Issues of Nonstandard Inference in Measurement Error Models

Kirill Evdokimov* Andrei Zeleneev†‡

November 18, 2019

[click here for the latest version]

Abstract

Models with errors-in-variables (EIV) often employ instrumental variable approaches to remove the EIV bias. This paper points out that in such models the issue of nonstandard inference can arise even when the instruments are strong. Moreover, this occurs at very important points of parameter space; for instance, when the coefficient on the mismeasured regressor in a nonlinear regression is close to zero. The root of the problem is weak identification of the nuisance parameters, such as the distribution of the measurement error or control variable. These parameters are weakly identified when the mismeasured variable has small effect on the outcomes. As a result, the estimators of the parameters of interest generally are not asymptotically normal and the standard tests and confidence sets can be invalid. We illustrate how this issue arises in several estimation approaches. This complication can be particularly problematic when the nuisance parameters are infinite-dimensional.

Making use of the specific structure of the EIV problem, the paper proposes simple approaches to conducting uniformly valid inference about the parameter of interest. The high-level conditions are illustrated by a detailed analysis of a semiparametric approach to EIV in the general moment condition settings.

Keywords: errors-in-variables, identification-robust inference, nonlinear models, GMM

*Universitat Pompeu Fabra and Barcelona GSE: kirill.evdokimov@upf.edu.

†Princeton University: zeleneev@princeton.edu.

‡We thank the participants of various seminars and conferences for helpful comments and suggestions. We are also grateful to the Gregory C. Chow Econometrics Research Program at Princeton University and the Department of Economics at the Massachusetts Institute of Technology for their hospitality and support. Evdokimov also acknowledges the financial support from the National Science Foundation via grant SES-1459993, and from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563).

1 Introduction

Measurement errors are present in many datasets. Correspondingly, the problem of errors-in-variables (EIV) has received a lot of attention in the econometrics literature. A variety of measurement error robust (MER) estimation approaches aim at removing the EIV bias. Among these approaches, the instrumental variables are the most common source of identification in practice. The importance of having strong instruments is also well recognised; when instruments are weak, even in large samples, the estimators may have non-gaussian distributions and the standard inference procedures generally fail.

This paper shows that for nonlinear models with EIV the issue of nonstandard inference arises even if the instruments are *strong*. This issue arises when the dependent variable is only weakly affected by the true mismeasured variables. This weak relationship then fails to provide the information necessary to separate the effects of the measurement error from the structural unobservables.

To illustrate this, consider a simple nonlinear regression model, where Y_i is the outcome, X_i^* is a covariate, and Z_i is an instrument. Suppose X_i^* is mismeasured, and we instead observe X_i:

$$\begin{aligned} Y_i &= \rho(\theta_{01} X_i^* + \theta_{02}) + U_i, \quad \mathbb{E}[U_i | X_i^*, Z_i] = 0, \qquad (1.1) \\ X_i &= X_i^* + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i] = 0, \end{aligned}$$

where $\rho(\cdot)$ is known nonlinear regression function, and $\theta_0 = (\theta_{01}, \theta_{02})'$ is the vector of unknown parameters of interest. For illustration, suppose that in addition,

$$X_i^* = \pi_0 Z_i + V_i, \quad V_i,\ \varepsilon_i, \text{ and } Z_i \text{ are mutually independent.} \qquad (1.2)$$

Suppose the instruments are strong, i.e., $|\pi_0|$ is bounded away from zero. Notice that we can write

$$X_i = \pi_0 Z_i + V_i + \varepsilon_i.$$

In order to remove the EIV bias in these settings, one needs to identify the distribution of ε_i or V_i (equivalently, X_i^*). The effects of these variables can be separated because Y_i depends on V_i through X_i^*, but is independent from ε_i. When $\theta_{01} = 0$, it is no longer possible to separate the impacts of V_i and ε_i.

A variety of MER estimators can be used to deal with the EIV bias in this example. Because $\rho(\cdot)$ is a nonlinear function, these MER methods estimate some nuisance parameters γ_0, such as the distribution (or some of its features, e.g., moments) of V_i and/or ε_i.[1] When $\theta_{01} = 0$, these nuisance parameters γ_0 are not identified. When θ_{01} is close to zero, γ_0 are weakly identified.

Section 2 illustrates how the issue of nonstandard inference manifests itself in several approaches to estimation of nonlinear models with EIV. The problem arises even in parametric models estimated by the Maximum Likelihood Estimator. When the nuisance parameters γ_0 are weakly or not identified, the estimators of the parameters of interest θ_0 are not approximately normally distributed

[1] Taken together, conditions (1.2) are excessively restrictive, so different MER approaches relax at least some of them. Different approaches involve estimation of different nuisance parameters. It is convenient to denote these nuisance parameters by γ_0, although the meaning and dimensionality of γ_0 will depend on the specific estimation method.

and may be inconsistent, and the standard inference procedures are generally invalid.[2,3]

This problem of nonstandard inference is not limited to the nonlinear regression models. For general models, we will use θ_{01} to denote the part of the vector of parameters θ_0 that controls identification of γ_0. The remaining components of θ_0 are denoted by θ_{02}. Let us denote the full set of parameters by $\beta_0 \equiv (\theta_0, \gamma_0)$.

Weak identification of the nuisance parameters is particularly problematic when the nuisance parameter γ_0 is infinite-dimensional. In this case, the derivations of the asymptotic properties of the estimators $\widehat{\theta}$ usually rely on the condition that the nonparametric estimator of γ_0 converges at a rate faster than $n^{-1/4}$. However, the latter condition is violated unless $\|\theta_{01}\|$ is sufficiently large.

We now turn to the constructive contributions of the paper. We leverage the particular structure of the EIV problem, which allows us to develop simple uniformly valid inference methods. These methods can be used together with any MER estimator that satisfies certain conditions on its behavior near $\theta_{01} = 0$. Importantly, we show that in these settings it is possible to avoid the more complicated methods of inference that are used in other settings with weakly identified nuisance parameters. Then, we provide a specific uniformly valid approach to estimation and inference in general moment condition problems with EIV.

The EIV problem has a particular structure that we exploit. Note that, on the one hand, when $\theta_{01} \approx 0$, the nuisance parameter γ_0 is weakly identified. On the other hand, the EIV bias (and the corresponding EIV bias correction) in this case is small, because it is typically proportional to θ_{01}.

First, we show that estimators similar to Non-Linear Instrumental Variables (NLIV) estimator in the nonlinear regression settings, have a "small bias" property when θ_{01} is small.[4] The idea is that for sufficiently small θ_{01} the model is approximately linear in $\theta_{01} X_i \approx 0$, so the instrumental variable regression addresses the EIV problem. As a result, we show that the NLIV-type estimators are asymptotically normal and unbiased as long as $\theta_{01} = o(n^{-1/4})$.[5]

At the same time, when θ_{01} is "sufficiently far" from zero, the MER approaches work as predicted by the "standard" asymptotic theory. For example, when the dimension of the nuisance parameter γ_0 is finite, some of the MER estimators can be shown to satisfy the assumptions of the general inference framework of Andrews and Cheng (2012, 2013, 2014) (henceforth AC12, AC13, and AC14). AC12 show that for such models, estimators $\widehat{\beta}$ have asymptotically normal distribution, and the standard tests about them are valid, as long as $\sqrt{n}\,\|\theta_{01}\| \to \infty$. The latter condition ensures that the amount of information about γ_0 increases with the sample size, so γ_0 can be consistently estimated. When the condition fails, $\widehat{\gamma}$ is inconsistent, and the distribution of $\widehat{\theta}$ is complicated (and depends on γ_0). To conduct uniformly valid inference about θ_0, AC12 provide procedures to numerically

[2]Throughout the paper, we assume the that the instruments are strong. More generally, for any MER approach, we focus on the settings in which the approach provides asymptotically normal and $\sqrt{n}$-consistent estimators of θ_0 when $\|\theta_{01}\|$ is bounded away from zero.

[3]Weak identification of γ_0 also complicates estimation and inference about average partial effects and other counterfactuals that involve expectations with respect to the distribution of X^*. Estimation of such quantities usually requires an explicit correction of the EIV bias, with the correction terms depending on γ_0.

[4]Even though NLIV estimator is well known to generally not be invalid in the EIV problem, i.e., when θ_{01} is not "small"; e.g., see Amemiya (1985).

[5]In contrast, naive MLE and NLLS estimators are asymptotically unbiased only if $\theta_{01} = o(n^{-1/2})$.

calculate the critical values for the standard test statistics. Confidence sets can then be constructed by the inversion of the tests.

Then, a test $\mathcal{T}_{\text{MER}}$ based on the MER approach can be combined with a NLIV-based test $\mathcal{T}_{\text{NLIV}}$ to provide inference that is uniformly valid for all values of θ_{01}. One simple test is the "Robust" test that rejects the null hypothesis only if both $\mathcal{T}_{\text{MER}}$ and $\mathcal{T}_{\text{NLIV}}$ reject it. This is a simple standard way of constructing a uniformly valid test, but such tests are conservative. The key for this and the following approaches to inference is that the validity regions of the MER and NLIV-based tests overlap.

We also provide more powerful "hybrid" tests, which switch between $\mathcal{T}_{\text{MER}}$ and $\mathcal{T}_{\text{NLIV}}$ depending on the estimated strength of identification of γ_0. For some MER approaches, it can be shown that $\widehat{\theta}$ is a uniformly $\sqrt{n}$-consistent estimator of θ_0 for all θ_{01}.[6] In such cases, we propose a hybrid test that coincides with $\mathcal{T}_{\text{MER}}$ when θ_{01} is "sufficiently large", with $\mathcal{T}_{\text{NLIV}}$ when θ_{01} is near zero, and smoothly links the two tests for the intermediate values of θ_{01}. Moreover, we construct a hybrid estimator that is uniformly asymptotically normal and unbiased. This is an exceptional situation, made possible by the availability of the NLIV-type estimators in the EIV settings. For example, no such uniformly asymptotically normal estimators and hybrid tests described above appear to be available for the general class of problems considered by AC12.

When the nuisance parameter γ_0 is infinite-dimensional the problem can be very complicated. (Semi-)Nonparametric EIV settings usually correspond to ill-posed inverse problems. Thus, we are dealing with a weakly identified infinite-dimensional nuisance parameter in an ill-posed inverse problem. Little is known about the properties of estimators in such problems. It is tempting to hope that the inference approaches, proven to work when the dimension of γ_0 is finite, will remain valid if the infinite-dimensional γ_0 is estimated using, e.g., a finite-dimensional sieve (series) approximation, as long as the sieve dimension grows sufficiently slowly with the sample size. However, one has to be cautious about this.

Next, we provide a uniformly valid estimation and inference approach for the general moment condition models with EIV. We show that the approach of Evdokimov and Zeleneev (2016), in a combination with the hybrid tests, provides uniformly valid approach to inference about parameters θ_0 and the marginal effects. We also show that their estimator remains $\sqrt{n}$-consistent uniformly over all values of θ_{01}, i.e., regardless of the identifiability of the nuisance parameters. We also verify the higher-level conditions in the example of nonlinear regression.

The issue of nonstandard inference, which this paper points out, is not limited to EIV models identified using instrumental variables. More generally, this issue is likely to arise when (i) the approach to EIV involves explicit estimation of some nuisance parameters γ_0, and (ii) estimation of these nuisance parameters relies on the effect of X_i^* on other variables, e.g., outcomes Y_i in the nonlinear regression. For instance, some estimators that rely on non-linearity or non-normality for identification can suffer from weak identification of γ_0.

[6]This is possible, because smaller values of θ_{01} reduce the amount of information about γ_0, but at the same time, for some MER approaches, reduce the impact of the estimation error $\widehat{\gamma} - \gamma_0$ on the estimator $\widehat{\theta}$.
In particular, MER estimators that satisfy the assumptions of AC12 are $\sqrt{n}$-consistent.

On the other hand, this issue of nonstandard inference does not arise in linear models estimated by IV regression, since one does not estimate the distribution of any unobservables in this case. This issue also does not arise when the nuisance parameter γ_0, e.g., the distribution of the measurement error, is identified using auxiliary information, such as repeated measurements (e.g., Section 2 of Hausman, Ichimura, Newey, and Powell (1991),[7] Li (2002), Schennach (2004)) or auxiliary/validation data (e.g., Chen, Hong, and Tamer (2005)). In these cases one does not rely on the effect of X_i^* on Y_i to identify the nuisance parameter γ_0.

Chen, Hong, and Nekipelov (2011) and Schennach (2013, 2016) provide excellent overviews of the MER approaches in nonlinear models.

Measurement Error Robust approaches that use instrumental variables for identification include Hausman et al. (1991); Hausman, Newey, and Powell (1995); Newey (2001); Schennach (2007); Hu and Schennach (2008); Wang and Hsiao (2011); Evdokimov and Zeleneev (2016). It is important to note that the regularity conditions in all of these papers do implicitly rule out the case of $\theta_{01} = 0$, e.g., because some rank conditions fail when $\theta_{01} = 0$. Our paper brings attention to the importance of complementing these analyses with a uniformly valid asymptotic analyses that allow for θ_{01} to be small.

When the dimension of the nuisance parameter γ_0 is finite, the settings fall into the general class of problems of inference in models with weakly identified parameters. Stock and Wright (2000), Kleibergen (2005), Guggenberger and Smith (2005), Guggenberger, Ramalho, and Smith (2012), Andrews and Mikusheva (2016), Andrews (2016) develop tools for uniformly valid inference on the entire vector of parameters β_0 in the general GMM setting. The approaches proposed in these papers allow for concentrating out strongly identified nuisance parameters. However, when the nuisance parameter γ_0 is weakly identified, one needs to combine the tests on the full vector β_0 with the projection methods (Dufour, 1989; Dufour and Jasiak, 2001; Dufour and Taamouti, 2005) in order to test hypotheses about and construct confidence sets for the parameter of interest θ_0. Such projection based tests are known to be asymptotically conservative and not efficient under strong identification. Another strand of the literature utilizes the Bonferroni correction method to construct uniformly valid subvector inference procedures when the nuisance parameter is (potentially) weakly identified (see, for example, Chaudhuri and Zivot, 2011; McCloskey, 2017; Andrews, 2017 among others). Notably, refining the approach of Chaudhuri and Zivot (2011), Andrews (2017) constructs identification robust subvector tests and confidence sets, which are asymptotically non-conservative and efficient under strong identification. Unfortunately, in some settings, these procedures may be difficult to implement and the high-level assumptions may be hard to verify. Another related stream of the literature featuring AC12, AC13, AC14, Cheng (2015), Cox (2017), Han and McCloskey (2019) exploit the knowledge of which parameters are (potentially) weakly identified and which parameters control their strength of identification. In this respect, their frameworks are similar to the EIV

[7]Hausman et al. (1991) includes two approaches to the EIV. In Section 2 they consider identification and estimation using repeated measurements. In Section 3 they consider instrumental variables strategy. As a result, the estimator from Section 2 does not suffer from nonstandard inference problem, but the estimator from Section 3 does.

setting, in which θ_{01} controls the strength of identification of γ_0. The inference procedures developed in these papers are asymptotically non-conservative and efficient under strong identification but their implementation involves computation of the least favorable critical values, which can be difficult especially when γ_0 is not a scalar. Other recent work on optimal testing with nuisance parameters also includes Elliott, Müller, and Watson (2015), Montiel Olea (2019), and Moreira and Moreira (2013), who construct tests maximizing weighted average power.

Little is known about inference in semiparametric models with weakly identified infinite-dimensional nuisance parameters. The approach of Newey (1994), who establishes asymptotic normality of the finite-dimensional parameters, cannot be used. Formally, the approach cannot be used because the estimators of the nuisance parameters do not converge at a rate faster than $n^{-1/4}$. More importantly, the estimators $\widehat{\theta}$ have non-gaussian distributions when $\theta_{01} \approx 0$, and hence one cannot hope to fix the problem by finding some alternative regularity conditions for Newey (1994). Uniformly valid inference about θ_0 and its components could possibly be provided by the projection method applied to a semi-nonparametric analog of the S-test, following Chen and Pouzo (2015) and Chernozhukov, Newey, and Santos (2015). However, this approach is likely to yield very conservative inferences.

The rest of the paper is organized as follows. Section 2 illustrates the problem of weak identification of the nuisance parameters for several MER estimators. Section 3.1 shows that the NLIV-type estimators have a low-bias property that other naive estimators do not have. Section 3.2 introduces several simple approaches to inference that are uniformly valid regardless of the strength of identification of the nuisance parameter γ_0. Section 4.1 introduces the Moderate Measurement Error estimator (MME) for general moment condition models with EIV. Section 4.2 provides a simple exposition of how the properties of such estimators change depending on the magnitude of $\sqrt{n}\|\theta_{01}\|$. Section 5 presents Monte Carlo experiments that illustrate the properties of the proposed inference procedures. Section 6 develops the large sample properties of the MME estimator. Section 7 formally establishes uniform validity of the proposed inference approaches. All proofs are collected in the Appendix.

Notation: All vectors are columns. For some generic parameter vector β and a vector valued function $a(x, s, \beta)$ and , let $a_i(\beta) \equiv a(X_i, S_i, \beta)$, $\overline{a}(\beta) = n^{-1}\sum_{i=1}^n a_i(\beta)$, $a(\beta) \equiv \mathbb{E}[a_i(\beta)]$. Let $\hat{\Omega}_{aa}(\beta) \equiv \overline{a}(\beta) = n^{-1}\sum_{i=1}^n a_i(\beta)a_i(\beta)'$ and $\Omega_{aa}(\beta) \equiv \mathbb{E}[a_i(\beta)a_i(\beta)']$. For the true value of the parameter β_{0n} we often write $\overline{a} \equiv \overline{a}(\beta_{0n})$, $a \equiv a(\beta_{0n})$, $\hat{\Omega}_{aa} \equiv \hat{\Omega}_{aa}(\beta_{0n})$, $\Omega_{aa} \equiv \Omega_{aa}(\beta_{0n})$. Let $\lambda_{\min}(M)$ and $\lambda_{max}(M)$ stand for the smallest and largest eigenvalues of a symmetric matrix M.

2 Illustration of the Issue

In this section we illustrate how the problem of weak identification of the nuisance parameters affects the properties of several EIV-robust estimators. To keep this section concise, we provide only provide brief descriptions of the estimators, leaving more detailed descriptions to later sections and the Appendix.

2.1 Semiparametric Control Variable Approach

Hausman et al. (1991); Newey (2001); Schennach (2007); Wang and Hsiao (2011) develop estimators for the NLR model. To simplify the presentation, we consider the model without additional covariates W_i. The papers make use the following assumptions:

$$Y_i = \rho\left(X_i^*, \theta_0\right) + U_i, \qquad \mathbb{E}\left[U_i | \mathcal{Z}_i, V_i\right] = 0, \tag{2.1}$$

$$X_i = X_i^* + \varepsilon_i, \qquad \mathbb{E}\left[\varepsilon_i | \mathcal{Z}_i, V_i, U_i\right] = 0, \tag{2.2}$$

$$X_i^* = \pi' \mathcal{Z}_i + V_i, \qquad V_i \perp \mathcal{Z}_i. \tag{2.3}$$

Only $(Y_i, X_i, \mathcal{Z}_i')$ are observed. Equations (2.1) and (2.2) are exogeneity conditions on U_i and ε_i. Note that the assumption on the measurement error ε_i is relatively weak, for example, it allows ε_i to be conditionally heteroskedastic. Condition (2.3) is restrictive, as it requires additivity and full independence of V_i. The location of V_i needs to be normalized, so one assumes $\mathbb{E}\left[V_i\right] = 0$. These assumptions do not impose any parametric restrictions on the distributions of the unobservables.

Under these assumptions, coefficients π can be estimated by the linear regression of X_i on $\mathcal{Z}_i$. For simplicity of exposition, let us view π as known and define $Z_i \equiv \pi' \mathcal{Z}_i$. Then, we can write

$$X_i^* = Z_i + V_i, \qquad \mathbb{E}\left[V_i\right] = 0.$$

The analysis in this framework was pioneered by the seminal papers of Hausman et al. (1991) and Newey (2001). Our exposition follows Newey (2001), who observes that the following conditional moment restrictions hold:

$$\begin{aligned}
\mathbb{E}\left[\left.\begin{pmatrix} Y_i \\ Y_i X_i \end{pmatrix}\right| Z_i = z\right] &= \mathbb{E}\left[\left.\begin{pmatrix} \rho\left(Z_i + V_i, \theta_0\right) + U_i \\ \left(\rho\left(Z_i + V_i, \theta_0\right) + U_i\right)\left(Z_i + V_i + \varepsilon_i\right) \end{pmatrix}\right| Z_i = z\right] \\
&= \int \rho\left(z + v, \theta_0\right) \begin{pmatrix} 1 \\ z + v \end{pmatrix} f_V(v)\, dv,
\end{aligned} \tag{2.4}$$

where $f_V(\cdot)$ is the (unknown) density of V_i. Here, the nuisance parameter $\gamma_0 \equiv f_V(\cdot)$ is infinite-dimensional.

Hausman et al. (1991) consider polynomial specifications of function $\rho(x; \theta)$, in which case the right-hand side of equation (2.4) simplifies and becomes a set of equations that contain θ_0 and the nuisance parameters $\gamma_0 \equiv \left(\mathbb{E}\left[V_i^2\right], \ldots, \mathbb{E}\left[V_i^{p_{\text{poly}}+1}\right]\right)'$, where p_{poly} is the order of the polynomial $\rho(x; \theta)$. Hausman et al. (1991) propose jointly estimating θ_0 and γ_0. To illustrate, consider the quadratic model: $\rho(x, \theta_0) = \theta_{01} x^2 + \theta_{02,1} x + \theta_{02,2}$. Then, equation (2.4) becomes

$$\mathbb{E}\left[\left.\begin{pmatrix} Y_i \\ Y_i X_i \end{pmatrix}\right| Z_i = z\right] = \theta_{01} \begin{pmatrix} z^2 + \mathbb{E}\left[V^2\right] \\ z^3 + \mathbb{E}\left[V^3\right] + 3zE\left[V^2\right] \end{pmatrix} + \theta_{02,1} \begin{pmatrix} z \\ z^2 + \mathbb{E}\left[V^2\right] \end{pmatrix} + \theta_{02,2} \begin{pmatrix} 1 \\ z \end{pmatrix}.$$

These conditional moment restrictions can be used to estimate (θ_0, γ_0).

Note that when $\theta_{01} = 0$ the nuisance parameter $\mathbb{E}[V_i^3]$ cannot be identified, and hence an estimator that attempts to jointly estimate (θ_0, γ_0) may have nonstandard properties.

Newey (2001) suggests that the conditional moment restrictions (2.4) can identify θ_0 for general nonlinear functional forms of ρ.[8] Newey (2001) proposes estimating θ_0 jointly with the infinite dimensional nuisance parameter f_V using a series estimator of f_V.

Suppose the regression function has the structure $\rho(x^*, \theta_0) \equiv \rho(\theta_{01}x^*, \theta_{02})$, e.g., as in the GLM model of equation (1.1). When $\theta_{01} = 0$, the moment conditions (2.4) become

$$\mathbb{E}\left[\left.\begin{pmatrix} Y_i \\ Y_i X_i \end{pmatrix}\right| Z_i = z\right]\Bigg|_{\theta_{01}=0} = \int \rho(0, \theta_{02}) \begin{pmatrix} 1 \\ z+v \end{pmatrix} f_V(v)\, dv = \rho(0, \theta_{02}) \begin{pmatrix} 1 \\ z \end{pmatrix}.$$

The right-hand side of the equation does not depend on f_V, i.e., the nuisance parameter f_V is not identified. Let us illustrate the properties of these estimators in finite samples.

Example 1 (Logit MC)**.** Consider the following design for Monte Carlo experiments. The model is logit binary choice:

$$\begin{aligned} P(Y_i = 1 | X_i^*) &= \rho(\theta_{01}X_i^* + \theta_{02}), \quad \rho(\cdot) \equiv 1/(1+\exp(-\cdot)) && (2.5)\\ X_i^* &= Z_i + V_i, \quad X_i = X_i^* + \varepsilon_i, && (2.6)\\ (Z_i, V_i, \varepsilon_i)' &\sim N\left((0,0,0)', Diag\left(\sigma_Z^2, \sigma_{V0}^2, \sigma_{\varepsilon 0}^2\right)\right). && (2.7) \end{aligned}$$

We take $(\theta_{02}, \sigma_Z^2, \sigma_{V0}^2, \sigma_{\varepsilon 0}^2) = (1,1,1,1)$ and $n = 1000$. In the regression of X_i on Z_i the average value of the F-statistic is 500.

To sidestep the issues of choosing the number of approximating series terms in nonparametric estimation of f_V, we consider a simplified version of Newey (2001) estimator. We assume that it is known that $V_i \sim N(0, \sigma_{V0}^2)$, so that f_V is known up to the scalar parameter σ_{V0}. Then, $(\theta', \sigma_V)'$ are estimated using 2-step GMM with the following moment conditions:

$$\psi_i(\theta, \sigma_V) \equiv \begin{pmatrix} \int \rho(\theta_1(Z_i + \sigma_V\eta) + \theta_2)(Z_i + \sigma_V\eta)\phi(\eta)\, d\eta - Y_i X_i \\ \int \rho(\theta_1(Z_i + \sigma_V\eta) + \theta_2)\phi(v)\, d\eta - Y_i \end{pmatrix} \otimes \begin{pmatrix} 1 \\ Z_i \\ Z_i^2 \\ Z_i^3 \end{pmatrix}, \qquad (2.8)$$

where $\phi(\eta) \equiv (2\pi)^{-1/2}\exp(\eta^2/2)$, and we treat Z_i as known for brevity. In addition, Newey (2001) uses simulation to compute the integrals. To avoid discussing the impact of simulation, in this simple model we instead use quadratures to compute the integral.

Figure 1 presents histograms of the distribution of $\hat{\theta}$ and $\hat{\sigma}_V$ for $\theta_{01} \in \{1, 0.05, 0.0\}$. When θ_{01} is large, the estimator $\hat{\theta}$ performs as predicted by the standard asymptotic theory: it has approximately normal distribution centered around the true value of the parameter. When θ_{01} is small, σ_V is not identified and $\hat{\theta}$ does not have a normal distribution. Note however that the distribution of $\hat{\theta}$ is still concentrated around the true parameter value.

[8] Schennach (2007) studies nonparametric identification of this model.

Next, consider the quadratic model. In the simulations, we take $(\theta_{02,1}, \theta_{02,2}, \sigma_U^2, \sigma_{\varepsilon 0}^2, \sigma_Z^2) = (0, 0, 1, 1, 1)$, $n = 1000$, and consider the estimator of Hausman et al. (1991). Figure 2 presents the results. The results are qualitatively similar to the Logit model estimator.

Remark 1. Since Hausman et al. (1991) consider only polynomial specifications of $\rho(\cdot)$, their assumptions only require the relevant implications of the conditions (2.1)-(2.3) for the polynomial settings.

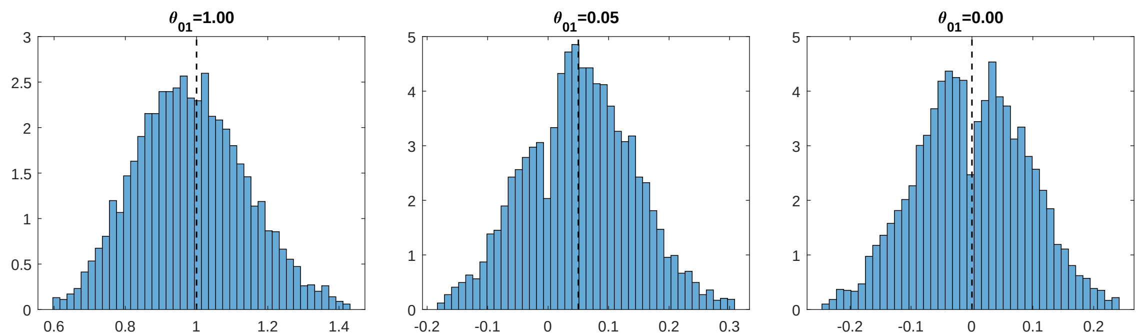


Figure 1: Logit Model estimated using Newey (2001) estimator.

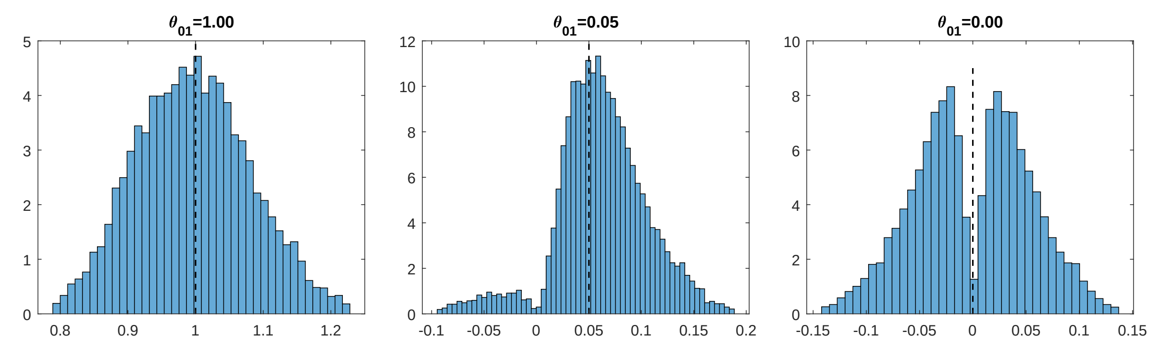


Figure 2: Polynomial Model.

2.2 Moderate Measurement Error Approach

Consider the Moderate Measurement Error (MME) approach of Evdokimov and Zeleneev (2016). We describe this approach in detail in Section 4. For now, let us note that MME approach assumes that the measurement error ε_i has features of classical measurement error in that $\mathbb{E}[\varepsilon_i^k | X_i^*] = \mathbb{E}[\varepsilon_i^k]$ for $k = 2, \ldots, K$, for some $K \geq 2$, and $\mathbb{E}[\varepsilon_i | X_i^*] = 0$. Also, it is assumed that the finite sample properties of the problem are approximated by viewing the magnitudes of $\mathbb{E}\left[|\varepsilon_i|^k\right]$ as moderate or small, relative to the sample size. The latter condition captures the situations where researchers expect the magnitude of the EIV bias to be comparable to the magnitude of the standard error (as opposed to the standard approach approach to measurement errors, which implies that in large samples the standard errors are negligible compared to the EIV bias). The MME estimator is then able to remove the EIV bias in the general moment conditions settings, including the nonlinear regression models. For the MME estimator, the nuisance parameters γ_0 are, essentially, the first few moments of the measurement error $\mathbb{E}[\varepsilon_i^2], \ldots, \mathbb{E}[\varepsilon_i^K]$, e.g., $K = 2$ or $K = 4$.

Note that the assumptions of the MME approach are substantially different from those in the previous section, aside from both methods assuming that $\mathbb{E}[\varepsilon_i | X_i^*] = 0$. Newey (2001) estimates the distribution of V_i (which requires additivity and independence of V_i), but does not impose any restrictions on the higher moments of ε_i. For example, the variance of the measurement error can depend on X_i^*. MME approach estimates moments of ε_i (which requires conditional moments of ε_i to not depend on X_i^*), but does not restrict or estimate the relationship between X_i^* and V_i, allowing $X_i^* = h(Z_i, V_i)$ for a general unknown function h and multivariate unobserved V_i. This in particular implies that the nuisance parameters γ_0 for the two approaches are conceptually different.

When X_i^* does not affect the outcomes Y_i, the moments $\mathbb{E}[\varepsilon_i^k]$ cannot be identified. Figure 2 presents the distribution of the MME estimator $\hat{\theta}$ (with $K = 4$) in the Logit Monte Carlo design of the previous section. The distributions of $\hat{\theta}_1$ are qualitatively similar to those in Section 2.1, despite the differences between the estimation approaches.

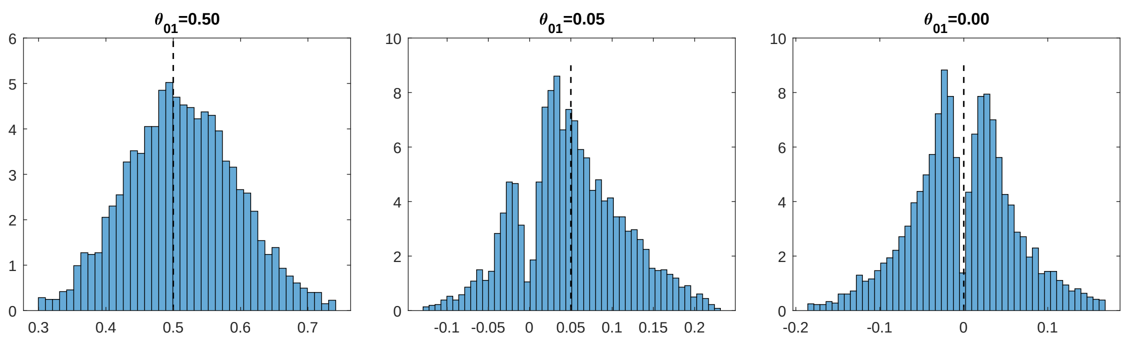


Figure 3: Logit Model estimated using MME Estimator.

2.3 MLE

Finally, to illustrate that the issue is not limited to semiparametric models, we consider MLE estimation of the linear regression model with gaussian structural and measurement errors:

$$\begin{cases} Y_i = \theta_{01} X_i^* + U_i, \\ X_i = X_i^* + \varepsilon_i, \\ X_i^* = Z_i + V_i; \end{cases} \qquad \text{where} \begin{pmatrix} U_i \\ V_i \\ \varepsilon_i \\ Z_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{U0}^2 & 0 & 0 & 0 \\ 0 & \sigma_{V0}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon 0}^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right).$$

We take $\sigma_{U0} = \sigma_{V0} = \sigma_{\varepsilon 0} = 1$, $n = 1000$, and estimate parameters $(\theta_{01}, \sigma_{U0}, \sigma_{V0}, \sigma_{\varepsilon 0})'$ by MLE. Here, the nuisance parameter is $\gamma_0 = (\sigma_{V0}^2, \sigma_{\varepsilon 0}^2)'$. Even though the sum $\sigma_{V0}^2 + \sigma_{\varepsilon 0}^2 = V[X_i - Z_i]$ is always identified, σ_{V0} and $\sigma_{\varepsilon 0}$ are not separately identified when $\theta_{01} = 0$.

The results of the Monte Carlo experiment are presented in Figure 4.

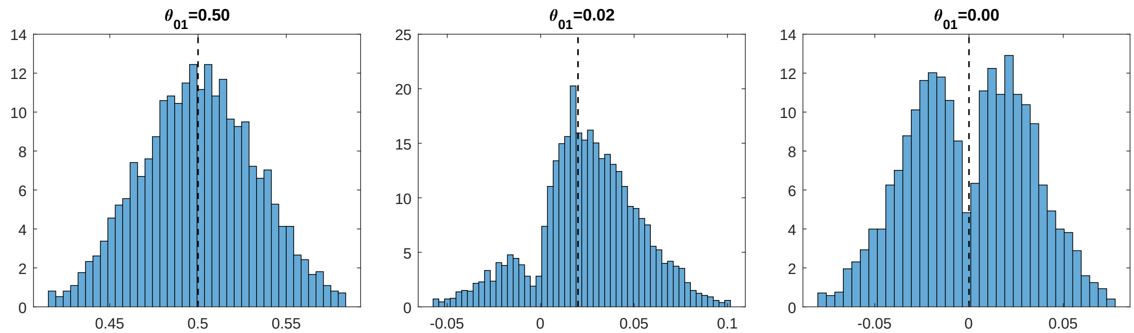


Figure 4: Linear Gaussian MLE

3 Simple Solutions

In Section 3.1 we consider the properties of “naive” or “uncorrected” estimators, such as NLLS, ML, GMM, and NLIV estimators. We focus on the case of θ_{01n} near zero. In this case, the impact of the measurement error and the EIV bias are shown to be small, although generally non-ignorable. Importantly, it turns out that NLIV(-type) estimators have much smaller biases than NLLS and MLE estimators when $|\theta_{01n}|$ is small. As a result, NLIV-type estimators are asymptotically unbiased for a wider range of values of θ_{01n} near zero.

On the other hand, when $|\theta_{01n}|$ is not too small, the usual MER estimators and inference methods may be expected to work in accordance with the existing pointwise asymptotic theory results. If the regions of validity of the NLIV-type and the MER methods overlap, the two methods can be linked to construct tests that are valid uniformly for all values of θ_{01n}. We provide several such approaches in Section 3.2.

3.1 Properties of Naive Estimators When θ_{01n} Is Small

In this section, we consider a general model of measurement error, assuming that

$$X_i = X_i^* + \varepsilon_i, \qquad \mathbb{E}\left[\varepsilon_i | X_i^*, S_i\right] = 0. \tag{3.1}$$

This condition is weaker than the assumptions made in, e.g., Newey (2001), Schennach (2007), or Evdokimov and Zeleneev (2016). Neither the additive control variable structure, nor the assumption of classical (or even conditionally homoskedastic) measurement error need to hold. This condition alone is not sufficient for identification or estimation of models with measurement errors. We will consider general moment condition models, using the following nonlinear regression example for illustration.

Example GLM. *We will pay particular attention to the GLM model*

$$Y_i = \rho\left(\theta_{01n} X_i^* + \theta_{02}' W_i\right) + U_i, \qquad \mathbb{E}[U_i | X_i^*, W_i, Z_i], \tag{3.2}$$

where $\rho(\cdot)$ is a smooth and typically monotone function. For example, in a binary choice model, $\rho(\cdot)$ is a CDF of some smooth distribution.

Many standard estimators correspond to the moment condition model with the moments of the form

$$g\left(X_i, W_i, Y_i, \theta\right) = \left(Y_i - \rho(\theta_1 X_i + \theta_2' W_i)\right) \eta\left(\theta_1 X_i + \theta_2' W_i\right) h_i, \tag{3.3}$$

where v is some function, and $h_i \equiv h\left(X_i, W_i, Z_i\right)$ is a vector.

For example, consider the NLLS estimator $\hat{\theta}_{\text{NLLS}} \equiv \operatorname{argmin}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \rho(\theta_1 X_i + \theta_2' W_i)\right)^2$ for estimation in the GLM Example. It is equivalent to the moment condition (3.3) with

$$\eta_{\text{NLLS}}\left(\cdot\right) \equiv \rho'\left(\cdot\right), \text{ and } h_{\text{NLLS},i} \equiv \left(X_i, W_i'\right)'.$$

Consider the binary choice model with $P\left(Y_i = 1 | X_i^*, W_i\right) = \rho(\theta_{01n} X_i^* + \theta_{02}' W_i)$, e.g., Probit or Logit. The score of the MLE estimator in this model has the form of equation (3.3) with

$$\eta_{\text{BinChMLE}}\left(\cdot\right) \equiv \frac{\rho'\left(\cdot\right)}{\rho\left(\cdot\right)\left(1 - \rho\left(\cdot\right)\right)}, \text{ and } h_{\text{BinChMLE},i} \equiv \left(X_i, W_i'\right)'.$$

In equation (3.3), X_i enters in two ways: as a product $\theta_1 X_i$, and by itself as a part of the instrument vector h_i. Thus, we consider general moment conditions of the form $q(\theta_1 x, x, \theta_2, s)$, which identify θ_{0n} if we have correctly measured X_i^*:

$$\mathbb{E}\left[q(\theta_{01n} X_i^*, X_i^*, \theta_{02}, S_i)\right] = 0. \tag{3.4}$$

We will consider moment restrictions q that satisfy the following additional condition:

$$\mathbb{E}\left[q(\theta_{01n} X_i^*, X_i, \theta_{02}, S_i)\right] = 0. \tag{3.5}$$

Notice that we have X_i instead of X_i^* in the second argument.

Condition (3.5) is unusual, but holds under weak conditions in the nonlinear regression settings. Consider the GLM model of equation (3.2). To verify condition (3.5), we need to make an additional assumption:

$$\mathbb{E}\left[U_i | X_i^*, W_i, \varepsilon_i\right] = 0.$$

The regression error term is mean independent of the measurement error. Some form of independence between the regression and measurement errors is usually necessary, and is made by Hausman et al. (1991), Newey (2001), Schennach (2007), Hu and Schennach (2008), and Evdokimov and Zeleneev (2016), among others. Then, for any function b, by the law of iterated expectations

$$\mathbb{E}\left[\left(Y_i - \rho(\theta_{01n} X_i^* + \theta_{02}' W_i)\right) b\left(X_i, W_i, \theta_{0n}\right)\right] = \mathbb{E}\left[\mathbb{E}\left[U_i | X_i^*, W_i, \varepsilon_i\right] b\left(X_i^* + \varepsilon_i, W_i, \theta_{0n}\right)\right] = 0,$$

so equation (3.5) holds. Note that the above arguments hold for all $\theta_{01n} \in \Theta$ and do not restrict θ_{01n} to be small.

Consider the GMM estimator with weighting matrix $\hat{\Xi}$

$$\hat{\theta}_q = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \overline{q}_n(\theta)' \hat{\Xi} \overline{q}_n(\theta). \tag{3.6}$$

Generally, estimators $\hat{\theta}_q$ are biased in the presence of the measurement error, since $\mathbb{E}\left[q(\theta_{01n}X_i, X_i, \theta_{02}, S_i)\right] \neq 0$. We will now consider the properties of $\hat{\theta}_q$ when θ_{01n} is small, i.e., $|\theta_{01n}| = o_n(1)$.

Remember that naive estimators typically do not suffer from the EIV bias when θ_{01n} is exactly zero. It turns out that all estimators with q satisfying equation (3.5) have this property. When $\theta_{01n} = 0$,

$$\mathbb{E}\left[q(\theta_{01n}X_i, X_i, \theta_{02}, S_i)\right]\big|_{\theta_{01n}=0} = \mathbb{E}\left[q(\theta_{01n}X_i^*, X_i, \theta_{02}, S_i)\right]\big|_{\theta_{01n}=0} = 0, \tag{3.7}$$

where the first equality holds because $\theta_{01n}X_i = 0 = \theta_{01}X_i^*$, and the second equality holds by equation (3.5).

Let $\overline{\theta}_{0n}$ denote the "pseudo-true" value to which $\hat{\theta}_q$ converges, i.e., $\hat{\theta}_q - \overline{\theta}_{0n} = o_{p,n}(1)$. Here $\overline{\theta}_{0n}$ is a function of the true parameter θ_{0n}. Equation (3.7) implies that when $\theta_{01n} = 0$, the pseudo-true and true values coincide, i.e., $\overline{\theta}_{0n} = \theta_{0n}$. Thus, typically, when θ_{01n} is close to 0, $\overline{\theta}_{0n}$ is close to θ_{0n}, i.e., $\overline{\theta}_{0n} - \theta_{0n} = o_n(1)$ when $\theta_{0n1} = o_n(1)$.

Denote $Q(\theta) \equiv \mathbb{E}\left[\nabla_\theta q_i(\theta)\right]$. Let $\Omega_{qq}(\theta)$ be the covariance matrix of $q_i(\theta)$, and let $Q \equiv Q(\theta_{0n})$ and $\Omega_{qq} \equiv \Omega_{qq}(\theta_{0n})$.

Theorem 1. *Suppose $|\theta_{01n}| = o_n(n^{-\delta})$ for a $\delta > 0$, and the moment conditions $q_i(\theta)$ satisfy equations (3.4) and (3.5). Suppose $\overline{\theta}_{0n} - \theta_{0n} = o_n(1)$, $\hat{\Xi} - \Xi = o_{p,n}(1)$, $\lambda_{\min}(Q'\Xi Q)$ is bounded away from zero. Finally, suppose the regularity conditions stated in Assumption* **??** *in the Appendix hold. Then*

$$\sqrt{n}\left(\hat{\theta}_q - \overline{\theta}_{0n}\right) \to_d N(0, \Sigma_q), \quad \Sigma_q \equiv (Q'\Xi Q)^{-1} Q'\Xi\Omega_{qq}\Xi Q (Q'\Xi Q)^{-1},$$

$$\begin{aligned}\overline{\theta}_{0n} &= \theta_{0n} + B_q\theta_{01n} + O_n\left(\theta_{01n}^2\sigma^2\right),\\ B_q &\equiv -(Q'\Xi Q)^{-1} Q'\Xi \times E\left[\nabla_{a_1} q(\theta_{01n}X_i^*, X_i, \theta_{02n}, S_i)\varepsilon_i\right],\end{aligned}$$

where $\nabla_{a_1} q$ is the derivative of function q with respect to the first argument.

The bias of the naive estimator $\hat{\theta}_q$ can be ignored when $\sqrt{n}\left(\overline{\theta}_{0n} - \theta_{0n}\right) = o_n(1)$, since in this case $\sqrt{n}\left(\hat{\theta}_q - \theta_{0n}\right) \to_d N(0, \Sigma_q)$ and the standard tests and confidence intervals are valid. The key implication of Theorem 1 is that B_q is generally of order 1, and hence the bias $\sqrt{n}B_q\theta_{01n}$ is negligible only if $\theta_{01n} = o\left(n^{-1/2}\right)$, i.e., is very small.

Remark 2. We stress that the general model we consider is *globally* misspecified, in the sense that the variance of the measurement error can be large, and that for most of the values of $\theta_{0n} \in \Theta$ the bias of $\hat{\theta}_q$ is of order one. The Theorem proves that if the moment conditions satisfy the additional property (3.5), the globally misspecified model behaves as if it were locally misspecified when θ_{01n} is small. Once this has been established, the derivation of the bias expression is standard.

Remark 3. Note that $\Omega_{qq}(\theta_{0n}) - \Omega_{qq}\left(\overline{\theta}_{0n}\right) = o_n(1)$ and $Q(\theta_{0n}) - Q\left(\overline{\theta}_{0n}\right) = o_n(1)$. Thus, $\lambda_{\min}(Q'\Xi Q)$ is bounded away from zero whenever $\lambda_{\min}\left(Q\left(\overline{\theta}_{0n}\right)' \Xi Q\left(\overline{\theta}_{0n}\right)\right)$ is. Matrices Ω_{qq} and Q

can be estimated by the standard sample Jacobian and covariance matrix estimators $\hat{\Omega}_{qq}\left(\hat{\theta}_q\right)$ and $\overline{Q}_n\left(\hat{\theta}_q\right)$.

Remark 4. The bias term B_q cannot be estimated without further assumptions because ε_i is not observed.

Remark 5. When the variance σ^2 of the measurement error is small, the bias term $B_q\theta_{01n}$ is of order $O\left(|\theta_{01n}|\,\sigma^2\right)$. We investigate the implications of this in Sections 4 and 6. There, we also consider general estimators that do not need to satisfy condition (3.5).

Non-Linear IV estimator Suppose we have an instrument, and consider Non-Linear IV (NLIV) estimators in the GLM model. NLIV corresponds to the moment conditions

$$g_{\mathrm{NLIV},i}\left(\theta\right) = \left(\rho(\theta_1 X_i + \theta_2' W_i) - Y_i\right)\varphi\left(Z_i, W_i, \theta\right)$$

where $\varphi\left(z, w, \theta\right)$ is a vector of length p or larger. Equations (3.4) and (3.5) are satisfied for the moments $q_i\left(\theta\right) \equiv g_{\mathrm{NLIV},i}\left(\theta\right)$, and Theorem 1 applies.

It turns out that this naive NLIV estimator has much lower EIV bias than NLLS and MLE when θ_{01n} is small. For the NLIV, the moment conditions depend on X_i only through the product $\theta_1 X_i$, i.e., moment function $q_{\mathrm{NLIV}}(\theta_1 x, x, \theta_2, s)$ does not depend on its second argument. As a result, $B_{q,\mathrm{NLIV}} = 0$, since

$$\mathbb{E}\left[\nabla_{a_1} q_{\mathrm{NLIV}}(\theta_{01n}X_i^*, X_i, \theta_{02n}, S_i)\varepsilon_i\right] = \mathbb{E}\left[\nabla_{a_1} q_{\mathrm{NLIV}}(\theta_{01n}X_i^*, \theta_{02n}, S_i)\mathbb{E}\left[\varepsilon_i | X_i^*, S_i\right]\right] = 0.$$

Thus, $\overline{\theta}_{0n}^{\mathrm{NLIV}} = \theta_{0n} + O_n\left(\theta_{01n}^2\sigma^2\right)$, and hence

$$\sqrt{n}\left(\hat{\theta}_{\mathrm{NLIV}} - \theta_{0n}\right) \to_d N\left(0, \Sigma_{\mathrm{NLIV}}\right), \quad \text{if } \theta_{01n} = o_n\left(n^{-1/4}\right).$$

To sum up, NLIV estimator stands out among the naive estimators that do not correct for the presence of measurement error. The bias of NLIV estimator is negligible when $\theta_{01n} = o_n\left(n^{-1/4}\right)$, i.e., for a wider range of the true parameter θ_{01n}.

3.2 Simple Robust Inference

3.2.1 The Setup

Given the properties of the naive estimators, we may be able to construct inference procedures that are uniformly valid for all values of θ_{01}. Suppose we have two estimators (or inference procedures): one that is valid for the "larger" values of $|\theta_{01n}|$ and one that is valid when $|\theta_{01n}|$ is small. For example:

- $\hat{\theta}_{\mathrm{L}}$ ("Larger" values of θ_{01n}), which satisfies $\sqrt{n}\hat{\Sigma}_{\mathrm{L}}^{-1/2}\left(\hat{\theta}_{\mathrm{L}} - \theta_{0n}\right) \to_d N\left(0, 1\right)$ for $n^{\omega_{\mathrm{L}}}\left|\theta_{01}\right| \to \infty$ for an $\omega_{\mathrm{L}} \in (0, 1/2)$.

- $\hat{\theta}_{\mathrm{NZ}}$ ("Near Zero"), which satisfies $\sqrt{n}\hat{\Sigma}_{\mathrm{NZ}}^{-1/2}\left(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}\right) \to_d N\left(0,1\right)$ for $n^{\omega_{\mathrm{NZ}}}\left|\theta_{01}\right| \to 0$ for an $\omega_{\mathrm{NZ}} > 0$.

Here $\hat{\Sigma}_{\mathrm{L}}$ and $\hat{\Sigma}_{\mathrm{NZ}}$ are estimators of the asymptotic variances of $\hat{\theta}_{\mathrm{L}}$ and $\hat{\theta}_{\mathrm{NZ}}$.

Condition $n^{\omega_{\mathrm{L}}}\left|\theta_{01n}\right| \to \infty$ defines a region of true parameters θ_{01n}, for which estimator $\hat{\theta}_{\mathrm{L}}$ has the standard behavior, i.e., is asymptotically normal and unbiased. Estimator $\hat{\theta}_{\mathrm{L}}$ corresponds to one of the Measurement-Error Robust (MER) estimators. Condition $n^{\omega_{\mathrm{NZ}}}\left|\theta_{01n}\right| \to 0$, i.e., $\left|\theta_{01n}\right| = o\left(n^{-\omega_{\mathrm{NZ}}}\right)$ defines a region of parameter values θ_{01n} for which $\hat{\theta}_{\mathrm{NZ}}$ estimator has the standard properties. This region in particular includes $\theta_{01n} = 0$. As we have seen above, naive estimators can satisfy the conditions imposed on $\hat{\theta}_{\mathrm{NZ}}$. For example, we can take $\hat{\theta}_{\mathrm{NZ}} \equiv \hat{\theta}_{\mathrm{NLIV}}$, with $\omega_{\mathrm{NZ}} = \frac{1}{4} - \delta$ for an arbitrarily small number $\delta > 0$.

In Section 2 we have seen that MER estimators fail to have asymptotically normal distribution when $\left|\theta_{01n}\right|$ is too small. In Section 3.1 we saw that the naive estimators remain asymptotically unbiased only if $\left|\theta_{01n}\right|$ is sufficiently small.

Suppose a researcher wants to test a hypothesis (or construct a confidence set) for a function of the parameter θ: $r(\cdot) : \Theta \to \mathbb{R}^q$, $q \leqslant p$. Most commonly, $r(\cdot)$ is a scalar function ($q = 1$) representing, for example, a particular component of θ_{0n} or a marginal effect implied by the model. Formally, we want to test

$$H_0 : r(\theta_{0n}) = v \qquad \text{vs} \qquad H_1 : r(\theta_{0n}) \neq v.$$

Let $\Phi\left(\cdot\right)$ denote the cumulative distribution function of the standard normal distribution.

Example (t-test of a linear hypothesis)**.** Consider a linear hypothesis about parameter θ_{0n}:

$$H_0 : \lambda'\theta_{0n} = v \qquad \text{vs} \qquad H_1 : \lambda'\theta_{0n} \neq v. \tag{3.8}$$

When $\left|\theta_{01n}\right|$ is sufficiently large, this null hypothesis can be tested with the standard t-statistic based on the estimator $\hat{\theta}_{\mathrm{L}}$:

$$t_{\mathrm{L}}\left(v\right) \equiv \frac{\sqrt{n}(\lambda'\hat{\theta}_{\mathrm{L}} - v)}{\sqrt{\lambda'\hat{\Sigma}_{\mathrm{L}}\lambda}}. \tag{3.9}$$

The p-value of this test is $p_{\mathcal{T}_{\mathrm{L}}} \equiv 2\Phi\left(-\left|t_{\mathrm{L}}\left(v\right)\right|\right)$, and the test with the significance level α can be written as

$$\phi_{\mathcal{T}_{\mathrm{L}}} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{L}}} < \alpha\},$$

where $\phi_{\mathcal{T}_{\mathrm{L}}}$ denotes the outcome of the test: it equals 1 if the null hypothesis is rejected and equals 0 otherwise.

When, $\left|\theta_{01n}\right|$ is sufficiently small, a valid test of the null hypothesis (3.8) can be constructed using the t-statistic $t_{\mathrm{NZ}}(v)$ defined as in equation (3.9) with $\hat{\theta}_{\mathrm{L}}$ and $\hat{\Sigma}_{\mathrm{L}}$ replaced by $\hat{\theta}_{\mathrm{NZ}}$ and $\hat{\Sigma}_{\mathrm{NZ}}$. Let us call this test $\mathcal{T}_{\mathrm{NZ}}$, and let us denote its p-value by $p_{\mathcal{T}_{\mathrm{NZ}}}$.

More broadly, for a given H_0, let $p_{\mathcal{T}_{\mathrm{L}}}$ and $p_{\mathcal{T}_{\mathrm{NZ}}}$ denote p-values of some tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ that are valid when $n^{\omega_{\mathrm{L}}}\left|\theta_{01n}\right| \to \infty$ and $n^{\omega_{\mathrm{NZ}}}\left|\theta_{01n}\right| \to 0$, respectively. The difficulty of the inference problem is that test $\mathcal{T}_{\mathrm{L}}$ is not valid when $|\theta_{01n}|$ is relatively "small", and test $\mathcal{T}_{\mathrm{NZ}}$ is not valid when $|\theta_{01n}|$ is relatively "large", so neither of the tests is valid uniformly for all values of the true (unknown) parameter θ_{01n}.

Remark 6. Our focus is on inference about a low-dimensional parameter.

3.2.2 Robust Test $\mathcal{T}_{\mathbf{Robust}}$

We can construct uniformly valid tests if the regions of validity of the two tests overlap, i.e., when

$$\omega_{\mathrm{NZ}} < \omega_{\mathrm{L}}. \tag{3.10}$$

When this condition is satisfied, for any value of θ_{01n}, at least one of the two tests $\mathcal{T}_{\mathrm{NZ}}$ and $\mathcal{T}_{\mathrm{L}}$ is valid. Thus, we define a valid test $\mathcal{T}_{\mathrm{Robust}}$ of level α as

$$\phi_{\mathcal{T}_{\mathrm{Robust}}} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{Robust}}} < \alpha\}, \qquad p_{\mathcal{T}_{\mathrm{Robust}}} \equiv \max\left\{p_{\mathcal{T}_{\mathrm{NZ}}}, p_{\mathcal{T}_{\mathrm{L}}}\right\}. \tag{3.11}$$

This test rejects H_0 if and only if both tests $\mathcal{T}_{\mathrm{NZ}}$ and $\mathcal{T}_{\mathrm{L}}$ reject the null hypothesis. Hence, H_0 is rejected only if it is rejected by at least one valid test.

Test $\mathcal{T}_{\mathrm{Robust}}$ is simple, but is usually conservative. For example, if we knew that $|\theta_{01n}|$ is small enough for test $\mathcal{T}_{\mathrm{NZ}}$ to be valid, we would have preferred $\mathcal{T}_{\mathrm{NZ}}$ over $\mathcal{T}_{\mathrm{Robust}}$, since the power of $\mathcal{T}_{\mathrm{NZ}}$ is higher. More generally, we can think of these settings as having three regions of $|\theta_{01n}|$:

Region	Condition	Valid Test(s)
NZ	$\|\theta_{01n}\| = o\left(n^{-\omega_{\mathrm{NZ}}}\right)$	$\mathcal{T}_{\mathrm{NZ}}$
INT	$\|\theta_{01n}\| = o\left(n^{-\omega_{\mathrm{NZ}}}\right)$ and $n^{\omega_{\mathrm{L}}}\|\theta_{01n}\| \to \infty$	$\mathcal{T}_{\mathrm{NZ}}$ and $\mathcal{T}_{\mathrm{L}}$
L	$n^{\omega_{\mathrm{L}}}\|\theta_{01n}\| \to \infty$	$\mathcal{T}_{\mathrm{L}}$

Regions "NZ" and "L" have been described above. In the intermediate region "INT" both tests are valid. This region is the overlap between "NZ" and "L" regions. To conduct uniformly valid inference, regions "NZ" and "L" need to overlap, i.e., region "INT" should not be empty.

3.2.3 Hybrid (Adaptive) Tests

When there is a way to consistently determine to which region parameter θ_{01n} belongs, we can provide more powerful inference procedures by adaptively combining different tests. Specifically, suppose we can find an identification-category-selection (ICS) statistic $\hat{A}_{\mathrm{ICS}}$ that allows us to determine the region of parameter θ_{01n}. For example, consider the statistic

$$\hat{A}_{\mathrm{ICS}} = \sqrt{n}\left|\hat{\theta}_{\mathrm{ICS},1}\right| \Big/ \sqrt{\hat{v}_{\mathrm{ICS},1}}\,, \tag{3.12}$$

where $\hat{\theta}_{\mathrm{ICS},1}$ is an estimator of θ_{01n}, and the scaling $\hat{v}_{\mathrm{ICS},1}$ aims to estimate the asymptotic variance of $\hat{\theta}_{\mathrm{ICS},1}$. Essentially, here $\hat{A}_{\mathrm{ICS}}$ is the absolute value of the t-statistic for testing the hypothesis $\theta_{01n} = 0$.

Large values of $\hat{A}_{\text{ICS}}$ can be treated as statistical evidence that $|\theta_{01n}|$ is "sufficiently large", so the test $\mathcal{T}_{\text{L}}$ is valid. Smaller values of $\hat{A}_{\text{ICS}}$ suggest that $|\theta_{01n}|$ may not be large enough, and a different test ($\mathcal{T}_{\text{Robust}}$ or even $\mathcal{T}_{\text{NZ}}$) should be used instead.

For this categorization approach to work, it is important that $\hat{\theta}_{\text{ICS},1}$ is a precise estimator of θ_{01n} for *all* values of $\theta_{01n} \in \Theta$. Whether we can find such an estimator $\hat{\theta}_{\text{ICS}}$ depends on the specific MER estimation approach. Naive estimators (e.g., $\hat{\theta}_{\text{NZ}}$) cannot be used as $\hat{\theta}_{\text{ICS}}$ estimators, because for large $|\theta_{01n}|$ their bias is of order 1. MER estimators ($\hat{\theta}_{\text{L}}$) are potential candidates for $\hat{\theta}_{\text{ICS}}$. Although MER estimators are not asymptotically normal for some of the values θ_{01n}, they could still be sufficiently precise for all values of θ_{01n}, to serve as $\hat{\theta}_{\text{ICS},1}$. For example, in Section 6 we show that the MME estimator $\hat{\theta}_{\text{MME}}$ satisfies $\hat{\theta}_{\text{MME}} - \theta_{0n} = O_p\left(n^{-1/2}\right)$ for all θ_{01n}, i.e., is uniformly $\sqrt{n}$-consistent. We also show that taking $\hat{v}_{\text{ICS},1} \equiv \hat{\Sigma}_{\text{MME},11}$ in equation (3.12) yields a valid category-selection statistic $\hat{A}_{\text{ICS}}$. We propose two types of hybrid tests based on $\hat{A}_{\text{ICS}}$.

Type-I hybrid test $\mathcal{T}_H^I$ is defined as $\phi^I_{\mathcal{T}_{\text{H}}} \equiv \mathbb{1}\{p^I_{\mathcal{T}_{\text{H}}} < \alpha\}$ with the p-value given by

$$p^I_{\mathcal{T}_{\text{H}}} \equiv (1-\hat{\lambda}_{\text{L}}) \max\{p_{\mathcal{T}_{\text{NZ}}}, p_{\mathcal{T}_{\text{L}}}\} + \hat{\lambda}_{\text{L}} p_{\mathcal{T}_{\text{L}}}, \qquad \hat{\lambda}_{\text{L}} \equiv \lambda(\hat{A}_{\text{ICS}} - \kappa_{\text{L},n}),$$

where $\lambda(z) = \mathbb{1}\{z > 0\}(1 - \exp(-cz))$ for some $c > 0$, and $\kappa_{\text{L},n}$ is a slowly growing threshold sequence that satisfies $n^{\omega_{\text{L}}}\left(n^{-1/2}\kappa_{\text{L},n}\right) \to \infty$. This condition implies that $|\theta_{01n}| \propto n^{-1/2}\kappa_{\text{L},n}$ falls into the "intermediate" region, where both $\mathcal{T}_{\text{L}}$ and $\mathcal{T}_{\text{NZ}}$ are valid. When $\hat{A}_{\text{ICS}}$ is much larger than $\kappa_{\text{L},n}$, $\hat{\lambda}_{\text{L}} \approx 1$ and $p^I_{\mathcal{T}_{\text{H}}} \approx p_{\mathcal{T}_{\text{L}}}$, i.e., type-I hybrid test $\mathcal{T}^I_{\text{H}}$ essentially behaves as test $\mathcal{T}_{\text{L}}$ for large $\hat{A}_{\text{ICS}}$. On the other hand, $\hat{\lambda}_{\text{L}} = 0$ when $\hat{A}_{\text{ICS}} \le \kappa_{\text{L},n}$, so for smaller $\hat{A}_{\text{ICS}}$, this hybrid test of type-I is equivalent to $\mathcal{T}_{\text{Robust}}$ test of equation (3.11).

Similar hybrid tests are often used to provide valid inference procedures in problems with weak identification of some of the parameters. For example, Andrews and Cheng (2012) develop a general framework for such problems and use $\hat{A}_{\text{ICS}}$ as in equation (3.12) to assess the strength of identification of some parameters. They use a single test statistic, but develop two critical values for it. The first critical value is valid only under (semi-)strong identification regime. The second value is robust, i.e., valid regardless of the strength of identification, but is conservative. The computation of such robust critical value is complicated, because it depends on the unknown weakly identified nuisance parameter. Andrews and Cheng (2012) then use the ICS statistic to smoothly link the two critical values to obtain an adaptive critical value for the test statistic. Our type-I hybrid test instead smoothly links the p-values to construct a test.[9] When the identification is strong ($|\theta_{01n}|$ is large in our settings), such adaptive tests are efficient, in the sense that they coincide with the non-conservative test $\mathcal{T}_{\text{L}}$. However, such tests are generally conservative when the identification is weaker.

Our settings differ from the typical setup with weakly identified parameters. We may have an alternative test $\mathcal{T}_{\text{NZ}}$ that is valid when the identification of the nuisance parameter is weak, and the

[9]Suppose $p_{\mathcal{T}_1}$ and $p_{\mathcal{T}_2}$ are the p-values of two valid tests of the same hypothesis. Take a $\lambda \in (0,1)$. Test $\phi_n = \mathbb{1}\{(1-\lambda)\, p_{\mathcal{T}_1} + \lambda p_{\mathcal{T}_2} < \alpha\}$ generally is not a valid test of size α, but test $\phi_n = \mathbb{1}\{(1-\lambda)\max\{p_{\mathcal{T}_1}, p_{\mathcal{T}_2}\} + \lambda p_{\mathcal{T}_2} < \alpha\}$ is.

usual test $\mathcal{T}_{\mathrm{L}}$ fails to deliver valid inference. As a result, in contrast to most of the literature we can introduce another type of hybrid tests.

Type-II hybrid test $\mathcal{T}_H^{II}$ is defined as $\phi_{\mathcal{T}_{\mathrm{H}}}^{II} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{H}}}^{II} < \alpha\}$ with the p-value given by

$$\begin{aligned} p_{\mathcal{T}_{\mathrm{H}}}^{II} &\equiv (1-\hat{\lambda}_{\mathrm{NZ}} - \hat{\lambda}_{\mathrm{L}}) \max\{p_{\mathcal{T}_{\mathrm{NZ}}}, p_{\mathcal{T}_{\mathrm{L}}}\} + \hat{\lambda}_{\mathrm{NZ}}\, p_{\mathcal{T}_{\mathrm{NZ}}} + \hat{\lambda}_{\mathrm{L}} p_{\mathcal{T}_{\mathrm{L}}}, \\ \hat{\lambda}_{\mathrm{NZ}} &\equiv \lambda(\kappa_{\mathrm{NZ},n} - \hat{A}_{\mathrm{ICS}}), \qquad \hat{\lambda}_{\mathrm{L}} \equiv \lambda(\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{L},n}). \end{aligned} \tag{3.13}$$

Type-II hybrid test $\mathcal{T}_{\mathrm{H}}^{II}$ uses two threshold sequences $\kappa_{\mathrm{L},n}$ and $\kappa_{\mathrm{NZ},n}$ that satisfy the following conditions: (i) $0 < \kappa_{\mathrm{NZ},n} \leq \kappa_{\mathrm{L},n}$, (ii) $n^{\omega_{\mathrm{L}}}\left(n^{-1/2}\kappa_{\mathrm{L},n}\right) \to \infty$, and (iii) $n^{\omega_{\mathrm{NZ}}}\left(n^{-1/2}\kappa_{\mathrm{NZ},n}\right) \to 0$. When $\kappa_{\mathrm{NZ},n} < \hat{A}_{\mathrm{ICS}}$, $\hat{\lambda}_{\mathrm{NZ}} = 0$ and test $\mathcal{T}_{\mathrm{H}}^{II}$ is identical to $\mathcal{T}_{\mathrm{H}}^{I}$. Test $\mathcal{T}_{\mathrm{H}}^{II}$ differs only when $\hat{A}_{\mathrm{ICS}}$ is sufficiently small so $\kappa_{\mathrm{NZ},n} > \hat{A}_{\mathrm{ICS}}$. The lower values of $\hat{A}_{\mathrm{ICS}}$ indicate that $|\theta_{01n}|$ is likely to be small, and $\mathcal{T}_{\mathrm{NZ}}$ is likely to be valid. When $\hat{A}_{\mathrm{ICS}}$ is substantially smaller than $\kappa_{\mathrm{NZ},n}$, $\hat{\lambda}_{\mathrm{NZ}} \approx 1$, $\hat{\lambda}_{\mathrm{L}} = 0$, and $p_{\mathcal{T}_{\mathrm{H}}}^{II} \approx p_{\mathcal{T}_{\mathrm{NZ}}}$, so test $\mathcal{T}_{\mathrm{H}}^{II}$ is nearly identical to test $\mathcal{T}_{\mathrm{NZ}}$. Thus, $\mathcal{T}_{\mathrm{H}}^{II}$ has higher power than $\mathcal{T}_{\mathrm{H}}^{I}$ when $|\theta_{01n}|$ is small. For the intermediate values of $\hat{A}_{\mathrm{ICS}} \in [\kappa_{\mathrm{NZ},n}, \kappa_{\mathrm{L},n}]$, $\hat{\lambda}_{\mathrm{NZ}} = \hat{\lambda}_{\mathrm{L}} = 0$ and $\mathcal{T}_{\mathrm{H}}^{II}$ coincides with $\mathcal{T}_{\mathrm{Robust}}$. This represents the uncertainty about whether $|\theta_{01n}|$ is large enough for $\mathcal{T}_{\mathrm{L}}$ to work, or small enough for $\mathcal{T}_{\mathrm{NZ}}$ to work, so $\mathcal{T}_{\mathrm{Robust}}$ is used.

Consider an oracle test that "knows" the true magnitude of θ_{01n} and chooses between non-conservative tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ accordingly. Type-II hybrid test $\mathcal{T}_{\mathrm{H}}^{II}$ is essentially equivalent to the oracle test both when the nuisance parameters are strongly identified *and* when they are weakly identified. This differs from most of the literature on inference in problems with weakly identified parameters, which typically only provides tests that are non-conservative when the identification is strong, but not when it is weak, i.e., tests similar to the hybrid test of type-I. For the intermediate values of $|\theta_{01n}|$, it is harder to determine whether $\mathcal{T}_{\mathrm{L}}$ or $\mathcal{T}_{\mathrm{NZ}}$ is valid, so test $\mathcal{T}_{\mathrm{H}}^{II}$ relies on the robust test $\mathcal{T}_{\mathrm{Robust}}$.

The choice of $\kappa_{\mathrm{NZ},n}$, $\kappa_{\mathrm{L},n}$, and $\lambda(z)$ is important in practice. We provide some details and recommendations in Section 7.

Remark 7. As usual, the corresponding type-I and type-II hybrid confidence sets can be constructed by the test inversion.

Remark 8. For the above ICS procedures to work, estimator $\hat{\theta}_{\mathrm{ICS},1}$ needs to be precise enough uniformly over all values of θ_{01n}. Some category-selection tests can be constructed without this condition. Consider $\hat{A}_{\min} \equiv \min\left\{\left|\hat{t}_{\mathrm{L}}\right|, \left|\hat{t}_{\mathrm{NZ}}\right|\right\}$ and $\hat{A}_{\max} \equiv \max\left\{\left|\hat{t}_{\mathrm{L}}\right|, \left|\hat{t}_{\mathrm{NZ}}\right|\right\}$, where $\hat{t}_{\mathrm{L}} \equiv \sqrt{n}\Sigma_{\mathrm{L},11}^{-1/2}\hat{\theta}_{\mathrm{L},1}$ and $\hat{t}_{\mathrm{NZ}} \equiv \sqrt{n}\Sigma_{\mathrm{NZ},11}^{-1/2}\hat{\theta}_{\mathrm{NZ},1}$ denote the t-statistics for testing $\theta_{01n} = 0$ based on $\hat{\theta}_{\mathrm{L}}$ and $\hat{\theta}_{\mathrm{NZ}}$, respectively.

Since either $\hat{t}_{\mathrm{L}}$ or $\hat{t}_{\mathrm{NZ}}$ must be valid for any value of θ_{01n}, large values of $\hat{A}_{\min}$ can be treated as statistical evidence that $|\theta_{01n}|$ is "sufficiently large", so the tests $\mathcal{T}_{\mathrm{L}}$ based on $\hat{\theta}_{\mathrm{L}}$ are valid. Hence, we can use test $\phi_{\mathcal{T}_{\mathrm{H}}}^{I}$ with $\hat{\lambda}_{\mathrm{L}} \equiv \lambda(\hat{A}_{\min} - \kappa_{\mathrm{L},n})$. Likewise, small values of $\hat{A}_{\max}$ provide evidence that $|\theta_{01n}|$ is relatively "small". Then we could consider an analog of hybrid test $\phi_{\mathcal{T}_{\mathrm{H}}}^{II}$ with $\hat{\lambda}_{\mathrm{NZ}} \equiv \lambda(\kappa_{\mathrm{NZ},n} - \hat{A}_{\max})$, and $\hat{\lambda}_{\mathrm{L}} \equiv \lambda(\hat{A}_{\min} - \kappa_{\mathrm{L},n})$. The values $\kappa_{\mathrm{L},n}$ and $\kappa_{\mathrm{NZ},n}$ may need to be altered to accommodate the changes in the category selection criteria. Since the regions, over which these alternative hybrid tests coincide with $\mathcal{T}_{\text{Robust}}$, are wider, these tests are likely to have lower power than the corresponding tests based on $\hat{A}_{\text{ICS}}$.

Remark 9. If one is interested in inference about θ_{01n} itself, a "null-imposed" category selection criterion can be used $\hat{A}_{\text{ICS}} \equiv \sqrt{n}\,|\theta_{01n}| \Big/ \sqrt{\hat{v}_{\text{ICS},1}}$.

4 Moderate Measurement Error Framework

4.1 Moment Conditions and Estimator

Suppose a researcher has a general semiparametric model that can be represented by a set of moment conditions

$$\mathbb{E}[g(X_i^*, S_i, \theta)] = 0 \text{ iff } \theta = \theta_{0n}, \tag{4.1}$$

for some m-dimensional moment function $g(\cdot)$, and a p-dimensional parameter of interest θ_{0n}. Variables X_i^* are mismeasured, while variables S_i are not. Were a random sample $\{(X_i^*, S_i)\}_{i=1}^n$ available, parameters θ_{0n} could have been estimated using standard GMM estimators, which would have been $\sqrt{n}$-consistent and asymptotically normal and unbiased. However, instead of X_i^* the researcher observes

$$X_i = X_i^* + \varepsilon_i,$$

the analogue of X_i^* contaminated with the measurement error ε_i.[10] If the researcher ignores the problem of measurement errors, and estimates model (4.1) with X_i in place of X_i^*, the standard GMM estimator suffers from the EIV bias, and the corresponding tests (t, Wald, etc.) and confidence sets are invalid.

Evdokimov and Zeleneev (2016) develop a framework for estimation of general semiparametric models with EIV. They make use of an alternative asymptotic approximation that models the magnitude of the measurement error as shrinking with the sample size. The rationale behind this approach is as follows. On the one hand, in many settings of interest viewing at least some of the variables as perfectly measured is implausible. On the other hand, the researchers may believe that the magnitude of the measurement errors is not too large, so that the signal X_i^* dominates the noise ε_i, and the magnitude of the EIV-bias of the naive estimator is thought to be somewhat comparable to its standard errors. To provide a better approximation of the finite sample properties of the estimators in such settings, it is useful to model the distribution of ε_i as drifting with the sample size in a particular way.

[10] Although we denote the mismeasured variables by X_i, these variables need not be covariates. For instance, in general nonlinear models, X_i may denote a mismeasured outcome variable, since measurement error in any variable generally leads to the EIV-bias.

Specifically, in the spirit of Amemiya (1985); Chesher (1991), suppose that the variance and the higher-order moments of ε_i (slowly) shrink towards zero as the sample size grows. Suppose $\mathbb{E}[\varepsilon_i] = 0$ and let $\sigma_n^2 \equiv \mathbb{E}[\varepsilon_i^2]$. For exposition, it will be convenient to assume that $\mathbb{E}[|\varepsilon_i/\sigma_n|^{K+1}]$ is bounded, so $\mathbb{E}[|\varepsilon_i|^\ell] \propto \sigma_n^\ell$.

When $\sigma_n^2 \to 0$ as $n \to \infty$, under some regularity conditions,

$$\begin{aligned}\mathbb{E}[g(X_i, S_i, \theta)] &= \mathbb{E}[g(X_i^*, S_i, \theta)] + \mathbb{E}\left[\sum_{k=1}^{K} \frac{\varepsilon_i^k}{k!} g_x^{(k)}(X_i^*, S_i, \theta)\right] + O(\mathbb{E}\left[|\varepsilon_{in}|^{K+1}\right]) \\ &= \mathbb{E}[g(X_i^*, S_i, \theta)] + \sum_{k=1}^{K} \frac{1}{k!}\mathbb{E}[\varepsilon_i^k], \mathbb{E}\left[g_x^{(k)}(X_i^*, S_i, \theta)\right] + O(\sigma_n^{K+1}), \end{aligned} \tag{4.2}$$

where we use the notation $g_x^{(k)}(x, s, \theta) \equiv \partial^k g(x, s, \theta) / \partial x^k$.

In particular this implies that

$$\mathbb{E}[g(X_i, S_i, \theta_{0n})] = O(\sigma_n^2), \tag{4.3}$$

and hence a naive GMM estimator that ignores the measurement error and uses X_i in place of X_i^* has bias of order σ_n^2. Thus, the naive estimator is asymptotically biased, unless the measurement error is very small (theoretically, unless $\sigma_n^2 = o_n\left(n^{-1/2}\right)$). Moreover, if σ_n^2 shrinks at a rate slower than $O_n\left(n^{-1/2}\right)$, the naive estimator is not $\sqrt{n}$-consistent. The tests based on the naive estimator (e.g., t-statistics) can then provide highly misleading results.

Evdokimov and Zeleneev (2016) consider estimation and inference in these settings using the Moderate Measurement Error (MME) framework that relies on the following key assumption:

Assumption MME (Moderate Measurement Errors). *(i) ε_i is independent from (X_i^*, S_i) and $\mathbb{E}[\varepsilon_i] = 0$; (ii) $\sigma_n^2 = o(n^{-1/(K+1)})$ for some integer $K \geqslant 2$, and $\mathbb{E}[|\varepsilon_i/\sigma_n|^L]$ is bounded for some $L \geq K+1$.*

Assumption MME (i) is the classical measurement error assumption. Assumption MME (ii) ensures that the bias due to the remainder in equation (4.2) is negligible.[11]

Thus, we can rearrange equation (4.2) as

$$\mathbb{E}[g(X_i^*, S_i, \theta)] = \mathbb{E}[g(X_i, S_i, \theta)] - \sum_{k=2}^{K} \frac{\mathbb{E}[\varepsilon_i^k]}{k!}\mathbb{E}\left[g_x^{(k)}(X_i^*, S_i, \theta)\right] + o(n^{-1/2}). \tag{4.4}$$

The left-hand side of this equation is exactly the moment condition (4.1) that we would like to use for estimation of θ_{0n}. The first term on the right-hand side involves only observed variables, and hence can be estimated by the sample average $\mathbb{E}_n\left[g(X_i, S_i, \theta)\right]$. The second term on the right-hand side can be viewed as the bias correction that removes the EIV-bias from $\mathbb{E}[g(X_i, S_i, \theta)]$.

Evdokimov and Zeleneev (2016) use representation (4.4) as a point of departure to jointly estimate parameters θ and the moments of the measurement error. The right-hand side of equa-

[11] Assumption MME (ii) can be replaced with a weaker condition $\mathbb{E}[|\varepsilon_i|^L] = o\left(n^{-1/2}\right)$, in which case one should define $\sigma_n \equiv \mathbb{E}\left[|\varepsilon_i|^L\right]^{1/L}$. We use the stated form of Assumption MME (ii) to simplify the exposition.

tion (4.4) could be viewed as a moment condition, except the expectations $\mathbb{E}\left[g_{xi}^{*(k)}(\theta)\right]$ cannot be immediately estimated from data, since X_i^* is not observed. To obtain a feasible moment condition, one can estimate $\mathbb{E}\left[g_{xi}^{*(k)}(\theta)\right]$ by $\mathbb{E}_n\left[g_{xi}^{(k)}(\theta)\right]$. As Evdokimov and Zeleneev (2016) point out, a naive application of this strategy generally does not work. $\mathbb{E}_n\left[g_{xi}^{(k)}(\theta)\right]$ is a biased estimator, since $\mathbb{E}\left[g_{xi}^{(k)}(\theta)\right]-\mathbb{E}\left[g_{xi}^{*(k)}(\theta)\right]=O\left(\sigma_n^2\right)$, and this bias is not negligible in general. When $K\geq 4$, one needs to bias correct the estimator of the bias correction term. However, the bias corrected correction term is itself a linear combination of the higher order derivatives of $g_{xi}^{(k)}$, and hence one can use the following *corrected moment function*:

$$\psi(X_i,S_i,\theta,\gamma)\equiv g(X_i,S_i,\theta)-\sum_{k=2}^{K}\gamma_k g_x^{(k)}(X_i,S_i,\theta), \tag{4.5}$$

where $\gamma=(\gamma_2,\ldots,\gamma_K)'$ is a $K-1$ dimensional vector of nuisance parameter to be estimated. Evdokimov and Zeleneev (2016) show that for some γ_{0n},

$$\mathbb{E}[\psi(X_i,S_i,\theta_{0n},\gamma_{0n})]=o\left(n^{-1/2}\right), \tag{4.6}$$

and hence the corrected moment conditions ψ can be used to estimate the true parameters θ_{0n} and γ_{0n}.[12] The values γ_{0kn} are related to the moments of ε_i as follows:

$$\gamma_{02n}=\sigma_n^2/2,\ \gamma_{03n}=\mathbb{E}\left[\varepsilon_i^3\right]/6, \text{ and } \gamma_{0kn}=\frac{\mathbb{E}\left[\varepsilon_i^k\right]}{k!}-\sum_{\ell=2}^{k-2}\frac{\mathbb{E}\left[\varepsilon_i^{k-\ell}\right]}{(k-\ell)!}\gamma_{0\ell n} \text{ for } k\geq 4.$$

It is important to note that $\gamma_{0kn}\neq\mathbb{E}\left[\varepsilon_i^k\right]/k!$ for $k\geq 4$, contrary to what equation (4.4) might suggest. For example, $\gamma_{04n}=\left(\mathbb{E}\left[\varepsilon_i^4\right]-6\sigma_n^4\right)/24$ is negative for many distributions, including Normal. The reason that generally $\gamma_{0kn}\neq\mathbb{E}\left[\varepsilon_i^k\right]/k!$ is that the estimators of the correction terms themselves need a correction, which is accounted for by the form of γ_{0kn}. Since there is a one-to-one relationship between γ_{0n} and the moments $\mathbb{E}\left[\varepsilon_i^\ell\right]$, parameter space Γ for γ_{0n} can incorporate restrictions that the moments must satisfy (e.g., $\sigma_n^2\geq 0$ and $\mathbb{E}\left[\varepsilon_i^4\right]\geq\sigma_n^4$). Such restrictions can increase the efficiency of the estimator and power of the tests.

Moderate Measurement Errors (MME) estimator is a GMM estimator (Hansen, 1982) based on the moment conditions (4.5):

$$(\hat{\theta},\hat{\gamma})\equiv\operatorname*{argmin}_{\theta\in\Theta,\gamma\in\Gamma}\hat{Q}\left(\theta,\gamma\right),\qquad \hat{Q}\left(\theta,\gamma\right)\equiv\overline{\psi}(\theta,\gamma)'\hat{\Xi}\overline{\psi}(\theta,\gamma), \tag{4.7}$$

where Θ and Γ are the parameter spaces for θ and γ, and $\hat{\Xi}$ is a weighting matrix.

MME jointly estimates the parameters of interest θ_{0n} and the nuisance parameters γ_{0n}, so it is convenient to consider the joint vector of parameters

$$\beta\equiv(\theta',\gamma')',\ \beta_{0n}\equiv(\theta_{0n}',\gamma_{0n}')',\ \hat{\beta}\equiv(\hat{\theta}',\hat{\gamma}')'.$$

[12]In the moment condition settings, having $o\left(n^{-1/2}\right)$ is equivalent to having 0 on the right-hand side of equation (4.6).

Let $\mathcal{B} \equiv \Theta \times \Gamma$. Then, equation (4.7) can be equivalently written as

$$\hat{\beta} = \operatorname*{argmin}_{\beta \in \mathcal{B}} \hat{Q}(\beta), \qquad \hat{Q}(\beta) \equiv \overline{\psi}(\beta)' \hat{\Xi} \overline{\psi}(\beta).$$

Under some regularity conditions, estimator $\hat{\beta}$ behaves as a standard GMM-type estimator: it is $\sqrt{n}$-consistent and asymptotically unbiased:

$$n^{1/2} \Sigma^{-1/2} (\hat{\beta} - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}), \tag{4.8}$$

$$\Sigma \equiv (\Psi' \Xi \Psi)^{-1} \Psi' \Xi \Omega_{\psi\psi} \Xi \Psi (\Psi' \Xi \Psi)^{-1}. \tag{4.9}$$

The asymptotic variance Σ takes the standard sandwich form, with $\Psi \equiv \mathbb{E}\left[\nabla_\beta \psi_i(\beta_{0n})\right]$, $\Omega_{\psi\psi} \equiv \mathbb{E}\left[\psi_i \psi_i'\right]$, and $\Xi = \operatorname{plim}_{n\to\infty} \hat{\Xi}$.

Thus, the MME approach addresses the EIV problem, and provides an asymptotically normal estimator $\hat{\theta}$ that can be used for drawing inference about θ_{0n} using standard tests and confidence intervals. Of course, for equation (4.8) to hold, one has to assume that the nuisance parameters γ_{0n} are identified, i.e., $|\theta_{01n}|$ is sufficiently large.

Remark 10. Although MME asymptotic approximation considers $\gamma_{02} = o_n(1)$, to avoid imposing arbitrary restrictions on γ_{0n}, the parameter set Γ is not shrinking with the sample size. In practice, we restrict it by assuming a lower bound on the the signal-to-noise ratio, e.g., $\sigma_n^2 / \sigma_{X^*}^2 \leq 1$.

Remark 11. No parametric assumptions are imposed on the distribution of ε_i.

Remark 12. Considering larger K allows σ_n^2 converging to zero at a slower rate, which in finite samples corresponds to the asymptotics providing good approximations for larger ranges of values of $\mathbb{E}\left[\varepsilon_i^2\right]$. Taking larger K increases the dimension of the nuisance parameter γ_{0n}, and requires having at least $m \geqslant p + K - 1$ moment conditions.

Remark 13. In contrast to most alternative approaches, in the MME framework, having a discrete instrument can be sufficient for identification and estimation of θ_{0n}.

Remark 14. For the weighting matrix, 2-step GMM estimator uses $\hat{\Xi}_{\text{GMM2}} \equiv \hat{\Omega}_{\psi\psi}^{-1}(\tilde{\theta}, \tilde{\gamma})$, where $(\tilde{\theta}, \tilde{\gamma})$ are some preliminary estimators. Although not indicated by the notation in equation (4.7), the weighting matrix $\hat{\Xi} \equiv \hat{\Xi}(\theta, \gamma)$ is allowed to be a function of θ and γ. For example, Continuously Updating GMM Estimator corresponds to $\hat{\Xi}_{\text{CUE}}(\theta, \gamma) \equiv \hat{\Omega}_{\psi\psi}^{-1}(\theta, \gamma)$. One may also consider the "regularized" versions of the weighting matrices $\hat{\Xi}_{\text{GMM2,R}} \equiv \hat{\Omega}_{\psi\psi}^{-1}(\tilde{\theta}, 0)$ and $\hat{\Xi}_{\text{CUE,R}}(\theta, \gamma) \equiv \hat{\Omega}_{\psi\psi}^{-1}(\theta, 0)$ that set $\gamma = 0$. Since γ_{0n} is assumed to be small, the regularized weighting matrices do cause a loss of efficiency, but can be useful when γ_{0n} is poorly identified.

Remark 15. Moment conditions ψ are linear in γ, and it is convenient to choose the weighting matrix $\hat{\Xi}(\theta, \gamma)$ that does not depend on γ, e.g., any of the weighting matrices in the previous remark except $\hat{\Xi}_{\text{CUE}}$. Then, $\hat{Q}_n(\theta, \gamma)$ is a quadratic form in γ, hence γ can be profiled out analytically. This reduces the optimization problem to minimizing $\hat{Q}_n(\theta, \hat{\gamma}(\theta))$ over $\theta \in \Theta$. Then, the dimension of the optimization parameter θ for the corrected moment condition problem remains the same as for the original estimation problem without the EIV correction.

Example GLM. *Consider the Generalized Linear Model (GLM) of equation (3.2). It can be written as a conditional moment restriction*

$$\mathbb{E}[(\rho(X_i^*, W_i, \theta_{0n}) - Y_i) \, | X_i^*, W_i, Z_i] = 0,$$

which can be converted into the unconditional moment restrictions $\mathbb{E}[g(X_i^*, S_i, \theta_{0n})] = 0$*:*

$$g(x, w, y, z, \theta) \equiv (\rho(x, w, \theta) - y)h(x, z, w),$$

where $S_i \equiv (W_i, Y_i, Z_i)$*. Here* $h(x, z, w)$ *is a vector of instrument functions, for example* $h(x, z, w) \equiv (x, x^2, z, xz, w')'$*.*

4.2 Weak Identification of γ_{0n}

Since the issue of weak identification of the distribution of ε_i is the feature of the EIV problem, MME estimator also suffers from it. Let us demonstrate how this issue of weak identification arises in the MME framework. The standard local identification condition for the GMM estimators is that the Jacobian matrix Ψ has full column rank, or, more precisely, that

$$\lambda_{\min}(\Psi'\Psi) \text{ is bounded away from zero.} \tag{4.10}$$

Identification of nonlinear models with measurement errors is an intricate question. Evdokimov and Zeleneev (2016) study identification in the MME framework in detail. For example, for condition (4.10) to hold, it may be necessary to have an instrumental variable. Condition (4.10) is necessary for a GMM estimator $\hat{\beta}$ to be $\sqrt{n}$-consistent and asymptotically normal. Note that we can split Ψ into two parts:

$$\Psi \equiv (\ \Psi_\theta\,,\ \Psi_\gamma\ )$$
$$\Psi_\theta \equiv \mathbb{E}\left[\nabla_\theta \psi_i(\theta_{0n}, \gamma_{0n})\right], \qquad \Psi_\gamma \equiv \mathbb{E}\left[\nabla_\gamma \psi_i(\theta_{0n}, \gamma_{0n})\right].$$

To simplify the exposition, let us focus on the case of $K = 2$ for the rest of this section. Then, $\gamma \equiv \gamma_2$,

$$\psi(X_i, S_i, \theta, \gamma) = g(X_i, S_i, \theta) - \gamma g_{xx}(X_i, S_i, \theta).$$

First, consider Ψ_θ. Let $G \equiv \mathbb{E}\left[\nabla_\theta g_i(\theta_{0n})\right]$ and $G_{xx} \equiv \mathbb{E}\left[\nabla_\theta g_{xxi}(\theta_{0n})\right]$.[13] Then $\Psi_\theta = G - \gamma_{0n} G_{xx} = G + o_n(1)$, since $\gamma_{0n} = o_n(1)$. It is natural to assume that Ghas full column rank, since this would have been the local identification condition for the GMM estimator of model (4.1) had X_i^* been observable. Thus, typically $\Psi_\theta = G + o_n(1)$ has full column rank.

The local identification of γ_{0n} is controlled by the Jacobian $\Psi_\gamma = g_{xx}(\theta_{0n}) \equiv \mathbb{E}\left[g_{xx}(X_i, S_i, \theta_{0n})\right]$. In the MME framework, the general problem of weak identification of the nuisance parameters considered in Section 2 manifests itself in the violation of the local identification condition (4.10) due to

$$\lambda_{\min}\left(\Psi_\gamma' \Psi_\gamma\right) \to 0 \text{ as } |\theta_{01n}| \to 0. \tag{4.11}$$

[13]For clarity, in this and the next subsections, for a function a, we often write a_x and a_{xx} in place of $a_x^{(1)}$ and $a_x^{(2)}$.

In particular, $\lambda_{\min}\left(\Psi_\gamma' \Psi_\gamma\right) = 0$ when $\theta_{01n} = 0$. The smaller $|\theta_{01n}|$ is, the less information about γ_{0n} we have. For small $|\theta_{01n}|$, estimator $\hat{\gamma}$ is imprecise or even inconsistent.

Example GLM (continued). *$g_{xx}(x, w, y, z, \theta) = \nabla_{xx}\left[\left(\rho\left(\theta_1 x + \theta_2' w\right) - y\right) h(x, z, w)\right]$, so*

$$g_{xx}(x, w, y, z, \theta) = (\rho(\theta, x, w) - y)\, h_{xx}(x, z, w) + 2\theta_1 \rho^{(1)}(\theta, x, w)\, h_x(x, z, w) + \theta_1^2 \rho^{(2)}(\theta, x, w)\, h(x, z, w), \tag{4.12}$$

where $\rho^{(k)}$ is $\partial^k / \partial a^k \rho(a)$. When $\theta_{01n} = 0$,

$$\Psi_\gamma = g_{xx}(\theta_{0n})|_{\theta_{01n}=0} = \mathbb{E}\left[\left(\rho\left(0 + \theta_{02}' W_i\right) - Y_i\right) h_{xx}(X_i, W_i, Z_i)\right] = 0, \tag{4.13}$$

and hence the local identification condition is violated. On the other hand, if the instruments are strong, condition (4.10) is satisfied as long as $\|\theta_{01n}\|$ is bounded away from zero.

Thus, in the GLM example, and any other model with the property (4.11), $\hat{\beta}$ is not $\sqrt{n}$-consistent and the asymptotically normal approximation (4.8) generally does not hold when $\theta_{01n} \to 0$. In the following sections we study the properties of the estimators $\hat{\theta}$ and $\hat{\gamma}$. Interestingly, we show that in many models we may expect $\hat{\theta}$ (but not $\hat{\gamma}$) to be $\sqrt{n}$-consistent, regardless of the magnitude of $|\theta_{01n}|$. At the same time, the asymptotic theory confirms the findings of Section 2 that in general $\hat{\theta}$ may not be approximately normally distributed when $|\theta_{01n}|$ is small. We then establish the uniform validity of the inference procedures considered in Section 3.2.

Next, we outline some of the main ideas and provide the roadmap of the analysis of the properties of $\hat{\theta}$ and $\hat{\gamma}$. The asymptotic theory is then formally developed in Section 6.

4.3 Properties of $\hat{\theta}$ and $\hat{\gamma}$: an Overview

Let us consider the local identification properties of the moment condition ψ. For simplicity, suppose $K = 2$ so γ is a scalar and

$$\overline{\psi}(\theta, \gamma) \equiv \overline{g}(\theta) - \gamma \overline{g}_{xx}(\theta). \tag{4.14}$$

Estimator $\hat{\theta}$ turns out to be consistent regardless of the strength of identification of γ_{0n}. To understand the properties of $\hat{\theta}$ and $\hat{\gamma}$, consider the approximation of the moment condition $\overline{\psi}(\theta, \gamma)$ in a shrinking neighborhood of θ_{0n}, i.e., for $\theta \in B_{\delta_n}(\theta_{0n})$ for some $\delta_n \to 0$. Then

$$\begin{aligned}
& \overline{\psi}(\theta, \gamma) \\
= \; & \overline{\psi}(\theta_{0n}, \gamma) + \overline{\Psi}_\theta(\theta_{0n}, \gamma)(\theta - \theta_{0n}) + o_{p,n}(\|\theta - \theta_{0n}\|) \\
= \; & \overline{\psi}(\theta_{0n}, \gamma_{0n}) + \overline{\Psi}_\theta(\theta_{0n}, \gamma_{0n})(\theta - \theta_{0n}) - (\gamma - \gamma_{0n})\left[\overline{g}_{xx}(\theta_{0n}) + \overline{G}_{xx}(\theta_{0n})(\theta - \theta_{0n})\right] + o_{p,n}(\|\theta - \theta_{0n}\|),
\end{aligned}$$

where the first equality follows by the Taylor expansion in θ around θ_{0n}, and the second equality follows from the linearity of $\overline{\psi}(\theta, \gamma)$ in γ and rearranging the terms using equation (4.14). Note that this approximation does not impose any restrictions on γ, other than $\|\gamma\|$ being bounded (since Γ is compact). By the Central Limit Theorem we expect $\sqrt{n}\overline{\psi} \to_d N(0, \Omega_{\psi\psi})$ and $\sqrt{n}\xi_{2n} = O_{p,n}(1)$

for $\xi_{2n} \equiv \overline{g}_{xx}(\theta_{0n}) - g_{xx}(\theta_{0n})$. Moreover, by the Law of Large Numbers $\overline{\Psi}_\theta - \Psi_\theta = o_{p,n}(1)$ and $\overline{G}_{xx} - G_{xx} = o_{p,n}(1)$, for some bounded Ψ_θ and G_{xx}. Thus, omitting the arguments of functions evaluated at $(\theta_{0n}, \gamma_{0n})$, for $\theta \in B_{\delta_n}(\theta_{0n})$ we have

$$\sqrt{n}\psi(\theta,\gamma) = \underbrace{\sqrt{n}\psi}_{N(0,\Omega_{\psi\psi})} + \underbrace{\Psi_\theta}_{\approx G}\sqrt{n}(\theta-\theta_{0n}) - (\gamma-\gamma_{0n})\left[\sqrt{n}g_{xx} + \underbrace{\sqrt{n}\xi_{2n}}_{O_{p,n}(1)} + G_{xx}\sqrt{n}(\theta-\theta_{0n})\right] + o_{p,n}\left(\sqrt{n}(\theta-\theta_{0n})\right). \tag{4.15}$$

The local identification of γ_{0n} and the properties of $\hat{\theta}$ and $\hat{\gamma}$ depend on the term $\sqrt{n}g_{xx}$. It helps to think of the following three scenarios:

Strong-ID: $\|g_{xx}\|$ is bounded away from zero;

Semi-Strong-ID: $\|g_{xx}\| \to 0$ "slowly" with the sample size, so $\sqrt{n}\|g_{xx}\| \to \infty$;

Weak-ID: $\|g_{xx}\| \to 0$ "quickly" with the sample size, so $\sqrt{n}\|g_{xx}\| \to C \in [0,\infty)$.

In the next section we will give more precise characterization for these scenarios, and study the properties of the estimators under each of them. Considering such three scenarios is typical in the literature addressing the settings, in which some parameters are weakly identified, for example, see Andrews, Cheng, and Guggenberger (2011).

The "Strong-ID" scenario corresponds to the standard GMM estimator settings with no concerns about weak identification of the nuisance parameter. When $\|\theta_{01n}\|$ is bounded away from zero, $\|g_{xx}\|$ is bounded away from zero. Then, the first term in the square brackets in equation (4.15) dominates, the second and third terms are negligible, and the asymptotically normal approximation (4.8) holds.

When $\|\theta_{01n}\|$ is close to zero, $\|g_{xx}\|$ can be close to zero and the second and third term in the square brackets may become non-negligible, generally causing $\hat{\theta}$ and $\hat{\gamma}$ to have nonstandard asymptotic distributions. It is useful to linearize $g_{xx} \equiv g_{xx}(\theta_{0n})$ around $\theta_{01n} = 0$, i.e., under some regularity conditions we can write

$$g_{xx}(\theta_{0n}) = A\theta_{01n} + o_n(|\theta_{01n}|),$$

for a vector $A \equiv A(\theta_{02n})$ that does not depend on θ_{01n}. When $A \neq 0$, $\sqrt{n}\|g_{xx}\| \propto \sqrt{n}\theta_{01n}$, and the "Semi-Strong-ID" scenario corresponds to $\sqrt{n}|\theta_{01n}| \to \infty$, i.e., $|\theta_{01n}|$ shrinking to zero at a slower than $n^{-1/2}$ rate.

In the "Semi-Strong-ID" scenario, $\sqrt{n}\|g_{xx}\| \to \infty$ and hence the second and third term in the square brackets are dominated by the first one. Since $\Psi = (G, g_{xx}) + o_n(1)$, the full rank condition on Ψ in this scenario is replaced with:

$$\text{matrix } (G, A) \text{ has full column rank.}$$

The first two scenarios have many features in common. In particular, the estimators $\left(\hat{\theta}, \hat{\gamma}\right)$ are both consistent and asymptotically normal, and $\hat{\theta}$ is $\sqrt{n}$-consistent. However, the rate of convergence of $\hat{\gamma}$ is $\sqrt{n}|\theta_{01n}|$, and is slower than $\sqrt{n}$ in the "Semi-Strong-ID" case. The first two scenarios are considered in Section 6.1.

The "Weak-ID" scenario corresponds to $\sqrt{n}|\theta_{01n}| = O_n(1)$. In the square brackets in equation (4.15), term $\sqrt{n}g_{xx}$ is bounded and hence the other two terms in the square brackets affect the properties of $\hat{\theta}$ and $\hat{\gamma}$. As a result, $\hat{\gamma}$ is inconsistent, and $\hat{\theta}$ does not have an asymptotically normal distribution. Notice that the term involving G_{xx} is now affecting identification of θ_{0n}. Section 6.2 below studies the "Weak-ID" scenario and proves that estimator $\hat{\theta}$ remains $\sqrt{n}$-consistent for arbitrarily small $|\theta_{01n}|$ if matrix $G - cG_{xx}$ has full column rank for all c.

Combining these conditions, if matrix $(G - cG_{xx}, A)$ has full column rank for all c, $\hat{\theta} - \theta_{0n} = O_{p,n}\left(n^{-1/2}\right)$ uniformly over all values of θ_{01n}.

Remark 16. We can write $A(\theta_{02n})$ as $A(\theta_{02n}) = \nabla_{\theta_{01n}} \int g_{xx}(x, s, (\theta_{01n}, \theta_{02n}))\, dF_{X,S}(x, s|\theta_{01n}, \theta_{02n})$. Note that θ_{01n} appears twice under the integral sign, so $A(\theta_{02n})$ is *not* equal to $\mathbb{E}\left[\nabla_{\theta_{01n}} g_{xx}(X_i, S_i, (\theta_{01n}, \theta_{02n}))\right]$. Note that $A(\theta_{02n})$ only appears in the regularity conditions, and is not used for estimation or inference.

Remark 17. It turns out that the identification conditions can be equivalently written with X_i^* in place of X_i in the expectations. We indicate expectations computed with X_i^* in place of X_i with superscript "$*$". For example, the condition for uniform $\sqrt{n}$-consistency of $\hat{\theta}$ can be equivalently stated as: matrix $(G^* - cG_{xx}^*, A^*)$ has full column rank for all c.

Example GLM (Continued)**.** *Consider $\theta_{01n} = o_n(1)$. From equation (4.12) we have*

$$\begin{aligned} g_{xx}^* &\equiv \mathbb{E}\left[g_{xx}(X_i^*, W_i, \theta_{0n})\right] = E_{\theta_{0n}}\left[2\theta_{01n}\rho^{(1)}(\theta_{01n}X_i^* + \theta_{02}'W_i)h_{xi}^*\right] + O\left(\|\theta_{01n}\|^2\right) \\ &= A^*\theta_{01n} + O\left(\|\theta_{01n}\|^2\right), \text{ where } A^* \equiv 2E\left[\rho^{(1)}(\theta_{02}'W_i)h_{xi}^*\right]. \end{aligned}$$

The strength of identification of γ_{0n} is proportional to $|\theta_{01n}|$. By a direct calculation, $G^* = \mathbb{E}\left[\rho^{(1)}(\theta_{02}'W_i)h_i \times (X_i, W_i')\right] + O_n(|\theta_{01n}|)$ and $G_{xx}^* = \mathbb{E}\left[(\rho^{(1)}(\theta_{02}'W_i)\,(X_i^* h_{xxi}^* + 2h_{xi}^*, W_i' h_{xxi}^*)\right] + O_n(|\theta_{01n}|)$.

For illustration, consider $h_i \equiv (X_i, W_i, Z_i)'$, $W_i = 1$, and, without loss of generality, $\mathbb{E}[X_i^*] = \mathbb{E}[Z_i] = 0$. Denote $\varrho_{02} \equiv \rho^{(1)}(\theta_{02})$. Then, up to the $o_n(1)$ terms,

$$G^* = \varrho_{02}\begin{pmatrix} \sigma_{X^*}^2 & 0 \\ 0 & 1 \\ \sigma_{ZX^*} & 0 \end{pmatrix}, \quad A^* = 2\varrho_{02}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad G_{xx}^* = \varrho_{02}\begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

where we have used equality $\rho^{(1)}(\theta_{01n}x + \theta_{02}'w) = \varrho_{02} + O_n(|\theta_{01n}|)$ repeatedly to simplify the expressions. The identification condition in the Semi-Strong-ID scenario is that matrix (G^*, A^*) has full column rank. This condition is satisfied as long as instrument Z_i is relevant, i.e., $\sigma_{ZX^*} \neq 0$. Moreover, when $\sigma_{ZX^*} \neq 0$, matrix $(G^* - cG_{xx}^*, A^*)$ has full column rank for all c. Thus, the local identification condition for uniform $\sqrt{n}$-normality is also satisfied.

As in the GLM example, verification of the regularity conditions is greatly simplified by the focus on $\theta_{01n} = o_n(1)$, since this leads to linearization of the moment conditions and their derivatives.

Remark 18. A reader familiar with the work of AC12 might be surprised by our linearizing the moment condition $\overline{\psi}(\theta,\gamma)$ in θ_1 around θ_{01n} and not around 0, as was suggested by AC12. The key assumption of AC12, stated in our notation, is that the sample criterion function does not depend on γ when evaluated at $\theta_1 = 0$, i.e., that $\hat{Q}\left((\theta_1,\theta_2),\gamma\right)\big|_{\theta_1=0}$ does not depend on γ.[14] This condition is generally violated for the MME estimator. For example, in the GLM model, $g_{xxi}\left((0,\theta_2)\right) = \left(\rho\left(\theta_2' W_i\right) - Y_i\right) h_x^{(2)}\left(X_i, W_i, Z_i\right)$ and hence $\overline{\psi}\left((0,\theta_2),\gamma\right)$ and $\hat{Q}\left((0,\theta_2),\gamma\right)$ depend on γ. As a result, our analysis differs substantially from that of AC12.

Remark 19. In general, parameter θ_1 that controls identification of γ_{0n} may be a vector. For example, suppose $\mathbb{E}\left[Y_{i\ell}|X_i^*, W_i\right] = \rho_\ell\left(\theta_{01,\ell}X_i^* + \theta_{02,\ell}' W_i\right)$ for $\ell \in \{1,2\}$, as in, e.g., multinomial choice model with 3 choices. Consider the moment conditions

$$g\left(X_i, S_i, \theta\right) \equiv \begin{pmatrix} \left(\rho_1(\theta_{1,1}X_i + \theta_{2,1}' W_i) - Y_{i1}\right) h_1\left(X_i, Z_i, W_i\right) \\ \left(\rho_2(\theta_{1,2}X_i + \theta_{2,2}' W_i) - Y_{i2}\right) h_2\left(X_i, Z_i, W_i\right) \end{pmatrix}.$$

Here, $\theta_1 \equiv \left(\theta_{1,1}, \theta_{1,2}\right)'$, since to identify γ_{0n} it is sufficient that at least one of the two coefficients on X_i is not zero. In other words, γ_{0n} is weakly identified only when $\|(\theta_{01,1}, \theta_{01,2})'\| \to 0$.

In the next section we provide a formal analysis of the MME estimator and inference procedures, in particular considering expansions of general order K and multivariate θ_1.

5 Finite Sample Experiments

To illustrate the finite sample properties of the proposed hybrid test, we consider the following logit model

$$\begin{gathered} Y_i = \mathbb{1}\{\theta_{01}X_i^* + \theta_{02}W_i + \theta_{03} - \zeta_i > 0\}, \quad \zeta_i \sim \mathit{Logistic}, \\ X_i^* = Z_i + V_i, \quad X_i = X_i^* + \varepsilon_i, \quad W_i = \rho X_i^*/(\sigma_Z^2 + \sigma_V^2) + \sqrt{1-\rho^2}\eta_i, \\ \zeta_i \perp (Z_i, V_i, \varepsilon_i, \eta_i)' \sim N\left((0,0,0,0)', \mathit{Diag}(\sigma_Z^2, \sigma_Z^2, \sigma_\varepsilon^2, \sigma_\eta^2)\right). \end{gathered}$$

We fix $(\theta_{02}, \theta_{03}, \rho, \sigma_Z^2, \sigma_V^2, \sigma_\varepsilon^2, \sigma_\eta^2) = (0, 1, 0.7, 1, 1, 1, 1)$ and $n = 1000$ and vary $\theta_{01} \in \{0, 0.15, 0.5, 1.0\}$, which controls the strength of identification of the nuisance parameter.

We focus on testing the null hypothesis $H_0 : \theta_{01n} = v$ against $H_1 : \theta_{01n} \neq v$ at the $\alpha = 5\%$ significance level. We evaluate the finite sample properties of the following tests. First, we consider the standard t-test based on the naive MLE estimator (t_{MLE}), which ignores the presence of the measurement error. Second, we consider the t-test based on the NLIV estimator (t_{NLIV}), which corresponds to the moment function $g_{\text{NLIV},i}(\theta) = \left(\rho(\theta_1 X_i + \theta_{02}W_i + \theta_{03}) - Y_i\right)(1, Z_i, W_i)'$, where ρ stands for

[14] This is Assumption A in AC12; see also the discussion immediately following Assumption C1 on page 2169 of their paper.

the logistic CDF. Third, we consider the t-statistic based on the MME estimator of Evdokimov and Zeleneev (2016) (t_{MME}). Specifically, following Evdokimov and Zeleneev (2016), we transform the original moment function $g_i(\theta) = (\rho(\theta_1 X_i + \theta_{02} W_i + \theta_{03}) - Y_i)((1, X_i, X_i^2) \otimes (1, Z_i), W_i)'$ into $\psi_i(\theta, \gamma) = g_i(\theta) - \gamma g_{xx,i}(\theta)$ using the simplest correction scheme with $K = 2$. Then, the considered MME estimator corresponds to the two-step GMM estimator based on the corrected moment function ψ_i. Forth, we consider the projection test based on the S-statistic of Stock and Wright (2000) using the corrected moment function $\psi_i(\theta, \gamma)$ as a benchmark identification robust test. Specifically, since conditional on the true value of $\gamma = \gamma_{0n}$, θ_{02} and θ_{03} are strongly identified, we take the $1 - \alpha$ quantile of χ_5^2 as the critical value of the S-test. Finally, we consider the type II hybrid test as in (3.13) (H_{II}), which uses t_{NLIV} as $\mathcal{T}_{\text{NZ}}$ and t_{MME} as $\mathcal{T}_{\text{L}}$ and

$$\hat{A}_{\text{ICS}} = \frac{|\hat{\theta}_{\text{MME},1}|}{\text{s.e.}\left(\hat{\theta}_{\text{MME},1}\right)}.$$

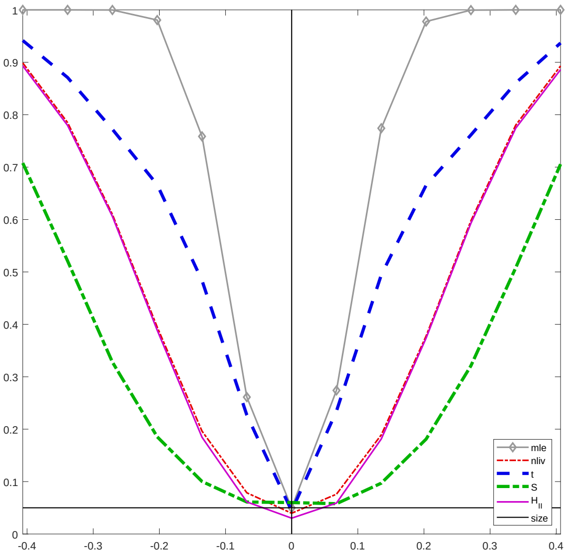
The thresholds are chosen as $\kappa_{\text{NZ},n} = 0.75 n^{1/10}$ and $\kappa_{\text{L},n} = 0.5 n^{1/5}$, and the weights are computed as $\hat{\lambda}_{\text{NZ}} = \lambda(\kappa_{\text{NZ},n} - \hat{A}_{\text{ICS}}; c_{\text{NZ}})$ and $\hat{\lambda}_{\text{L}} = \lambda(\hat{A}_{\text{ICS}} - \kappa_{\text{L},n}; c_{\text{L}})$, where $\lambda(z; c) = \mathbb{1}\{z \geqslant 0\}(1 - \exp(-cz))$. Specifically, we report the results for $c_{\text{NZ}} = 2$ and $c_{\text{L}} = 2$.[15]

Figure 5 below presents the rejection rates (based on 5000 replications) of the considered tests for θ_{01} equal 0 (γ_{0n} is not identified), 0.15 (γ_{0n} is weakly identified), and 0.5 and 1.0 (γ_{0n} is strongly identified). Since the MLE estimator suffers from the EIV bias unless $\theta_{01} = 0$, t_{MLE} is heavily size distorted even when $\theta_{01} = 0.15$ and wrongly rejects the true null hypothesis (almost) 100% of the time when $\theta_{01} \in \{0.5, 1.0\}$. The MME estimator effectively removes the EIV bias, and the corresponding test t_{MME} controls size in the strong identification scenarios. However, t_{MME} fails to control size when identification of γ_{0n} is weak ($\theta_{01} = 0.15$). As predicted in Section 3, the EIV bias of the NLIV estimator is substantially smaller relative to the bias of the MLE estimator when θ_{01} is not "too large", and t_{NLIV} controls size in all designs (however, its power is very low when $\theta_{01} = 1.0$). The hybrid test H_{II} controls size in all designs. It behaves as t_{NLIV} when $\theta_{01} \in \{0, 0.15\}$ (weak identification of γ_{0n}) and as t_{MME} when $\theta_{01} \in \{0.5, 1.0\}$ (strong identification of γ_{0n}). Notice that switching to t_{NLIV} under weak identification of γ_{0n} is also advantageous in terms of the power: both t_{NLIV} and H_{II} considerably overperform the projection S-test serving as a benchmark identification robust test.

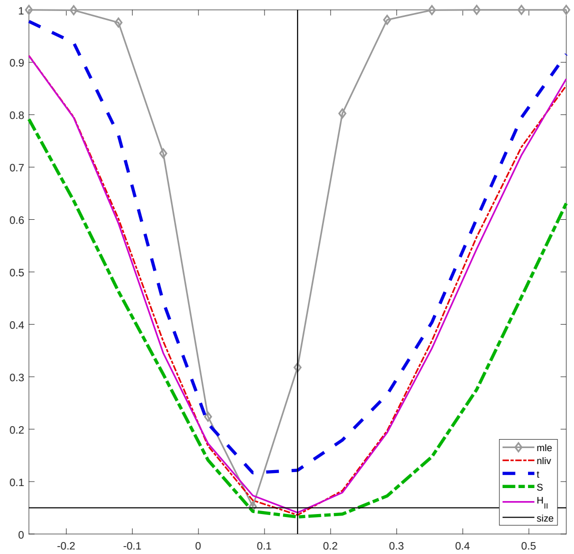
6 Large Sample Theory in the MME Framework

In this section, we formally introduce the MME framework and study asymptotic properties of the MME estimator. In particular, we provide a set of low-level regularity conditions that ensure validity of the results, which were informally introduced in the previous section. The outline of the section is as follows. First, we establish asymptotic normality of the MME estimator under semi-strong and strong identification of γ_{0n}. Second, we demonstrate that the MME estimator is also uniformly $\sqrt{n}$-consistent regardless the strength of γ_{0n} identification. Finally, we also study

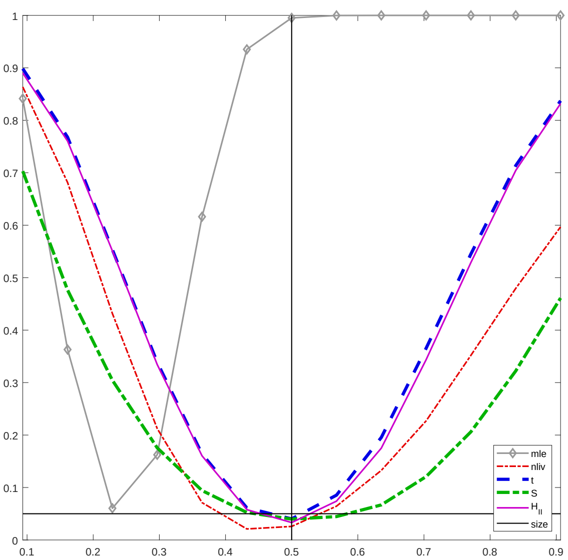
[15] The rejection rates of H_{II} are similar for $c_{\text{NZ}}, c_{\text{L}} \in \{1, 2, 4\}$.



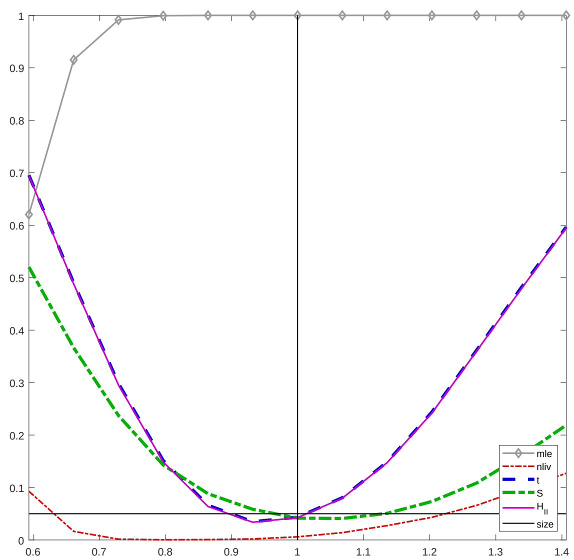
(a) $\theta_{01} = 0$



(b) $\theta_{01} = 0.15$



(c) $\theta_{01} = 0.5$



(d) $\theta_{01} = 1.0$

Figure 5: Simulated rejection probabilities of the t_{MLE} (solid gray w/ diamonds), t_{NLIV} (dashed red), t_{MME} (loosely dashed thick blue), projection S (dashed thick green), and H_{II} (solid purple) tests against the hypothesized value v for $\theta_{01} \in \{0, 0.15, 0.5, 1.0\}$. The results are based on 5000 replications with sample size $n = 1000$. The nominal level of the tests is 5%.

the asymptotic properties of the uncorrected estimator, which ignores the potential presence of the measurement error. In particular, we show that the uncorrected estimator is asymptotically normal and unbiased when $\|\theta_{01n}\|$ is sufficiently "small".

Before proceeding with the asymptotic analyses, let us introduce the following notations. We use Υ_n to denote the distribution of $(X_i^*, S_i, \varepsilon_i)$ in a sample of size n. Note that every Υ_n is implicitly associated with a unique θ_{0n} identified by the moment restrictions (4.1). We want to allow for drifting Υ_n in order to (i) address the uniformity issue by letting θ_{01n} change arbitrarily with n; (ii) incorporate the MME framework, under which the distribution of $\varepsilon_i \equiv \varepsilon_{in}$ necessarily drifts. Note that we also let the other features of the distribution of (X_i^*, S_i) drift but suppress additional indexing by n for simplicity of notation.

Notation. C denotes a generic constant uniform over Υ_n or, when specified, over a particular set of Υ_n. For example, $\mathbb{E}\left[\|g(X_i^*, S_i, \theta_{0n})\|\right] < C$ should be read as $\sup_{\Upsilon_n} \mathbb{E}_{\Upsilon_n}[\|g(X_i^*, S_i, \theta_{0n})\|] < C$, where the supremum is taken over all possible Υ_n (effectively, over all possible distributions of (X_i^*, S_i) and the corresponding θ_{0n}) and $\mathbb{E}_{\Upsilon_n}$ denotes the expectation under Υ_n. Analogously to the standard $o(\cdot)$, $O(\cdot)$, $o_p(\cdot)$, and $O_p(\cdot)$ notations, we use $o_n(\cdot)$, $O_n(\cdot)$, $o_{p,n}(\cdot)$, and $O_{p,n}(\cdot)$, which also should be treated as uniform in Υ_n. For example, $\eta_n = o_{p,n}(1)$ means that $\forall \epsilon > 0$ $\limsup_{n\to\infty} \sup_{\Upsilon_n} \mathbb{P}_{\Upsilon_n}(||\eta_n|| > \epsilon) = 0$, and $\eta_n = O_{p,n}(1)$ means that $\forall \epsilon > 0$ $\exists C > 0$: $\limsup_{n\to\infty} \sup_{\Upsilon_n} \mathbb{P}_{\Upsilon_n}(||\eta_n|| > C) < \epsilon$.

Also, analogously to the previously introduced notations, we let $a_i^*(\beta) \equiv a(X_i^*, S_i, \beta)$ and $a^*(\beta) \equiv \mathbb{E}[a_i^*(\beta)]$. Similarly, let $\Omega_{aa}^*(\beta) \equiv \mathbb{E}[a_i^*(\beta)a_i^*(\beta)']$. However, note that, unlike before, we use $a^* \equiv a(\theta_{0n}, 0)$ and $\Omega_{aa}^* \equiv \Omega_{aa}^*(\theta_{0n}, 0)$, so it plugs in $\theta = \theta_{0n}$ and $\gamma = 0$ (if γ is a part of β).

6.1 Asymptotic normality of the MME estimator under semi-strong and strong identification

In this subsection, we provide a set of low-level conditions, which ensure asymptotic normality of the MME estimator under semi-strong and strong identification of γ_{0n}.

Assumption 1 (DGP)**.** $\{(X_i^*, S_i')\}_{i=1}^n$ are i.i.d. and satisfy the moment restrictions (4.1).

Assumption 2 (ME)**.** For each n the measurement errors $\{\varepsilon_{in}\}_{i=1}^n$ are i.i.d., and Assumption MME holds with $L = 2M$ for some integer $M \geqslant K + 1$.

Assumption 1 and 2 formally reintroduce the MME framework. Assumption 2 also strengthens previously introduced Assumption MME: it requires $|\varepsilon_{in}/\sigma_n|$ to have $2M$ moments for some $M \geqslant K + 1$.

Assumption 3 (Parameter space)**.** $\Theta \subset \mathbb{R}^p$ and $\Gamma \subset \mathbb{R}^{K-1}$ are compact. In addition, Θ is convex, $\theta_{0n} \in \Theta$ and $\gamma_{0n} \in \Gamma$, and θ_{0n} is bounded away from the boundary of Θ.

Assumption 3 is an analogue of the standard parameter space assumption. The standard assumption $\theta_{0n} \in \operatorname{int}(\Theta)$ should be strengthened since we allow θ_{0n} to vary with the sample size: we want to rule out sequences of θ_{0n}, which approach the boundary of Θ arbitrarily closely. Unlike θ_{0n}, γ_{0n} is potentially allowed to be on (or arbitrarily close to) the boundary of Γ.

Let $\mathcal{X} \subseteq \mathbb{R}$ be some closed convex set containing the union of supports of X_i^* and X_i, and $\mathcal{S} = \operatorname{supp}(S_i)$

Assumption 4 (Moment function).

(i) For every $s \in \mathcal{S}$, $G_x^{(K)}(x,s,\theta)$ exists and is continuous on $\mathcal{X} \times \Theta$;

(ii) There exist functions b_1, b_2, b_{G1}, b_{G2} and a $\delta > 0$ such that for all $x, x' \in \mathcal{X}$, $(s,\theta) \in \mathcal{S} \times \Theta$

$$\left\| g_x^{(K)}(x',s,\theta) - g_x^{(K)}(x,s,\theta) \right\| \leqslant b_1(x,s,\theta)|x'-x| + b_2(x,s,\theta)|x'-x|^{M-K} \tag{6.1}$$

and for all $x, x' \in \mathcal{X}$, $(s,\theta) \in \mathcal{S} \times B_\delta(\theta_0)$

$$\left\| G_x^{(K)}(x',s,\theta) - G_x^{(K)}(x,s,\theta) \right\| \leqslant b_{G1}(x,s,\theta)|x'-x| + b_{G2}(x,s,\theta)|x'-x|^{M-K} \tag{6.2}$$

Assumption 4 is a smoothness requirement. Condition (ii) allows to bound the effect of the measurement error on the moment function and its Jacobian. Specifically, Eq. (6.1) ensures that the remainder in (4.2) is $O(\sigma_n^{K+1})$ and, hence, helps to establish that the corrected moment function ψ satisfy the moment restrictions in the form of (4.6).

Remark 20. If $g(x,s,\theta)$ is a sufficiently smooth function, there are simple conditions, which guarantee that condition (ii) is satisfied. For example, for (6.1) to be satisfied, it is sufficient to require $g_x^{(J)}(x,s,\theta)$ to be bounded on $\mathcal{X}$, where $K < J \leqslant M$. Indeed,

$$\begin{aligned}\left\| g_x^{(K)}(x',s,\theta) - g_x^{(K)}(x,s,\theta) \right\| \leqslant & \sum_{j=1}^{J-K-1} \frac{1}{j!} \left\| g_x^{(K+j)}(x,s,\theta) \right\| |x'-x|^j \\ & + \frac{1}{(J-K)!} \left\| g_x^{(J)}(\tilde{x},s,\theta) \right\| |x'-x|^{J-K},\end{aligned}$$

for some $\tilde{x}$, which lies between x and x'. Hence, one may take

$$b_1(x,s,\theta) = b_2(x,s,\theta) = \sum_{j=1}^{J-K-1} \frac{1}{j!} \left\| g_x^{(K+j)}(x,s,\theta) \right\| + \frac{1}{(J-K)!} \sup_{x \in \mathcal{X}} \left\| g_x^{(J)}(x,s,\theta) \right\|.$$

Therefore Assumption 4 (ii) is satisfied if (i) $g_x^{(J)}(x,s,\theta)$ exists on $\mathcal{X}$ for some $K < J \leqslant M$ and $\sup_{x\in\mathcal{X}} \left\| g_x^{(J)}(x,s,\theta) \right\| < \infty$ for every $(s,\theta) \in \mathcal{S} \times \Theta$ (Condition (6.1)); (ii) $G_x^{(J_G)}(x,s,\theta)$ exists on $\mathcal{X}$ for some $K < J_G \leqslant M$ and $\sup_{x\in\mathcal{X}} \left\| G_x^{(J_G)}(x,s,\theta) \right\| < \infty$ for every $(s,\theta) \in \mathcal{S} \times B_\delta(\theta_{0n})$ (Condition (6.2)).

Assumption 5.

(i) for some $\eta > 0$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left(\sum_{k=0}^{K} \left(\left\| g_{xi}^{(k)*}(\theta) \right\|^{1+\eta} + \left\| G_{xi}^{(k)*}(\theta) \right\| \right) + b_{1i}^*(\theta) + b_{2i}^*(\theta) \right) + \|g_i^*(\theta_{0n})\|^{2+\eta} + \|G_i^*(\theta_{0n})\|^{1+\eta} \right] < C;$$

(ii) for some $C > 0$, $C \leq \lambda_{\min}\left(\Omega_{gg}^*\right)$ for $\Omega_{gg}^* \equiv \mathbb{E}\left[g_i^*(\theta_{0n}) g_i^*(\theta_{0n})'\right]$;

(iii) for some $\delta > 0$, $\operatorname{vec}(G_i^*(\theta))$ is a.s. differentiable with respect to (w.r.t.) θ on $B_\delta(\theta_{0n})$ and

$$\mathbb{E}\left[\sup_{\theta\in B_\delta(\theta_{0n})}\left(\sum_{k=0}^{K}\left\|g_{xi}^{(k)*}(\theta)\right\|^2+b_{1i}^{*2}(\theta)+b_{2i}^{*2}(\theta)+b_{G1i}^{*}(\theta)+b_{G2i}^{*}(\theta)\right.\right.$$
$$\left.\left.+\left\|g_i^*(\theta)\right\|\left\|G_i^*(\theta)\right\|+\left\|\nabla_\theta \mathrm{vec}\left(G_i^*(\theta)\right)\right\|\right)\right]<C;$$

(iv) $\sup_{\beta\in\mathcal{B}}\left\|\hat{\Xi}(\beta)-\Xi(\beta)\right\|=o_{p,n}(1)$, where $\Xi(\beta)$ is a symmetric matrix satisfying $0<1/C<\inf_{\beta\in\mathcal{B}}\lambda_{\min}(\Xi(\beta))\leqslant\lambda_{\max}(\Xi(\beta))<C$, and, for some $\delta>0$, $\sup_{\beta\in B_\delta(\beta_0)}\left\|\nabla_\beta vec(\hat{\Xi}(\beta))\right\|=O_{p,n}(1)$.

Assumption 5 is a collection of low-level regularity conditions. Note that Condition (iv) allows the weighting matrix $\hat{\Xi}$ to be potentially a function of β, covering the standard continuously updated GMM with $\hat{\Xi}_{\mathrm{CUE}}(\beta)=\left(\hat{\Omega}_{\psi\psi}(\beta)\right)^{-1}$. Note that, if $\hat{\Xi}$ does not depend on β, Condition (iv) simplifies to the standard consistency requirement: $\hat{\Xi}=\Xi+o_{p,n}(1)$, where Ξ is a symmetric matrix with the eigenvalues bounded from zero and above.

Remark 21. Thanks to the MME framework and Assumption 4, the effect of the measurement error is localized. Hence, the asymptotic properties of the estimators and tests depend on the distribution of (X_i^*, S_i) only, and the effect of ε_{in} (i.e., the magnitude of the EIV bias) is fully captured by its (first K) moments. As a result, we can formulate the regularity conditions in terms of the true data (X_i^*, S_i), which greatly increases transparency of the assumptions and simplifies the exposition.

Assumptions 1-5 are the basic set of primitive conditions specifying the framework that we will rely on throughout the rest of the paper. Next, we consider the sufficient conditions for asymptotic normality of $\hat{\theta}$.

First, we discuss the issue of local identification. As pointed out in the previous section, local identification of β_{0n} is controlled by the Jacobian of the moment condition $\Psi=(\Psi_\theta,\Psi_\gamma)$. Note that, in the MME framework, the effect of ε_{in} on Ψ vanishes asymptotically, and we have $\Psi_\theta=G^*+o_n(1)$ and $\Psi_\gamma=\Psi_\gamma^*+o_n(1)$, where $\Psi_\gamma^*=\left(-g_x^{(2)*},\ldots,-g_x^{(K)*}\right)$. Hence, for β_{0n} to be locally identified, the limiting Jacobian

$$\Psi^*=\left(G^*,-g_x^{(2)*},\ldots,-g_x^{(k)*}\right),$$

must have full column rank, or, alternatively, we require $\lambda_{\min}(\Psi^{*\prime}\Psi^*)>C>0$. In this case, the standard (pointwise) GMM asymptotic theory applies: the distribution of $\hat{\beta}$ is given by (4.8) and the standard GMM tests and confidence intervals are valid. However, as we have seen, $g_x^{(k)*}$ may be equal to 0 and, consequently, the rank condition could be violated for some DGPs of interest. To address this issue and facilitate the analysis, we assume that identifiability of γ_{0n} is controlled by θ_{01n} in the following sense.

Assumption 6 (Strong local ID)**.** For all $\delta>0$, $\exists C_\delta>0$ such that, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\|\geqslant\delta$, $\lambda_{\min}(\Psi^{*\prime}\Psi^*)>C_\delta$.

Assumption 7 (Semi-strong local ID)**.**

(i) For $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant \delta_n$ with any $\delta_n \downarrow 0$, $g_x^{(k)*} = A_k^* \theta_{01n} + o_n(\|\theta_{01n}\|)$ for some uniformly bounded matrices A_k^* for $k \in \{2, \ldots, K\}$.

(ii) There exists $\delta_0 > 0$ such that for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant \delta_0$: $\lambda_{\min}(\Psi_A^{*\prime}\Psi_A^*) > C_A > 0$ for $\Psi_A^* \equiv (G^*, A_2^*, \ldots, A_K^*)$;

(iii) $g_x^{(k)} - g_x^{(k)*} = o_n(\|\theta_{01n}\|)$ for $k \in \{2, \ldots, K\}$.

Assumption 6 defines the strong identification part of the true parameter space: once $\|\theta_{01n}\|$ is bounded away from 0, the standard local identification condition is satisfied and the textbook GMM asymptotic theory applies: $\hat{\beta}$ is asymptotically normal and follows (4.8). Its asymptotic variance Σ can be consistently estimated, and the standard GMM tests (e.g., Wald, LM, and QLR) are uniformly valid over this part of the true parameter space. Note that Assumption 6 also ensures that, if $\|\theta_{01n}\|$ is bounded away from zero, then so are $\left\|g_x^{(k)*}\right\|$, $k \in \{2, \ldots, K\}$.

Assumption 7 (i) introduces $\theta_{01n} = 0$ as a point of γ_{0n} identification failure and specifies the behavior of $g_x^{(k)*}$, $k \in \{2, \ldots, K\}$, in its neighborhood: when $\|\theta_{01n}\|$ goes to zero, these derivatives can be approximated by linear functions of θ_{01n}. This assumption allows us to explicitly specify the parts of the true parameter space, for which γ_{0n} is weakly identified. For example, DGP sequences satisfying $n^{1/2}\|\theta_{01n}\| \to C$ with fixed $C \in \mathbb{R}$ are classical weak identification sequences like the ones defined in Andrews and Cheng (2012). Under sequences of this type, the nuisance parameter γ_{0n} is weakly identified and cannot be consistently estimated and, as a result, the estimator $\hat{\beta}_{0n}$ has a non-standard asymptotic distribution.

At the same time, Assumption 7 (ii) allows to restore consistency of $\hat{\gamma}$ and asymptotic normality of $\hat{\beta}_{0n}$ for DGP sequences satisfying $\|\theta_{01n}\| \to 0$ and $n^{1/2}\|\theta_{01n}\| \to \infty$. Such DGP sequences are called semi-strong identification sequences (Andrews and Cheng, 2012). Under these sequences, $\|\theta_{01n}\|$ is allowed to converge to 0, the point of γ_{0n} identification failure, but slowly enough so $\hat{\gamma}$ is still consistent for γ_{0n}. However, unlike in the standard strong identification scenario, the rate of convergence of $\hat{\gamma}$ is no longer $n^{1/2}$ but $n^{1/2}\|\theta_{01n}\|$, so $n^{1/2}\|\theta_{01n}\|(\hat{\gamma} - \gamma_{0n}) = O_{p,n}(1)$.[16] Moreover, once scaled appropriately, the joint asymptotic normality of $\hat{\theta}$ and $\hat{\gamma}$ still holds. Specifically, the distribution of $n^{1/2}\left((\hat{\theta} - \theta_{0n})', \|\theta_{01n}\|(\hat{\gamma} - \gamma_{0n})'\right)'$ can be (asymptotically) approximated by a standard normal with a well defined variance-covariance matrix.

Note that Assumption 7 (ii) is an analogue of the standard local identification condition. Indeed, once $\|\theta_{01n}\|$ converges to zero, the limiting Jacobian has the following approximation:

$$\Psi^* = (G^*, -A_2^*\theta_{01n}, \ldots, -A_K^*\theta_{01n}) + o_n(\|\theta_{01n}\|).$$

Then, after accounting for the slower rate of convergence for $\hat{\gamma}$, the rescaled Jacobian takes the (asymptotic) form of

$$\left(G^*, -A_2^*\frac{\theta_{01n}}{\|\theta_{01n}\|}, \ldots, -A_K^*\frac{\theta_{01n}}{\|\theta_{01n}\|}\right).$$

[16]This reduces to the standard $\sqrt{n}$-consistency of $\hat{\gamma}$ if $\|\theta_{01n}\|$ is bounded away from zero.

Assumption 7 (ii) guarantees that the rescaled Jacobian delivers local identification of the normalized parameter vector regardless of the direction of $\frac{\theta_{01n}}{\|\theta_{01n}\|}$ (which is also important unless θ_{01n} is scalar).

Assumption 7 (iii) is a weak regularity condition. Intuitively, it ensures that, once $\|\theta_{01n}\|$ is small, not only the $g_x^{(k)*}$ linearization reminder is $o_n(\|\theta_{01n}\|)$, but also the difference between $g_x^{(k)}$ and $g_x^{(k)*}$ is $o_n(\|\theta_{01n}\|)$. Consequently, we also have $g_x^{(k)} = A_k^* \theta_{01n} + o_n(\|\theta_{01n}\|)$ for $k \in \{2, \ldots, K\}$.

Assumption 7 is not restrictive and can be verified in many models of interest. For example, in the appendix, we verify it for the generalized linear model.

Assumptions 6 and 7 ensure local identification of θ_{0n} under both strong and semi-strong identification of γ_{0n}. The following assumption is their natural extension to global terms.

Assumption 8 (Global Identification)**.** There exist functions $\zeta_\theta(\cdot), \zeta_\gamma(\cdot)$: $\mathbb{R}_+ \to \mathbb{R}_+$ satisfying:

(i) for all $\{\Upsilon_n\}$, for all $\theta \in \Theta$, $\gamma \in \Gamma$,

$$\|\psi^*(\theta, \gamma)\| \geqslant \zeta_\theta(\|\theta - \theta_{0n}\|) + \zeta_\gamma(\|\theta_{01n}\| \, \|\gamma - \gamma_{0n}\|) + r_n(\theta, \gamma),$$

where $\sup_{\theta \in \Theta, \gamma \in \Gamma} r_n(\theta, \gamma) = o_n(1)$;

(ii) for any $\eta > 0$, $\inf_{\kappa \geqslant \eta} \zeta_\theta(\kappa) > 0$ and $\inf_{\kappa \geqslant \eta} \zeta_\gamma(\kappa) > 0$.

Assumption 8 ensures global identification of θ_{0n} (and γ_{0n} when $\|\theta_{01n}\|$ is bounded away from zero). Condition (i) provides a bound on the norm of the population moment $\|\psi^*(\theta, \gamma)\|$ in the form of the sum of $\zeta_\theta(\|\theta - \theta_{0n}\|)$ and $\zeta_\gamma(\|\theta_{01n}\| \, \|\gamma - \gamma_{0n}\|)$ (and an asymptotically negligible reminder). Then Condition (ii) becomes an identification condition: $\|\psi^*(\theta, \gamma)\|$ can only be small when both $\|\theta - \theta_{0n}\|$ and $\|\theta_{01n}\| \, \|\gamma - \gamma_{0n}\|$ are small. Again, to account for a possible identification failure, $\|\gamma - \gamma_{0n}\|$ is multiplied by the factor $\|\theta_{01n}\|$, which controls the strength of identification of γ_{0n}.

The following theorem establishes asymptotic normality of the MME estimator under both strong and semi-strong identification of γ_{0n}.

Theorem 2 (Asymptotic Normality)**. (i)** *Suppose* $0_{K-1} \in \operatorname{int}(\Gamma)$. *Also suppose* $\hat{Q}(\hat{\theta}, \hat{\gamma}) \leqslant \inf_{\theta \in \Theta, \gamma \in \Gamma} \hat{Q}(\theta, \gamma) + o_{p,n}(1)$ *and* $\nabla_\beta \hat{Q}(\hat{\beta}) = o_{p,n}(n^{-1/2})$. *Then, under Assumptions 1-6, 8, for all* $\{\Upsilon_n\}$ *satisfying* $\|\theta_{01n}\| \geqslant \delta$ *for any fixed* $\delta > 0$, *the asymptotic distribution of* $\hat{\beta}$ *is given by* (4.8),

(ii) *Moreover, if we suppose* $\hat{Q}(\hat{\theta}, \hat{\gamma}) \leqslant \inf_{\theta \in \Theta, \gamma \in \Gamma} \hat{Q}(\theta, \gamma) + O_{p,n}(n^{-1})$ *and* $\nabla_\beta \hat{Q}(\hat{\beta}) = o_{p,n}(n^{-1})$, *then, under additional Assumption 7, the same result still holds for all* $\{\Upsilon_n\}$ *satisfying* $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ *for any fixed* $C > 0$ *and* $0 < \omega_{\mathrm{L}} < 1/2$.

Part (i) is a standard result: asymptotic normality of the MME estimator $\hat{\theta}$ under "textbook" conditions was also previously demonstrated in Evdokimov and Zeleneev (2016). Once $\|\theta_{01n}\|$ is assumed to be bounded away from zero, the classical local identification condition is satisfied, and the standard GMM asymptotic approximations and tests are (uniformly) valid. Note that, to ensure that $\hat{\theta}$ is asymptotically normal, we need to make an additional assumption on Γ to avoid the parameter on the boundary problem. Since $\gamma_{0n} \to 0$, it is sufficient to require $0_{K-1} \in \operatorname{int}(\Gamma)$ in order to make sure that γ_{0n} is bounded away from the boundary of the optimization parameter space Γ.

Part (ii) extends this result to the semi-strong identification region of the true parameter space. It turns out that the same asymptotic approximation (4.8) remains valid even when $\|\theta_{01n}\|$ is allowed to shrink towards zero, provided that the speed of that convergence is not too fast (slower than the $\sqrt{n}$ rate). This means that, once we restrict Υ_n to satisfy $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ with any fixed $\omega_{\mathrm{L}} \in (0, 1/2)$, the standard tests still remain uniformly valid.

Testing a linear hypothesis about β_{0n}

Consider a linear hypothesis about β_{0n}: $H_0 : \lambda'\beta_{0n} = v$ against $H_a : \lambda'\beta_{0n} \neq v$ for some fixed non-zero $\lambda \in \mathbb{R}^{p+K-1}$ and hypothesized $v \in \mathbb{R}$. Theorem 2 suggests that H_0 can be tested by using the standard t-statistic

$$t \equiv \frac{n^{1/2}(\lambda'\hat{\beta} - v)}{\sqrt{\lambda'\hat{\Sigma}\lambda}},$$

where $\hat{\Sigma}$ is a standard estimator of the asymptotic variance given by

$$\hat{\Sigma} = (\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}_{\psi\psi}\hat{\Xi}\hat{\Psi}(\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1}, \qquad \hat{\Psi} = \overline{\Psi}(\hat{\beta}), \quad \hat{\Xi} = \hat{\Xi}(\hat{\beta}). \tag{6.3}$$

Under the null, the t-statistic converges in distribution to $N(0, 1)$ for all DGP sequences allowed in Theorem 2, which is formalized by the following lemma:

Lemma 1. **(i)** *For any fixed $\lambda \in \mathbb{R}^{p+K-1}$, $\lambda \neq 0$, under hypotheses of Theorem 2 Part (i), for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta$ for any fixed $\delta > 0$,*

$$\frac{n^{1/2}\lambda'(\hat{\beta} - \beta_{0n})}{\sqrt{\lambda'\hat{\Sigma}\lambda}} \xrightarrow{d} N(0, 1),$$

where $\hat{\Sigma}$ is given by (6.3).

(ii) *Under hypotheses of Theorem 2 Part (ii), the same result still holds for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ for any fixed $C > 0$ and $0 < \omega_{\mathrm{L}} < 1/2$.*

Hence, under hypotheses of Theorem 2, the standard t-test is uniformly valid over strong and semi-strong identification parts of the true parameter space.

Remark 22. When $\|\theta_{01n}\|$ is bounded away from zero (strong identification), the asymptotic variance Σ given in (4.9) has eigenvalues uniformly bounded from zero and above and can be consistently estimated by $\hat{\Sigma}$. However, once $\|\theta_{01n}\|$ approaches zero, some components (corresponding to $\hat{\gamma}$) of Σ become unbounded. In this case, $\hat{\Sigma}$ is no longer guaranteed to be consistent for Σ (however, in the appendix, we show that, after a proper rescaling, it still converges to a well defined limiting object). Nonetheless, we still have $n^{1/2}\hat{\Sigma}^{-1/2}(\hat{\beta} - \beta_{0n}) \to N(0, I_{p+K-1})$ and, as demonstrated by Lemma 1, the standard t-statistic based on $\hat{\Sigma}$ is still uniformly valid under semi-strong identification for testing a linear hypothesis about β_{0n}. Unfortunately, this result cannot be straightforwardly generalized to a nonlinear hypothesis about β_{0n} if some components of γ_{0n} are non-trivially involved. However, Σ_θ, the $\hat{\theta}$ corresponding submatrix of Σ, still has eigenvalues bounded from zero and above and is consistently estimated by $\hat{\Sigma}_\theta$ under semi-strong identification. Hence, if one is interested in inference on θ_{0n} (or a non-linear function of it), which is typically the parameter of interest, the standard tests based on $\hat{\Sigma}_\theta$ remain uniformly valid under semi-strong identification of γ_{0n}. For example, in the next section, we establish this result for the standard Wald test based on $\hat{\theta}$ and $\hat{\Sigma}_\theta$.

6.2 Uniform square-root-n consistency

In this section, we demonstrate uniform $\sqrt{n}$-consistency of the MME estimator $\hat{\theta}$ over a wide range of DGPs without assuming any particular strength of identification of γ_{0n}. The following additional assumption helps to establish the result.

Assumption 9 (Uniform $\sqrt{n}$-consistency)**.**

(i) For some $\eta > 0$, for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant \delta_n$ with any $\delta_n \downarrow 0$, $g_x^{(k)*} = A_k^* \theta_{01n} + O_n(\|\theta_{01n}\|^{1+\eta})$ for some uniformly bounded matrices A_k^* for $k \in \{2, \dots, K\}$;

There exists $\delta_0 > 0$ such that for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant \delta_0$:

(ii) for some $\delta > 0$, $\operatorname{vec}\left(G_{xi}^{(k)*}(\theta)\right)$ is a.s. differentiable w.r.t. θ on $\theta \in B_\delta(\theta_{0n})$ for $k \in \{2, \dots, K\}$ and

$$\mathbb{E}\left[\sum_{k=2}^{K}\left(\left\|G_{xi}^{(k)*}(\theta_{0n})\right\|^{1+\eta} + \sup_{\theta \in B_\delta(\theta_{0n})}\left\|\nabla_\theta \operatorname{vec}\left(G_{xi}^{(k)*}(\theta)\right)\right\|\right)\right] < C;$$

(iii) $\inf_{\gamma\in\Gamma} \lambda_{\min}\left(\Psi_A^*(\gamma)'\Psi_A^*(\gamma)\right) > C > 0$, where $\Psi_A^*(\gamma) \equiv (\Psi_\theta^*(\theta_{0n}, \gamma), A_2^*, \dots, A_K^*)$;

(iv) $g_x^{(k)} - g_x^{(k)*} = O_n(\sigma_n \|\theta_{01n}\|)$ for $k \in \{2, \dots, K\}$.

Assumption 9 is a stronger version of Assumption 7. Since Assumption 7 only ensures asymptotic normality of $\hat{\theta}$ under semi-strong identification, it is needed to be strengthened to ensure uniform $\sqrt{n}$-consistency of the MME estimator over the *entire* true parameter space. Condition (i) is analogous to Assumption 7 (i) but requires the linearization remainder to be $O_n(\|\theta_{01n}\|^{1+\eta})$ instead if $o_n(\|\theta_{01n}\|)$. Similarly, Condition (iv) corresponds to Assumption 7 (iii) but the requirement imposed on the reminder is strengthened from $o_n(\|\theta_{01n}\|)$ to $o_n(\sigma_n \|\theta_{01n}\|)$.

Condition (iii) is a stronger analogue of Assumption 7 (ii). Indeed, $\Psi_A^*(0) = \Psi_A^*$, so Assumption 7 (ii) automatically follows from Condition (iii). It turns out that, under weak identification of γ_{0n} (i.e. when $n^{1/2}\|\theta_{01n}\|$ is bounded), a weaker requirement $\inf_{\gamma\in\Gamma}\lambda_{\min}(\Psi_\theta^*(\gamma)'\Psi_\theta^*(\gamma)) > C > 0$ ensures $\sqrt{n}$-consistency of $\hat{\theta}$. However, in order to ensure uniform $\sqrt{n}$-normality over the *entire* true parameter space, it needs to be strengthened to cover both weak and semi-strong identification regimes. Condition (iii) addresses this issue by essentially merging the requirement $\inf_{\gamma\in\Gamma}\lambda_{\min}(\Psi_\theta^*(\gamma)'\Psi_\theta^*(\gamma)) > C > 0$ (needed under weak identification) with Assumption 7 (ii) (needed under semi-strong identification).

Finally, Condition (ii) is an additional weak smoothness requirement. Assumption 9 is not restrictive and can be verified in many cases of empirical interest. For example, in the appendix, we provide a number of basic low-level conditions, which make sure that Assumption 9 is satisfied for the generalized linear model (GLM).

Theorem 3 (Uniform $\sqrt{n}$-consistency)**.** *Suppose* $\hat{Q}(\hat{\theta},\hat{\gamma}) \leqslant \inf_{\theta\in\Theta,\gamma\in\Gamma} \hat{Q}(\theta,\gamma) + O_{p,n}(n^{-1})$*. Then, under Assumptions 1-6, 8, 9,* $\hat{\theta} - \theta_{0n} = O_{p,n}(n^{-1/2})$.

Remark 23. Note that, for $\|\theta_{01n}\| \leqslant Cn^{-1/2}$, $\hat{\gamma}$ is inconsistent. In this case, the standard two-step GMM optimal weighting matrix estimator $\hat{\Xi}_{\mathrm{GMM2}} = \hat{\Omega}_{\psi\psi}(\hat{\theta}_p, \hat{\gamma}_p)^{-1}$ is also no longer consistent since a preliminary step estimator $\hat{\gamma}_p$ is not consistent. Hence, under weak identification, $\hat{\Xi}_{\mathrm{GMM2}}$ violates Assumption 5 (iv) and may behave unstably. To address this issue, we propose using the following regularized two-step optimal weighting matrix estimator

$$\hat{\Xi}_{\mathrm{GMM2,R}} \equiv \hat{\Omega}_{\psi\psi}(\hat{\theta}_p, 0)^{-1},$$

where $\hat{\theta}_p$ is a preliminary first step MME estimator. Instead of using an estimate of γ_{0n}, the regularized estimator $\hat{\Xi}_{\mathrm{GMM2,R}}$ sets $\gamma = 0$, which is the limiting point of γ_{0n} under the MME asymptotics. Since Theorem 3 ensures $\hat{\theta}_p = \theta_{0n} + O_{p,n}(n^{-1/2})$, we have $\hat{\Xi}_{\mathrm{GMM2,R}} \xrightarrow{p} \Omega^*_{gg} \equiv \mathbb{E}\left[g_i^* g_i^{*\prime}\right]$, which is optimal under semi-strong and strong identification. Similarly, we introduce a regularized version of the CUE weighting matrix estimator given by

$$\hat{\Xi}_{\mathrm{CUE,R}}(\theta) \equiv \hat{\Omega}_{\psi\psi}(\theta, 0)^{-1}.$$

Unlike the standard CUE weighting matrix $\hat{\Xi}_{\mathrm{CUE}}(\theta,\gamma) \equiv \hat{\Omega}_{\psi\psi}(\theta,\gamma)^{-1}$, $\hat{\Xi}_{\mathrm{CUE,R}}(\theta)$ is not a function of γ, which is instead fixed at 0. Similarly, under the MME asymptotics, $\hat{\Xi}_{\mathrm{CUE,R}}(\theta)$ is (asymptotically) optimal under semi-strong and strong identification and yet provides stability under weak identification.

Remark 24. Also note that the proof of Theorem 3 allows to relax Assumption 5 (iv). A sufficient condition needed to be imposed on the weighting matrix is

$$0 < 1/C < \inf_{\theta\in\Theta,\gamma\in\Gamma} \lambda_{\min}(\hat{\Xi}(\theta,\gamma)) \leqslant \sup_{\theta\in\Theta,\gamma\in\Gamma} \leqslant \lambda_{\max}(\hat{\Xi}(\theta,\gamma)) < C$$

with probability approaching one. To get uniform $\sqrt{n}$-consistency of $\hat{\theta}$, we do not have to require consistency of the weighting matrix like Assumption 5 (iv) does. Hence, despite being inconsistent under weak identification, using the standard (non-regularized) estimator $\hat{\Xi}_{\mathrm{GMM2}}$ does not threaten validity of Theorem 3 provided that the aforementioned condition is satisfied. Specifically, the following condition ensures that $\lambda_{\min}(\hat{\Omega}_{\psi,\psi}$ is bounded away from zero with probability approaching one despite $\hat{\gamma}$ being inconsistent:

$$\inf_{\gamma\in\Gamma} \lambda_{\min}(\Omega^*_{\psi\psi}(\theta_{0n},\gamma)) > C > 0, \qquad \text{for all } \{\Upsilon_n\} : \|\theta_{01n}\| \leqslant \delta_0 \text{ for some } \delta_0 > 0. \tag{6.4}$$

Also note that Assumption 5 (iii) and boundedness of $\Gamma \ni \gamma$ already imply that $\sup_{\gamma\in\Gamma} \lambda_{\max}(\Omega^*_{\psi\psi}(\theta_{0n},\gamma)) < C$. Under these conditions, the eigenvalues $\hat{\Xi}_{\mathrm{GMM2}}$ are bounded away from zero and above with probability approaching one.

Remark 25. Similarly to the regularized weighting matrices, we also introduce a regularized version of the asymptotic variance estimator $\hat{\Sigma}$, which sets $\gamma = 0$ instead of plugging $\hat{\gamma}$. Specifically, we introduce

$$\hat{\Sigma}_{\mathrm{R}} \equiv (\hat{\Psi}_{\mathrm{R}}'\hat{\Xi}\hat{\Psi}_{\mathrm{R}})^{-1}\hat{\Psi}_{\mathrm{R}}'\hat{\Xi}\hat{\Omega}_{\psi\psi,\mathrm{R}}\hat{\Xi}\hat{\Psi}_{\mathrm{R}}(\hat{\Psi}_{\mathrm{R}}'\hat{\Xi}\hat{\Psi}_{\mathrm{R}})^{-1},$$
$$\hat{\Psi}_{\mathrm{R}} \equiv \overline{\Psi}(\hat{\theta}, 0), \qquad \hat{\Omega}_{\psi\psi,\mathrm{R}} \equiv \hat{\Omega}_{\psi\psi}(\hat{\theta}, 0).$$

Since, under the MME asymptotics $\gamma_{0n} \to 0$, the regularized estimator of the asymptotic variance $\hat{\Sigma}_{\mathrm{R}}$ is a valid alternative to $\hat{\Sigma}$.

6.3 Uncorrected estimator

Theorem 3 establishes uniform $\sqrt{n}$-consistency of the MME estimator $\hat{\theta}$ regardless the strength of identification of the nuisance parameter γ_{0n}. However, when γ_{0n} is weakly identified (for example, under DGP sequences satisfying $n^{1/2}\|\theta_{01n}\| \to C$ with some finite C), the asymptotic normality of $\hat{\theta}$ breaks down. In this case, the tests based on the asymptotic approximation (4.8) fail to control size and/or suffer from power loss. Our objective is to develop inference tools, which are valid under weak identification of the nuisance parameter.

The inference procedure we propose is based on the following insight. Consider a DGP sequence with $\|\theta_{01n}\|$ shrinking towards 0, the point of γ_{0n} identification failure. Then Assumption 7 (i) implies that $\mathbb{E}[g(X_i, S_i, \theta_{0n})] = O_n(\sigma_n^2\|\theta_{01n}\|)$. Hence, if $\|\theta_{01n}\|$ goes to 0 fast enough in the sense $\sigma_n^2\|\theta_{01n}\| = o_n(n^{-1/2})$, the uncorrected moment conditions satisfy $\mathbb{E}[g(X_i, S_i, \theta_{0n})] = o_n(n^{-1/2})$. In this case, unlike in the strong identification scenario (when $\|\theta_{01n}\|$ is bounded away from zero), the measurement error correction is no longer needed: the asymptotic distribution of the GMM estimator based on the uncorrected set of moment is not affected by the measurement error.

Specifically, we define the uncorrected estimator as

$$\hat{\theta}_{\mathrm{U}} = \operatorname*{argmin}_{\theta\in\Theta} \hat{Q}_{\mathrm{U}}(\theta),$$

where

$$\hat{Q}_{\mathrm{U}}(\theta) = \overline{g}_{\mathrm{U}}(\theta)'\hat{\Xi}_{\mathrm{U}}(\theta)\overline{g}_{\mathrm{U}}(\theta).$$

Note that the uncorrected estimator $\hat{\theta}_{\mathrm{U}}$ is based on a potentially different set of moments (and a weighting matrix) compared to the MME estimator $\hat{\theta}$.

Theorem 4. *Suppose that (i) $\mathbb{E}[g_{\mathrm{U}}(X_i^*, S_i, \theta_{0n})] = 0$ and $\lambda_{\min}(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^*) > C > 0$; (ii) $\hat{\theta}_{\mathrm{U}} = \theta_{0n} + o_{p,n}(1)$; (iii) Assumptions 4 and 5 are satisfied with g_{U} and $\hat{\Xi}_{\mathrm{U}}$ as g and $\hat{\Xi}$, respectively; (iv) $g_{\mathrm{U}x}^{(k)*} = A_{\mathrm{U}k}^*\theta_{01n} + O_n(\|\theta_{01n}\|^2)$ for some uniformly bounded $A_{\mathrm{U}k}^*$ for $k \in \{2, \ldots, K\}$; (v) $\hat{Q}_{\mathrm{U}}(\hat{\theta}_{\mathrm{U}}) \leqslant \inf_{\theta\in\Theta}\hat{Q}(\theta_{\mathrm{U}}) + o_{p,n}(1)$ and $\nabla_\theta\hat{Q}_{\mathrm{U}}(\hat{\theta}_{\mathrm{U}}) = o_{p,n}(n^{-1/2})$. Then, under Assumptions 1 and 2, we have*

$$n^{1/2}\Sigma_{\mathrm{U}}^{-1/2}(\hat{\theta}_{\mathrm{U}} - \overline{\theta}_{\mathrm{U}0n}) \xrightarrow{d} N(0, I_p), \qquad \Sigma_{\mathrm{U}} \equiv (G_{\mathrm{U}}\Xi_{\mathrm{U}}G_{\mathrm{U}}')^{-1}G_{\mathrm{U}}'\Xi_{\mathrm{U}}\Omega_{g_{\mathrm{U}}g_{\mathrm{U}}}\Xi_{\mathrm{U}}G_{\mathrm{U}}(G_{\mathrm{U}}'\Xi_{\mathrm{U}}G_{\mathrm{U}})^{-1},$$

where

$$\overline{\theta}_{\mathrm{U}0n} = \theta_{0n} + B_{\mathrm{U}}^{*}\theta_{01n} + O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2),$$
$$B_{\mathrm{U}}^{*} \equiv -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}\sum_{k=2}^{K}\frac{\mathbb{E}\left[\varepsilon_{in}^{k}\right]}{k!}A_{\mathrm{U}k}^{*},$$

provided that $n^{1/2}\left(\sum_{k=2}^{K}\frac{\mathbb{E}[\varepsilon_{in}^{k}]}{k!}A_{\mathrm{U}k}^{*}\right)\theta_{01n} = O_n(1)$ *and* $n^{1/2}\sigma_n^2\left\|\theta_{01n}\right\|^2 = O_n(1)$.

Theorem specifies the asymptotic distribution of the uncorrected estimator $\hat{\theta}_{\mathrm{U}}$ and provides an explicit expression for its asymptotic bias due to the presence of the measurement error. The asymptotic bias of $\hat{\theta}_{\mathrm{U}}$ consists of two parts: the main linear part $B_{\mathrm{U}}^{*}\theta_{01n}$ and the quadratic reminder $O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2)$. Also note that the last hypothesis of the theorem requires $B_{\mathrm{U}}^{*}\theta_{01n} + O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2) = O_n(n^{-1/2})$, so the expression for the asymptotic bias is guaranteed to be valid only if it is not "too large".

Remark 26. The rest of the hypotheses of the theorem are standard. Condition (iv) is similar to Assumption 7 (i). However, it strengthens the requirement on the linearization remainder from $o_n(\left\|\theta_{01n}\right\|)$ to $O_n(\left\|\theta_{01n}\right\|^2)$. This results in an explicit bound on the asymptotic bias remainder equal to $O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2)$.

Note that, if $B_{\mathrm{U}}^{*}\theta_{01n}+O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2) = o_n(n^{-1/2})$, $\hat{\theta}_{\mathrm{U}}$ is asymptotically unbiased. As announced before, the leading bias term $B_{\mathrm{U}}^{*}\theta_{01n}$ can be bounded as $O_n(\sigma_n^2\left\|\theta_{01n}\right\|)$, and this requirement reduces to $\sigma_n^2\left\|\theta_{01n}\right\| = o(n^{-1/2})$. However, in some special cases (e.g., NLIV), we may have $A_{\mathrm{U}k}^{*} = 0$, $k \in \{2,\ldots,K\}$. In this case, this requirement weakens to $\sigma_n^2\left\|\theta_{01n}\right\|^2 = o_n(n^{-1/2})$, allowing for a substantially larger range of $\left\|\theta_{01n}\right\|$.

Corollary 1. *Suppose that the hypotheses of Theorem 4 are satisfied. Then, for* $\{\Upsilon_n\}$*, satisfying* $\left\|\theta_{01n}\right\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$ *for any fixed* $C \geqslant 0$ *and* $\omega_{\mathrm{NZ}} \geqslant 1/2 - 1/(K+1)$*, we have*

$$n^{1/2}\Sigma_{\mathrm{U}}^{-1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n}) \xrightarrow{d} N(0, I_p). \tag{6.5}$$

Moreover, if $A_{\mathrm{U}k}^{*} = 0$ *for* $k \in \{2,\ldots,K\}$*, then the former statement can be strengthen to* $\omega_{\mathrm{NZ}} \geqslant 1/4 - 1/(2K+2)$.

Corollary 1 provides the bounds on the possible value of ω_{NZ}, under which $\hat{\theta}_{\mathrm{U}}$ is asymptotically unbiased, and the standard tests based on (6.5) and the asymptotic variance estimator

$$\hat{\Sigma}_{\mathrm{U}} = (\hat{G}_{\mathrm{U}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}}')^{-1}\hat{G}_{\mathrm{U}}'\hat{\Xi}_{\mathrm{U}}\hat{\Omega}_{g_{\mathrm{U}}g_{\mathrm{U}}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}}(\hat{G}_{\mathrm{U}}'\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}})^{-1} \tag{6.6}$$

are valid.

7 Uniformly Valid Inference and Adaptive Estimation

7.1 Asymptotic properties of the hybrid tests

In this section, we formally study asymptotic properties of the proposed hybrid tests in the context of MME framework. We want to test

$$H_0 : r(\theta_{0n}) = v \qquad \text{vs} \qquad H_1 : r(\theta_{0n}) \neq v. \tag{7.1}$$

We start with two tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$, which have correctly defined p-values $p_{\mathcal{T}_{\mathrm{L}}}$ and $p_{\mathcal{T}_{\mathrm{NZ}}}$ and take the form $\phi_{\mathcal{T}} = \mathbb{1}\{p_{\mathcal{T}} < \alpha\}$ for $\mathcal{T} \in \{\mathcal{T}_{\mathrm{L}}, \mathcal{T}_{\mathrm{NZ}}\}$, where α denotes the nominal level of the tests. We also suppose that there exist some ω_{L} and ω_{NZ} such that (i) $0 < \omega_{\mathrm{NZ}} < \omega_{\mathrm{L}} < 1/2$; (ii) $\mathcal{T}_{\mathrm{L}}$ is (uniformly) valid for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$; (iii) $\mathcal{T}_{\mathrm{NZ}}$ is (uniformly) valid for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{\omega_{\mathrm{NZ}}}$. Formally, $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ satisfy

$$\limsup_{n\to\infty} \sup_{\Upsilon_n : \|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}} \mathbb{E}_{\Upsilon_n}[\phi_{\mathcal{T}_{\mathrm{L}}}] \leqslant \alpha, \tag{7.2}$$

$$\limsup_{n\to\infty} \sup_{\Upsilon_n : \|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}} \mathbb{E}_{\Upsilon_n}[\phi_{\mathcal{T}_{\mathrm{NZ}}}] \leqslant \alpha, \tag{7.3}$$

for all $C > 0$.

Theorem 2 specifies the asymptotic distribution of $\hat{\theta}$ under semi-strong and strong identification of γ_{0n}. This result, along with the asymptotic variance estimator $\hat{\Sigma}_{\mathrm{L}}$ defined in (6.3), can be used to construct a test $\mathcal{T}_{\mathrm{L}}$, which would satisfy (7.2) with any $\omega_{\mathrm{L}} \in (0, 1/2)$.

Similarly, using the result of Corollary 1, it is straightforward to construct a test of (7.1) based on the uncorrected estimator $\hat{\theta}_{\mathrm{U}}$, its asymptotic distribution (6.5) and the asymptotic variance estimator $\hat{\Sigma}_{\mathrm{U}}$ as in (6.6). Such a test would satisfy (7.3) with any $\omega_{\mathrm{NZ}} \geqslant 1/2 - 1/(K+1)$ (or even any $\omega_{\mathrm{NZ}} \geqslant 1/4 - 1/(2K+2)$ for certain uncorrected estimators, e.g. NLIV).

Example (Wald tests)**.** As an example of $\mathcal{T}_{\mathrm{L}}$, one can take a simple Wald test based on $\hat{\theta}$. Specifically, let $R(\theta) \equiv \nabla_\theta r(\theta)$ and $\hat{R} \equiv R(\hat{\theta})$. Then, the $\hat{\theta}$ based Wald statistics takes the form of

$$\mathcal{W}_{\mathrm{L}} = n(r(\hat{\theta}) - v)' \left(\hat{R}\hat{\Sigma}_\theta \hat{R}'\right)^{-1} (r(\hat{\theta}) - v),$$

and the outcome and the p-value of $\mathcal{T}_{\mathrm{L}}$ can be taken as

$$\phi_{\mathcal{T}_{\mathrm{L}}} = \mathbb{1}\{\mathcal{W}_{\mathrm{L}} > \chi^2_{d_r, 1-\alpha}\} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{L}}} < \alpha\}, \qquad p_{\mathcal{T}_{\mathrm{L}}} = 1 - F_{\chi^2_{d_r}}(\mathcal{W}_{\mathrm{L}}), \tag{7.4}$$

where $d_r \equiv \dim(r(\theta_{0n}))$, and $\chi^2_{d_r,1-\alpha}$ and $F_{\chi^2_{d_r}}$ denote the $1-\alpha$ quantile and the CDF of a $\chi^2_{d_r}$, respectively. As mentioned before, such a test satisfies (7.2) with any $\omega_{\mathrm{L}} \in (0, 1/2)$.

$\mathcal{T}_{\mathrm{NZ}}$ can be constructed in the same way but with $\hat{\theta}_{\mathrm{U}}$ taking place of $\hat{\theta}$. Specifically, the corresponding test statistic is given by

$$\mathcal{W}_{\mathrm{U}} = n(r(\hat{\theta}_{\mathrm{U}}) - v)' \left(\hat{R}_{\mathrm{U}}\hat{\Sigma}_{\mathrm{U}} \hat{R}_{\mathrm{U}}'\right)^{-1} (r(\hat{\theta}_{\mathrm{U}}) - v)$$

with $\hat{R}_{\mathrm{U}} \equiv R(\hat{\theta}_{\mathrm{U}})$. The outcome and the p-value of the test analogously take the form of

$$\phi_{\mathcal{T}_{\mathrm{U}}} = \mathbb{1}\{\mathcal{W}_{\mathrm{U}} > \chi^2_{d_r, 1-\alpha}\} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{U}}} < \alpha\}, \qquad p_{\mathcal{T}_{\mathrm{U}}} = 1 - F_{\chi^2_{d_r}}(\mathcal{W}_{\mathrm{U}}). \tag{7.5}$$

Such a test would satisfy (7.3) with any $\omega_{\mathrm{L}} \geqslant 1/2 - 1/(K+1)$.

The tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ are adaptively combined based on the inferred strength of identification measured by the following identification-category-selection statistic:

$$\hat{A}_{\mathrm{ICS}} = (n\hat{\theta}_1' \hat{V}_{11}^{-1} \hat{\theta}_1 / p_1)^{1/2}, \tag{7.6}$$

where $\hat{\theta}_1$ is the MME estimator of θ_{01n}, $p_1 \equiv \dim(\theta_1)$, and $\hat{V}_{11}$ is a symmetric positive definite scaling matrix, which needs to satisfy the following regularity conditions.

Assumption 10. $\hat{V}_{11}$ is a symmetric positive definite matrix and

(i) $0 < 1/C < \lambda_{\min}(\hat{V}_{11})$ for some $C > 0$ with probability approaching one uniformly;

(ii) for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ with any fixed $C > 0$, $\lambda_{\max}(\hat{V}_{11})$ is uniformly bounded with probability approaching one.

Condition (i) ensures that $\hat{V}_{11}^{-1}$ is uniformly bounded with probability one, and Condition (ii) ensures that, with probability approaching one, $\lambda_{\min}(\hat{V}_{11}^{-1})$ is uniformly bounded away from below under semi-strong and strong identification. Since, for the MME estimator, we have $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$ (Theorem 3), these conditions ensure that (i) for some $\overline{C} > 0$, $\hat{A}_{\mathrm{ICS}} < \overline{C}\,\|\theta_{01n}\|$ with probability approaching one; (ii) for some $\underline{C} > 0$, $\hat{A}_{\mathrm{ICS}} > \underline{C}\,\|\theta_{01n}\|$ with probability approaching one for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$. These conditions together ensure that $\hat{A}_{\mathrm{ICS}}$ is "small" under weak identification (when $\|\theta_{01n}\|$ is relatively "small") and "large" under strong identification (when $\|\theta_{01n}\|$ is relatively "large").

Although, for example, an identity matrix as $\hat{V}_{11}$ clearly satisfies Assumption 10, it is desired to choose $\hat{V}_{11}$ in a data dependent way, so it represents a measure of sample uncertainty about θ_{01n}. Then, two natural choices of $\hat{V}_{11}$ are $\hat{\Sigma}_{\theta_1}$ and $\hat{\Sigma}_{\mathrm{R},\theta_1}$, which are the θ_1-corresponding submatrices of $\hat{\Sigma}$ and $\hat{\Sigma}_{\mathrm{R}}$, respectively.[17] Both $\hat{\Sigma}_{\theta_1}$ and $\hat{\Sigma}_{\mathrm{R},\theta_1}$ are consistent estimators of the asymptotic variance of $\hat{\theta}_1$ under semi-strong and strong identification. However, their properties under weak identification are less clear. The following lemma guarantees that both the estimators satisfy Assumption 10.

Lemma 2. *Under Assumptions 2-9, $\hat{\Sigma}_{\mathrm{R},\theta_1}$ satisfies Assumption 10 with any $\omega_{\mathrm{L}} \in (0, 1/2)$. If, in addition,* (6.4) *is also satisfied, then $\hat{\Sigma}_{\theta_1}$ also satisfies Assumption 10 with any $\omega_{\mathrm{L}} \in (0, 1/2)$.*

Remark 27. Unlike the regularized estimator $\hat{\Sigma}_{\mathrm{R}}$, the non-regularized estimator $\hat{\Sigma}$ depends on $\hat{\gamma}$, which is not consistent under weak identification. As a result, in this case, $\hat{\Sigma}$ is also random. To address this issue, Lemma 2 introduces an additional assumption to ensure that $\hat{\Sigma}$ does not become singular in this case.

Recall that the p-value the type-I hybrid test is given by

$$p^I_{\mathcal{T}_{\mathrm{H}}} = (1 - \hat{\lambda}_{\mathrm{L}}) \max\{p_{\mathcal{T}_{\mathrm{NZ}}}, p_{\mathcal{T}_{\mathrm{L}}}\} + \hat{\lambda}_{\mathrm{L}} p_{\mathcal{T}_{\mathrm{L}}}, \qquad \hat{\lambda}_{\mathrm{L}} = \lambda_{\mathrm{L}}(\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{L},n}). \tag{7.7}$$

[17] With $V_{11} = \hat{\Sigma}_{\theta_1}$, $\hat{A}_{\mathrm{ICS}}$ is exactly the ICS statistic of Andrews and Cheng (2012).

Then, the hybrid test rejects the null whenever its p-value is less than α, where α denotes its nominal level. So, the outcome of $\mathcal{T}_{\mathrm{H}}^{I}$ is given by

$$\phi_{\mathcal{T}_{\mathrm{H}}}^{I} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{H}}}^{I} < \alpha\}. \tag{7.8}$$

First, we establish the asymptotic properties of the type-I hybrid tests under the following high-level condition.

Assumption 11. There exist some ω_{L} and ω_{NZ}, $0 < \omega_{\mathrm{NZ}} < \omega_{\mathrm{L}} < 1/2$, such that:

(i) under the null, the tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ satisfy (7.2) and (7.3) (with any fixed $C > 0$), respectively;

(ii) Assumption 10 is satisfied;

(iii) $n^{\omega_{\mathrm{L}}}(n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$ as $n \to \infty$;

(iv) $\lambda_{\mathrm{L}} : \mathbb{R} \to [0,1]$ is continuous and weakly increasing, $\lambda_{\mathrm{L}}(z) = 0$ for $z \leqslant 0$, and $\lim_{z\to+\infty} \lambda_{\mathrm{L}}(z) = 1$.

Theorem 5. *Suppose that $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$. Then, under Assumption 11:*
(i) *the hybrid test $\mathcal{T}_{\mathrm{H}}^{I}$ given by (7.8) is asymptotically uniformly valid, i.e., under the null,*

$$\limsup_{n\to\infty} \sup_{\Upsilon_n} \mathbb{E}_{\Upsilon_n}[\phi_{\mathcal{T}_{\mathrm{H}}}^{I}] \leqslant \alpha;$$

(ii) *moreover, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\|/(n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$, $\left|p_{\mathcal{T}_{\mathrm{H}}}^{I} - p_{\mathcal{T}_{\mathrm{L}}}\right| = o_{p,n}(1)$ (uniformly over the hypothesized value v in (7.1)).*

The first part of Theorem 5 establishes (asymptotic) uniform validity of the type-I hybrid test. The second part says that $\mathcal{T}_{\mathrm{H}}^{I}$ becomes essentially equivalent to $\mathcal{T}_{\mathrm{L}}$ whenever $\|\theta_{01n}\|$ is substantially larger than $n^{-1/2}\kappa_{\mathrm{L},n}$. Note that, the slower $\kappa_{\mathrm{L},n}$ increases as $n \to \infty$, the more powerful the hybrid test becomes since it switches to $\mathcal{T}_{\mathrm{L}}$ more aggressively. Therefore, if the test $\mathcal{T}_{\mathrm{L}}$ has certain optimality properties under semi-strong and strong identification, then also the test $\mathcal{T}_{\mathrm{H}}^{I}$ does (under semi-strong and strong identification DGP sequences satisfying $\|\theta_{01n}\|/(n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$). In particular, if $n^{-1/2}\kappa_{\mathrm{L},n} \to 0$ as $n \to \infty$, then $\mathcal{T}_{\mathrm{H}}^{I}$ is guaranteed to be asymptotically equivalent to $\mathcal{T}_{\mathrm{L}}$ under strong identification (whenever $\|\theta_{01n}\|$ is bounded away from zero).

Also recall that, for the type-II hybrid test $\mathcal{T}_{\mathrm{H}}^{II}$, the p-value is given by

$$\begin{aligned} p_{\mathcal{T}_{\mathrm{H}}}^{II} &= (1 - \hat{\lambda}_{\mathrm{NZ}} - \hat{\lambda}_{\mathrm{L}}) \max\{p_{\mathcal{T}_{\mathrm{NZ}}}, p_{\mathcal{T}_{\mathrm{L}}}\} + \hat{\lambda}_{\mathrm{NZ}} p_{\mathcal{T}_{\mathrm{NZ}}} + \lambda_{\mathrm{L}} p_{\mathcal{T}_{\mathrm{L}}}, \\ \hat{\lambda}_{\mathrm{NZ}} &\equiv \lambda_{\mathrm{NZ}}(\kappa_{\mathrm{NZ},n} - \hat{A}_{\mathrm{ICS}}), \qquad \hat{\lambda}_{\mathrm{L}} \equiv \lambda_{\mathrm{L}}(\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{L},n}). \end{aligned} \tag{7.9}$$

Similarly, the outcome of $\mathcal{T}_{\mathrm{H}}^{II}$ is given by

$$\phi_{\mathcal{T}_{\mathrm{H}}}^{II} = \mathbb{1}\{p_{\mathcal{T}_{\mathrm{H}}}^{II} < \alpha\}. \tag{7.10}$$

The type-II hybrid test $\mathcal{T}_{\mathrm{H}}^{II}$ is a more aggressive version of $\mathcal{T}_{\mathrm{H}}^{I}$: it is equivalent to $\mathcal{T}_{\mathrm{H}}^{I}$ when $\hat{A}_{\mathrm{ICS}} \geqslant \kappa_{\mathrm{NZ},n}$ but starts switching from the robust critical value $\max\{p_{\mathcal{T}_{\mathrm{NZ}}}, p_{\mathcal{T}_{\mathrm{L}}}\}$ to non-conservative $p_{\mathcal{T}_{\mathrm{NZ}}}$

when $\hat{A}_{\text{ICS}} < \kappa_{\text{NZ},n}$ (i.e. when $\|\theta_{01n}\|$ appear to be relatively "small"). Similarly, we establish asymptotic properties of $\mathcal{T}_{\text{H}}^{II}$ under the following high-level assumption.

Assumption 12. Suppose that Assumption 11 is satisfied. In addition, suppose that:

(i) for $0 < \kappa_{\text{NZ},n} < \kappa_{\text{L},n}$, we have $n^{\omega_{\text{NZ}}}(n^{-1/2}\kappa_{\text{NZ},n}) \to 0$ as $n \to \infty$;

(ii) $\lambda_{\text{NZ}} : \mathbb{R} \to [0,1]$ is continuous and weakly increasing, $\lambda_{\text{NZ}}(z) = 0$ for $z \leqslant 0$, and $\lim_{z\to+\infty} \lambda_{\text{NZ}}(z) = 1$.

Theorem 6. *Suppose that $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$. Then, under Assumption 12:*
(i) *the hybrid test $\mathcal{T}_{\text{H}}^{II}$ given by* (7.10) *is asymptotically uniformly valid, i.e., under the null,*

$$\limsup_{n\to\infty} \sup_{\Upsilon_n} \mathbb{E}_{\Upsilon_n}[\phi_{\mathcal{T}_{\text{H}}}^{II}] \leqslant \alpha;$$

(ii) *moreover, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\|/(n^{-1/2}\kappa_{\text{L},n}) \to \infty$, $\left|p_{\mathcal{T}_{\text{H}}}^{II} - p_{\mathcal{T}_{\text{L}}}\right| = o_{p,n}(1)$ (uniformly over the hypothesized value v in* (7.1)*);*
(iii) *finally, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\|/(n^{-1/2}\kappa_{\text{NZ},n}) \to 0$, $\left|p_{\mathcal{T}_{\text{H}}}^{II} - p_{\mathcal{T}_{\text{NZ}}}\right| = o_{p,n}(1)$ (uniformly over the hypothesized value v in* (7.1)*).*

Similarly to $\mathcal{T}_{\text{H}}^{I}$, the type-II hybrid test is (i) asymptotically uniformly valid and (ii) essentially equivalent to $\mathcal{T}_{\text{L}}$ when $\|\theta_{01n}\|/(n^{-1/2}\kappa_{\text{NZ},n})$ is large (i.e. under certain semi-strong and strong identification DGP sequences). However, it also becomes equivalent to $\mathcal{T}_{\text{NZ}}$ (and inherits its optimality properties), when $\|\theta_{01n}\|$ is much smaller than $n^{-1/2}\kappa_{\text{NZ},n}$. Similarly, the faster $\kappa_{\text{NZ},n}$ grows, the more powerful $\mathcal{T}_{\text{H}}^{II}$ becomes. In particular, if $\kappa_{\text{NZ},n} \to \infty$ as $n \to \infty$, then $\mathcal{T}_{\text{H}}^{II}$ is guaranteed to be asymptotically equivalent to $\mathcal{T}_{\text{NZ}}$ under weak identification of γ_{0n} (i.e. whenever $n^{1/2}\|\theta_{01n}\| \leqslant C$).

Note that, under the provided conditions, both the hybrid tests are theoretically demonstrated to be asymptotically uniformly valid. At the same time, by design, $\mathcal{T}_{\text{H}}^{II}$ is less conservative than $\mathcal{T}_{\text{H}}^{I}$ and has more attractive power properties (both asymptotically and in finite samples). Hence, judging by the asymptotic properties only, it may be appealing to conclude that the type-II hybrid test $\mathcal{T}_{\text{H}}^{II}$ is simply superior to the type-I hybrid test $\mathcal{T}_{\text{H}}^{I}$. We want to stress that, although asymptotic analysis provides valuable guidance on which test should be and, more importantly, which tests should not be used in practice, one needs to take these results with a grain of salt. That being said, we want to emphasize that, while $\mathcal{T}_{\text{H}}^{II}$ has better power properties, the type-I hybrid test $\mathcal{T}_{\text{H}}^{I}$ has less tuning parameters (and hence is less sensitive to their sometimes arbitrary choice) and, being more conservative, is more likely to provide credible inference in finite samples.

So far, we have provided the high-level conditions under which the hybrid tests (based on some abstract tests $\mathcal{T}_{\text{L}}$ and $\mathcal{T}_{\text{NZ}}$) are demonstrated to be (asymptotically) uniformly valid. However, the general results of Theorems 5 and 6 can also be straightforwardly applied in the MME framework. For example, the following theorem establishes the same asymptotic properties of $\mathcal{T}_{\text{H}}^{I}$ and $\mathcal{T}_{\text{H}}^{II}$ based on the standard Wald tests given by (7.4) and (7.5) as $\mathcal{T}_{\text{L}}$ and $\mathcal{T}_{\text{NZ}}$, respectively.

Theorem 7. *Suppose that the hypotheses of Theorems 2, 3, and 4 are satisfied. Also, suppose that (i) $\hat{V}_{11} = \hat{\Sigma}_{\text{R},\theta_1}$ or $\hat{V}_{11} = \hat{\Sigma}_{\theta_1}$ and* (6.4) *holds; (ii) $n^{\omega_{\text{L}}}(n^{-1/2}\kappa_{\text{L},n}) \to \infty$ as $n \to \infty$ for* *some* $\omega_{\mathrm{L}} \in (1/2 - 1/(K+1), 1/2)$*; (iii)* λ_{L} *satisfies Assumption 11* (iv)*; (iv)* $r(\theta)$ *is continuously differentiable on* Θ *and* $\lambda_{\min}(RR') > C > 0$ *with* $R \equiv R(\theta_{0n})$*. Finally, suppose that* $\mathcal{T}_{\mathrm{L}}$ *and* $\mathcal{T}_{\mathrm{NZ}}$ *are given by* (7.4) *and* (7.5)*, respectively. Then, the type-I hybrid test* (7.8) *satisfies the assertions of Theorem 5. Suppose that, in addition, we also have (v)* $n^{\omega_{\mathrm{NZ}}}(n^{-1/2}\kappa_{\mathrm{NZ},n}) \to 0$ *as* $n \to \infty$ *for some* ω_{NZ} *satisfying* $1/2 - 1/(K+1) \leqslant \omega_{\mathrm{NZ}} < \omega_{\mathrm{L}} < 1/2$*, (vi)* λ_{NZ} *satisfies Assumption 12* (ii)*. Then, the type-II hybrid test* (7.10) *satisfies the assertions of Theorem 6.*

Remark 28. As pointed out before (see, e.g. Corollary 1), for certain estimators, e.g. NLIV, the requirement in (v) can be substantially weakened to $1/4 - 1/(2K+2) \leqslant \omega_{\mathrm{NZ}} < \omega_{\mathrm{L}} < 1/2$, which allows for substantially larger range of $\kappa_{\mathrm{NZ},n}$.

Remark 29. Although the suggested hybrid tests are similar to the procedures proposed in Andrews and Cheng (2012), we want to stress some differences. Andrews and Cheng (2012) derive the asymptotic distribution of a test statistic under certain weak, semi-strong, and strong identification DGP sequences. Then, to run a test or to construct a confidence set, they propose a number of ways to compute an identification robust critical value, which should be used (instead of the standard critical value) to ensure uniform validity. Some of the suggested approaches (like in Sections 5.2 and 5.3 in Andrews and Cheng (2012)) adaptively compute the critical value depending on the inferred strength of identification measured by the same ICS statistic $\hat{A}_{\mathrm{ICS}}$. This differs from the procedure suggested in this paper: while Andrews and Cheng (2012) use the same test statistic and adjust the critical value, our tests continuously switch between $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{NZ}}$ depending on the value of $\hat{A}_{\mathrm{ICS}}$. The second difference is that, unlike the approaches proposed in Andrews and Cheng (2012), the type-II hybrid test $\mathcal{T}_{\mathrm{H}}^{II}$ not only reduces to the proper test $\mathcal{T}_{\mathrm{L}}$ under strong identification but are also asymptotically equivalent to $\mathcal{T}_{\mathrm{NZ}}$ under weak identification (when $n^{1/2}\|\theta_{01n}\| \leqslant C$). This means that we can perfectly discriminate between weak and strong identification sequences without violating uniform validity of the tests. It becomes possible because the validity regions of the test $\mathcal{T}_{\mathrm{NZ}}$ and $\mathcal{T}_{\mathrm{L}}$ have a non-trivial overlap when $\omega_{\mathrm{NZ}} < \omega_{\mathrm{L}}$ (which is demonstrated to be the case in the MME framework).

7.2 Specific choices of the basic tests

To construct any of the hybrid tests we described above, one needs to choose the tests $\mathcal{T}_{\mathrm{NZ}}$ and $\mathcal{T}_{\mathrm{L}}$. For example, as Theorem 4 suggests, the standard Wald test based on the uncorrected estimator $\hat{\theta}_{\mathrm{U}}$ is a natural candidate for $\mathcal{T}_{\mathrm{NZ}}$. With a proper choice of moments, such a test can achieve local semiparametric efficiency for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$. Moreover, in a fully parametric setting, $\hat{\theta}_{\mathrm{U}}$ can be replaced by the ML estimator $\hat{\theta}_{\mathrm{MLE}}$ (with the score constructing a set of just identified moments). Then, a standard test based on $\hat{\theta}_{\mathrm{MLE}}$ will be efficient in this setting. Finally, as mentioned before, the NLIV estimator $\hat{\theta}_{\mathrm{NLIV}}$ is a more conservative choice of $\hat{\theta}_{\mathrm{U}}$, which should provide a better finite sample coverage but less informative inference.

The choice of $\mathcal{T}_{\mathrm{L}}$ is more intricate. One of the possibilities is build such a test on the MME estimator $\hat{\theta}$. Under the hypotheses of Theorem 2, $\hat{\theta}$ is asymptotically normal and can be used to construct a standard Wald test. However, one of the standard premises needed to ensure asymptotic normality requires the true parameter to be in the interior of the parameter space. Specifically, Theorem 2 requires $0_{K-1} \in \Gamma$, where 0_{K-1} is the limiting point of γ_{0n} under moderate measurement error asymptotics. This restrictions results in Γ to be chosen suboptimally. For example, consider the simplest correction scheme with $K = 2$. In this case, $\gamma_{0n} = \sigma_n^2/2$ and, hence, it is a priori known that γ_{0n} is non-negative. This suggests that it is reasonable to choose $\Gamma = [0, \bar{\gamma}]$ for some $\bar{\gamma} > 0$. Such a choice, however, would necessarily violate the requirement of Theorem 2: the nuisance parameter γ_{0n} can be on or arbitrary close to the boundary of Γ, and, consequently, asymptotic normality of the MME estimator $\hat{\theta}$ would be threatened. It means that, in order to secure asymptotic normality of the estimator, one should choose $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for some $\underline{\gamma} < 0$, with $|\underline{\gamma}|$ being sufficiently large, for the finite sample distribution of $\hat{\theta}$ not being distorted by the potential boundary problem. Hence, either (i) Γ is chosen conservatively or (ii) asymptotic normality of $\hat{\theta}$ is not ensured.[18] Clearly, in the first scenario, useful information about the structure of γ_{0n} is ignored. This necessarily leads to less informative inference, especially when γ_{0n} is close to 0 (relative to the finite sample standard error). The same parameter on the boundary issue also applies in more general setup (when $K > 2$).

Instead of modifying Γ a priori, it is still possible to exploit the natural restrictions on γ_{0n} after an asymptotically normal estimator (of both θ_{0n} and γ_{0n}) is obtained. One of the solutions is to use the conditional likelihood ratio (CLR) test of Ketz (2018) specifically designed for subvector inference with some of the parameters being on or close to the parameter space boundary. Although the CLR test requires a numerical simulation of the critical values, (i) it is computationally inexpensive even for $K > 2$ and (ii) the p-value of the test becomes directly available after the simulation. Moreover, Ketz (2018) shows that the CLR test enjoys certain optimality properties if the parameter of interest ($r(\theta_{0n})$ in our case) is scalar. Specifically, Ketz (2018) establishes that, in the limiting Gaussian shift experiment, the CLR test is admissible and, building on the result of Montiel Olea (2019), almost WAP (weighted average power) maximizing subject to the similarity constraint. In the simulation study, Ketz (2018) demonstrates that the CLR test has competitive power properties compared to the optimal test of Elliott et al. (2015) and the WAP-similar test of Montiel Olea (2019). Consequently, the CLR test can be recommended as $\mathcal{T}_{\mathrm{L}}$.

7.3 Adaptive Hybrid Estimation

The idea behind construction of the hybrid tests naturally extends to estimation: the MME estimator $\hat{\theta}$ and the uncorrected estimator $\hat{\theta}_{\mathrm{U}}$ can be adaptively combined based on the inferred strength of identification. Specifically, we propose using the following convex combination of $\hat{\theta}$ and $\hat{\theta}_{\mathrm{U}}$ as an adaptive estimator:

$$\hat{\theta}_{\mathrm{A}} = \hat{\Lambda}_n \hat{\theta} + (1 - \hat{\Lambda}_n)\hat{\theta}_{\mathrm{U}}, \qquad \hat{\Lambda}_n \equiv \Lambda\left(\frac{\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{NZ},n}}{\kappa_{\mathrm{L},n} - \kappa_{\mathrm{NZ},n}}\right). \tag{7.11}$$

[18] Note that, for estimation purposes, we recommend imposing restrictions on Γ.

Here $\hat{A}_{\text{ICS}}$ is the same ICS statistic used to infer the strength of identification, and $\Lambda : \mathbb{R} \to [0,1]$ is a weakly increasing function satisfying $\Lambda(z) = 0$ for all $z \leqslant 0$, $\Lambda(z) = 1$ for all $z \geqslant 1$ and $|\Lambda(z') - \Lambda(z)| \leqslant \overline{\Lambda}\,|z' - z|$ for all $z', z \in \mathbb{R}$. As before, $\kappa_{\text{NZ},n}$ and $\kappa_{\text{L},n} > \kappa_{\text{NZ},n}$ are the thresholds, which need to grow at certain rates as the sample size grows.

Note that $\hat{\theta}_{\text{A}} = \hat{\theta}_{\text{U}}$ whenever $\hat{A}_{\text{ICS}} \leqslant \kappa_{\text{NZ},n}$: in this case the value of $\|\theta_{01n}\|$ is sufficiently small, so the asymptotic approximation $\hat{\theta}_{\text{U}} \overset{a}{\sim} N(\theta_{0n}, \Sigma_{\text{U}}/n)$ applies (Theorem 4). Similarly, $\hat{\theta}_{\text{A}} = \hat{\theta}$ if $\hat{A}_{\text{ICS}} \geqslant \kappa_{\text{L},n}$: $\|\theta_{01n}\|$ is large enough to guarantee that $\hat{\theta} \overset{a}{\sim} N(\theta_{0n}, \Sigma_{\theta}/n)$ is an adequate approximation (Theorem 2). Finally, if $\hat{A}_{\text{ICS}} \in (\kappa_{\text{NZ},n}\kappa_{\text{L},n})$, then both the approximations are valid and we have

$$\begin{pmatrix} \hat{\theta} \\ \hat{\theta}_{\text{U}} \end{pmatrix} \overset{a}{\sim} N\left(\begin{pmatrix} \theta_{0n} \\ \theta_{0n} \end{pmatrix}, \begin{pmatrix} \Sigma_{\theta}/n & \Sigma_{\theta\theta_{\text{U}}}/n \\ \Sigma'_{\theta\theta_{\text{U}}}/n & \Sigma_{\text{U}}/n \end{pmatrix} \right),$$

and, consequently, a convex combination of $\hat{\theta}$ and $\hat{\theta}_{\text{U}}$ is also (approximately) normally distributed. Specifically, the asymptotic variance of the adaptive estimator $\hat{\theta}_{\text{A}}$ can be approximated by

$$\hat{\Sigma}_{\text{A}} = \hat{\Lambda}_n^2 \hat{\Sigma}_{\theta} + \hat{\Lambda}_n (1 - \hat{\Lambda}_n)(\hat{\Sigma}_{\theta\theta_{\text{U}}} + \hat{\Sigma}'_{\theta\theta_{\text{U}}}) + (1 - \hat{\Lambda}_n)^2 \hat{\Sigma}_{\text{U}}. \tag{7.12}$$

Here $\hat{\Sigma}_{\theta}$ is the θ-corresponding submatrix of $\hat{\Sigma}$ given in (6.3), $\hat{\Sigma}_{\text{U}}$ as in (6.6), and $\hat{\Sigma}_{\theta\theta_{\text{U}}}$ is a standard estimator of the asymptotic covariance of two GMM estimators $\hat{\theta}$ and $\hat{\theta}_{\text{U}}$ (an example is provided in the appendix). The following theorem shows that approximation $\hat{\theta}_{\text{A}} \overset{a}{\sim} N(\theta_{0n}, \hat{\Sigma}_{\text{A}}/n)$ is (asymptotically) uniformly valid irrespectively of the strength of identification and, hence, can be used to draw uniformly valid inference on θ_{0n}.

Theorem 8. *Suppose that the hypotheses of Theorems 2, 3, and 4 are satisfied. Also, suppose that (i) $\hat{V}_{11} = \hat{\Sigma}_{\text{R},\theta_1}$ or $\hat{V}_{11} = \hat{\Sigma}_{\theta_1}$ and (6.4) holds; (ii) $\kappa_{\text{NZ},n} < \kappa_{\text{L},n}$ satisfy $n^{\omega_{\text{L}}-1/2}\kappa_{\text{NZ},n} \to \infty$ and $n^{\omega_{\text{NZ}}-1/2}\kappa_{\text{L},n} \to 0$ for some $1/2 - 1/(K+1) \leqslant \omega_{\text{NZ}} < \omega_{\text{L}} < 1/2$. Then, we have, for any $\{\Upsilon_n\}$,*

$$n^{1/2}\hat{\Sigma}_{\text{A}}^{-1/2}(\hat{\theta}_{\text{A}} - \theta_{0n}) \xrightarrow{d} N(0, I_p),$$

where $\hat{\theta}_{\text{A}}$ and $\hat{\Sigma}_{\text{A}}$ are given by (7.11) and (7.12).

Remark 30. Theorem 8 is, of course, also valid if, instead of $\hat{\Sigma}_{\theta}$ and $\hat{\Sigma}_{\theta\theta_{\text{U}}}$, one uses their regularized analogues $\hat{\Sigma}_{\text{R},\theta}$ and $\hat{\Sigma}_{\text{R},\theta\theta_{\text{U}}}$.

Remark 31. Theorem 8 establishes uniform asymptotic normality of the adaptive estimator based on the MME estimator $\hat{\theta}$ and the uncorrected estimator $\hat{\theta}_{\text{U}}$ in the MME context. In the appendix, we also provide a general result, which can be used to establish asymptotic normality of hybrid estimators based on some estimators $\hat{\theta}_{\text{L}}$ and $\hat{\theta}_{\text{NZ}}$ in a general context, i.e. when the variance of the measurement error is potentially "large".

References

Amemiya, Y. (1985): "Instrumental variable estimator for the nonlinear errors-in-variables model," *Journal of Econometrics*, 28, 273–289.

Andrews, D. W. K. (2017): "Identification-Robust Subvector Inference," *Cowles Foundation Discussion Paper.*

Andrews, D. W. K. and X. Cheng (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

——— (2013): "Maximum likelihood estimation and uniform inference with sporadic identification failure," *Journal of Econometrics*, 173, 36–56.

——— (2014): "GMM Estimation And Uniform Subvector Inference With Possible Identification Failure," *Econometric Theory*, 30, 287–333.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Papers 1813, Cowles Foundation for Research in Economics, Yale University.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84, 2155–2182.

Andrews, I. and A. Mikusheva (2016): "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

Battistin, E. and A. Chesher (2014): "Treatment effect estimation with covariate measurement error," *Journal of Econometrics*, 178, 707–715.

Chaudhuri, S. and E. Zivot (2011): "A new method of projection-based inference in GMM with weakly identified nuisance parameters," *Journal of Econometrics*, 164, 239–251.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49, 901–937.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *The Review of Economic Studies*, 72, 343–366.

Chen, X. and D. Pouzo (2015): "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013–1079.

Cheng, X. (2015): "Robust inference in nonlinear models with mixed identification strength," *Journal of Econometrics*, 189, 207–228.

Chernozhukov, V., W. K. Newey, and A. Santos (2015): "Constrained Conditional Moment Restriction Models," .

Chesher, A. (1991): "The effect of measurement error," *Biometrika*, 78, 451–462.

Cox, G. (2017): "Weak Identification in a Class of Generically Identified Models with an Application to Factor Models," Working Paper.

Dufour, J.-M. (1989): "Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions," *Econometrica: Journal of the Econometric Society*, 335–355.

Dufour, J.-M. and J. Jasiak (2001): "Finite sample limited information inference methods for structural equations and models with generated regressors," *International Economic Review*, 42, 815–844.

Dufour, J.-M. and M. Taamouti (2005): "Projection-based statistical inference in linear structural models with possibly weak instruments," *Econometrica*, 73, 1351–1365.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Evdokimov, K. S. and A. Zeleneev (2016): "Simple Estimation of Semiparametric Models with Measurement Errors," *Working Paper, Princeton University.*

Guggenberger, P., J. J. Ramalho, and R. J. Smith (2012): "GEL statistics under weak identification," *Journal of Econometrics*, 170, 331–349.

Guggenberger, P. and R. J. Smith (2005): "Generalized empirical likelihood estimators and tests under partial, weak, and strong identification," *Econometric Theory*, 21, 667–709.

Han, S. and A. McCloskey (2019): "Estimation and inference with a (nearly) singular Jacobian," *Quantitative Economics*, 10, 1019–1068.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hausman, J. A., H. Ichimura, W. K. Newey, and J. L. Powell (1991): "Identification and estimation of polynomial errors-in-variables models," *Journal of Econometrics*, 50, 273–295.

Hausman, J. A., W. K. Newey, and J. L. Powell (1995): "Nonlinear errors in variables Estimation of some Engel curves," *Journal of Econometrics*, 65, 205–233.

Hu, Y. and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

Ketz, P. (2018): "Subvector inference when the true parameter vector may be near or at the boundary," *Journal of Econometrics*, 207, 285–306.

KLEIBERGEN, F. (2005): "Testing Parameters in GMM Without Assuming that They are Identified," *Econometrica*, 73, 1103–1123.

LI, T. (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110, 1–26.

MCCLOSKEY, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200, 17–35.

MONTIEL OLEA, J.-L. (2019): "Admissible, Similar Tests: A Characterization," *Forthcoming in Econometric Theory*.

MOREIRA, H. A. AND M. J. MOREIRA (2013): "Contributions to the theory of optimal tests," .

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349.

——— (2001): "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *The Review of Economics and Statistics*, 83, 616–627.

REIERSØL, O. (1950): "Identifiability of a Linear Relation between Variables Which Are Subject to Error," *Econometrica*, 18, 375–389.

SCHENNACH, S. M. (2004): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33–75.

——— (2007): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75, 201–239.

——— (2013): "Measurement Error in Nonlinear Models – A Review," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, 296–337.

——— (2016): "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 8, 341–377.

STOCK, J. H. AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

WANG, L. AND C. HSIAO (2011): "Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models," *Journal of Econometrics*, 165, 30–44.

A.1 Proofs: "Standard" conditions

A.1.1 UC of the sample moment

Lemma A.1. *Under Assumptions 1-5,* $\sup_{\theta\in\Theta,\gamma\in\Gamma}\left\|\overline{\psi}(\theta,\gamma)-\psi^*(\theta,\gamma)\right\| = o_{p,n}(1)$.

Proof. For brevity, denote $\beta \equiv (\theta',\gamma')'$ and $\mathcal{B} = \Theta\times\Gamma$. By the triangle inequality,

$$\sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}(\beta)-\psi^*(\beta)\right\| \leqslant \sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}^*(\beta)-\psi^*(\beta)\right\| + \sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}(\beta)-\overline{\psi}^*(\beta)\right\|. \tag{A.1}$$

First, we argue that $\sup_{\beta\in\mathcal{B}}||\overline{\psi}^*(\beta)-\psi^*(\beta)|| = o_{p,n}(1)$. Note that

$$\sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}^*(\beta)-\psi^*(\beta)\right\| \leqslant \sup_{\theta\in\Theta}\left\|\overline{g}^*(\theta)-g^*(\theta)\right\| + \sum_{k=2}^{K}\sup_{\gamma\in\Gamma}|\gamma_k|\sup_{\theta\in\Theta}\left\|\overline{g}_x^{(k)*}(\theta)-g_x^{(k)*}(\theta)\right\|,$$

so, since Γ is bounded (Assumption 3), it suffices to show that

$$\sup_{\theta\in\Theta}\left\|\overline{g}_x^{(k)*}(\theta)-g_x^{(k)*}(\theta)\right\| = o_{p,n}(1) \tag{A.2}$$

for $k\in\{0,2,\dots,K\}$. This result is ensured by Lemma A.11. Indeed, note that the conditions of Lemma A.11 are satisfied with $g_{xi}^{(k)*}(\theta)$ taking place of $\eta_i(\theta)$ for $k\in\{0,2,\dots,K\}$.

For the second term in (A.1), by the triangle inequality,

$$\sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}(\beta)-\overline{\psi}^*(\beta)\right\| \leqslant \sup_{\theta\in\Theta}\left\|\overline{g}(\theta)-\overline{g}^*(\theta)\right\| + \sum_{k=2}^{K}\sup_{\gamma\in\Gamma}|\gamma_k|\sup_{\theta\in\Theta}\left\|\overline{g}_x^{(k)}(\theta)-\overline{g}_x^{(k)*}(\theta)\right\|.$$

By Assumption 4,

$$\begin{aligned}\left\|\overline{g}_n(\theta)-\overline{g}_n^*(\theta)\right\| &= \left\|\sum_{k=1}^{K-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{k!}g_x^{(k)}(X_i^*,S_i,\theta)\varepsilon_i^k + \frac{1}{K!}\frac{1}{n}\sum_{i=1}^{n}g_x^{(K)}(\tilde{X}_i^*,S_i,\theta)\varepsilon_i^K\right\| \\ &\leqslant \sum_{k=1}^{K}\frac{1}{k!}\frac{1}{n}\sum_{i=1}^{n}\left\|g_x^{(k)}(X_i^*,S_i,\theta)\right\||\varepsilon_i|^k \\ &\quad + \frac{1}{K!}\left(\frac{1}{n}\sum_{i=1}^{n}b_1(X_i^*,S_i,\theta)|\varepsilon_i|^{K+1} + \frac{1}{n}\sum_{i=1}^{n}b_2(X_i^*,S_i,\theta)|\varepsilon_i|^M\right).\end{aligned}$$

where $\tilde{X}_i$ lies in between of X_i^* and X_i and is allowed to be component specific. By Assumption 2, Assumption 5 (i), Markov's inequality, and $\mathbb{E}[|\varepsilon_i|^k] = o_n(1)$ for $k\in\{1,\dots,M\}$, each term following the inequality sign is $o_{p,n}(1)$ (uniformly in $\theta\in\Theta$). By a nearly identical argument,

$$\sup_{\theta\in\Theta}\left\|\overline{g}_x^{(k)}(\theta)-\overline{g}_x^{(k)*}(\theta)\right\| = o_{p,n}(1)$$

for $k\in\{2,\dots,K\}$. As a result, $\sup_{\beta\in\mathcal{B}}\left\|\overline{\psi}(\beta)-\overline{\psi}^*(\beta)\right\| = o_{p,n}(1)$, and the conclusion of the Lemma follows from equation (A.1). Q.E.D.

A.1.2 Generic local consistency

Lemma A.2 (Generic consistency result).

Suppose Assumption 1 holds, and that for some $\delta > 0$, (i) $\eta(x,s,\theta)$ is Lipschitz continuous of order J in x on $B_\delta(\theta_{0n})$, i.e. there is a collection of functions $c_j(x,s,\theta)$ for $j \in \{1,\dots,J\}$ such that, for every $x, x' \in \mathcal{X}$ and $(s,\theta) \in \mathcal{S} \times B_\delta(\theta_{0n})$, $\|\eta(x',s,\theta) - \eta(x,s,\theta)\| \leqslant \sum_{j=1}^{J} c_j(x,s,\theta)\,|x'-x|^j$; (ii) $\mathbb{E}\left[\sup_{\theta \in B_\delta(\theta_{0n})} \|c_j(X_i^*, S_i, \theta)\|\right] < C$ for $j \in \{1,\dots,J\}$; (iii) $\mathrm{vec}\,(\eta(X_i^*, S_i, \theta))$ is a.s. differentiable w.r.t. θ on $B_\delta(\theta_{0n})$; (iv) $\mathbb{E}\left[\sup_{\theta\in B_\delta(\theta_{0n})} \|\nabla_\theta \mathrm{vec}\,(\eta(X_i^*, S_i, \theta))\|\right] < C$; (v) for some $\eta > 0$, $\mathbb{E}\left[\|\eta(X_i^*, S_i, \theta_{0n})\|^{1+\eta}\right] < C$; (vi) $\varepsilon_i \perp (X_i^*, S_i)$ and $\mathbb{E}\left[|\varepsilon_i|^J\right] = o_n(1)$. Then, for any $\tilde{\theta} = \theta_{0n} + o_{p,n}(1)$,

$$n^{-1}\sum_{i=1}^{n} \eta(X_i, S_i, \tilde{\theta}) = \mathbb{E}\left[\eta(X_i^*, S_i, \theta_{0n})\right] + o_{p,n}(1).$$

In addition, $\mathbb{E}\left[\eta(X_i, S_i, \theta_0)\right]$ exists and satisfies $\mathbb{E}\left[\eta(X_i, S_i, \theta_0)\right] = \mathbb{E}\left[\eta(X_i^*, S_i, \theta_0)\right] + o_n(1)$.

Proof. First, we argue that $\sup_{\theta\in B_\delta(\theta_{0n})} \|\overline{\eta}(\theta) - \overline{\eta}^*(\theta)\| = o_{p,n}(1)$. Indeed,

$$\sup_{\theta\in B_\delta(\theta_{0n})} \|\overline{\eta}(\theta) - \overline{\eta}^*(\theta)\| \leqslant \sum_{j=1}^{J} n^{-1} \sum_{i=1}^{n} \sup_{\theta\in B_\delta(\theta_{0n})} c_j^*(\theta)|\varepsilon_i|^j = O_{p,n}\left(\sum_{j=1}^{J} \mathbb{E}\left[|\varepsilon_i|^j\right]\right) = o_{p,n}(1),$$

where the inequality follows from condition (i), the first equality holds by condition (ii) and Markov's inequality, and the last equality holds by condition (vi). Note that, since $\tilde{\theta} - \theta_{0n} = o_{p,n}(1)$ implies $\tilde{\theta} \in B_\delta(\theta_{0n})$ with probability approaching one uniformly, we have established that $\overline{\eta}(\tilde{\theta}) = \overline{\eta}^*(\tilde{\theta}) + o_{p,n}(1)$.

The next step is to demonstrate $\overline{\eta}^*(\tilde{\theta}) - \overline{\eta}^*(\theta_{0n}) = o_{p,n}(1)$. Again, using $\tilde{\theta} \in B_\delta(\theta_{0n})$ with probability approaching one uniformly, we get

$$\left\|\overline{\eta}^*(\tilde{\theta}) - \overline{\eta}^*(\theta_{0n})\right\| \leqslant n^{-1}\sum_{i=1}^{n} \sup_{\theta\in B_\delta(\theta_{0n})} \|\nabla_\theta \mathrm{vec}\,(\eta(X_i^*, S_i, \theta))\| \left\|\tilde{\theta} - \theta_{0n}\right\| = o_{p,n}(1),$$

where the inequality follows from condition (iii), and the equality is ensured by condition (iv) along with $\left\|\tilde{\theta} - \theta_{0n}\right\| = o_{p,n}(1)$. So, we have proved that $\overline{\eta}(\tilde{\theta}) = \overline{\eta}^*(\theta_{0n}) + o_{p,n}(1)$. Combining with the previously obtained results, we obtain $\overline{\eta}(\tilde{\theta}) = \overline{\eta}^*(\theta_{0n}) + o_{p,n}(1)$. Finally, condition (v) ensures that $\overline{\eta}^*(\theta_{0n}) = \eta^*(\theta_{0n}) + o_{p,n}(1)$, which completes the proof.

Finally,

$$\eta(X_i, S_i, \theta_0) = \eta(X_i^*, S_i, \theta_0) + r(X_i, X_i^*, S_i, \theta_0),$$

where

$$\|r(X_i, X_i^*, S_i, \theta_0)\| \leqslant \sum_{j=1}^{J} c_j(X_i^*, S_i, \theta_0)\,|\epsilon_i|^j.$$

Since the expectation of the right hand side of the last inequality exists and is equal to $o_n(1)$, we conclude $\mathbb{E}\left[\eta(X_i, S_i, \theta_0)\right] = \mathbb{E}\left[\eta(X_i^*, S_i, \theta_0)\right] + o_n(1)$. Q.E.D.

A.1.3 Consistency of (some) sample analogues

Consider the following sample analogues based (conditional on θ and γ) estimators:

$$\hat{\Psi}(\theta,\gamma) \equiv n^{-1}\sum_{i=1}^{n}\Psi_i(\theta,\gamma),$$
$$\hat{\Omega}_{\psi\psi}(\theta,\gamma) \equiv n^{-1}\sum_{i=1}^{n}\psi_i(\theta,\gamma)\psi_i(\theta,\gamma)'.$$

Lemma A.3. *Suppose $\tilde{\theta}-\theta_{0n}=o_{p,n}(1)$ and $\tilde{\gamma}=o_{p,n}(1)$. Then, under Assumptions 1, 2, 4, 5, $\hat{\Psi}(\tilde{\theta},\tilde{\gamma}) \equiv \overline{\Psi}(\tilde{\theta},\tilde{\gamma}) = \Psi^* + o_{p,n}(1)$, $\hat{\Omega}_{\psi\psi}(\tilde{\theta},\tilde{\gamma}) \equiv n^{-1}\sum_{i=1}^{n}\psi_i(\tilde{\theta},\tilde{\gamma})\psi_i(\tilde{\theta},\tilde{\gamma})' = \Omega^*_{gg}$, and $\hat{\Xi}(\tilde{\theta},\tilde{\gamma}) = \Xi + o_{p,n}(1)$. In addition, $\Psi=\Psi^*+o_n(1)$ and $\Omega_{\psi\psi}=\Omega^*_{gg}+o_{p,n}(1)$, where $\Psi \equiv \mathbb{E}\left[\Psi_i(\theta_{0n},\gamma_{0n})\right]$ and $\Omega_{\psi\psi} \equiv \mathbb{E}\left[\psi_i(\theta_{0n},\gamma_{0n})\psi_i(\theta_{0n},\gamma_{0n})'\right]$.*

Proof of Lemma A.3. We start with showing that $\overline{\Psi}_\theta(\tilde{\theta},\tilde{\gamma}) = G^* + o_{p,n}(1)$. $\overline{\Psi}_\theta(\tilde{\theta},\tilde{\gamma}) = \overline{G}(\tilde{\theta}) - \sum_{k=2}^{K}\tilde{\gamma}_k\overline{G}_x^{(k)}(\tilde{\theta})$. Since $\tilde{\theta}-\theta_{0n}=o_{p,n}(1)$, $\tilde{\theta}\in B_\delta(\theta_{0n})$ with probability approaching one uniformly. Hence, with probability approaching one uniformly, by Assumptions, 2, 4 (i) and (iii),

$$\begin{aligned}\left\|\overline{G}_x^{(k)}(\tilde{\theta})\right\| \leqslant & \sum_{l=k}^{K}\frac{1}{(l-k)!}n^{-1}\sum_{i=1}^{n}\sup_{\theta\in B_\delta(\theta_{0n})}\left\|G_{xi}^{(l)*}(\theta)\right\||\varepsilon_i|^{l-k} \\ &+\frac{1}{(K-k)!}\left(n^{-1}\sum_{i=1}^{n}\sup_{\theta\in B_\delta(\theta_{0n})}b^{\theta*}_{G1i}(\theta)\,|\varepsilon_i|^{K+1-k}+n^{-1}\sum_{i=1}^{n}\sup_{\theta\in B_\delta(\theta_{0n})}b^{\theta*}_{G2i}(\theta)\,|\varepsilon_i|^{M-k}\right) \\ =& O_{p,n}(1) \end{aligned} \tag{A.3}$$

for $k\in\{2,\ldots,K\}$. This, along with $\tilde{\gamma}=o_{p,n}(1)$, implies that $\overline{\Psi}_\theta(\tilde{\theta},\tilde{\gamma})=\overline{G}(\tilde{\theta})+o_{p,n}(1)$. The next step is to show that $\overline{G}(\tilde{\theta})=G^*+o_{p,n}(1)$. To prove that, we inspect that the hypotheses of Lemma A.2 are satisfied for $\eta(\cdot)=G(\cdot)$ and directly apply the results. Condition (i) with $J=M$ is guaranteed by Assumption 4. Conditions (ii)-(v) are ensured by Assumptions 5 (i) and (iii). So, we conclude that $\overline{\Psi}_\theta(\tilde{\theta},\tilde{\gamma})=G^*+o_{p,n}(1)$. By a very similar argument, $\overline{\Psi}_{\gamma_k}(\tilde{\theta})=-\overline{g}_x^{(k)}(\tilde{\theta})=-g_x^{(k)*}+o_{p,n}(1)$, so $\overline{\Psi}(\tilde{\theta},\tilde{\gamma})=\Psi^*+o_{p,n}(1)$.

We also want to argue $\Psi\equiv\mathbb{E}\left[\Psi_i(\theta_{0n},\gamma_{0n})\right]$ exists and satisfies $\Psi=\Psi^*+o_n(1)$. Note that Lemma A.2 also guarantees that $G=G^*+o_n(1)$ and $g_x^{(k)}=g_x^{(k)*}+o_n(1)$ for $k\in\{0,\ldots,K\}$. Hence, we have already established $\Psi_\gamma=\Psi^*_\gamma+o_n(1)$, so it is sufficient to show that Ψ_θ exists and satisfies $\Psi_\theta=\Psi^*_\theta+o_n(1)$. Since $\Psi_{\theta i}=G_i-\sum_{k=2}^{K}\gamma_{0kn}G_{xi}^{(k)}$ and $\gamma_{0n}=o_n(1)$, it would be sufficient to show that $\mathbb{E}\left[G_{xi}^{(k)}\right]$ exists and is (uniformly) bounded. This can be straightforwardly expected utilizing the same expansion of $G_{xi}^{(k)}$ around $G_{xi}^{(k)*}$ as in (A.3).

Now we want to show that

$$\hat{\Omega}(\tilde{\theta},\tilde{\gamma})=n^{-1}\sum_{i=1}^{n}\left(g_i(\tilde{\theta})-\sum_{k=2}^{K}\tilde{\gamma}_k g_{xi}^{(k)}(\tilde{\theta})\right)\left(g_i(\tilde{\theta})-\sum_{k=2}^{K}\tilde{\gamma}_k g_{xi}^{(k)}(\tilde{\theta})\right)'=\Omega^*_{gg}+o_{p,n}(1).$$

The first step is to show that $n^{-1}\sum_{i=1}^{n}g_{xi}^{(k)}(\tilde{\theta})g_{xi}^{(k')}(\tilde{\theta})=O_{p,n}(1)$ for any $k,k'\in\{0,\ldots,K\}$. This is ensured by Assumptions 2, 4, 5 (iii), and uniform consistency of $\tilde{\theta}$ to θ_{0n} (to show that one simply needs to expand $g_{xi}^{(k)}(\tilde{\theta})$ and $g_{xi}^{(k')}(\tilde{\theta})$ around $g_{xi}^{(k)*}(\tilde{\theta})$ and $g_{xi}^{(k')*}(\tilde{\theta})$ respectively and bound the sum as in (A.3)). This, along with $\tilde{\gamma}=o_{p,n}(1)$, implies that $\hat{\Omega}(\tilde{\theta},\tilde{\gamma})=n^{-1}\sum_{i=1}^{n}g_i(\tilde{\theta})g_i(\tilde{\theta})'+o_{p,n}(1)$.

To argue that $n^{-1}\sum_{i=1}^n g_i(\tilde{\theta})g_i(\tilde{\theta})' = \Omega_{gg}^* + o_{p,n}(1)$, we invoke the result of Lemma A.2 with $\eta(\cdot) = g(\cdot)g(\cdot)'$. Note that Assumptions 4 imply that $\eta(\cdot)$ is Lipschitz continuous of order $2M$, so condition (i) of Lemma A.2 is satisfied. Conditions (ii)-(v) are ensured by Assumption 5 (i) and (iii). As a result, we conclude $\hat{\Omega}(\tilde{\theta}, \tilde{\gamma}) = \Omega_{gg}^* + o_{p,n}(1)$.

We also want to show $\Omega_{\psi\psi} \equiv \mathbb{E}\left[\psi_i(\theta_{0n}, \gamma_{0n})\psi_i(\theta_{0n}, \gamma_{0n})\right]$ exists and satisfies $\Omega_{\psi\psi} = \Omega_{gg}^* + o_n(1)$. Again, Lemma A.2 invoked with $\eta(\cdot) = g(\cdot)g(\cdot)$ guarantees that $\Omega_{gg} \equiv \mathbb{E}\left[g_i(\theta_{0n})g_i(\theta_{0n})'\right] = \Omega_{gg}^* + o_n(1)$. Since

$$\psi_i(\theta_{0n}, \gamma_{0n})\psi_i(\theta_{0n}, \gamma_{0n})' = \left(g_i(\theta_{0n}) - \sum_{k=2}^{K} \tilde{\gamma}_{0kn} g_{xi}^{(k)}(\theta_{0n})\right)\left(g_i(\theta_{0n}) - \sum_{k=2}^{K} \tilde{\gamma}_{0kn} g_{xi}^{(k)}(\theta_{0n})\right)'$$

$\gamma_{0n} = o_n(1)$, it is sufficient to show $\mathbb{E}\left[g_{xi}(\theta_{0n})^{(k)} g_{xi}(\theta_{0n})^{(k')}\right]$ exists and is (uniformly) bounded for all $k, k' \in \{0, \ldots, K\}$. Again, this can be established by utilizing the same argument as used above to verify that $n^{-1}\sum_{i=1}^n g_{xi}^{(k)}(\tilde{\theta})g_{xi}^{(k')}(\tilde{\theta}) = O_{p,n}(1)$. Thus, we have shown $\Omega_{\psi\psi} = \Omega_{gg}^* + o_n(1)$.

The next statement is $\hat{\Xi}(\tilde{\theta}, \tilde{\gamma}) = \Xi + o_{p,n}(1)$. By uniform consistency, $\tilde{\beta} \in B_\delta(\beta_0)$ with probability approaching one uniformly. Hence, with probability approaching one uniformly, by Assumption 5 (iv), $\|\hat{\Xi}(\tilde{\beta}) - \hat{\Xi}(\beta_0)\| \leqslant \sup_{\beta \in B_\delta}\|\nabla_\beta \hat{\Xi}(\beta)\| \|\tilde{\beta} - \beta_0\|$. As a result, since $\sup_{\beta \in B_\delta}\|\nabla_\beta \hat{\Xi}(\beta)\| = O_{p,n}(1)$ and $\|\tilde{\beta} - \beta_0\| = o_{p,n}(1)$, $\hat{\Xi}(\tilde{\beta}) = \hat{\Xi}(\beta_0) + o_{p,n}(1)$. At the same time, also by Assumption 5 (iv), $\hat{\Xi}(\beta_0) - \Xi(\beta_0) = \hat{\Xi}(\beta_0) - \Xi = o_{p,n}(1)$, which delivers the desired result.

Q.E.D.

A.1.4 CLT for the corrected moment

Lemma A.4. *Under Assumptions 1, 2, 4, 5, for every $\{\Upsilon_n\}$, $n^{1/2}\Omega_{gg}^{*-1/2}\overline{\psi}(\theta_{0n}, \gamma_{0n}) = n^{1/2}\Omega_{gg}^{*-1/2}\overline{g}(\theta_{0n}) + o_{p,n}(1) \xrightarrow{d} N(0, I_m)$ for some $\gamma_{0n} = o_n(1)$.*

Proof of Lemma A.4. Recall $\beta_{0n} = (\theta_{0n}', \gamma_{0n}')'$. We also put $\gamma_{0n} = (\gamma_{02n}, \ldots, \gamma_{0Kn})'$, and $\gamma_{0kn} = a_{kn}\sigma_n^k$ for some $a_n = (a_{2n}, \ldots, a_{Kn})' \in \mathbb{R}^{K-1}$. Making use of Assumption 4 (i),

$$\begin{aligned}
n^{1/2}\overline{\psi}_n(\beta_{0n}) = {} & n^{-1/2}\sum_{i=1}^n g(X_i^*, S_i, \theta_{0n}) + n^{-1/2}\sum_{i=1}^n g_x^{(1)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n) \\
& + \sum_{k=2}^K \frac{\sigma_n^k}{k!} n^{-1/2}\sum_{i=1}^n g_x^{(k)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n)^k \\
& - \sum_{k=2}^K \gamma_{0kn}\sum_{l=k}^K \frac{\sigma_n^{l-k}}{(l-k)!} n^{-1/2}\sum_{i=1}^n g_x^{(l)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n)^{l-k} \\
& + \frac{\sigma_n^K}{K!} n^{-1/2}\sum_{i=1}^n (g_x^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n}))(\varepsilon_{in}/\sigma_n)^K \\
& - \sum_{k=2}^K \gamma_{0kn}\frac{\sigma_n^{K-k}}{(K-k)!} n^{-1/2}\sum_{i=1}^n (g_x^{(K)}(\tilde{X}_{ki}, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n}))(\varepsilon_{in}/\sigma_n)^{K-k}
\end{aligned}$$

where, as usual, $\tilde{X}_i$ and $\tilde{X}_{ki}$ for $k \in \{2, \ldots, K\}$ lie between X_i^* and X_i. Plugging $\gamma_{0kn} = a_{kn}\sigma_n^k$ and rearranging the terms results in

$$
\begin{aligned}
n^{1/2}\overline{\psi}_n(\beta_{0n}) &= n^{-1/2}\sum_{i=1}^{n} g(X_i^*, S_i, \theta_{0n}) + n^{-1/2}\sum_{i=1}^{n} g_x^{(1)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n) \\
&+ \sum_{k=2}^{K} \sigma_n^k n^{-1/2} \sum_{i=1}^{n} g_x^{(k)}(X_i^*, S_i, \theta_{0n}) \left(\frac{1}{k!}(\varepsilon_{in}/\sigma_n)^k - \sum_{l=2}^{k} \frac{1}{(k-l)!}(\varepsilon_{in}/\sigma_n)^{k-l} a_{ln} \right) \\
&+ \frac{\sigma_n^K}{K!} n^{-1/2} \sum_{i=1}^{n} \left(g_x^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n}) \right) (\varepsilon_{in}/\sigma_n)^K \\
&- \sum_{k=2}^{K} a_{kn} \frac{\sigma_n^K}{(K-k)!} n^{-1/2} \sum_{i=1}^{n} \left(g_x^{(K)}(\tilde{X}_{ki}, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n}) \right) (\varepsilon_{in}/\sigma_n)^{K-k}.
\end{aligned}
$$

Note that, by Assumption 5 (i) and (ii), $\mathbb{E}\left[\left\|\Omega_{gg}^{*-1/2} g(X_i^*, S_i, \theta_{0n})\right\|^{2+\delta}\right] < C$. By that and Assumptions 1, for every $\{\Upsilon_n\}$,

$$
n^{-1/2}\Omega_{gg}^{*-1/2} \sum_{i=1}^{n} g(X_i^*, S_i, \theta_{0n}) \xrightarrow{d} N(0, I_m).
$$

To complete the proof, it is sufficient to show that all the remaining terms are $o_{p,n}(1)$. By Assumptions 1, 2, 5 (iii), Chebyshev's inequality guarantees

$$
\sigma_n n^{-1/2} \sum_{i=1}^{n} g_x^{(1)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n) = o_{p,n}(1).
$$

Next, Evdokimov and Zeleneev (2016) demonstrate that the system of equations

$$
\mathbb{E}\left[\frac{1}{k!}(\varepsilon_{in}/\sigma_n)^k - \sum_{l=2}^{k} \frac{1}{(k-l)!}(\varepsilon_{in}/\sigma_n)^{k-l} a_{ln} \right] = 0, \quad k \in \{2, \dots, K\} \tag{A.4}
$$

has a unique solution $\|a_n\| = O_n(1)$. Hence, for a properly chosen uniformly bounded a_n (and $\gamma_{0n} \to 0$), the system (A.4) is satisfied. Then, for such a choice of a_n (and γ_{0n}), Chebyshev's inequality coupled with Assumptions 1, 2, 5 (iii) guarantees that

$$
\sigma_n^k n^{-1/2} \sum_{i=1}^{n} g_x^{(k)}(X_i^*, S_i, \theta_{0n}) \left(\frac{1}{k!}(\varepsilon_{in}/\sigma_n)^k - \sum_{l=2}^{k} \frac{1}{(k-l)!}(\varepsilon_{in}/\sigma_n)^{k-l} a_{ln} \right) = o_{p,n}(1)
$$

for $k \in \{2, \dots, K\}$.

Next, by Assumption 4 (ii),

$$
\begin{aligned}
&\frac{\sigma_n^K}{K!} n^{-1/2} \left\| \sum_{i=1}^{n} \left(g_x^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n}) \right) (\varepsilon_{in}/\sigma_n)^K \right\| \\
\leqslant &\frac{1}{K!} \left(n^{1/2}\sigma_n^{K+1} n^{-1} \sum_{i=1}^{n} b_1(X_i^*, S_i, \theta_{0n}) |\varepsilon_{in}/\sigma_n|^{K+1} + n^{1/2}\sigma_n^M n^{-1} \sum_{i=1}^{n} b_2(X_i^*, S_i, \theta_{0n}) |\varepsilon_{in}/\sigma_n|^M \right).
\end{aligned}
$$

By Assumption 2, $n^{1/2}\sigma_n^{K+1} = o_n(1)$ and $n^{1/2}\sigma_n^M = o_n(1)$, and, by Assumptions 1, 2 and Assumption 5 (i),

$$n^{-1}\sum_{i=1}^n b_1(X_i^*, S_i, \theta_{0n})|\varepsilon_{in}/\sigma_n|^{K+1} = O_{p,n}(1),$$
$$n^{-1}\sum_{i=1}^n b_2(X_i^*, S_i, \theta_{0n})|\varepsilon_{in}/\sigma_n|^{M} = O_{p,n}(1)$$

Hence, as a result,

$$\frac{\sigma_n^K}{K!}n^{-1/2}\sum_{i=1}^n \left(g_x^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n})\right)(\varepsilon_{in}/\sigma_n)^K = o_{p,n}(1).$$

Applying the same reasoning and recalling $||a_n|| = O_n(1)$,

$$\sum_{k=2}^K a_{kn}\frac{\sigma_n^K}{(K-k)!}n^{-1/2}\sum_{i=1}^n \left(g_x^{(K)}(\tilde{X}_{ki}, S_i, \theta_{0n}) - g_x^{(K)}(X_i^*, S_i, \theta_{0n})\right)(\varepsilon_{in}/\sigma_n)^{K-k} = o_{p,n}(1),$$

and, recalling Assumption 5 (ii), we conclude

$$n^{1/2}\Omega_{gg}^{*-1/2}\overline{\psi}(\theta_{0n}, \gamma_{0n}) = n^{-1/2}\Omega^{-1/2}\sum_{i=1}^n g(X_i^*, S_i, \theta_{0n}) + o_{p,n}(1) \xrightarrow{d} N(0, I_m).$$

Finally, note that, since $\|a_n\| = O_n(1)$ and $\sigma_n = o_n(1)$, $\gamma_{0n} = o_n(1)$. Q.E.D.

A.1.5 Consistency under strong and semi-strong identification

Lemma A.5. *Suppose $\hat{Q}(\hat{\theta}, \hat{\gamma}) \leqslant \inf_{\theta\in\Theta,\gamma\in\Gamma}\hat{Q}(\theta, \gamma) + o_{p,n}(1)$. Then, under Assumptions 1-8, $\hat{\theta} - \theta_{0n} = o_{p,n}(1)$. Moreover, under additional Assumption 6, for any fixed $\delta > 0$, and for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta$, $\hat{\gamma} - \gamma_{0n} = o_{p,n}(1)$.*

Proof of Lemma A.5. First, by Lemma A.4, $\overline{\psi}(\theta_{0n}, \gamma_{0n}) = O_{p,n}(n^{-1/2})$. Hence, by Assumption 5 (iv), we have $\inf_{\theta\in\Theta,\gamma\in\Gamma}\hat{Q}(\theta, \gamma) \leqslant O_{p,n}(n^{-1})$. Hence, since $\hat{Q}(\hat{\theta}, \hat{\gamma}) \leqslant \inf_{\theta\in\Theta,\gamma\in\Gamma}\hat{Q}(\theta, \gamma) + o_{p,n}(1)$, using Assumption 5 (iv) again, we conclude

$$\overline{\psi}(\hat{\theta}, \hat{\gamma}) = o_{p,n}(1). \tag{A.5}$$

Also, Lemma A.1 guarantees $\left\|\overline{\psi}(\hat{\theta}, \hat{\gamma}) - \psi^*(\hat{\theta}, \hat{\gamma})\right\| = o_{p,n}(1)$. This and (A.5) together imply $\psi^*(\hat{\theta}, \hat{\gamma}) = o_{p,n}(1)$. Finally, using Assumption 8 (i), we conclude

$$\zeta_\theta\left(\left\|\hat{\theta} - \theta_{0n}\right\|\right) + \zeta_\gamma(\|\theta_{01n}\|\,\|\hat{\gamma} - \gamma_{0n}\|) = o_{p,n}(1).$$

Then, Assumption 8 (ii) ensures $\left\|\hat{\theta} - \theta_{0n}\right\| = o_{p,n}(1)$ (completes the proof of the first statement) and $\|\theta_{01n}\|\,\|\hat{\gamma} - \gamma_{0n}\| = o_{p,n}(1)$. Moreover, if $\|\theta_{01n}\|$ is bounded away from zero, we automatically obtain $\hat{\gamma} - \gamma_{0n} = o_{p,n}(1)$, which completes the proof of the second part. Q.E.D.

A.1.6 Proof of Theorem 2 Part (i): asymptotic normality under strong identification

Proof of Theorem 2 Part (i). Lemma A.5 guarantees $\hat{\beta} - \beta_{0n} = o_{p,n}(1)$, so, for any fixed $\delta > 0$, $\hat{\beta} \in B_\delta(\beta_{0n})$ with probability approaching one uniformly, and, so we can consider a local expansion of $\overline{\psi}(\hat{\beta})$ around β_{0n}. At the same time, by the hypothesis of the theorem, $\nabla_\beta \hat{Q}(\hat{\beta}) = \hat{D}(\hat{\beta})\hat{\Xi}(\hat{\beta})\overline{\psi}(\hat{\beta}) = o_{p,n}(n^{-1/2})$, where the j-th column of $\hat{D}(\hat{\beta})$ is given by $\hat{D}_j(\hat{\beta}) \equiv \overline{\Psi}_j(\hat{\beta}) + \frac{1}{2}\hat{\Xi}^{-1}(\hat{\beta})\nabla_{\beta_j}\hat{\Xi}(\hat{\beta})\overline{\psi}(\hat{\beta})$ with $\overline{\Psi}_j(\hat{\beta})$ denoting the j-th column of $\overline{\Psi}(\hat{\beta})$ for $j \in \{1, \dots, p+K-1\}$. Hence,

$$\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\left(\overline{\psi}(\beta_{0n}) + \overline{\Psi}(\tilde{\beta})(\hat{\beta} - \beta_{0n})\right) = o_{p,n}(n^{-1/2}), \tag{A.6}$$

where, as usual, $\tilde{\beta}$ lies in between of β_{0n} and $\hat{\beta}$. Since $\gamma_{0n} = o_n(1)$ (see Lemma A.4), $\tilde{\beta} = \beta_0 + o_{p,n}(1)$, so, by Lemma A.3, $\overline{\Psi}(\tilde{\beta}) = \Psi^* + o_{p,n}(1)$. At the same time, by Lemma A.4, $\overline{\psi}(\beta_{0n}) = O_{p,n}(n^{-1/2})$. Hence, since $\hat{\beta} - \beta_{0n} = o_{p,n}(1)$, we conclude that $\overline{\psi}(\hat{\beta}) = \overline{\psi}(\beta_{0n}) + \overline{\Psi}(\tilde{\beta})(\hat{\beta} - \beta_{0n}) = o_{p,n}(1)$. Also note that, by Assumption 5 (iv), $\hat{\Xi}(\hat{\beta})^{-1} = O_{p,n}(1)$ and $\nabla_{\beta_j}\hat{\Xi}(\hat{\beta}) = O_{p,n}(1)$. Therefore, $\hat{D}_j(\hat{\beta}) = \overline{\Psi}_j(\hat{\beta}) + o_{p,n}(1)$, and, invoking Lemma A.3, we conclude $\hat{D}(\hat{\beta}) = \Psi^* + o_{p,n}(n^{-1/2})$. Finally, also by Lemma A.3, $\hat{\Xi}(\hat{\beta}) = \Xi + o_{p,n}(1)$, and we conclude $\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\Psi}(\tilde{\beta}) = \Psi^{*\prime}\Xi\Psi^* + o_{p,n}(1)$. Moreover, by Assumption 6, $\left(\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\Psi}(\tilde{\beta})\right)^{-1}$ is invertible with probability approaching one, and, moreover, $\left(\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\Psi}(\tilde{\beta})\right)^{-1} = (\Psi^{*\prime}\Xi\Psi^*)^{-1} + o_{p,n}(1)$. Therefore, rearranging the terms in (A.6) gives

$$\begin{aligned}
n^{1/2}(\hat{\beta} - \beta_{0n}) &= -\left(\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\Psi}(\tilde{\beta})\right)^{-1}\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})n^{1/2}\overline{\psi}(\beta_{0n}) + o_{p,n}(1) \\
&= -(\Psi^{*\prime}\Xi\Psi^*)\Psi^{*\prime}\Xi n^{1/2}\overline{\psi}(\beta_{0n}) + o_{p,n}(1) \\
&= -(\Psi^{*\prime}\Xi\Psi^*)^{-1}\Psi^{*\prime}\Xi n^{1/2}\overline{g}(\theta_{0n}) + o_{p,n}(1),
\end{aligned}$$

where the last equality follows from Lemma A.4. Finally, we conclude that, for any $\{\Upsilon_n\}$, $n^{1/2}\Sigma^{*-1/2}(\hat{\beta} - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1})$, where

$$\Sigma^* = (\Psi^{*\prime}\Xi\Psi^*)^{-1}\Psi^{*\prime}\Xi\Omega^*_{gg}\Xi\Psi^*(\Psi^{*\prime}\Xi\Psi^*)^{-1}$$

has eigenvalues, which are uniformly bounded from below and above. Moreover, Lemma A.2 also guarantees that $\Psi = \Psi^* + o_n(1)$, $\Omega_{\psi\psi} = \Omega^*_{gg} + o_n(1)$, and, as a result, $\Sigma = \Sigma^* + o_n(1)$. This allows us to conclude that $n^{1/2}\Sigma^{-1/2}(\hat{\beta} - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1})$ too. Q.E.D.

A.2 Proof of Theorem 2 Part (ii): asymptotic normality under semi-strong identification

First, we provide high-level conditions (Assumption A.1), under which we show asymptotic normality of the estimator under semi-strong identification (Theorem A.1). Then we show Assumption A.1 is implied under the hypotheses of Theorem 2 Part (ii).

Notation. Let S be a $(p+K-1) \times (p+K-1)$ matrix defined as

$$S \equiv \begin{pmatrix} I_p & 0_{p\times(K-1)} \\ 0_{(K-1)\times p} & S_\gamma \end{pmatrix},$$

where S_γ is a $(K-1)\times(K-1)$ diagonal matrix with $\left(\left\|g_x^{(2)*}\right\|,\ldots,\left\|g_x^{(K)*}\right\|\right)'$ on its diagonal. Also let $S_n \equiv n^{1/2}S$.

A.2.1 Semi-strong ID asymptotic normality under high-level conditions

Assumption A.1 (Semi-strong and strong ID)**.** For $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ for any fixed $C>0$ and $0<\omega_{\mathrm{L}}<1/2$:

(i) $n^{1/2}\left\|g_x^{(k)*}\right\| \to \infty$ uniformly for $k \in \{2,\ldots,K\}$;

(ii) $\left(g_x^{(k)} - g_x^{(k)*}\right)/\left\|g_x^{(k)*}\right\| = o_n(1)$ for $k \in \{2,\ldots,K\}$;

(iii) $\liminf_{n\to\infty}\inf_{\Upsilon_n}\lambda_{\min}(B'B) > C > 0$ with $B \equiv \left(G^*, -g_x^{(2)*}/\left\|g_x^{(2)*}\right\|,\ldots,-g_x^{(K)*}/\left\|g_x^{(K)*}\right\|\right)$.

Theorem A.1 (Semi-strong and strong ID asymptotic normality)**.** *Suppose* $\gamma_0 = 0 \in \mathrm{int}\,(\Gamma)$. *Suppose also* $\hat{Q}(\hat\theta,\hat\gamma) \leqslant \inf_{\theta\in\Theta,\gamma\in\Gamma}\hat{Q}(\theta,\gamma) + o_{p,n}(n^{-1})$ *and* $\nabla_\beta\hat{Q}(\hat\beta) = o_{p,n}(n^{-1})$. *Finally, suppose* $\hat\theta - \theta_{0n} = o_{p,n}(1)$. *Then, under Assumptions 1-5, A.1, for all* $\{\Upsilon_n\}$ *satisfying* $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ *for any fixed* $C>0$ *and* $0<\omega_{\mathrm{L}}<1/2$,

$$\underline{\Sigma}^{-1/2}S_n(\hat\beta - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}), \tag{A.1}$$

and

$$\underline{\Sigma} = (B'\Xi B)^{-1}B'\Xi\Omega_{gg}^*\Xi B(B'\Xi B)^{-1}. \tag{A.2}$$

Moreover, it is also true that

$$\Sigma^{*-1/2}(\hat\beta - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}) \tag{A.3}$$

where

$$\Sigma^* = (\Psi^{*\prime}\Xi\Psi^*)^{-1}\Psi^{*\prime}\Xi\Omega_{gg}^*\Xi\Psi^*(\Psi^{*\prime}\Xi\Psi^*)^{-1},$$

and

$$\Sigma^{-1/2}(\hat\beta - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}). \tag{A.4}$$

Proof of Theorem A.1. **1.** The first step is to show $n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_k - \gamma_{0kn}) = O_{p,n}(1)$ for $k \in$

$\{2,\ldots,K\}$, so $\hat\gamma-\gamma_{0n}=o_{p,n}(1)$ (Assumption A.1 (i)). Note that

$$
\begin{aligned}
n^{1/2}\overline{\psi}(\hat\theta,\hat\gamma) =& n^{1/2}\overline{\psi}+\overline{\Psi}_\theta(\tilde\theta,\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n})-\sum_{k=2}^{K}(\hat\gamma_k-\gamma_{0kn})\left(n^{1/2}g_x^{(k)}+n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)\right)\\
=& n^{1/2}\overline{\psi}+\overline{\Psi}_\theta(\tilde\theta,\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n})-\sum_{k=2}^{K}(\hat\gamma_k-\gamma_{0kn})\left(n^{1/2}g_x^{(k)}+n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)\right),\\
=& n^{1/2}\overline{\psi}+\overline{\Psi}_\theta(\tilde\theta,\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n})-\sum_{k=2}^{K}\frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_k-\gamma_{0kn})\\
&-\sum_{k=2}^{K}n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_k-\gamma_{0kn})\left(\frac{g_x^{(k)}-g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}+\frac{n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)}{n^{1/2}\left\|g_x^{(k)*}\right\|}\right).
\end{aligned}
$$

Note that $\overline{\Psi}_\theta(\tilde\theta,\hat\gamma)=\overline{G}(\tilde\theta)-\sum\hat\gamma_k\overline{G}_x^{(k)}(\tilde\theta)$. Lemma A.5 establishes $\hat\theta-\theta_{0n}=o_{p,n}(1)$ and, consequently, $\tilde\theta-\theta_{0n}=o_{p,n}(1)$. Then, as shown in Lemma A.3, $\overline{G}(\tilde\theta)=G^*+o_{p,n}(1)$, so

$$
\begin{aligned}
n^{1/2}\overline{\psi}(\hat\theta,\hat\gamma) =& n^{1/2}\overline{\psi}+G^*n^{1/2}(\hat\theta-\theta_{0n})-\sum_{k=2}^{K}\frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_k-\gamma_{0kn})-\sum_{k=2}^{K}\hat\gamma_k\overline{G}_x^{(k)}(\tilde\theta)n^{1/2}(\hat\theta-\theta_{0n})\\
&-\sum_{k=2}^{K}n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_k-\gamma_{0kn})\left(\frac{g_x^{(k)}-g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}+\frac{n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)}{n^{1/2}\left\|g_x^{(k)*}\right\|}\right)+o_{p,n}(n^{1/2}(\hat\theta-\theta_0)).
\end{aligned}
$$

We want to show that $\hat b_n\equiv S_n(\hat\beta-\beta_{0n})=O_{p,n}(1)$.

$$
\begin{aligned}
n^{1/2}\overline{\psi}(\hat\theta,\hat\gamma) =& n^{1/2}\overline{\psi}+B\hat b_n-\sum_{k=2}^{K}\hat\gamma_k\overline{G}_x^{(k)}(\tilde\theta)n^{1/2}(\hat\theta-\theta_{0n})\\
&-\sum_{k=2}^{K}n^{1/2}\left\|g_x^{(k)*}\right\|(\hat\gamma_{kn}-\gamma_{0kn})\left(\frac{g_x^{(k)}-g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}+\frac{n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)}{n^{1/2}\left\|g_x^{(k)*}\right\|}\right)+o_{p,n}(\hat b_n).
\end{aligned}
\tag{A.5}
$$

By Assumption 5 (iii), $\mathbb{E}\left[\left\|g_{xi}^{(k)}(\theta_{0n})\right\|^2\right]<C$, so $n^{1/2}\left(\overline{g}_x^{(k)}-g_x^{(k)}\right)=O_{p,n}(1)$, and hence by Assumptions A.1 (i) and (ii),

$$
\frac{g_x^{(k)}-g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|}+\frac{O_{p,n}(1)}{n^{1/2}\left\|g_x^{(k)*}\right\|}=o_{p,n}(1).
$$

Moreover, since $\gamma_{0k}=o_n(1)$ and $\|\hat\gamma_k-\gamma_{0k}\|=o_{p,n}\left(\left\|\hat b_n\right\|\right)$, and, as shown in Lemma A.3, $\overline{G}_x^{(k)}(\tilde\theta)=$

$O_{p,n}(1)$,

$$\sum_{k=2}^{K} \hat{\gamma}_k \overline{G}_x^{(k)}(\tilde{\theta}) n^{1/2}(\hat{\theta} - \theta_{0n}) = \sum_{k=2}^{K} (\hat{\gamma}_k - \gamma_{0k}) \overline{G}_x^{(k)}(\tilde{\theta}) n^{1/2}(\hat{\theta} - \theta_{0n}) + o_{p,n}\left(\|\hat{b}_n\|\right)$$
$$= o_{p,n}\left(\|\hat{b}_n\| + \|\hat{b}_n\|^2\right).$$

Finally, recall $n^{1/2}\overline{\psi} = O_{p,n}(1)$, and, hence, by the triangle inequality,

$$\|n^{1/2}\overline{\psi}(\hat{\theta}, \hat{\gamma})\| \geqslant \|B\hat{b}_n\| - o_{p,n}\left(\|\hat{b}_n\| + \|\hat{b}_n\|^2\right) - O_{p,n}(1)$$
$$\geqslant C_B \|\hat{b}_n\| - o_{p,n}\left(\|\hat{b}_n\| + \|\hat{b}_n\|^2\right) - O_{p,n}(1),$$

where in the second inequality $C_B > 0$ by Assumption A.1 (iii).

At the same time, since $\hat{Q}(\hat{\theta}, \hat{\gamma}) < \inf_{\theta \in \Theta, \gamma \in \Gamma} \hat{Q}(\theta, \gamma) + o_{p,n}(n^{-1})$, we have

$$C_B \|\hat{b}_n\| - o_{p,n}\left(\|\hat{b}_n\| + \|\hat{b}_n\|^2\right) - O_{p,n}(1) \leqslant \|n^{1/2}\overline{\psi}(\hat{\theta}, \hat{\gamma})\| \leqslant O_{p,n}(1),$$

which implies $\hat{b}_n = O_{p,n}(1)$. Hence, $\hat{\theta} - \theta_0 = O_{p,n}(n^{-1/2})$, and $n^{1/2}\left\|g_x^{(k)*}\right\|(\hat{\gamma}_k - \gamma_{0kn}) = O_{p,n}(1)$ for $k \in \{2, \ldots, K\}$, so $\hat{\gamma} - \gamma_{0n} = o_{p,n}(1)$.

2. As in the proof of Theorem 2 Part (i), the FOC takes the form of

$$\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\psi}(\hat{\beta}) = o_{p,n}(n^{-1}).$$

Multiplying both sides by S^{-1} and noting that $n^{-1/2}S^{-1} = o_n(1)$ (by Assumption A.1 (i)) gives

$$S^{-1}\hat{D}(\hat{\beta})'\hat{\Xi}(\hat{\beta})\overline{\psi}(\hat{\beta}) = o_{p,n}(n^{-1/2}).$$

Now we argue that

$$\hat{D}(\hat{\beta})S^{-1} = B + o_{p,n}(1). \tag{A.6}$$

Recall that the j-th column of $\hat{D}(\hat{\beta})$ is given by $\hat{D}_j(\hat{\beta}) \equiv \overline{\Psi}_j(\hat{\beta}) + \frac{1}{2}\hat{\Xi}^{-1}(\hat{\beta})\nabla_{\beta_j}\hat{\Xi}(\hat{\beta})\overline{\psi}(\hat{\beta})$ and note that $\overline{\psi}(\hat{\beta})S^{-1} = n^{1/2}\overline{\psi}(\hat{\beta})n^{-1/2}S^{-1} = o_{p,n}(1)$ since $n^{1/2}\overline{\psi}(\hat{\beta}) = O_{p,n}(1)$. Then, $\hat{D}(\hat{\beta})S^{-1} = \overline{\Psi}(\hat{\beta})S^{-1} + o_{p,n}(1)$, where

$$\overline{\Psi}(\hat{\beta}) = \left[\overline{\Psi}_\theta(\hat{\beta}), -\overline{g}_x^{(2)}(\hat{\theta}), \ldots, -\overline{g}_x^{(K)}(\hat{\theta})\right].$$

Note that for the first p columns of $\hat{D}(\hat{\beta})S^{-1}$ we have $\left(\overline{\Psi}(\hat{\beta})S^{-1}\right)_{1:p} = G^* + o_{p,n}(1) = B_{1:p} + o_{p,n}(1)$, since $\hat{\beta} - \beta_{0n} = o_{p,n}(1)$. Then to establish that $\overline{\Psi}(\hat{\beta})S^{-1} = B + o_{p,n}(1)$, we need to verify that $\overline{g}_x^{(k)}(\hat{\theta})/\left\|g_x^{(k)*}\right\| = g_x^{(k)*}/\left\|g_x^{(k)*}\right\| + o_{p,n}(1)$ for $k \in \{2, \ldots, K\}$. First, $\overline{g}_x^{(k)}(\hat{\theta}) = g_x^{(k)} + O_{p,n}(n^{-1/2})$. Indeed,

$$n^{1/2}(\overline{g}_x^{(k)}(\hat{\theta}) - g_x^{(k)}) = n^{1/2}(\overline{g}_x^{(k)} - g_x^{(k)}) + \overline{G}_x^{(k)}(\tilde{\theta})n^{1/2}(\hat{\theta} - \theta_{0n}) = O_{p,n}(1),$$

since $n^{1/2}(\hat{\theta}-\theta_{0n}) = O_{p,n}(1)$ and $\overline{G}_x^{(k)}(\tilde{\theta}) = O_{p,n}(1)$. Then, using Assumptions A.1 (i) and (ii),

$$\begin{aligned}
\frac{\overline{g}_x^{(k)}(\hat{\theta})}{\left\|g_x^{(k)*}\right\|} &= \frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} + \frac{g_x^{(k)} - g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} + \frac{O_{p,n}(n^{-1/2})}{\left\|g_x^{(k)*}\right\|} \\
&= \frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} + o_{p,n}(1).
\end{aligned}$$

Therefore, $\overline{\Psi}(\hat{\beta})S^{-1} = B + o_{p,n}(1)$, and equation (A.6) holds.

In the analysis of remainders in equation (A.5) we have established that $n^{1/2}\overline{\psi}(\hat{\beta}) = n^{1/2}\overline{\psi} + BS_n(\hat{\beta}-\beta_{0n}) + o_{p,n}(1)$, hence, the FOC simplifies as

$$B'\Xi(n^{1/2}\overline{\psi} + BS_n(\hat{\beta}-\beta_{0n})) = o_{p,n}(1),$$

By Assumption A.1 (iii), $B'\Xi B$ is invertible and its inverse is uniformly bounded, so

$$S_n(\hat{\beta}-\beta_{0n}) = -(B'\Xi B)^{-1}B'\Xi n^{1/2}\overline{\psi} + o_{p,n}(1).$$

Finally, Lemma A.4 guarantees that $n^{1/2}\Omega^{-1/2}\overline{\psi} \xrightarrow{d} N(0, I_m)$, so

$$\underline{\Sigma}^{-1/2}S_n(\hat{\beta}-\beta_{0n}) \xrightarrow{d} N(0, I_p).$$

Then, we also want to show that

$$n^{1/2}\Sigma^{*-1/2}(\hat{\beta}-\beta_{0n}) \xrightarrow{d} N(0, I_p). \tag{A.7}$$

Indeed,

$$\begin{aligned}
n^{1/2}\Sigma^{*-1/2}(\hat{\beta}-\beta_{0n}) &= \Sigma^{*-1/2}S^{-1}\underline{\Sigma}^{1/2}\underline{\Sigma}^{-1/2}S_n(\hat{\beta}-\beta_{0n}) \\
&= O\underline{\Sigma}^{-1/2}S_n(\hat{\beta}-\beta_{0n}),
\end{aligned}$$

where $O \equiv \Sigma^{*-1/2}S^{-1}\underline{\Sigma}^{1/2}$ is an orthogonal matrix: recall $S\Sigma^*S = \underline{\Sigma}$, then

$$\begin{aligned}
OO' &= \Sigma^{*-1/2}S^{-1}\underline{\Sigma}S^{-1}\Sigma^{*-1/2} \\
&= \Sigma^{*-1/2}\Sigma^*\Sigma^{*-1/2} \\
&= I_p,
\end{aligned}$$

and, hence, the desired result holds.

Finally, we argue that

$$n^{1/2}\Sigma^{-1/2}(\hat{\beta}-\beta_{0n}) \xrightarrow{d} N(0, I_p). \tag{A.8}$$

First, note that $S\Sigma S = \underline{\Sigma} + o_n(1) = S\Sigma^*S + o_n(1)$ (see Lemma A.7 below). Let $Q \equiv \Sigma^{-1/2}\Sigma^{*1/2}$, and note that

$$n^{1/2}\Sigma^{-1/2}\left(\beta-\beta_0\right) = Qn^{1/2}\Sigma^{*-1/2}\left(\beta-\beta_0\right).$$

We argue that $Q'Q = I + o_n(1)$, and $Q' = Q^{-1} + o_n(1)$. Indeed

$$Q'Q = \Sigma^{*1/2}\Sigma^{-1}\Sigma^{*1/2} = \Sigma^{*1/2} S (S\Sigma S)^{-1} S\Sigma^{*1/2}.$$

Let $\Lambda \equiv \Sigma^{*1/2} S$. Note that Λ has full rank, and $\Lambda'\Lambda = S\Sigma^* S$ has eigenvalues bounded away from zero and infinity. Thus,

$$\begin{aligned} Q'Q &= \Lambda\left(\Lambda'\Lambda + o_{p,n}(1)\right)^{-1}\Lambda' = \Lambda\left\{(\Lambda'\Lambda)^{-1} + o_{p,n}(1)\right\}\Lambda' \\ &= \Lambda(\Lambda'\Lambda)^{-1}\Lambda' + o_{p,n}(1) = I + o_{p,n}(1). \end{aligned}$$

Also, $Q' = Q^{-1} + o_{p,n}(1)$, since Q has singular values bounded away from zero and infinity. This, along with (A.7), implies that (A.8) holds.

Q.E.D.

A.2.2 Verification of Assumption A.1

Second, we show that high-level Assumption A.1 is implied by the following assumption.

Lemma A.6. *Under Assumptions 1-5, Assumptions 6 and 7 imply Assumption A.1.*

Proof of Lemma A.6. Verification of Assumption A.1 (i)*:* Assumption 7 (i) implies that, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \downarrow 0$,

$$g_x^{(k)*} = \left(A_k^* \frac{\theta_{01n}}{\|\theta_{01n}\|} + o_n(1)\right)\|\theta_{01n}\|,$$

and Assumption 7 (ii) guarantees that there exists $\tilde{C} > 0$ such that

$$\left\| A_k^* \frac{\theta_{01n}}{\|\theta_{01n}\|} \right\| > \tilde{C}$$

for $k \in \{2, \ldots, K\}$ and for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_0$ for any fixed $C > 0$ and $0 < \omega_{\mathrm{L}} < 1/2$. Therefore, there exist $\delta_1 \in (0, \delta_0)$ and $C_1 > 0$, such that

$$\left\|g_x^{(k)*}\right\| \geqslant C_1 \|\theta_{01n}\|$$

for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_1$. Hence, for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_1$, we establish $n^{1/2}\left\|g_x^{(k)*}\right\| \to \infty$. Also note that Assumption 6 guarantees that there exists $\tilde{C} > 0$ such that for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta_1$, $\left\|g_x^{(k)*}\right\| > \tilde{C}$. As a result, $n^{1/2}\left\|g_x^{(k)*}\right\| \to \infty$ for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta_1$. Hence, Assumption A.1 (i) is satisfied.

Verification of Assumption A.1 (ii)*:* As established before, there exist $\delta_1 > 0$ and $C_1 > 0$, such that $\left\|g_x^{(k)*}\right\| \geqslant C_1 \|\theta_{01n}\|$ for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_1$. Therefore, using 7 (iii), we conclude

$$\frac{g_x^{(k)} - g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} = o_n(1)$$

for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_1$. Also recall Assumption 6 guarantees that there exists $\tilde{C} > 0$ such that for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta_1$, $\left\|g_x^{(k)*}\right\| > \tilde{C}$. Then, since $\|\theta_{01n}\|$ is bounded, 7 (iii) also implies

$$\frac{g_x^{(k)} - g_x^{(k)*}}{\left\| g_x^{(k)*} \right\|} = o_n(1)$$

for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta_1$.

Verification of Assumption A.1 (iii)*:* First note that, since, for $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_1$, $\left\| g_x^{(k)*} \right\| \geqslant C_1 \|\theta_{01n}\|$ with $C_1 \geqslant 0$, Assumption 7 (i) establishes

$$\frac{g_x^{(k)*}}{\left\| g_x^{(k)*} \right\|} = \frac{A_k^* \theta_{01n}}{\left\| g_x^{(k)*} \right\|} + o_n(1)$$

for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \downarrow 0$. Hence, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \downarrow 0$,

$$B = B_A + o_n(1) \tag{A.9}$$

where

$$B_A \equiv \left(G^*, -\frac{A_2^* \theta_{01n}}{\left\| g_x^{(2)*} \right\|}, \ldots, -\frac{A_K^* \theta_{01n}}{\left\| g_x^{(K)*} \right\|} \right).$$

Note that Assumption 7 (ii) guarantees that there exists $\tilde{C} > 0$

$$\lambda_{\min} \left(\tilde{B}_A' \tilde{B}_A \right) > \tilde{C}, \tag{A.10}$$

where

$$\tilde{B}_A \equiv \left(G^*, -A_2^* \theta_{01n} / \|\theta_{01n}\|, \ldots, -A_K^* \theta_{01n} / \|\theta_{01n}\| \right).$$

for all $\{\Upsilon_n\}$ with $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_0$. Also note that Assumption 7 (i) guarantees that there exist $\delta_2 \in (0, \delta_1)$ and $C_2 > 0$ such that

$$\left\| g_x^{(k)*} \right\| / \|\theta_{01n}\| \leqslant C_2 \quad \text{for} \quad k \in \{2, \ldots, K\}$$

for $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_2$. Hence, for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_2$,

$$0 < C_1 \leqslant \left\| g_x^{(k)*} \right\| / \|\theta_{01n}\| \leqslant C_2 \quad \text{for} \quad k \in \{2, \ldots, K\}. \tag{A.11}$$

Note that

$$B_A = \left(G^*, -\frac{A_2^* \theta_{01n} / \|\theta_{01n}\|}{\left\| g_x^{(2)*} \right\| / \|\theta_{01n}\|}, \ldots, -\frac{A_K^* \theta_{01n} / \|\theta_{01n}\|}{\left\| g_x^{(K)*} \right\| / \|\theta_{01n}\|} \right),$$

and, consequently, by combining (A.10) and (A.11), we can establish that there exists $C_{B_A} > 0$ such that

$$\lambda_{\min}(B_A' B_A) > C_{B_A}$$

for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_2$. Finally, this, along with (A.9), guarantees that there exists $\delta_3 \in (0, \delta_2)$, such that

$$\lambda_{\min}(B'B) > C_{B_A}$$

for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant \delta_3$. Finally, Assumption 6 ensures that $\exists C_{\delta_3} > 0$ such that

$$\lambda_{\min}(B'B) > C_{\delta_3}$$

for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta_3$. Therefore

$$\lambda_{\min}(B'B) > \min\{C_{B_A}, C_{\delta_3}\}$$

for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\|$, which completes the proof. Q.E.D.

A.2.3 Proof of Theorem 2 Part (ii)

Proof of Theorem 2 Part (ii). Follows from Theorem A.1 and Lemma A.6. Q.E.D.

A.2.4 On asymptotic variance under semi-strong identification

Lemma A.7. *Suppose that the hypotheses of Theorem A.1 are satisfied. Then $S\hat{\Sigma}S = \underline{\Sigma} + o_{p,n}(1)$ and $S\Sigma S = \underline{\Sigma} + o_n(1)$.*

Proof of Lemma A.7. Finally, we want to argue that $S\hat{\Sigma}S = \underline{\Sigma} + o_{p,n}(1)$, where

$$\hat{\Sigma} = (\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}_{\psi\psi}\hat{\Xi}\hat{\Psi}(\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1},$$

and $\hat{\Psi} \equiv \hat{\Psi}(\hat{\beta}) \equiv \overline{\Psi}(\hat{\beta})$. Then

$$S\hat{\Sigma}S = (S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Psi}S^{-1})^{-1}S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}\hat{\Xi}\hat{\Psi}S^{-1}(S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Psi}S^{-1})^{-1} \tag{A.12}$$

Note that we have already shown in the proof (of Theorem A.1) above that $\hat{\Psi}S^{-1} = \overline{\Psi}(\hat{\beta})S^{-1} = B + o_{p,n}(1)$, so $S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Psi}S^{-1} = B'\Xi B + o_{p,n}(1)$ with $\lambda_{\min}(B'\Xi B)$ bounded away from zero (Assumption A.1 (iii)). Combined with $\hat{\Omega}_{gg} = \Omega_{gg} + o_n(1)$ (Lemma (A.3)), this delivers the result for $\hat{\Sigma}$.

A similar reasoning can be used to show $S\Sigma S = \underline{\Sigma} + o_n(1)$. Indeed, since we have

$$\begin{aligned} \frac{g_x^{(k)}}{\left\|g_x^{(k)*}\right\|} &= \frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} + \frac{g_x^{(k)} - g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} \\ &= \frac{g_x^{(k)*}}{\left\|g_x^{(k)*}\right\|} + o_n(1) \end{aligned}$$

and $\Psi_\theta = \Psi_\theta^* + o_n(1)$ (see Lemma A.3), we again obtain $\Psi S^{-1} = B + o_n(1)$. Combining this with $\Omega_{\psi\psi} = \Omega_{gg}^* + o_n(1)$ (Lemma A.3) and the same representation as in (A.12) delivers the result for Σ. Q.E.D.

A.2.5 Proof of Lemma 1

Proof of Lemma 1. Let

$$\mu = \frac{\underline{\Sigma}^{1/2} S^{-1}\lambda}{\left\|\underline{\Sigma}^{1/2} S^{-1}\lambda\right\|},$$

so $\|\mu\| = 1$.

$$\begin{aligned}
\frac{n^{1/2}\lambda'(\hat\beta-\beta_{0n})}{\sqrt{\lambda'\hat\Sigma\lambda}} &= \frac{\lambda' S^{-1}\underline{\Sigma}^{1/2}\underline{\Sigma}^{-1/2} S_n(\hat\beta-\beta_{0n})}{\sqrt{\lambda' S^{-1}\underline{\Sigma}^{1/2}\left(\underline{\Sigma}^{-1/2} S\hat\Sigma S\underline{\Sigma}^{-1/2}\right)\underline{\Sigma}^{1/2} S^{-1}\lambda}}\\
&= \frac{\mu'\underline{\Sigma}^{-1/2} S_n(\hat\beta-\beta_{0n})}{\sqrt{\mu'(I_{p+K-1}+o_{p,n}(1))\mu}}\\
&= \mu'\underline{\Sigma}^{-1/2} S_n(\hat\beta-\beta_{0n}) + o_{p,n}(1),
\end{aligned}$$

where the second equality holds since $S\hat\Sigma S = \underline{\Sigma} + o_{p,n}(1)$ (Lemma A.7) and the eigenvalues of $\underline{\Sigma}$ are bounded away from zero, and the third equality holds since $\underline{\Sigma}^{-1/2} S_n(\hat\beta-\beta_{0n}) = O_{p,n}(1)$. Finally, since $\|\mu\| = 1$ and $\underline{\Sigma}^{-1/2} S_n(\hat\beta-\beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1})$, the desired result follows. Q.E.D.

A.3 Proof of Theorem 3

A.3.1 Proof of Lemma A.8

Lemma A.8 (Uniform $\sqrt{n}$-consistency)**.** *Suppose $\hat Q(\hat\theta,\hat\gamma) \leqslant \inf_{\theta\in\Theta,\gamma\in\Gamma}\hat Q(\theta,\gamma) + O_{p,n}(n^{-1})$. Then, under Assumptions 1-8, 9, there exits $\tilde\omega_0 \in (0,1/2)$ such that for all $\{\Upsilon_n\}$, satisfying $\|\theta_{01n}\| \leqslant Cn^{-\tilde\omega_0}$ for any fixed $C \geqslant 0$, $\hat\theta - \theta_{0n} = O_{p,n}(n^{-1/2})$.*

Proof of Lemma A.8. Since $\hat\theta - \theta_{0n} = o_{p,n}(1)$ (Lemma A.5), we can linearize the sample moment function again:

$$\begin{aligned}
n^{1/2}\overline{\psi}(\hat\theta,\hat\gamma) =& n^{1/2}\overline{\psi} + \overline{\Psi}_\theta(\tilde\theta,\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n}) - \sum_{k=2}^{K}(\hat\gamma_k-\gamma_{0kn})\left(n^{1/2}g_x^{(k)} + n^{1/2}\left(\overline{g}_x^{(k)} - g_x^{(k)}\right)\right)\\
=& n^{1/2}\overline{\psi} + \Psi^*_\theta(\theta_{0n},\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n}) - \sum_{k=2}^{K}(\hat\gamma_k-\gamma_{0kn})\left(n^{1/2}g_x^{(k)} + n^{1/2}\left(\overline{g}_x^{(k)} - g_x^{(k)}\right)\right) + o_{p,n}(n^{1/2}(\hat\theta-\theta_{0n}))
\end{aligned}$$

where Assumption 9 (ii) helps to establish $\sup_{\gamma\in\Gamma}\left\|\overline{\Psi}_\theta(\tilde\theta,\gamma) - \Psi^*_\theta(\theta_{0n},\gamma)\right\| = o_{p,n}(1)$.

Combining Assumption 9 (i) and 9 (iv), we obtain $g_x^{(k)} = A_k^*\theta_{01n} + O_n(\sigma_n\|\theta_{01n}\| + \|\theta_{01n}\|^{1+\eta})$. Recall $\sigma_n = o_n(n^{-\frac{1}{2(K+1)}})$. Then if we choose any $\tilde\omega_0 \in \left(\max\{\frac{1}{2(1+\eta)}, \frac{1}{2} - \frac{1}{2(K+1)}\}, \frac{1}{2}\right)$, $n^{1/2}O_n(\sigma_n\|\theta_{01n}\| + \|\theta_{01n}\|^2) = o_n(1)$ for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{-\tilde\omega_0}$. Hence, for these DGPs, $n^{1/2}g_x^{(k)} = n^{1/2}A_k^*\theta_{01n} + o_n(1)$ for $k \in 2,\ldots,K$. At the same time, $n^{1/2}\overline{\psi} = O_{p,n}(1)$ (Lemma A.4) and, as argued in the proof Theorem A.1, $n^{1/2}\left(\overline{g}_x^{(k)} - g_x^{(k)}\right) = O_{p,n}(1)$. Hence, since Γ is bounded,

$$n^{1/2}\overline{\psi}(\hat\theta,\hat\gamma) = \Psi^*_\theta(\theta_{0n},\hat\gamma)n^{1/2}(\hat\theta-\theta_{0n}) - \sum_{k=2}^{K}(\hat\gamma_k-\gamma_{0kn})n^{1/2}A_k^*\theta_{01n} + o_{p,n}(n^{1/2}(\hat\theta-\theta_{0n})) + O_{p,n}(1).$$

Denote $M_A^* = I_m - \Lambda_A^*(\Lambda_A^{*\prime}\Lambda_A^*)^{-1}\Lambda_A^{*\prime}$, where $\Lambda_A^* = (A_2^*, \ldots, A_K^*)$. Then

$$n^{1/2} M_A^* \overline{\psi}(\hat{\theta}, \hat{\gamma}) = M_A^* \Psi_\theta^*(\theta_{0n}, \hat{\gamma}) n^{1/2}(\hat{\theta} - \theta_{0n}) + o_{p,n}(n^{1/2}(\hat{\theta} - \theta_{0n})) + O_{p,n}(1).$$

Note that, by Assumption 9 (iii), there exists $C > 0$ such that

$$\left\| M_A^* \Psi_\theta^*(\theta_{0n}, \hat{\gamma}) n^{1/2}(\hat{\theta} - \theta_{0n}) \right\| \geqslant C n^{1/2} \left\| \hat{\theta} - \theta_0 \right\|,$$

so

$$n^{1/2} \left\| \overline{\psi}(\hat{\theta}, \hat{\gamma}) \right\| \geqslant n^{1/2} \left\| M_A^* \overline{\psi}(\hat{\theta}, \hat{\gamma}) \right\| \geqslant C n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| + o_{p,n}(n^{1/2}(\hat{\theta} - \theta_{0n})) + O_{p,n}(1).$$

At the same time, since $n^{1/2} \left\| \overline{\psi}(\theta_{0n}, \gamma_{0n}) \right\| = O_{p,n}(1)$, by the hypothesis of the theorem and Assumption 5 (iv),

$$n^{1/2} \left\| \overline{\psi}(\hat{\theta}, \hat{\gamma}) \right\| \leqslant O_{p,n}(1).$$

Thus,

$$C n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| + o_{p,n}(n^{1/2}(\hat{\theta} - \theta_{0n})) + O_{p,n}(1) \leqslant O_{p,n}(1),$$

which implies $\hat{\theta} - \theta_{0n} = O_{p,n}(n^{-1/2})$. Q.E.D.

A.3.2 Proof of Theorem 3

Proof of Theorem 3. We want to show $\hat{\theta} - \theta_{0n} = O_{p,n}(n^{-1/2})$, i.e. that for any $\epsilon > 0$, there exists $C_\epsilon > 0$ such that

$$\limsup_{n\to\infty} \sup_{\Upsilon_n} \mathbb{P}_{\Upsilon_n} \left(n^{1/2} \left\| \hat{\theta} - \theta_{0n} \right\| > C_\epsilon \right) < \epsilon. \tag{A.1}$$

Take $\tilde{\omega}_0$ as in Lemma A.8 and fixed $C > 0$. Then, by Lemma A.8, for any $\epsilon > 0$, there exists C_ϵ^0 such that

$$\limsup_{n\to\infty} \sup_{\Upsilon_n : \|\theta_{01n}\| \leqslant C n^{-\tilde{\omega}_0}} \mathbb{P}_{\Upsilon_n} \left(n^{1/2} \left\| \hat{\theta} - \theta_{0n} \right\| > C_\epsilon^0 \right) < \epsilon.$$

At the same time, since $\tilde{\omega}_0 < 1/2$, by Theorem 2, for any $\epsilon > 0$, there exists C_ϵ^s such that

$$\limsup_{n\to\infty} \sup_{\Upsilon_n : \|\theta_{01n}\| \geqslant C n^{-\tilde{\omega}_0}} \mathbb{P}_{\Upsilon_n} \left(n^{1/2} \left\| \hat{\theta} - \theta_{0n} \right\| > C_\epsilon^s \right) < \epsilon.$$

Hence, (A.1) holds with $C_\epsilon = \max\{C_\epsilon^0, C_\epsilon^s\}$. Q.E.D.

A.4 Proof of Theorem 4

Proof of Theorem 4. The first step is to show that $n^{1/2}\overline{g}_{\mathrm{U}}(\theta_{0n}) = n^{1/2}\overline{g}_{\mathrm{U}}^*(\theta_{0n}) + n^{1/2} b_n + o_{p,n}(1)$ for the moment bias b_n defined later. By the standard expansion

$$
\begin{aligned}
n^{1/2}\overline{g}_{\mathrm{U}}(\theta_{0n}) =& n^{-1/2}\sum_{i=1}^{n} g_{\mathrm{U}}(X_i^*, S_i, \theta_{0n}) + n^{-1/2}\sum_{i=1}^{n} g_{\mathrm{U}x}^{(1)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n) \\
&+ \sum_{k=2}^{K} \frac{\sigma_n^k}{k!} n^{-1/2} \sum_{i=1}^{n} g_{\mathrm{U}x}^{(k)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n)^k \\
&+ \frac{\sigma_n^K}{K!} n^{-1/2} \sum_{i=1}^{n} (g_{\mathrm{U}x}^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_{\mathrm{U}x}^{(K)}(X_i^*, S_i, \theta_{0n}))(\varepsilon_{in}/\sigma_n)^K.
\end{aligned}
$$

As established in the proof of Lemma A.4,

$$
\begin{aligned}
n^{-1/2}\sum_{i=1}^{n} g_{\mathrm{U}x}^{(1)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n) &= o_{p,n}(1), \\
\frac{\sigma_n^K}{K!} n^{-1/2} \sum_{i=1}^{n} (g_{\mathrm{U}x}^{(K)}(\tilde{X}_i, S_i, \theta_{0n}) - g_{\mathrm{U}x}^{(K)}(X_i^*, S_i, \theta_{0n}))(\varepsilon_{in}/\sigma_n)^K &= o_{p,n}(1).
\end{aligned}
$$

Let us denote $\xi_{kin} \equiv g_{\mathrm{U}x}^{(k)}(X_i^*, S_i, \theta_{0n})(\varepsilon_{in}/\sigma_n)^k$ for $k \in \{2, \ldots, K\}$. Note that, using the hypothesis of the theorem,

$$
\xi_{kn} \equiv \mathbb{E}\left[\xi_{kin}\right] = A_{\mathrm{U}k}^* \theta_{01n} \mathbb{E}\left[(\varepsilon_{in}/\sigma_n)^k\right] + O_n(\|\theta_{01n}\|^2).
$$

So, we have

$$
n^{1/2}\overline{g}_{\mathrm{U}}(\theta_{0n}) = n^{1/2}\overline{g}_n^*(\theta_{0n}) + n^{1/2} b_n + \sum_{k=2}^{K} \frac{\sigma_n^k}{k!} n^{-1/2} \sum_{i=1}^{n} (\xi_{kin} - \xi_{kn}) + o_p(1).
$$

where

$$
b_n \equiv \sum_{k=2}^{K} \frac{\sigma_n^k}{k!} A_{\mathrm{U}k}^* \theta_{01n} \mathbb{E}\left[(\varepsilon_{in}/\sigma_n)^k\right] + O_n(\sigma_n^2 \|\theta_{01n}\|^2).
$$

Moreover, we have

$$
\mathbb{V}\left[\frac{\sigma_n^k}{k!} n^{-1/2} \sum_{i=1}^{n} (\xi_{kin} - \xi_{kn})\right] = o_n(1),
$$

and, hence, by Chebyshev's inequality,

$$
\frac{\sigma_n^k}{k!} n^{-1/2} \sum_{i=1}^{n} (\xi_{kin} - \xi_{kn}) = o_{p,n}(1)
$$

for $k \in \{2, \ldots, K\}$. As a result, we conclude

$$
n^{1/2}\overline{g}_{\mathrm{U}}(\theta_{0n}) = n^{1/2}\overline{g}_{\mathrm{U}}^*(\theta_{0n}) + n^{1/2} b_n + o_{p,n}(1). \tag{A.1}
$$

The rest of the proof is standard and follows the corresponding lines of the proof of Theorem

2 Part (i). We start with $\nabla_\theta \hat{Q}_{\mathrm{U}}(\hat{\theta}_{\mathrm{U}}) = o_{p,n}(n^{-1/2})$, expand $\overline{g}_{\mathrm{U}}(\hat{\theta}_{\mathrm{U}})$ around θ_{0n}, exploit uniform consistency of $\hat{\theta}_{\mathrm{U}}$ along with the result of Lemma A.3, Assumptions 5 (iv) and Condition (ii) of the theorem, to show that

$$
\begin{aligned}
n^{1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n}) &= -\left((G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}} + o_{p,n}(1)\right) n^{1/2}\overline{g}_{\mathrm{U}}(\theta_{0n}) \\
&= -\left((G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}} + o_{p,n}(1)\right)\left(n^{1/2}\overline{g}_{\mathrm{U}}^{*}(\theta_{0n}) + n^{1/2}b_n + o_{p,n}(1)\right) \\
&= -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}n^{1/2}\overline{g}_{\mathrm{U}}^{*}(\theta_{0n}) + n^{1/2}b_{\theta n} + o_{p,n}(n^{1/2}b_n) + o_{p,n}(1),
\end{aligned}
$$

where

$$
b_{\theta_n} \equiv -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}b_n = -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}\sum_{k=2}^{K}\frac{\sigma_n^k}{k!}A_{\mathrm{U}k}^{*}\theta_{01n}\mathbb{E}\left[(\varepsilon_{in}/\sigma_n)^k\right] + O_n(\sigma_n^2\left\|\theta_{01n}\right\|^2).
$$

Rearranging the terms gives

$$
\begin{aligned}
n^{1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n} - b_{\theta n}) &= -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}n^{1/2}\overline{g}_{\mathrm{U}}^{*}(\theta_{0n}) + o_{p,n}(n^{1/2}b_{\theta n}) + o_{p,n}(1) \\
&= -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}n^{1/2}\overline{g}_{\mathrm{U}}^{*}(\theta_{0n}) + o_{p,n}(1),
\end{aligned}
$$

where the second equality is due to $n^{1/2}b_n = O_n(1)$ (guaranteed by $n^{1/2}\left(\sum_{k=2}^{K}\frac{\mathbb{E}[\varepsilon_{in}^k]}{k!}A_{\mathrm{U}k}^{*}\right)\theta_{01n} = O_n(1)$ and $n^{1/2}\sigma_n^2\left\|\theta_{01n}\right\|^2 = O_n(1)$). This, along with $n^{1/2}\Omega_{g_{\mathrm{U}}g_{\mathrm{U}}}^{*-1/2}\overline{g}_{\mathrm{U}}^{*}(\theta_{0n}) \xrightarrow{d} N(0, I_m)$, ensures

$$
n^{1/2}\Sigma_{\mathrm{U}}^{*-1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n} - b_{\theta n}) \xrightarrow{d} N(0, I_p),
$$

where

$$
\Sigma_{\mathrm{U}}^{*} \equiv (G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}\Omega_{g_{\mathrm{U}}g_{\mathrm{U}}}^{*}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*}(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^{*})^{-1}
$$

has eigenvalues uniformly bounded from below and above. Finally, Lemma A.3 ensures that $G_{\mathrm{U}} = G_{\mathrm{U}}^{*} + o_n(1)$ and $\Omega_{g_{\mathrm{U}}g_{\mathrm{U}}} = \Omega_{g_{\mathrm{U}}g_{\mathrm{U}}}^{*} + o_n(1)$, so we also have $\Sigma_{\mathrm{U}} = \Sigma_{\mathrm{U}}^{*} + o_p(1)$ and

$$
n^{1/2}\Sigma_{\mathrm{U}}^{-1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n} - b_{\theta n}) \xrightarrow{d} N(0, I_p),
$$

which completes the proof, since $\overline{\theta}_{\mathrm{U}0n} = \theta_{0n} + b_{\theta n}$. Q.E.D.

A.5 Proofs concerning hybrid inference

A.5.1 Proof of Lemma 2

Proof of Lemma 2. We show that Assumption 10 is satisfied for $\hat{\Sigma}_{\theta_1}$. First, we inspect Condition (ii). Take any fixed $\tilde{C} > 0$ and $\omega_{\mathrm{L}} \in (0, 1/2)$ as the condition requires. In this case, we are in the semi-strong identification regime and Lemma A.7 applies. Specifically, it ensures $S\hat{\Sigma}S = \underline{\Sigma} + o_{p,n}(1)$. Also note that $P_\theta S = P_\theta$ and $SP_\theta' = P_\theta'$, where $P_\theta = (I_p, 0_{p\times(K-1)})$. Then, taking the θ-corresponding submatrices gives

$$
\hat{\Sigma}_\theta = P_\theta\hat{\Sigma}P_\theta' = P_\theta S\hat{\Sigma}SP_\theta' = P_\theta\underline{\Sigma}P_\theta' + o_{p,n}(1). \tag{A.1}
$$

Since $\lambda_{\max}(\underline{\Sigma}) < C$, we also have that $\lambda_{\max}(\hat{\Sigma}_{\theta}) < C$ with probability approaching one. Finally, this also implies that $\lambda_{\max}(\hat{\Sigma}_{\theta_1}) < C$ with probability approaching one. Hence, Condition (ii) is satisfied.

Then we verify Condition (i). Take $\delta_0 > 0$ as in Assumption 9. Note that Condition (i) is trivially satisfied for $\|\theta_{01n}\| \geqslant \delta_0$, since, in this case, $\hat{\Sigma}_{\theta_1} = \Sigma^*_{\theta_1} + o_{p,n}(1)$, and the minimal eigenvalue of $\Sigma^*_{\theta_1}$ is bounded away from zero (the standard strong identification asymptotics applies). Then we verify Condition (i) for $\|\theta_{01n}\| \leqslant \delta_0$. By construction,

$$\hat{\Sigma} \geqslant (\hat{\Psi}'\hat{\Omega}_{\psi\psi}^{-1}\hat{\Psi})^{-1}.$$

Since $\hat{\Sigma}_{\theta_1}$ is a submatrix of $\hat{\Sigma}$, it is sufficient to demonstrate that $\lambda_{\min}(\hat{\Psi}'\hat{\Omega}_{\psi\psi}^{-1}\hat{\Psi})^{-1} > 1/C > 0$, or, equivalently, that $\lambda_{\max}(\hat{\Psi}'\hat{\Omega}_{\psi\psi}^{-1}\hat{\Psi}) < C$ with probability approaching one. Recall $\hat{\Psi} = \left(\overline{\Psi}_{\theta}(\hat{\theta},\hat{\gamma}), -g_x^{(2)}(\hat{\theta}), \ldots, -g_x^{(K)}(\hat{\theta})\right)$. In the proof of Lemma A.8, we have shown that $\sup_{\gamma\in\Gamma}\left\|\overline{\Psi}_{\theta}(\hat{\theta},\gamma) - \Psi^*_{\theta}(\theta_{0n},\gamma)\right\|$ $o_p(1)$, where $\Psi^*_{\theta}(\theta_{0n},\gamma)$ is uniformly bounded (this is ensured for $\|\theta_{01n}\| \leqslant \delta_0$). Also, as shown in the proof of Lemma A.3, $\overline{g}_x^{(k)}(\hat{\theta}) = g_x^{(k)*} + o_{p,n}(1)$, where, again, $g_x^{(k)*}$ are uniformly bounded, for $k \in \{2,\ldots,K\}$. Hence, for some $C > 0$, we have $\left\|\hat{\Psi}\right\| \leqslant C$ with probability approaching one uniformly. Then, using the same argument as in the proof of Lemma A.3, we have

$$\sup_{\gamma\in\Gamma}\left\|\hat{\Omega}_{\psi\psi}(\hat{\theta},\gamma)^{-1} - \Omega^*_{\psi\psi}(\theta_{0n},\gamma)\right\| = o_{p,n}(1).$$

Combining this fact with $\inf_{\gamma\in\Gamma}\lambda_{\min}(\Omega^*_{\psi\psi}(\theta_{0n},\gamma)) > C > 0$ (which is a hypothesis of the lemma), ensures that, for some $C > 0$, we have $\lambda_{\max}(\hat{\Omega}_{\psi\psi}^{-1}) < C$ with probability approaching one. Combining this with the bound on $\left\|\hat{\Psi}\right\|$, we conclude that, for some $C > 0$, we have $\lambda_{\max}(\hat{\Psi}'\hat{\Omega}_{\psi\psi}^{-1}\hat{\Psi}) < C$ with probability approaching one, which ensures that with probability approaching one we have $\lambda_{\min}(\hat{\Sigma}_{\theta}) > 1/C > 0$ with probability approaching one. Finally, note that this also implies $\lambda_{\min}(\hat{\Sigma}_{\theta_1}) \geqslant \lambda_{\min}(\hat{\Sigma}_{\theta}) > 1/C > 0$ with probability approaching one. Since Condition (i) is satisfied for both $\|\theta_{01n}\| \geqslant \delta_0$ and $\|\theta_{01n}\| \leqslant \delta_0$, it is trivially satisfied uniformly over the entire parameter space.

A logically very similar but technically simpler (since $\hat{\gamma}$ is no longer involved) argument can be invoked to show that $\hat{\Sigma}_{R,\theta_1}$ also satisfies Assumption 10. Q.E.D.

A.5.2 Proofs of Theorems 5 and 6

First, we prove Theorem 6. Then Theorem 5 follows immediately.

Proof of Theorem 6. To prove Part (i), we need to show that for any sequence of DPGs $\{\Upsilon_n\}$, $\limsup_{n\to\infty}\mathbb{E}_{\Upsilon_n}[\phi_{\mathcal{T}_{\mathrm{H}}}] \leqslant \alpha$. Take any fixed $C > 0$ and $\omega^* \in [\omega_{\mathrm{NZ}}, \omega_{\mathrm{L}}]$. Then, any sequence $\{\Upsilon_n\}$ can be split into two subsequences: the elements of the first subsequence are those which satisfy $\|\theta_{01n}\| \leqslant Cn^{-\omega^*}$, and the elements of the second subsequence are those which satisfy the complement requirement $\|\theta_{01n}\| > Cn^{-\omega^*}$. As usual, let $\{n_{1j}\}_{j=1}^{\infty}$ and $\{n_{2j}\}_{j=1}^{\infty}$ refer to the elements of these subsequences. To complete the proof it is sufficient to show that

$$\limsup_{j\to\infty}\mathbb{E}_{\Upsilon_{n_{1j}}}[\phi_{\mathcal{T}_{\mathrm{H}}}] \leqslant \alpha, \tag{A.2}$$

$$\limsup_{j\to\infty}\mathbb{E}_{\Upsilon_{n_{2j}}}[\phi_{\mathcal{T}_{\mathrm{H}}}] \leqslant \alpha. \tag{A.3}$$

We start with the first subsequence. Note that by $\hat{\theta}_1 - \theta_{01n} = O_{p,n}(n^{-1/2})$ and by Assumptions

10 (i) and 11 (iii), $\limsup_{j\to\infty} \mathbb{P}_{\Upsilon_{n_{1j}}}(\hat{A}_{\mathrm{ICS}} > \kappa_{\mathrm{L},n_{1j}}) = 0$. Moreover, the test $\mathcal{T}_{\mathrm{NZ}}$ controls size under any sequence of the first type, so

$$\limsup_{j\to\infty} \mathbb{E}_{\Upsilon_{n_{1j}}}[\phi_{\mathcal{T}_{\mathrm{NZ}}}] \leqslant \alpha.$$

Also note that, by construction, $\phi^{II}_{\mathcal{T}_{\mathrm{H}}} \leqslant \phi_{\mathcal{T}_{\mathrm{NZ}}} + \mathbb{1}\{\hat{A}_{\mathrm{ICS}} > \kappa_{\mathrm{L},n_{1j}}\}$. Hence,

$$\mathbb{E}_{\Upsilon_{n_{1j}}}[\phi_{\mathcal{T}^{II}_{\mathrm{H}}}] \leqslant \mathbb{E}_{\Upsilon_{n_{1j}}}[\phi_{\mathcal{T}_{\mathrm{NZ}}}] + \mathbb{P}_{\Upsilon_{n_{1j}}}(\hat{A}_{\mathrm{ICS}} > \kappa_{\mathrm{L},n_{1j}}).$$

Taking $\limsup_{j\to\infty}$ of both sided ensures that (A.2) holds.

For the second subsequence note again that $\hat{\theta}_1 - \theta_{01n} = O_{p,n}(n^{-1/2})$ and Assumptions 10 (ii) and 12 (i) guarantee that $\limsup_{j\to\infty} \mathbb{P}_{\Upsilon_{n_{2j}}}(\hat{A}_{\mathrm{ICS}} < \kappa_{\mathrm{NZ},n_{2j}}) = 0$. Also, the test $\mathcal{T}_{\mathrm{L}}$ controls size under any sequence of the second type:

$$\limsup_{j\to\infty} \mathbb{E}_{\Upsilon_{n_{2j}}}[\phi_{\mathcal{T}_{\mathrm{L}}}] \leqslant \alpha.$$

Finally, note $\phi^{II}_{\mathcal{T}_{\mathrm{H}}} \leqslant \phi_{\mathcal{T}_{\mathrm{L}}} + \mathbb{1}\{\hat{A}_{\mathrm{ICS}} < \kappa_{\mathrm{NZ},n_{2j}}\}$. Hence,

$$\mathbb{E}_{\Upsilon_{n_{2j}}}[\phi_{\mathcal{T}_{\mathrm{H}}}] \leqslant \mathbb{E}_{\Upsilon_{n_{2j}}}[\phi_{\mathcal{T}_{s}}] + \mathbb{P}_{\Upsilon_{n_{2j}}}(\hat{A}_{\mathrm{ICS}} < \kappa_{\mathrm{NZ},n_{2j}}),$$

and taking $\limsup_{j\to\infty}$ of both sides ensures that (A.3) holds. Hence (A.2) and (A.3) are demonstrated to hold, the proof of the first statement is complete.

Now we prove Part (ii). First, note that $\|\theta_{01n}\| / (n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$, along with Assumption 11 (iii), implies that $\|\theta_{01n}\| \geqslant C n^{-\omega_{\mathrm{L}}}$ for some (actually, any fixed) $C > 0$. Hence, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| / (n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$, Assumption 10 (ii) applies. Combining $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$ with Assumption 10 (ii), we conclude that, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| / (n^{-1/2}\kappa_{\mathrm{L},n}) \to \infty$, $\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{L},n} \to \infty$ and $\lambda_{\mathrm{L}}(\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{L},n}) \to 1$ as $n \to \infty$ with probability approaching one uniformly. Hence, $\sup_v \left| p_{\mathcal{T}^{II}_{\mathrm{H}}} - p_{\mathcal{T}_{\mathrm{L}}} \right| = o_{p,n}(1)$, where uniformity over v follows trivially since the p-values are bounded.

Finally, we prove Part (iii). For $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| / (n^{-1/2}\kappa_{\mathrm{NZ},n}) \to 0$, $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$ and Assumption 10 (ii) together ensure that $\kappa_{\mathrm{NZ},n} - \hat{A}_{\mathrm{ICS}} \to \infty$ and $\lambda_{\mathrm{NZ}}(\kappa_{\mathrm{NZ},n} - \hat{A}_{\mathrm{ICS}}) \to 1$ as $n \to \infty$ with probability approaching one uniformly. Hence, $\sup_v \left| p^{II}_{\mathcal{T}_{\mathrm{H}}} - p_{\mathcal{T}_{\mathrm{NZ}}} \right| = o_{p,n}(1)$, where uniformity over v follows trivially since the p-values are bounded. Q.E.D.

Proof of Theorem 5. Note that uniformly validity of $\mathcal{T}^{II}_{\mathrm{H}}$ (Theorem 6 Part (i)) immediately implies that $\mathcal{T}^{I}_{\mathrm{H}}$, which is more conservative than $\mathcal{T}^{II}_{\mathrm{H}}$, is also uniformly valid, so Part (i) of Theorem 5 trivially holds.

The proof of Part (ii) is exactly the same as the proof of Part (ii) of Theorem 6. Q.E.D.

A.5.3 Proof of Theorem 7

Proof of Theorem 7. We prove the second part of the theorem. The proof of the first part is completely analogous and follows from second.

We just need to verify that the hypotheses of Theorem 6 are satisfied. The requirement of $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$ is trivially satisfied for the MME estimator $\hat{\theta}$ (Theorem 3). Assumption 11 (ii) is ensured to be satisfied by Lemma 2. Assumptions 11 (iii), (iv) and 12 (i) and (ii) are explicitly made by the theorem.

We are left to show that Assumption 11 (i) holds with ω_{L} and ω_{NZ} given in the text of the theorem. We start with $\mathcal{T}_{\mathrm{L}}$. Since $\omega_{\mathrm{L}} < 1/2$, the assertion of Theorem A.1 applies, so for any $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$, we have

$$n^{1/2}\underline{\Sigma}_\theta^{-1/2}(\hat{\theta} - \theta_{0n}) \xrightarrow{d} N(0, I_p), \tag{A.4}$$

where Σ_θ is the θ-corresponding submatrix of $\underline{\Sigma}$, i.e. $\Sigma_\theta = P_\theta \underline{\Sigma} P_\theta'$, where $P_\theta = (I_p, I_{p\times(K-1)})$. Also, in the proof of Theorem 2, we have established that $\hat{\Sigma} = \underline{\Sigma}_\theta + o_{p,n}(1)$, and the eigenvalues of $\underline{\Sigma}_\theta$ are bounded from zero and above. Next, since R is continuous on Θ, which is compact, it is also uniformly continuous. Then, since $\hat{\theta} - \theta_{0n} = O_{p,n}(n^{-1/2})$, we have $\hat{R} = R + o_{p,n}(1)$. Combining these consistency results for $\hat{\Sigma}_\theta$ and $\hat{R}$ with $\lambda_{\min}(RR') > C > 0$, we conclude that

$$(\hat{R}\hat{\Sigma}_\theta \hat{R}')^{-1} = (R\underline{\Sigma}_\theta R)^{-1} + o_{p,n}(1).$$

Finally, under the null, we have

$$n^{1/2}(r(\hat{\theta}) - v) = n^{1/2}(r(\hat{\theta}) - \theta_{0n}) = n^{1/2}R(\tilde{\theta})(\hat{\theta} - \theta_{01n}) = Rn^{1/2}(\hat{\theta} - \theta_{0n}) + o_{p,n}(1),$$

where the last equality is due $R(\tilde{\theta}) = R + o_{p,n}(1)$ and $n^{1/2}(\hat{\theta} - \theta_{0n}) = O_{p,n}(1)$. Combining these results,

$$\begin{aligned}
\mathcal{W}_{\mathrm{L}} &= n^{1/2}(\hat{\theta} - \theta_{0n})' R'(R\underline{\Sigma}_\theta R')^{-1} R n^{1/2}(\hat{\theta} - \theta_{0n}) + o_{p,n}(1) \\
&= n^{1/2}(\hat{\theta} - \theta_{0n})' \underline{\Sigma}_\theta^{-1/2} \underline{\Sigma}_\theta^{1/2} R'(R\underline{\Sigma}_\theta R')^{-1} R \underline{\Sigma}_\theta^{1/2} \underline{\Sigma}_\theta^{-1/2} n^{1/2}(\hat{\theta} - \theta_{0n}) + o_{p,n}(1) \\
&\xrightarrow{d} \chi^2_{d_r},
\end{aligned}$$

where the convergence in distribution is due (A.4). Since, for any $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$, we have $\mathcal{W}_{\mathrm{L}} \to \chi^2_{d_r}$, (7.2) holds.

To demonstrate that (7.3) holds, note that, since $\omega_{\mathrm{NZ}} \geqslant 1/2 - 1/(K+1)$, the assertion of Corollary 1 holds, and we have, for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$

$$n^{1/2}\Sigma_{\mathrm{U}}^{-1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n}) \xrightarrow{d} N(0, I_p),$$

and $\hat{\Sigma}_{\mathrm{U}} = \Sigma_{\mathrm{U}} + o_{p,n}(1)$. Then, a nearly identical argument can be invoked to demonstrate that, for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$, $\mathcal{W}_{\mathrm{U}} \to \chi^2_{d_r}$, and, consequently, (7.3) also holds. Q.E.D.

A.5.4 Proof of Theorem 8

First we prove a general statement about uniform asymptotic normality of an adaptive estimator. Then we verify that this result applies to the setting of Theorem 8.

A.5.4.1 General adaptive estimation

We consider the following adaptive estimator

$$\hat{\theta}_{\mathrm{A}} = \hat{\theta}_{\mathrm{NZ}} + \hat{\Lambda}_n(\hat{\theta}_{\mathrm{L}} - \hat{\theta}_{\mathrm{NZ}}), \qquad \hat{\Lambda}_n \equiv \Lambda\left(\frac{\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{NZ},n}}{\kappa_{\mathrm{L},n} - \kappa_{\mathrm{NZ},n}}\right),$$

where, as before,

$$\hat{A}_{\mathrm{ICS}} = (n\hat{\theta}_1'\hat{V}_{11}^{-1}\hat{\theta}_1/p_1)^{1/2}.$$

Assumption A.2. There exist some $0 < \omega_{\mathrm{NZ}} < \omega_{\mathrm{L}} < 1/2$ such that:

(i) for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$ for any fixed $C > 0$, we have

$$n^{1/2}\Sigma_{\mathrm{L}}^{-1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) \xrightarrow{d} N(0, I_p) \qquad \text{and} \qquad \hat{\Sigma}_{\mathrm{L}} = \Sigma_{\mathrm{L}} + o_{p,n}(1),$$

where Σ_{L} is a symmetric matrix with $0 < 1/C_\lambda < \lambda_{\min}(\Sigma_{\mathrm{L}}) \leqslant \lambda_{\max}(\Sigma_{\mathrm{L}}) < C_\lambda$;

(ii) for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$ for any fixed $C > 0$, we have

$$n^{1/2}\Sigma_{\mathrm{NZ}}^{-1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}) \xrightarrow{d} N(0, I_p) \qquad \text{and} \qquad \hat{\Sigma}_{\mathrm{NZ}} = \Sigma_{\mathrm{NZ}} + o_{p,n}(1),$$

where Σ_{NZ} is a symmetric matrix with $0 < 1/C_\lambda < \lambda_{\min}(\Sigma_{\mathrm{NZ}}) \leqslant \lambda_{\max}(\Sigma_{\mathrm{NZ}}) < C_\lambda$;

(iii) moreover, for all $\{\Upsilon_n\}$ satisfying $Cn^{-\omega_{\mathrm{L}}} \leqslant \|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$ for any fixed $C > 0$, for any fixed nonzero $c = (c_1, c_2)' \in \mathbb{R}^2$, we have

$$n^{1/2}\Sigma_c^{-1/2}\left(c_1\hat{\theta}_{\mathrm{L}} + c_2\hat{\theta}_{\mathrm{NZ}} - (c_1 + c_2)\theta_{0n}\right) \xrightarrow{d} N(0, I_p), \quad \Sigma_c \equiv (c_1^2\Sigma_{\mathrm{L}} + c_1c_2(\Sigma_{\mathrm{L,NZ}} + \Sigma_{\mathrm{L,NZ}}') + c_2^2\Sigma_{\mathrm{NZ}}),$$

and $\hat{\Sigma}_{\mathrm{L,NZ}} = \Sigma_{\mathrm{L,NZ}} + o_p(1)$, where $\Sigma_{\mathrm{L,NZ}}$ is uniformly bounded;

(iv) $\hat{\theta}_1 = \theta_{01n} + O_{p,n}(n^{-1/2})$;

(v) $0 < 1/C < \lambda_{\min}(\hat{V}_{11})$ for some $C > 0$ with probability approaching one uniformly;

(vi) for all $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$, we have $\hat{V}_{11} = V_{11} + o_{p,n}(1)$, where V_{11} is a symmetric matrix with $0 < 1/C < \lambda_{\min}(V_{11}) \leqslant \lambda_{\max}(V_{11}) < C$;

(vii) $\Lambda : \mathbb{R} \to [0, 1]$ is a weakly increasing function satisfying $\Lambda(z) = 0$ for all $z \leqslant 0$, $\Lambda(z) = 1$ for all $z \geqslant 1$ and $|\Lambda(z') - \Lambda(z)| \leqslant \overline{\Lambda}\,|z' - z|$ for all $z', z \in \mathbb{R}$;

(viii) the thresholds $\kappa_{\mathrm{NZ},n} < \kappa_{\mathrm{L},n}$ satisfy $n^{\omega_{\mathrm{L}}-1/2}\kappa_{\mathrm{NZ},n} \to \infty$ and $n^{\omega_{\mathrm{NZ}}-1/2}\kappa_{\mathrm{L},n} \to 0$.

Lemma A.9. *Under Assumption A.2, we have*

$$n^{1/2}\hat{\Sigma}_{\mathrm{A}}^{-1/2}(\hat{\theta}_{\mathrm{A}} - \theta_{0n}) \xrightarrow{d} N(0, I_p), \tag{A.5}$$

where

$$\hat{\Sigma}_{\mathrm{A}} = \hat{\Lambda}_n^2\hat{\Sigma}_{\mathrm{L}} + \hat{\Lambda}_n(1 - \hat{\Lambda}_n)(\hat{\Sigma}_{\mathrm{L,NZ}} + \hat{\Sigma}_{\mathrm{L,NZ}}') + (1 - \hat{\Lambda}_n)^2\hat{\Sigma}_{\mathrm{NZ}}.$$

Proof of Lemma A.9. Take any $\{\Upsilon_n\}$, and split it into 3 subsequences: the elements of the first satisfy satisfy $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{L}}}$, the elements of the second satisfy $Cn^{-\omega_{\mathrm{L}}} < \|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$, and the elements of the third satisfy $\|\theta_{01n}\| > Cn^{-\omega_{\mathrm{NZ}}}$, for some fixed $C > 0$. Then, it is sufficient to argue that (A.5) holds for all of the three subsequences.

We start with the first subsequence with elements satisfying $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{L}}}$. Using Conditions (iv) and (v), we have $\hat{A}_{\mathrm{ICS}} = (n\theta_{01n}'\hat{V}_{11}^{-1}\theta_{01n})^{1/2} + o_p(1)$. Combining these conditions with Condition (viii), we conclude that, in this case, we have $\hat{A}_{\mathrm{ICS}} - \kappa_{\mathrm{NZ},n} < 0$ with probability approaching one. Hence, by Condition (vii), we have $\hat{\theta}_{\mathrm{A}} = \hat{\theta}_{\mathrm{NZ}}$ and $\hat{\Sigma}_{\mathrm{A}} = \Sigma_{\mathrm{NZ}}$ with probability approaching one, so

$$n^{1/2}\hat{\Sigma}_{\mathrm{A}}^{-1/2}(\hat{\theta}_{\mathrm{A}} - \theta_{0n}) = n^{1/2}\hat{\Sigma}_{\mathrm{NZ}}^{-1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n})$$

holds with probability approaching one. Finally note that, since $\omega_N Z < \omega_{\mathrm{L}}$, Assumption A.2 (ii) guarantees that for the elements of the first subsequence we have

$$n^{1/2}\hat{\Sigma}_{\mathrm{NZ}}^{-1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}) \xrightarrow{d} N(0, I_p).$$

we conclude that also (A.5) holds.

Now we consider the second subsequence with elements satisfying $Cn^{-\omega_{\mathrm{L}}} < \|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$. Note for the elements of this subsequence the joint convergence condition (iii) applies. First, note that, in this case, Conditions (iv) and (vi) guarantee that $\hat{A}_{\mathrm{ICS}} = A + o_{p,n}(1)$, where $A \equiv (n\theta_{01n}' V_{11}^{-1}\theta_{01n})$. Then Condition (vii), along with $\kappa_{\mathrm{L},n} - \kappa_{\mathrm{NZ},n} \to \infty$, guarantees

$$\hat{\Lambda}_n - \Lambda_n = o_{p,n}(1),$$

where

$$\Lambda_n = \Lambda\left(\frac{A - \kappa_{\mathrm{NZ},n}}{\kappa_{\mathrm{L},n} - \kappa_{\mathrm{NZ},n}}\right).$$

Hence, we also have

$$\begin{aligned} n^{1/2}(\hat{\theta}_{\mathrm{A}} - \theta_{0n}) &= \hat{\Lambda}_n n^{1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) + (1 - \hat{\Lambda}_n)n^{1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}) \\ &= \Lambda_n n^{1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) + (1 - \Lambda_n)n^{1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}) + o_{p,n}(1), \end{aligned}$$

where the second equality also uses $n^{1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) = O_{p,n}(1)$ and $n^{1/2}(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n}) = O_p(1)$. In addition,

$$\hat{\Sigma}_{\mathrm{A}} = \Lambda_n^2\Sigma_{\mathrm{L}} + \Lambda_n(1 - \Lambda_n)(\Sigma_{\mathrm{L,NZ}} + \Sigma_{\mathrm{L,NZ}}') + (1 - \Lambda_n)^2\Sigma_{\mathrm{NZ}} + o_{p,n}(1).$$

Combining these results, we obtain

$$\begin{aligned} n^{1/2}\hat{\Sigma}_{\mathrm{A}}^{-1/2}(\hat{\theta}_{\mathrm{A}} - \theta_{0n}) =& (\Lambda_n^2\Sigma_{\mathrm{L}} + \Lambda_n(1 - \Lambda_n)(\Sigma_{\mathrm{L,NZ}} + \Sigma_{\mathrm{L,NZ}}') + (1 - \Lambda_n)^2\Sigma_{\mathrm{NZ}})^{-1/2} \\ &\times n^{1/2}(\Lambda_n(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) + (1 - \Lambda_n)(\hat{\theta}_{\mathrm{NZ}} - \theta_{0n})) + o_{p,n}(1) \xrightarrow{d} N(0, I_p), \end{aligned}$$

where the convergence in distribution follows from Condition (iii).

Finally, we consider the third subsequence with elements satisfying $\|\theta_{01n}\| > Cn^{-\omega_{\mathrm{NZ}}}$. In this case, we again have $\hat{A}_{\mathrm{ICS}} = A + o_{p,n}(1)$. Then, using Condition (viii), we also have $\hat{A}_{\mathrm{ICS}} > \kappa_{\mathrm{L},n}$ with probability approaching one. Hence, with probability approaching one, we have (by Condition (vii)) $\hat{\theta}_{\mathrm{A}} = \theta_{\mathrm{L}}$ and $\hat{\Sigma}_{\mathrm{A}} = \hat{\Sigma}_{\mathrm{L}}$. So, with probability approaching one, we have

$$n^{1/2}\Sigma_{\mathrm{A}}^{-1/2}(\hat{\theta}_{\mathrm{A}} - \theta_{0n}) = n^{1/2}\hat{\Sigma}_{\mathrm{L}}^{-1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}).$$

Finally, since $\omega_{\mathrm{NZ}} < \omega_{\mathrm{L}}$, Condition (i) guarantees that for the elements of the third subsequence we have

$$n^{1/2}\hat{\Sigma}_{\mathrm{L}}^{-1/2}(\hat{\theta}_{\mathrm{L}} - \theta_{0n}) \xrightarrow{d} N(0, I_p).$$

Hence, we conclude that (A.5) also holds.

Since we have shown that (A.5) holds for all of the considered subsequences of $\{\Upsilon_n\}$, for any arbitrary $\{\Upsilon_n\}$, the proof is complete. Q.E.D.

A.5.4.2 An example of the asymptotic covariance estimator and Proof of Theorem 8

The asymptotic covariance between $\hat{\theta}$ and $\hat{\theta}_{\mathrm{U}}$ can be estimated by

$$\hat{\Sigma}_{\theta\theta_{\mathrm{U}}} \equiv P_\theta(\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}_{\psi g_{\mathrm{U}}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}}(\hat{G}_{\mathrm{U}}'\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}})^{-1}, \qquad \hat{\Omega}_{\psi g_{\mathrm{U}}} \equiv n^{-1}\sum_{i=1}^{n}\psi_i(\hat{\beta})g_{ui}(\hat{\theta}_{\mathrm{U}}),$$

where $P_\theta = (I_p, 0_{p\times(K-1)})$, $p \equiv \dim(\theta_{0n})$. Similarly, to the regularized estimators Σ_R and $\Sigma_{\mathrm{U,R}}$, one can also consider a regularized version of $\hat{\Sigma}_{\theta,u}$.

Proof of Theorem 8. The proof is essentially a verification of Assumption A.2.

Verification of Condition (i). Recall that in the proof of Theorem A.1 we have

$$S_n(\hat{\beta} - \beta_{0n}) = -(B'\Xi B)^{-1}B'\Xi n^{1/2}\overline{\psi} + o_{p,n}(1).$$

So,

$$n^{1/2}(\hat{\theta} - \theta_{0n}) = -P_\theta(B'\Xi B)^{-1}B'\Xi n^{1/2}\overline{\psi} + o_{p,n}(1). \tag{A.6}$$

Hence,

$$n^{1/2}\underline{\Sigma}_{\theta}^{-1/2}(\hat{\theta} - \theta_{0n}) \xrightarrow{d} N(0, I_p),$$

where $\underline{\Sigma}_\theta = P_\theta\underline{\Sigma}P_\theta'$ has eigenvalues bounded from below and above. Then, note that Lemma A.7 ensures $S\hat{\Sigma}S = \underline{\Sigma} + o_{p,n}(1)$. Also note that $P_\theta S = P_\theta$ and $SP_\theta' = P_\theta'$. Then, taking the θ-corresponding submatrices gives

$$\hat{\Sigma}_\theta = P_\theta\hat{\Sigma}P_\theta' = P_\theta S\hat{\Sigma}SP_\theta' = P_\theta\underline{\Sigma}P_\theta' + o_{p,n}(1).$$

So, Condition (i) is satisfied with $\hat{\theta}_{\mathrm{L}} = \hat{\theta}$, $\Sigma_{\mathrm{L}} = P_\theta\underline{\Sigma}P_\theta'$ and $\hat{\Sigma}_{\mathrm{L}} = \hat{\Sigma}_\theta$.

Verification of Condition (ii). Corollary 1 ensures that the required condition is satisfied with $\hat{\theta}_{\mathrm{NZ}} = \theta_{\mathrm{U}}$ and $\Sigma_{\mathrm{NZ}} = \Sigma_{\mathrm{U}}^*$ (note that Lemma A.3 guarantees that $\Sigma_{\mathrm{U}}^* = \Sigma_{\mathrm{U}} + o_n(1)$) Also, note that Lemma A.3 guarantees that $\hat{\Sigma}_{\mathrm{U}} = \Sigma_{\mathrm{U}}^* + o_{p,n}(1)$.

Verification of Condition (iii).

In the proof of Theorem 4, we have established (A.1). Note that, for $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$, we also have $n^{1/2}b_n = o_n(1)$ (no asymptotic bias in the uncorrected moments). Hence, (A.1) simplifies as

$$n^{1/2}\overline{g}_{\mathrm{U}} = n^{1/2}\overline{g}_{\mathrm{U}}^* + o_{p,n}(1),$$

and, hence,

$$n^{1/2}(\hat{\theta}_{\mathrm{U}} - \theta_{0n}) = -(G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}G_{\mathrm{U}}^*)^{-1}G_{\mathrm{U}}^{*\prime}\Xi_{\mathrm{U}}\overline{g}_{\mathrm{U}}^* + o_{p,n}(1). \tag{A.7}$$

By the same argument, for $\|\theta_{01n}\| \leqslant Cn^{-\omega_{\mathrm{NZ}}}$, we also have

$$\overline{\psi} = \overline{g}^* + o_{p,n}(1).$$

In this case, (A.6) simplifies as

$$n^{1/2}(\hat{\theta} - \theta_{0n}) = -P_\theta (B'\Xi B)^{-1} B'\Xi n^{1/2} \overline{g}^* + o_{p,n}(1).$$

Then, combining this with (A.7), we obtain

$$n^{1/2}(c_1\hat{\theta} + c_2\hat{\theta}_{\mathrm{U}} - (c_1 + c_2)\theta_{0n}) = -c_1 P_\theta (B'\Xi B)^{-1} B'\Xi n^{1/2}\overline{g}^* - c_2 (G^{*\prime}_{\mathrm{U}}\Xi_{\mathrm{U}} G^*_{\mathrm{U}})^{-1} G^{*\prime}_{\mathrm{U}}\Xi_{\mathrm{U}}\overline{g}^*_{\mathrm{U}} + o_{p,n}(1).$$

In this case, we put

$$\Sigma_c = c_1^2 \Sigma_{\mathrm{L}} + c_1 c_2 (\Sigma_{\mathrm{L,NZ}} + \Sigma'_{\mathrm{L,NZ}}) + c_2^2 \Sigma_{\mathrm{NZ}},$$

with $\Sigma_{\mathrm{L}} = P_\theta \underline{\Sigma} P'_\theta$, $\Sigma_{\mathrm{NZ}} = \Sigma^*_{\mathrm{U}}$, and

$$\Sigma_{\mathrm{L,NZ}} = P_\theta (B'\Xi B)^{-1} B'\Xi \Omega^*_{gg_{\mathrm{U}}} \Xi_{\mathrm{U}} G^*_{\mathrm{U}} (G^{*\prime}_{\mathrm{U}} \Xi_{\mathrm{U}} G^*_{\mathrm{U}})^{-1}, \qquad \Omega^*_{gg_{\mathrm{U}}} \equiv \mathbb{E}\left[g^*_i g^{*\prime}_{ui}\right].$$

Then, we have

$$n^{1/2}\Sigma_c^{-1/2}(c_1\hat{\theta} + c_2\hat{\theta}_{\mathrm{U}} - (c_1 + c_2)\theta_{0n}) \xrightarrow{d} N(0, I_p).$$

Finally, we need to show that $\hat{\Sigma}_{\theta\theta_{\mathrm{U}}} = \Sigma_{\mathrm{L,NZ}} + o_p(1)$. Note that, again using $P_\theta = P_\theta S^{-1}$, we have

$$\begin{aligned}\hat{\Sigma}_{\theta,u} &\equiv P_\theta(\hat{\Psi}'\hat{\Xi}\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}_{\psi g_{\mathrm{U}}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}}(\hat{G}'_{\mathrm{U}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}})^{-1}, \\ &= P_\theta(S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Psi}S^{-1})^{-1}S^{-1}\hat{\Psi}'\hat{\Xi}\hat{\Omega}_{\psi g_{\mathrm{U}}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}}(\hat{G}'_{\mathrm{U}}\hat{\Xi}_{\mathrm{U}}\hat{G}_{\mathrm{U}})^{-1}.\end{aligned}$$

Then, $\hat{\Psi}S^{-1} = B + o_{p,n}(1)$ (see, for example, the proof of Lemma A.7). Also, as usual, $\hat{\Xi} = \Xi + o_{p,n}(1)$, $\hat{\Xi}_{\mathrm{U}} = \Xi_{\mathrm{U}} + o_{p,n}(1)$, and $\hat{G}_{\mathrm{U}} = G^*_{\mathrm{U}} + o_{p,n}(1)$. Finally, replicating (a part of the) proof of Lemma A.3, we also have $\hat{\Omega}_{\psi g_{\mathrm{U}}} = \Omega^*_{gg_{\mathrm{U}}} + o_{p,n}(1)$. Hence, Condition (iii) is satisfied.

Verification of Condition (v). Ensured by Lemma 2 with $\hat{V}_{11} = \hat{\Sigma}_{\mathrm{R},\theta_1}$ or $\hat{V}_{11} = \hat{\Sigma}_{\theta_1}$.

Verification of Condition (vi). For example, take $\hat{V}_{11} = \hat{\Sigma}_\theta$. Note that, above we have shown, $\hat{\Sigma}_\theta = P_\theta \underline{\Sigma} P'_\theta + o_{p,n}(1)$ whenever $\|\theta_{01n}\| \geqslant Cn^{-\omega_{\mathrm{L}}}$. Again, note that the eigenvalues of $P_\theta \underline{\Sigma} P'_\theta$ are (uniformly) bounded from below and above. So, Condition (vi) is satisfied with $V_{11} = P_\theta \underline{\Sigma} P'_\theta$. By the same reasoning, Condition (vi) is also satisfied with $\hat{V}_{11} = \hat{\Sigma}_{\mathrm{R},\theta}$.

Verification of Condition (iv). Explicitly stated as properties of Λ.

Verification of Condition (viii). Explicitly assumed as one of the hypotheses of Theorem 8. Q.E.D.

A.6 Auxiliary Lemmas: Uniform ULLN

Lemma A.10 (Uniform-ULLN). *Suppose that (i) $\{Z_i\}_{i=1}^n$ are independently and identically distributed according to law $F \in \mathcal{F}$; (ii) $\Theta \in \mathbb{R}^p$ is bounded; (iii) for some (measurable) function $M(Z_i)$ and for all $\delta > 0$, $\|\eta(Z_i, \theta) - \eta(Z_i, \theta')\| \leqslant M(Z_i)\delta$ for all $\theta, \theta' \in \Theta : \|\theta - \theta'\| \leqslant \delta$ a.s. for all $F \in \mathcal{F}$; (iv) for some $\eta > 0$, $\sup_{F\in\mathcal{F}} \mathbb{E}_F \left[\|\eta(Z_i, \theta)\|^{1+\eta}\right] < C$ for every $\theta \in \Theta$; (v) $\sup_{F\in\mathcal{F}} \mathbb{E}_F[M(Z_i)] < C$.*

Then, for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F \left(\sup_{\theta\in\Theta} \left\| n^{-1} \sum_{i=1}^{n} \eta(Z_i,\theta) - \mathbb{E}_F[\eta(Z_i,\theta)] \right\| > \epsilon \right) = 0.$$

Proof. Denote $\eta_i(\theta) = \eta(Z_i, \theta)$ and $q_{i,F}(\theta) = \eta_i(\theta) - \eta_F(\theta)$, where $\eta_F(\theta) \equiv \mathbb{E}_F[\eta_i(\theta)]$. We need to show that

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F \left(\sup_{\theta\in\Theta} \|\overline{q}_F(\theta)\| > \epsilon \right) = 0,$$

where $\overline{q}_F(\theta) \equiv n^{-1} \sum_{i=1}^{n} (\eta_i(\theta) - \eta_F(\theta))$. First, (iv) guarantees uniform pointwise convergence of $\overline{q}_F(\theta)$ towards zero, i.e. for every $\epsilon > 0$,

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F \left(\|\overline{q}_F(\theta)\| > \epsilon \right) = 0 \tag{A.1}$$

for every $\theta \in \Theta$. Second, for all $\delta > 0$,

$$\begin{aligned} \sup_{\|\theta-\theta'\|\leqslant\delta} \|\overline{q}_F(\theta) - \overline{q}_F(\theta')\| &\leqslant \sup_{\|\theta-\theta'\|\leqslant\delta} \left(\|\overline{\eta}_F(\theta) - \overline{\eta}_F(\theta')\| + \|\eta_F(\theta) - \eta_F(\theta')\| \right) && \text{(A.2)} \\ &\leqslant \sup_{\|\theta-\theta'\|\leqslant\delta} \|\overline{\eta}_F(\theta) - \overline{\eta}_F(\theta')\| + \sup_{\|\theta-\theta'\|\leqslant\delta} \|\eta_F(\theta) - \eta_F(\theta')\| . && \text{(A.3)} \end{aligned}$$

where $\sup_{\|\theta-\theta'\|\leqslant\delta}$ abbreviates $\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|\leqslant\delta}$. Applying (iii) gives

$$\sup_{\|\theta-\theta'\|\leqslant\delta} \|\overline{\eta}_F(\theta) - \overline{\eta}_F(\theta')\| \leqslant n^{-1} \sum_{i=1}^{n} M(Z_i)\delta \quad \text{a.s.}$$

Similarly, using (iii),

$$\begin{aligned} \sup_{\|\theta-\theta'\|\leqslant\delta} \|\eta_F(\theta) - \eta_F(\theta')\| &\leqslant \sup_{\|\theta-\theta'\|\leqslant\delta} \mathbb{E}_F \left[\|\eta(Z_i,\theta) - \eta(Z_i,\theta')\| \right] \\ &\leqslant \mathbb{E}_F[M(Z_i)]\delta. \end{aligned}$$

Hence,

$$\sup_{\|\theta-\theta'\|\leqslant\delta} \|\overline{q}_F(\theta') - \overline{q}_F(\theta)\| \leqslant n^{-1} \sum_{i=1}^{n} M_q(Z_i)\delta \quad \text{a.s.}, \tag{A.4}$$

where $M_q(Z_i) \equiv M(Z_i) + \mathbb{E}_F[M(Z_i)]$. Note that, combining (v) with Markov inequality, we conclude that $n^{-1} \sum_{i=1}^{n} M_q(Z_i)$ is uniformly tight, i.e., for every $\epsilon > 0$, there exists $C_\epsilon > 0$ such that

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F \left(n^{-1} \sum_{i=1}^{n} M_q(Z_i) > C_\epsilon \right) < \epsilon. \tag{A.5}$$

Therefore, by combining (A.4) and (A.5), we conclude that for any $\epsilon > 0$ there exists $\delta_\epsilon = \epsilon / C_\epsilon$

such that

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}_F\left(\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|<\delta_\epsilon}\|\overline{q}_F(\theta')-\overline{q}_F(\theta')\|>\epsilon\right)<\epsilon. \tag{A.6}$$

The rest of the proof is standard. Since Θ is bounded, it can be covered by a finite number of balls of radius δ_ϵ, $\{B_{\delta_\epsilon}(\theta_j)\}_{j=1}^J$.

$$\begin{aligned}\mathbb{P}_F\left(\sup_{\theta\in\Theta}\|\overline{q}_F(\theta)\|>2\epsilon\right)&=\mathbb{P}_F\left(\max_j\sup_{\theta\in B_{\delta_\epsilon}(\theta_j)}\|\overline{q}_F(\theta)\|>2\epsilon\right)\\&\leqslant\mathbb{P}_F\left(\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|<\delta_\epsilon}\|\overline{q}_F(\theta')-\overline{q}_F(\theta')\|>\epsilon\right)+\mathbb{P}_F\left(\max_j\|\overline{q}(\theta_j)\|>\epsilon\right)\\&\leqslant\mathbb{P}_F\left(\sup_{\theta,\theta'\in\Theta:\|\theta-\theta'\|<\delta_\epsilon}\|\overline{q}_F(\theta')-\overline{q}_F(\theta')\|>\epsilon\right)+\sum_{j=1}^J\mathbb{P}_F\left(\|\overline{q}(\theta_j)\|>\epsilon\right),\end{aligned}$$

where $\theta\in B_{\delta_\epsilon}(\theta_j)$ abbreviates $\theta\in\Theta:\theta\in B_{\delta_\epsilon}(\theta_j)$. Then taking $\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}$ of the both sides gives

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}_F\left(\sup_{\theta\in\Theta}\|\overline{q}_F(\theta)\|>2\epsilon\right)<\epsilon,$$

where we used (A.6) and (A.1). Since $\epsilon>0$ was arbitrary chose, this completes the proof. Q.E.D.

Lemma A.11 (Uniform-ULLN for a differentiable function)**.** *Suppose that (i) $\{Z_i\}_{i=1}^n$ are independently and identically distributed according to law $F\in\mathcal{F}$; (ii) $\Theta\in\mathbb{R}^p$ is bounded and convex; (iii) (a vector function) $\eta(Z_i,\theta)$ is differentiable on Θ with probability one for every $F\in\mathcal{F}$; (iv) for some $\eta>0$, $\sup_{F\in\mathcal{F}}\mathbb{E}_F\left[\|\eta_i(\theta)\|^{1+\eta}\right]<C$ for every $\theta\in\Theta$; (v) $\sup_{F\in\mathcal{F}}\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|\nabla_\theta\eta_i(\theta)\|\right]<C$. Then, for any $\epsilon>0$,*

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}_F\left(\sup_{\theta\in\Theta}\left\|n^{-1}\sum_{i=1}^n\eta(Z_i,\theta)-\mathbb{E}_F[\eta(Z_i,\theta)]\right\|>\epsilon\right)=0.$$

Proof. We just need to verify that Conditions (iii) and (v) of Lemma A.10 are satisfied. First, since Θ is convex and $\eta(Z_i,\theta)$ is a.s. differentiable,

$$\|\eta(Z_i,\theta)-\eta(Z_i,\theta')\|\leqslant\sup_{\theta\in\Theta}\|\nabla_\theta\eta(Z_i,\theta)\|\,\|\theta-\theta'\|\quad\text{a.s.}$$

Hence, Condition (iii) of Lemma A.10 is satisfied with $M(Z_i)=\sup_{\theta\in\Theta}\|\nabla_\theta\eta(Z_i,\theta)\|$. Then (v) immediately implies Condition (v) of Lemma A.10. Q.E.D.

A.7 Verification of Assumptions for GLM

In this Appendix we provide verify that the higher-level assumptions hold for the Generalized Linear Model (GLM). Consider GLM of the form

$$Y_i=\rho(\theta_{01n}X_i^*+\theta_{02}'W_i)+U_i,\quad\mathbb{E}\left[U_i|X_i^*,W_i,Z_i\right]=0,$$

and the moment function

$$g(x,s,\theta) = (\rho(\theta_1 x + \theta_2' w) - y)h(x,w,z).$$

where $r = (w', y, z')'$, $S_i = (W_i', Y_i, Z_i')'$, and $\theta = (\theta_1, \theta_2')'$. In this case,

$$g_x^{(k)}(x,s,\theta) = (\rho(\theta_1 x + \theta_2' w) - y)h_x^{(k)}(x,w,z) + \sum_{j=1}^{k} \binom{k}{j} \theta_1^j \rho^{(j)}(\theta_1 x + \theta_2' w) h_x^{(k-j)}(x,w,z), \quad \text{(A.1)}$$

and

$$g_x^{(k)*} \equiv \mathbb{E}\left[g_x^{(k)}(X_i^*, S_i, \theta_{0n})\right] = \mathbb{E}\left[\sum_{j=1}^{k} \binom{k}{j} \theta_{01n}^j \rho^{(j)}(\theta_{01n} X_i^* + \theta_{02n}' W_i) h_x^{(k-j)}(X_i^*, W_i, Z_i)\right]. \quad \text{(A.2)}$$

Also, let $H(x,w,z) \equiv h(x,w,z) \cdot (x, w')$. Then

$$G(x,s,\theta) = \rho^{(1)}(\theta_1 x + \theta_2' w) H(x,w,z)$$

and

$$G^* = \mathbb{E}\left[\rho^{(1)}(\theta_{01n} X_i^* + \theta_{02n}' W_i) H(x,w,z)\right].$$

Finally,

$$G_x^{(k)}(x,s,w) = \sum_{j=0}^{k} \binom{k}{j} \theta_1^j \rho^{(j+1)}(\theta_1 x + \theta_2' w) H_x^{(k-j)}(x,w,z), \quad \text{(A.3)}$$

with

$$H_x^{(k-j)}(x,w,z) = h_x^{(k-j)}(x,w,z) \cdot (x, w') + (k-j) h_x^{(k-j-1)}(x,w,z) \cdot (1, 0').$$

The following set of regularity conditions is sufficient to verify most of the needed high-level assumptions. Specifically, it ensures that the following assumptions hold: Assumptions 4, 5 (i)-(iii), 6, 7, 9.

Assumption A.3.

(i) $\rho(\cdot)$ is at least $K+2$ times differentiable (on its domain) and $h_x^{(K+1)}(x,w,z)$ exist (on $\mathcal{X}$) for all $w \in \mathcal{W}$ and $z \in \mathcal{Z}$;

(ii) for some integer J, $K+2 \leqslant J \leqslant M$, we have: (a) $\sup_{x\in\mathcal{X}} \left\| h_x^{(J)}(x,w,z) \right\| < \infty$ for all $w \in \mathcal{W}$ and $z \in \mathcal{Z}$; (b) for all non-negative integers k_1 and k_2 such that $k_1 + k_2 = J$, $\sup_{x\in\mathcal{X}} \left\| \rho^{(k_1)}(\theta_1 x + \theta_2' w) h_x^{(k_2)}(x,w,z) \right\| < d(w,z,\theta)$ and $\sup_{x\in\mathcal{X}} \left\| \rho^{(k_1+1)}(\theta_1 x + \theta_2' w) h_x^{(k_2)}(x,w,z) x \right\| < d(w,z,\theta)$ for some function $d(w,z,\theta)$, for all $\theta \in \Theta$, $w \in \mathcal{W}$, $z \in \mathcal{Z}$;

(iii) for some positive integer $J_\rho \leqslant M+2$, we have $\left\| \rho^{(J_\rho)} \right\|_\infty < \infty$;

(iv) for some $\delta_0 > 0$ and $C > 0$, for all $\{\Upsilon_n\}$ satisfying $|\theta_{01n}| \leqslant \delta_0$, we have $\lambda_{\min}(B^{*\prime} B^*) > C$, where

$$B^* \equiv \mathbb{E}\left[\rho_i^{(1)*}\left(h_i^* \cdot (X_i^*, W_i'), h_{xi}^{(1)*}, \dots, h_{xi}^{(K-1)*}\right)\right];$$

(v) for some $\delta_0 > 0$ and $C > 0$, for all $\{\Upsilon_n\}$ satisfying $|\theta_{01n}| \leqslant \delta_0$, we have $\lambda_{\min}(B^{*\prime}(\gamma)B^*(\gamma)) > C$, where

$$B^*(\gamma) \equiv \mathbb{E}\left[\rho_i^{(1)*}\left(\left(h_i^* - \sum_{k=2}^{K}\gamma_k h_{xi}^{(k)*}\right) \cdot (X_i^*, W_i'), h_{xi}^{(1)*}, \dots, h_{xi}^{(K-1)*}\right)\right];$$

(vi) for all $\delta > 0$, $\exists C_\delta > 0$ such that, for $\{\Upsilon_n\}$ satisfying $\|\theta_{01n}\| \geqslant \delta$, $\lambda_{\min}(\Psi^{*\prime}\Psi^*) > C_\delta$;

(vii) for some $\eta > 0$, $\mathbb{E}\left[|U_i|^{4+\eta}\right] < C$ and $\mathbb{E}\left[\|h(X_i^*, W_i, Z_i)\|^{4+\eta}\right] < C$;

(viii) $\lambda_{\min}(\mathbb{E}\left[U_i^2 h(X_i^*, W_i, Z_i) h(X_i^*, W_i, Z_i)'\right]) > C > 0$;

(ix)

$$\begin{aligned}\mathbb{E}\Bigg[X_i^{*4} + \|W_i\|^4 + \left(\sup_{x\in\mathcal{X}}\left\|h_x^{(J)}(x, W_i, Z_i)\right\|\right)^4 + \sum_{j=0}^{J-1}\left(\left\|h_x^{(j)}(X_i^*, W_i, Z_i)\right\|^4\left(1 + |X_i^*|^4 + \|W_i\|^4\right)\right) \\ + \sup_{\theta\in\Theta}\left(\sum_{j=0}^{\max\{J, J_\rho - 1\}}\rho^{(j)}(\theta_1 X_i^* + \theta_2' W_i)^4 + d(Z_i, W_i, \theta)^2\right)\Bigg] < C.\end{aligned}$$

Verification of Assumption 4. Under assumed smoothness, Assumption 4 (i) trivially follows from Assumption A.3. Assumption 4 (ii) is satisfied since, under Assumption A.3 (ii), we have

$$\sup_{x\in\mathcal{X}}\left\|g_x^{(J)}(x, s, \theta)\right\| < \infty, \quad \sup_{x\in\mathcal{X}}\left\|G_x^{(J)}(x, s, \theta)\right\| < \infty$$

for all $s \in \mathcal{S}$ and $\theta \in \Theta$ (by inspection, using formulas (A.1) and (A.3)). Then, the reasoning provided in Remark 20 guarantees that Assumption 4 (ii) holds.

Verification of Assumptions 7 (i) *and 9* (i)*.* (A.2) suggests that $g_x^{(k)*} = A_k^*\theta_{01n} + O_n(\|\theta_{01n}\|^2)$, with

$$A_k^* = k\mathbb{E}\left[\rho^{(1)}(\theta_{01n}X_i^* + \theta_{02n}'W_i)h_x^{(k-1)}(X_i^*, W_i, Z_i)\right].$$

Also note that, under additional (weak) smoothness conditions, one can also take

$$A_k^* = k\mathbb{E}\left[\rho^{(1)}(\theta_{02n}'W_i)h_x^{(k-1)}(X_i^*, W_i, Z_i)\right].$$

Verification of Assumptions 7 (iii) *and 9* (iv)*.* We will argue that $g_x^{(k)*} - g_x^{(k)} = O_n(\sigma_n\|\theta_{01n}\|)$, which covers the both. Using formulas (A.1) and (A.2), we obtain

$$\begin{aligned}\left\|g_x^{(k)} - g_x^{(k)*}\right\| \leqslant{}& \left\|\mathbb{E}\left[\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - Y_i)h_x^{(k)}(X_i, W_i, Z_i)\right]\right\| \\ &+ \sum_{j=1}^{k}\binom{k}{j}|\theta_{01n}|^j\Big\|\mathbb{E}[\rho^{(j)}(\theta_{01n}X_i + \theta_{02n}'W_i)h_x^{(k-j)}(X_i, W_i, Z_i) \\ &\qquad - \rho^{(j)}(\theta_{01n}X_i^* + \theta_{02n}'W_i)h_x^{(k-j)}(X_i^*, W_i, Z_i)]\Big\|. \end{aligned} \tag{A.4}$$

First, we argue that

$$\left\|\mathbb{E}\left[\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - Y_i)h_x^{(k)}(X_i, W_i, Z_i)\right]\right\| = O_n(\sigma_n^2 \left|\theta_{01n}\right|).$$

Note that, since the measurement error is classical, we also have $\mathbb{E}\left[U_i | X_i^*, W_i, Z_i, \varepsilon_{in}\right] = 0$. Hence, $\mathbb{E}\left[U_i h(X_i, W_i, Z_i)\right] = 0$ and, consequently,

$$\begin{aligned}&\left\|\mathbb{E}\left[\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - Y_i)h_x^{(k)}(X_i, W_i, Z_i)\right]\right\| \\ &\qquad = \left\|\mathbb{E}\left[\left(\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - \rho(\theta_{01n}X_i^* + \theta_{02n}'W_i)\right) h_x^{(k)}(X_i, W_i, Z_i)\right]\right\|.\end{aligned}$$

Using Assumption A.3 (iii), we have

$$\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - \rho(\theta_{01n}X_i^* + \theta_{02n}'W_i) = \sum_{j=1}^{J_\rho - 1} \frac{\theta_{01n}^j}{j!}\rho^{(j)}(\theta_{01n}X_i^* + \theta_{02n}W_i)\varepsilon_i^j + r_\rho(X_i, X_i^*, W_i, \theta_{0n}),$$

where

$$r_\rho(X_i, X_i^*, W_i, \theta_{0n}) \equiv \frac{\theta_{01n}^{J_\rho}}{J_\rho!}\rho^{(J_\rho)}(\theta_{01n}\tilde{X}_i + \theta_{02}W_i)\varepsilon_i^{J_\rho}, \tag{A.5}$$

for some $\tilde{X}_i$ between X_i and X_i^*. Then, Assumption A.3 (iii) guarantees that

$$\left\|r_\rho(X_i, X_i^*, W_i, \theta_{0n})\right\| \leqslant \frac{\left|\theta_{01n}\right|^{J_\rho}}{J_\rho!}\left\|\rho^{(J_\rho)}\right\|_\infty \left|\varepsilon_i\right|^{J_\rho}.$$

Similarly, using Assumption A.3 (ii), we have

$$h_x^{(k)}(X_i, W_i, Z_i) = \sum_{j=k}^{J-1} \frac{1}{(j-k)!}h_x^{(j)}(X_i^*, W_i, Z_i)\varepsilon_i^{j-k} + r_h(X_i, X_i^*, W_i, Z_i, \theta_{0n}), \tag{A.6}$$

where

$$r_h(X_i, X_i^*, W_i, \theta_{0n}) = \frac{1}{(J-k)!}h_x^{(J)}(\tilde{X}_i, W_i, Z_i)\varepsilon_i^{J-k},$$

so

$$\left\|r_h(X_i, X_i^*, W_i, \theta_{0n})\right\| \leqslant \frac{1}{J!}\sup_{x\in\mathcal{X}}\left\|h_x^{(J)}(x, W_i, Z_i)\right\| \left|\varepsilon_i\right|^J.$$

Finally, multiplying expansions (A.5) and (A.6) and bounding the expectation of their product give us

$$\left\|\mathbb{E}\left[\left(\rho(\theta_{01n}X_i + \theta_{02n}'W_i) - \rho(\theta_{01n}X_i^* + \theta_{02n}'W_i)\right) h_x^{(k)}(X_i, W_i, Z_i)\right]\right\| = O_n(\sigma_n \left|\theta_{01n}\right|). \tag{A.7}$$

Note that $J \leqslant M$ and $J_\rho \leqslant M + 2$ guarantee that these product does involve powers of $|\varepsilon_{in}|$ larger than $2M$, for all $k \in \{2, \ldots, K\}$. To complete the proof then it is sufficient to show that the second term in (A.4) is also $O_n(\sigma_n \left\|\theta_{01n}\right\|)$. This would follow from

$$\mathbb{E}\left[\rho^{(j)}(\theta_{01n}X_i + \theta_{02n}'W_i)h_x^{(k-j)}(X_i, W_i, Z_i) - \rho^{(j)}(\theta_{01n}X_i^* + \theta_{02n}'W_i)h_x^{(k-j)}(X_i^*, W_i, Z_i)\right] = O(\sigma_n)$$

for all $k \in \{2, \dots, K\}$ and $j \in \{1, \dots, k\}$. Let $\zeta_{kj}(x, s, \theta) \equiv \rho^{(j)}(\theta_1 x + \theta_2' w) h_x^{(k-j)}(x, w, z)$. Then we want to show that

$$\mathbb{E}\left[\zeta_{kj}(X_i, S_i, \theta_{0n}) - \zeta_{kj}(X_i^*, S_i, \theta_{0n})\right] = O_n(\sigma_n). \tag{A.8}$$

This is straightforward to verify: Assumption A.3 (ii) ensures that the $J-k$-th derivative of $\zeta_{kj}(x, s, \theta)$ with respect to x exists and satisfies

$$\sup_{x \in \mathcal{X}} \left\| \zeta_{kj,x}^{(J-k)}(x, s, \theta) \right\| < \infty$$

for all $s \in \mathcal{S}$ and $\theta \in \Theta$. Hence, the standard expansion of $\zeta_{kj}(X_i^*, W_i, \theta_{0n})$ around X_i^* applies:

$$\zeta_{kj}(X_i, S_i, \theta_{0n}) - \zeta_{kj}(X_i^*, S_i, \theta_{0n}) = \sum_{\ell}^{J-k-1} \frac{1}{\ell!} \zeta_{kj,x}^{(\ell)}(X_i^*, S_i, \theta_{0n}) \varepsilon_{in}^{\ell} + r_{\zeta,kj}(X_i, X_i^*, S_i, \theta_{0n}), \tag{A.9}$$

where the remainder can be bounded as

$$\|r_{\zeta,kj}\| \leqslant \frac{1}{(J-k)!} \sup_{x \in \mathcal{X}} \left\| \zeta_{kj,x}^{(J-k)}(x, S_i, \theta_{0n}) \right\| |\varepsilon_{in}|^{J-k}.$$

Then (A.9) ensures that (A.8) holds. Finally, the bounds (A.4), (A.7), and (i) together ensure that $g_x^{(k)*} - g_x^{(k)} = O_n(\sigma_n \|\theta_{01n}\|)$ for all $k \in \{2, \dots, K\}$. Note that Assumptions A.3 vii and ix ensure that all needed moments exist.

Verification of Assumption 7 (ii). Note that before, we have shown that it is possible to take $A_k^* = k a_k^*$, where $a_k^* = \mathbb{E}\left[\rho_i^{(1)*} h_{xi}^{(k-1)*}\right]$. Hence, Assumption A.3 (iv) ensures that Assumption 7 (ii) holds.

Verification of Assumption 9 (iii). First, note that $\Psi_A^*(0) = \Psi_A^*$. We have already verified that Assumption 7 (ii) holds, so $\lambda_{\min}(\Psi_A^{*\prime}\Psi_A^*) > C > 0$ for all $\{\Upsilon_n\}$ satisfying $|\theta_{01n}| \leqslant \delta_0$. Hence, by continuity or $\Psi_A^*(\gamma)$ it necessarily follows that there exist some $\tilde{\Gamma} \ni 0$ and $\tilde{C} > 0$ such that $\inf_{\gamma \in \tilde{\Gamma}} \lambda_{\min}(\Psi_A^{*\prime}(\gamma)\Psi_A^*(\gamma)) > \tilde{C} > 0$. Consequently, we conclude that Assumption 9 (iii) follows from Assumption 7 (ii) whenever $\Gamma \in \tilde{\Gamma}$.

To extend Γ beyond $\tilde{\Gamma}$, note that, using (A.3), we obtain

$$\begin{aligned}
\Psi^*(\theta_{0n}, \gamma) &= G^* - \sum_{k=2}^{K} \gamma_k G_x^{(k)*} \\
&= G^* - \sum_{k=2}^{K} \gamma_k \mathbb{E}\left[\rho_i^{(1)*} H_{xi}^{(k)*}\right] + O_n(|\theta_{01n}|) \\
&= \mathbb{E}\left[\rho_{xi}^{(1)*}\left(H_i^* - \sum_{k=2}^{K} \gamma_k H_{xi}^{(k)*}\right)\right] + O_n(|\theta_{01n}|) \\
&= \mathbb{E}\left[\rho_{xi}^{(1)*}\left(h_i^* \cdot (X_i, W_i') - \sum_{k=2}^{K} \gamma_k \left(h_{xi}^{(k)*} \cdot (X_i, W_i') + k h_{xi}^{(k-1)*} \cdot (1, 0')\right)\right)\right] + O_n(|\theta_{01n}|) \\
&= \mathbb{E}\left[\rho_{xi}^{(1)*}\left(\left(h_i^* - \sum_{k=2}^{K} \gamma_k h_{xi}^{(k)*}\right) \cdot (X_i, W_i') - \sum_{k=2}^{K} \gamma_k h_{xi}^{(k-1)*} \cdot (1, 0')\right)\right] + O_n(|\theta_{01n}|).
\end{aligned}$$

Hence, recalling $A_k^* = ka_k^*$ with $a_k^* = \mathbb{E}\left[\rho_i^{(1)*} h_{xi}^{(k-1)*}\right]$, we have

$$\Psi_A^*(\gamma) = \mathbb{E}\Bigg[\rho_{xi}^{(1)*}\Bigg(\Bigg(h_i^* - \sum_{k=2}^{K}\gamma_k h_{xi}^{(k)*}\Bigg)\cdot(X_i, W_i') - \sum_{k=2}^{K}\gamma_k h_{xi}^{(k-1)*}\cdot(1, 0'),$$

$$2h_{xi}^{(1)*}, \ldots, Kh_{xi}^{(K-1)*}\Bigg)\Bigg] + O_n(|\theta_{01n}|).$$

Finally, since (i) $\sum_{k=2}^{K}\gamma_k h_{xi}^{(k-1)*}$ is already in the span of the last $K-1$ columns of $\Psi_A^*(\gamma)$ and γ is bounded and (ii) without loss of generality δ_0 can be taken that small so $O_n(|\theta_{01n}|)$ is negligible, Assumption (A.3) (v) ensures that Assumption 9 (iii) holds.
Verification of Assumption 6. First, we have shown that Assumptions 7 (i) and (ii) hold. These Assumptions together necessarily imply that there exist some $\overline{\delta} > 0$ such that Assumption 6 also necessarily holds provided that we restrict the set of possible $\{\Upsilon_n\}$ to $|\theta_{01n}| \leqslant \overline{\delta}$. In other words, Assumptions 7 (i) and (ii) imply that the standard local identification condition (Assumption 6) should also hold once true true parameter space is restricted to $|\theta_{01n}| \leqslant \overline{\delta}$. However, to extend this for a wider range DGPs, we necessarily need assume that this condition is also globally satisfied (Assumption A.3 (vi)).
Verification of Assumptions 5 (i)-(iii) *and 9* (ii). Assumption 5 (ii) follows directly from Assumption A.3 (viii). Condition $\mathbb{E}\left[\|g_i^*(\theta_{0n})\|^{2+\eta}\right]$ follows from Assumption A.3 (vii). Assumptions A.3 (vii) and (ix) together imply

$$\mathbb{E}\Bigg[\sup_{\theta\in\Theta}\Bigg(\sum_{k=0}^{K}\left(\left\|g_{xi}^{(k)*}(\theta)\right\|^2 + \left\|G_{xi}^{(k)*}(\theta)\right\|^2 + \left\|\nabla_\theta \mathrm{vec}\left(G_{xi}^{(k)*}(\theta)\right)\right\|\right)$$

$$+ b_{1i}^{*2}(\theta) + b_{2i}^{*2}(\theta) + b_{G1i}^*(\theta) + b_{G2i}^*(\theta)\Bigg)\Bigg] < C;$$

which covers the rest of Assumptions 5 (i), (iii), and 9 (ii).