and Robert A Miller (2020), “Identifying dynamic discrete choice models off short panels.” *Journal of Econometrics*, 215 (2), 473–485.
- Armstrong, Timothy B and Michal Kolesár (2021), “Sensitivity analysis using approximate moment condition models.” *Quantitative Economics*, 12 (1), 77–108.
- Aruoba, S. Boragan, Pablo Cuba-Borda, and Frank Schorfheide (2018), “Macroeconomic dynamics near the ZLB: A tale of two countries.” *The Review of Economic Studies*, 85 (1), 87–118.
- Barwick, Panle Jia and Parag A Pathak (2015), “The costs of free entry: an empirical study of real estate agents in greater boston.” *The RAND Journal of Economics*, 46 (1), 103–145.
- Beresteanu, Arie, Paul B Ellickson, Sanjog Misra, and James C Reeder III (2024), “The dynamics of retail oligopoly.” *Available at SSRN 4784452*.
- Bianchi, Francesco and Leonardo Melosi (2018), “Constrained discretion and central bank transparency.” *The Review of Economics and Statistics*, 100 (1), 187–202.
- Bonhomme, Stéphane and Martin Weidner (2022), “Minimizing sensitivity to model misspecification.” *Quantitative Economics*, 13 (3), 907–954.
- Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt (2015), “Understanding the Great Recession.” *American Economic Journal: Macroeconomics*, 7 (1), 110–167.
- Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt (2016), “Unemployment and business cycles.” *Econometrica*, 84 (4), 1523–1569.
- Christiano, Lawrence J., Roberto Motto, and Massimo Rostagno (2014), “Risk shocks.” *American Economic Review*, 104 (1), 27–65.
- Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin (2021), “Involuntary unemployment and the business cycle.” *Review of Economic Dynamics*, 39, 26–54.
- Cocci, Matthew D and Mikkel Plagborg-Møller (2021), “Standard errors for calibrated parameters.” *arXiv preprint arXiv:2109.08109*.
- Cole, Stephen J. and Fabio Milani (2019), “The misspecification of expectations in New Keynesian models: A DSGE-VAR approach.” *Macroeconomic Dynamics*, 23 (3), 974–1007.
- Collard-Wexler, Allan (2011), “Productivity dispersion and plant selection in the ready-mix concrete industry.” *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-25*.

- Collard-Wexler, Allan (2013), “Demand fluctuations in the ready-mix concrete industry.” *Econometrica*, 81 (3), 1003–1037.
- Conley, Timothy G, Christian B Hansen, and Peter E Rossi (2012), “Plausibly exogenous.” *Review of Economics and Statistics*, 94 (1), 260–272.
- Cúrdia, Vasco, Andrea Ferrero, Ging Cee Ng, and Andrea Tambalotti (2015), “Has U.S. monetary policy tracked the efficient interest rate?” *Journal of Monetary Economics*, 70, 72–83.
- Das, Sanghamitra, Mark J Roberts, and James R Tybout (2007), “Market entry costs, producer heterogeneity, and export dynamics.” *Econometrica*, 75 (3), 837–873.
- Davies, R. B. (1980), “Algorithm AS 155: The distribution of a linear combination of chi-squared random variables.” *Applied Statistics*, 29 (3), 323–333.
- Del Negro, Marco and Frank Schorfheide (2004), “Priors from general equilibrium models for VARs.” *International Economic Review*, 45 (2), 643–673.
- Del Negro, Marco and Frank Schorfheide (2009), “Monetary policy analysis with potentially misspecified models.” *American Economic Review*, 99 (4), 1415–1450.
- Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters (2007), “On the fit of New Keynesian models.” *Journal of Business & Economic Statistics*, 25 (2), 123–143.
- Escanciano, Juan Carlos and Lin Zhu (2013), “Set inferences and sensitivity analysis in semiparametric conditionally identified models.” Technical report, cemmap working paper.
- Forneron, Jean-Jacques (2024), “Detecting identification failure in moment condition models.” *Journal of Econometrics*, 238 (1), 105552.
- Hazell, Jonathon, Juan Herreno, Emi Nakamura, and Jón Steinsson (2022), “The slope of the phillips curve: evidence from us states.” *The Quarterly Journal of Economics*, 137 (3), 1299–1344.
- Honore, Bo, Thomas Jørgensen, and Aureo de Paula (2020), “The informativeness of estimation moments.” *Journal of Applied Econometrics*, 35 (7), 797–813.
- Igami, Mitsuru (2017), “Estimating the innovator’s dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998.” *Journal of Political Economy*, 125 (3), 798–847.
- Iskrev, Nikolay (2010), “Local identification in dsge models.” *Journal of Monetary Economics*, 57 (2), 189–202.
- Iskrev, Nikolay (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, Thomas H (2023), “Sensitivity to calibrated parameters.” *Review of Economics and Statistics*, 105 (2), 474–481.
- Justiniano, Alejandro and Giorgio E. Primiceri (2008), “The time-varying volatility of macroeconomic fluctuations.” *American Economic Review*, 98 (3), 604–641.
- Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2010), “Investment shocks and business cycles.” *Journal of Monetary Economics*, 57 (2), 132–145.
- Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2011), “Investment shocks and the relative price of investment.” *Review of Economic Dynamics*, 14 (1), 102–121.
- Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2013), “Is there a trade-off between inflation and output stabilization?” *American Economic Journal: Macroeconomics*, 5 (2), 1–31.
- Kalouptside, Myrto, Yuichi Kitamura, Lucas Lima, and Eduardo Souza-Rodrigues (2026), “Counterfactual analysis for structural dynamic discrete choice models.” *Review of Economic Studies*, rdag039.
- Kalouptside, Myrto, Paul T Scott, and Eduardo Souza-Rodrigues (2021), “Identification of counterfactuals

- in dynamic discrete choice models." *Quantitative Economics*, 12 (2), 351–403.
- Kitamura, Yuichi, Taisuke Otsu, and Kirill Evdokimov (2013), "Robustness, infinitesimal neighborhoods, and moment restrictions." *Econometrica*, 81 (3), 1185–1201.
- Kocięcki, Andrzej and Marcin Kolasa (2023), "A solution to the global identification problem in dsge models." *Journal of Econometrics*, 236 (2), 105477.
- Komunjer, Ivana and Serena Ng (2011), "Dynamic identification of dynamic stochastic general equilibrium models." *Econometrica*, 79 (6), 1995–2032.
- Kryukov, Yaroslav (2010), "Dynamic r&d and the effectiveness of policy intervention in the pharmaceutical industry." *manuscript. Tepper School of Business. Carnegie Mellon University.*
- Lau, Chun Pong (2024), "Sensitivity analysis for dynamic discrete choice models." *arXiv preprint arXiv:2408.16330*.
- Leeper, Eric M., Nora Traum, and Todd B. Walker (2017), "Clearing up the fiscal multiplier morass." *American Economic Review*, 107 (8), 2409–2454.
- Lin, Haizhen (2015), "Quality choice and market structure: A dynamic analysis of nursing home oligopolies." *International Economic Review*, 56 (4), 1261–1290.
- Lubik, Thomas A. and Frank Schorfheide (2004), "Testing for indeterminacy: An application to U.S. monetary policy." *American Economic Review*, 94 (1), 190–217.
- Lubik, Thomas and Frank Schorfheide (2005), "A bayesian look at new open economy macroeconomics." *NBER macroeconomics annual*, 20, 313–366.
- Milani, Fabio (2007), "Expectations, learning and macroeconomic persistence." *Journal of Monetary Economics*, 54 (7), 2065–2082.
- Milani, Fabio (2008), "Learning, monetary policy rules, and macroeconomic stability." *Journal of Economic Dynamics and Control*, 32 (10), 3148–3165.
- Milani, Fabio (2011), "Expectation shocks and learning as drivers of the business cycle." *The Economic Journal*, 121 (552), 379–401.
- Milani, Fabio (2017), "Sentiment and the U.S. business cycle." *Journal of Economic Dynamics and Control*, 82, 289–311.
- Nakamura, Emi and Jón Steinsson (2018), "High-frequency identification of monetary non-neutrality: the information effect." *The Quarterly Journal of Economics*, 133 (3), 1283–1330.
- Newey, Whitney K (1985), "Generalized method of moments specification testing." *Journal of econometrics*, 29 (3), 229–256.
- Newey, Whitney K and Daniel McFadden (1994), "Large sample estimation and hypothesis testing." *Handbook of econometrics*, 4, 2111–2245.
- Norets, Andriy and Xun Tang (2014), "Semiparametric inference in dynamic binary choice models." *Review of Economic Studies*, 81 (3), 1229–1262.
- Qu, Zhongjun and Denis Tkachenko (2012), "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models." *Quantitative Economics*, 3 (1), 95–132.
- Qu, Zhongjun and Denis Tkachenko (2017), "Global identification in dsge models allowing for indeterminacy." *The Review of Economic Studies*, 84 (3), 1306–1345.
- Rabanal, Pau (2007), "Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model." *Journal of Economic Dynamics and Control*, 31 (3), 906–937.
- Rabanal, Pau and Juan F Rubio-Ramírez (2005), "Comparing new keynesian models of the business cycle:

- A bayesian approach." *Journal of Monetary Economics*, 52 (6), 1151–1166.
- Robin, Jean-Marc and Richard J. Smith (2000), "Tests of rank." *Econometric Theory*, 16 (2), 151–175.
- Ryan, Stephen P (2012), "The costs of environmental regulation in a concentrated industry." *Econometrica*, 80 (3), 1019–1061.
- Santos, Carlos Daniel (2017), "Sunk costs of r&d, trade and productivity: The moulds industry case." *The Economic Journal*, 127 (603), 1626–1664.
- Schorfheide, Frank (2000), "Loss function-based evaluation of dsge models." *Journal of Applied Econometrics*, 15 (6), 645–670.
- Smets, Frank and Raf Wouters (2003), "An estimated dynamic stochastic general equilibrium model of the euro area." *Journal of the European Economic Association*, 1 (5), 1123–1175.
- Smets, Frank and Rafael Wouters (2007), "Shocks and frictions in us business cycles: A bayesian dsge approach." *American economic review*, 97 (3), 586–606.
- Suzuki, Junichi (2013), "Land use regulation as a barrier to entry: evidence from the texas lodging industry." *International Economic Review*, 54 (2), 495–523.
- Sweeting, Andrew (2013), "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry." *Econometrica*, 81 (5), 1763–1803.
- Varela, Mauricio J (2018), "The costs of growth: accelerated growth and crowd-out in the mexican super-market industry." *International Journal of Industrial Organization*, 61, 1–52.