Econometric Theory, **37**, 2021, 633–663.
doi:10.1017/S0266466620000250

IDENTIFICATION OF LINEAR REGRESSIONS WITH ERRORS IN ALL VARIABLES*

DAN BEN-MOSHE
The Hebrew University of Jerusalem

This paper analyzes the classical linear regression model with measurement errors in all the variables. First, we provide necessary and sufficient conditions for identification of the coefficients. We show that the coefficients are not identified if and only if an independent normally distributed linear combination of regressors can be transferred from the regressors to the errors. Second, we introduce a new estimator for the coefficients using a continuum of moments that are based on second derivatives of the log characteristic function of the observables. In Monte Carlo simulations, the estimator performs well and is robust to the amount of measurement error and number of mismeasured regressors. In an application to firm investment decisions, the estimates are similar to those produced by a generalized method of moments estimator based on third to fifth moments.

1. INTRODUCTION

The efficacy of the classical linear regression model, which stands at the heart of applied economics, can be compromised by measurement errors. Despite their pervasiveness in data and the resulting endogeneity problems, measurement errors are mostly ignored in practice.

The linear errors-in-variables model generalizes the linear regression model in order to take measurement errors into account.[1] For the single regressor errors-in-variables model, the conditions under which the coefficient can be identified and estimated are well understood. Surprisingly, these conditions are not known for the multiple regressor errors-in-variables model. This is the problem the present paper addresses.

*The paper is loosely based on the second chapter of my PhD thesis. I am grateful to Rosa Matzkin and Jinyong Hahn for their generous support, advice, and guidance. I thank the editor Liangjun Su as well as Xavier D'Haultfœuille, Jinyong Hahn, Allyn Jackson, Arthur Lewbel, Rosa Matzkin, Yannay Spitzer, and Elie Tamer for reading various parts of drafts and providing feedback. I benefited from comments at Aarhus, Boston College, CREST, Harvard, Hebrew University of Jerusalem, and Penn State, and discussions with David Genesove, Patrik Guggenberger, Kei Hirano, Stefan Hoderlein, Max Kasy, Saul Lach, Áureo de Paula, Anna Simoni, James Stock, Daniel Wilhelm, and Martin Weidner. Address correspondence to Dan Ben-Moshe, The Hebrew University of Jerusalem, Jerusalem, Israel; e-mail: danbm@huji.ac.il.

[1] See Aigner et al. (1984), Buonaccorsi (2010), Carroll et al. (2006), Chen, Hong, and Nekipelov (2011), Cheng and Van Ness (1999), Fuller (2009), Gillard (2010), and Schennach (2016) for reviews of measurement error models.

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

In the case of the single regressor errors-in-variables model, Reiersøl (1950) shows that the coefficient is not identified if and only if (i) the regressor is normal and (ii) either one of the errors can be expressed as the sum of two independent variables, one of which is normal. Hence, identification requires some nongaussianity assumption and consistent estimation is usually based on higher moments. Accurate estimation of the coefficient based on higher moments, however, requires large samples and is particularly sensitive to outliers. Common practice is to ignore higher moments and use the standard ordinary least squares (OLS) estimator, based on second moments, which bounds the unknown coefficient and identifies its sign.

The general conditions for identification of the multiple regressor errors-in-variables model are not known. The main contribution of this paper is to provide necessary and sufficient conditions for identification of the coefficients. The coefficients are not identified by first and second moments, so that jointly normal observables do not identify the coefficients (see, e.g., Koopmans, 1937; Reiersøl, 1945). Further, unlike the single regressor model, first and second moments do not in general bound the coefficients, unless they all lie in the same orthant (see Klepper and Leamer, 1984). The literature has based identification on nongaussianity assumptions on the unobserved regressors. Cragg (1997), Dagenais and Dagenais (1997), Erickson and Whited (2002), Geary (1941), Lewbel (2012), and Pal (1980) identify the coefficients using rank conditions on third and higher moments. Bekker (1986), Kapteyn and Wansbeek (1983), and Willassen (1979) identify the coefficients using the assumption that no linear combinations of the regressors are normal.

We show that the model may still be identified even if the unobserved regressors are jointly normal. For nonidentification, a linear combination of regressors must be not only normal but also transferable to the errors. Specifically, we show that the coefficients are not identified if and only if (i) there is a normally distributed linear combination of unobserved regressors that is (ii) independent of some other linear combinations of regressors and (iii) some errors can be expressed as the sum of two independent random variables, one of which is normal.

Our identification strategy uses characteristic functions to take advantage of the model's linearity in the regressors and errors. The literature that identifies measurement error models that are nonlinear in the regressors uses additional information such as instruments or repeated measurements (see, e.g., Hausman et al., 1991; Hu and Sasaki, 2015; Schennach, 2007). A notable exception is Schennach and Hu (2013) who show identification of a nonparametric function with a scalar mismeasured regressor. Extending identification to nonlinear functions of multidimensional mismeasured regressors, without any additional information, would be a new result.

The second contribution of this paper is to introduce a new estimator for the coefficients that uses information from all higher moments and that is not based on a solution to a system of higher moment equations which, as mentioned earlier,

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

requires large samples for accurate estimation.[2] Although identification is not constructive (i.e., identification does not produce a closed form mathematical expression for the unknown coefficients), we will nevertheless be able to mimic the steps in the identification procedure to construct an estimator using a continuum of moment equations based on second-order partial derivatives of the log characteristic function of the observables. The estimator is root-n consistent and asymptotically normal.

The identification and estimation results allow all the regressors to have measurement errors. If there are some regressors that are known to be perfectly measured then there are other ways to identify the coefficients. For example, Lewbel (1997) constructs instruments using the perfectly measured regressors and Ben-Moshe, D'Haultfœuille, and Lewbel (2017) project off the perfectly measured regressors and use them as instruments. The simulations and application focus on the more empirically relevant setting in which some of the regressors are mismeasured but others are not. We estimate the coefficients by first partialling out the perfectly measured regressors and then applying our results to the residuals of regressions of the outcome variable and mismeasured regressors on the perfectly measured regressors. In the finite sample simulations, the estimator performs well with small bias and decreasing standard deviation at root-n rate. The coverage probabilities of the 95% confidence interval are close to the nominal level suggesting that inference using the estimator is trustworthy. The estimates are robust to the amount of measurement error, the number of mismeasured regressors, and the distribution of the regressors.

Using data on firm investment decisions, the estimator produces results that are similar to the generalized method of moments (GMM) estimator based on third, fourth, and fifth moments. According to the neoclassical theory of firm investment, these decisions should depend only on marginal returns to capital (Abel, 1983; Abel and Eberly, 1994; Brainard and Tobin, 1968; Hayashi, 1982; Kaldor, 1966; Mussa, 1977; Tobin, 1969). Empirical studies, however, show that measures of marginal returns to capital explain very little about firm investment decisions (see, e.g., Blanchard, Lopez-de Silanes, and Shleifer, 1994), and have spurred a long-standing debate on why the theory is not confirmed by the data. Erickson and Whited (2000) and Erickson, Jiang, and Whited (2014) provide evidence that a linear regression with measurement error in marginal returns to capital corroborates the theory. However, most of the literature contends that financial constraints need to be incorporated into the theory (see, e.g., Fazzari, Hubbard, and Petersen, 1988; Gilchrist and Himmelberg, 1995). These financial constraints are unobservable and often measured by cash flows. We show that the estimated coefficients are mostly insignificant in a linear regression with measurement errors in both marginal returns to capital and in financial constraints. This suggests that

[2] In addition to estimation methods based on higher moments, estimation can be based on, for example, distributional assumptions, instrumental variables, multiple measurements, small error variance, and so forth (e.g., Bickel and Ritov, 1987; Chesher, 1991; Hausman et al., 1991; Li, 2002; Schennach, 2004; Spiegelman, 1979).

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

either the theory is missing something important or the linear errors-in-variables model is inappropriate.

The paper is organized as follows. Section 2 contains the assumptions and identification results. Section 3 presents a root-n consistent and asymptotically normal estimator for the coefficients. In finite sample simulations in Section 4 and in data on firm investment decisions in Section 5, we compare our integrated generalized method of moments (IGMM) estimator to a GMM estimator based on third to fifth moments and the OLS estimator. All the proofs are contained in the Appendix.

Notation. The main tool of analysis is the log characteristic function $\varphi_{X^*}(\boldsymbol{u}) = \ln(E[e^{i\boldsymbol{u}'X^*}])$, where $\varphi_{X^*}(\boldsymbol{u})$ completely defines the p-dimensional random vector $\boldsymbol{X}^*$, $\boldsymbol{u}'$ is the transpose of $\boldsymbol{u} \in \mathbb{R}^p$, $i = \sqrt{-1}$ is the imaginary unit, $\ln(\cdot)$ is the natural log restricted to its principal branch, and $E[\cdot]$ is the expectation operator. The log characteristic function of a jointly normal random vector $\boldsymbol{X}^* \sim N(\boldsymbol{\mu}_{X^*}, \boldsymbol{\Sigma}_{X^*})$ with mean $\boldsymbol{\mu}_{X^*}$ and covariance matrix $\boldsymbol{\Sigma}_{X^*}$ is $\varphi_{X^*}(\boldsymbol{u}) = i\boldsymbol{\mu}'_{X^*}\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma}_{X^*}\boldsymbol{u}$. We use the convention that $\boldsymbol{\Sigma}_{X^*}$ can be zero (i.e., a normal random vector can be degenerate, unless otherwise stated). The matrix $\boldsymbol{\Sigma}_{X^*}$ is positive definite if $\boldsymbol{u}'\boldsymbol{\Sigma}_{X^*}\boldsymbol{u} > 0$ for all $\boldsymbol{u} \neq \boldsymbol{0}$, sometimes written as $\boldsymbol{\Sigma}_{X^*} \succ 0$. The log characteristic function is sometimes a polynomial in $\boldsymbol{u}$ of degree k denoted by $P_k(\boldsymbol{u})$, where P_0 is some generic constant.

Matrices and vectors are bolded. The concatenation of the matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ is denoted by $(\boldsymbol{A}_1, \boldsymbol{A}_2)$. The $p \times p$ identity matrix is denoted by $\boldsymbol{I}_p$, the standard basis vector that has a 1 in the tth coordinate and 0's elsewhere is denoted by $\boldsymbol{e}_t = (0, \ldots, 0, 1, 0, \ldots, 0)'$, and a matrix of 0's is denoted by $\boldsymbol{0}$. The dimensions of $\boldsymbol{e}_t$ and $\boldsymbol{0}$ are the ones needed to make sense of an expression. The euclidean norm of a vector $\boldsymbol{u}$ is $||\boldsymbol{u}|| = \sqrt{\boldsymbol{u}'\boldsymbol{u}}$ and the operator norm of a matrix $\boldsymbol{A}$ is $||\boldsymbol{A}|| = sup_{||\boldsymbol{u}||=1}||\boldsymbol{A}\boldsymbol{u}||$.

2. IDENTIFICATION

We provide necessary and sufficient conditions for identification of the coefficients in the classical linear errors-in-variables model, without any additional information such as instrumental variables or repeated measurements,

$$\begin{cases} \boldsymbol{X} &= \boldsymbol{X}^* + \boldsymbol{\eta}, \\ Y &= \boldsymbol{\beta}'\boldsymbol{X}^* + \varepsilon, \end{cases} \tag{1}$$

where $\boldsymbol{X}$, $\boldsymbol{X}^*$, and $\boldsymbol{\eta}$ are real p-dimensional random vectors and Y and ε are real scalar random variables. The observable outcome Y is a linear combination of the unobservable regressors $\boldsymbol{X}^*$ and the unobservable structural error ε while $\boldsymbol{X}$ contains observable measurements of $\boldsymbol{X}^*$ contaminated with unobservable measurement errors $\boldsymbol{\eta}$. The vector $\boldsymbol{\beta} \in \mathbb{R}^p$ contains unknown nonzero coefficients of interest. We assume that there does not exist a nonzero $\boldsymbol{b} \in \mathbb{R}^p$ such that $\boldsymbol{b}'\boldsymbol{X}^*$ is degenerate; otherwise there is multicollinearity and the model is not identified even without measurement errors.

We now introduce the distributional assumptions.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Assumption 2.1. Let $\boldsymbol{X}^*$ and $(\boldsymbol{\eta}, \varepsilon)$ be independent and either let

(i) $\eta_1, \ldots, \eta_p$, and ε be mutually independent or
(ii) $(\boldsymbol{\eta}, \varepsilon) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with positive definite covariance matrix $\boldsymbol{\Sigma}$.

The assumption states that the regressors $\boldsymbol{X}^*$ and the errors $(\boldsymbol{\eta}, \varepsilon)$ are independent and that either the errors are mutually independent or the errors are jointly normal with arbitrary covariance matrix. If the regressors and any of the errors are allowed to be dependent then the coefficients are not identified without additional information. If we assume that the regressors and the errors are independent but impose no further assumptions on the errors then the log characteristic function of the errors satisfies a partial differential equation with many functional solutions. The techniques in this paper identify the coefficients when the log characteristic function of the errors takes the form $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = P_a(\boldsymbol{s}) + \sum_{k=1}^{p+1} h_k(s_k)$, $\boldsymbol{s} \in \mathbb{R}^{p+1}$, for some polynomial P_a of finite degree a and continuous functions $h_1, \ldots, h_{p+1}$. In the case of mutually independent errors, $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = \varphi_\varepsilon(s_{p+1}) + \sum_{k=1}^{p} \varphi_{\eta_k}(s_k)$ and in the case of jointly normal errors, $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = P_2(\boldsymbol{s})$.

Next we define what it means for a random variable to be divisible by a nondegenerate normal distribution and discuss how this concept will be used to construct observationally equivalent models.

Definition 2.1. *A random variable is divisible by a nondegenerate normal distribution if its log characteristic function φ can be expressed as $\varphi(s) = \widetilde{\varphi}(s) - \frac{1}{2}\sigma^2 s^2$, for $s \in \mathbb{R}$ and $\sigma^2 > 0$, where $-\frac{1}{2}\sigma^2 s^2$ is the log characteristic function of a nondegenerate zero-mean normal distribution and $\widetilde{\varphi}$ is the log characteristic function of some other distribution. Equivalently, the random variable can be expressed as the sum of two independent random variables, one of which is nondegenerate normal.*

Suppose that ε is divisible by a nondegenerate normal distribution so that $\varepsilon = \widetilde{\varepsilon} + Z$ for some independent random variables $\widetilde{\varepsilon}$ and $Z \sim N(0, \sigma^2)$. Derivatives of the log characteristic function $\varphi_\varepsilon(s) = \varphi_{\widetilde{\varepsilon}}(s) - \frac{1}{2}\sigma^2 s^2$ evaluated at 0 relate the moments of ε and $\widetilde{\varepsilon}$ producing $E[\varepsilon^j] = E[\widetilde{\varepsilon}^j]$, for all $j \neq 2$, and $\mathrm{Var}(\widetilde{\varepsilon}) = \mathrm{Var}(\varepsilon) - \sigma^2$. Hence, ε is divisible by a nondegenerate normal distribution when there exists an $\widetilde{\varepsilon}$ that has the same moments as ε except for a smaller variance.

We will construct an observationally equivalent model by assuming that some unobservables are divisible by normal distributions and adding these normal distributions to other unobservables. Equivalently, as suggested in the previous paragraph, we will decrease the variances of some unobservables, increase the variances of other unobservables but leave the observable covariances unchanged. The assumption of divisibility by a nondegenerate normal distribution guarantees the validity of decreasing a variance of a random variable leaving it otherwise the same.

The following assumption is on the validity of transferring nondegenerate normal random variables between the regressors and errors.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Assumption 2.2. There exist $\boldsymbol{B} \in \mathbb{R}^{p\times(p-1)}$ and $\boldsymbol{b} \in \mathbb{R}^p$ such that,

(i) $\boldsymbol{b}'\boldsymbol{X}^*$ is nondegenerate normal,
(ii) $\boldsymbol{b}'\boldsymbol{X}^*$ and $\boldsymbol{B}'\boldsymbol{X}^*$ are independent,
(iii) $(\boldsymbol{b},\boldsymbol{B})$ is nonsingular,
(iv) if $\text{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*) < 0$, then ε is divisible by a nondegenerate normal distribution,
(v) if $\frac{\text{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*)}{b_k-\beta_k} < 0$, then η_k is divisible by a nondegenerate normal distribution, for $k = 1,\ldots,p$.

Assumptions 2.2(i)–(iii) state that the regressors can be decomposed into sums of independent components spanning a p dimensional space and one of which is normal. Assumptions 2.2(iv) and (v) state that if a covariance, depending on $\boldsymbol{b}$, $\boldsymbol{\beta}$, and $\boldsymbol{X}^*$, is negative then an error is divisible by a normal distribution.

The assumption is either satisfied for no values of $\boldsymbol{b}$ or infinitely many values of $\boldsymbol{b}$, some of which have arbitrarily small $||\boldsymbol{b}||$, because if $\boldsymbol{b}$ satisfies the assumption then so does $\delta\boldsymbol{b}$, for any $\delta > 0$. Equivalently, after defining $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{b}$, the assumption is either satisfied for no values of $\widetilde{\boldsymbol{\beta}}$ or infinitely many values of $\widetilde{\boldsymbol{\beta}}$, some of which are arbitrarily close to $\boldsymbol{\beta}$, because if $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{b}$ satisfies the assumption then so does $\breve{\boldsymbol{\beta}} = \delta\widetilde{\boldsymbol{\beta}} + (1-\delta)\boldsymbol{\beta} = \boldsymbol{\beta} - \delta\boldsymbol{b}$.

Most of the literature identifies the coefficients by assuming that no linear combinations of the regressors are normal or imposes rank conditions on third or higher moment equations (Assumption 2.2(i) does not hold). In the latter case, identification (and estimation) is based on the solution to a system of moment equations.

For any $\boldsymbol{b}$, we provide a $\boldsymbol{B}$ (with two restrictions on covariances) such that $(\boldsymbol{b},\boldsymbol{B})$ is nonsingular and $\boldsymbol{b}'\boldsymbol{X}^*$ and $\boldsymbol{B}'\boldsymbol{X}^*$ are uncorrelated.

LEMMA 2.1. *Suppose,*

$$\boldsymbol{B} = \begin{pmatrix} Cov(X_2^*,\boldsymbol{b}'\boldsymbol{X}^*) & \ldots & Cov(X_p^*,\boldsymbol{b}'\boldsymbol{X}^*) \\ -Cov(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -Cov(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*) \end{pmatrix}.$$

If $\sum_{k=1}^p b_k Cov(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*) \neq 0$ *and* $Cov(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*) \neq 0$, *then* $(\boldsymbol{b},\boldsymbol{B})$ *is nonsingular and* $Cov(\boldsymbol{b}'\boldsymbol{X}^*,\boldsymbol{B}'\boldsymbol{X}^*) = \boldsymbol{0}$.

Even if $\boldsymbol{b}'\boldsymbol{X}^*$ and $\boldsymbol{B}'\boldsymbol{X}^*$ are uncorrelated, they may still be dependent. Bekker (1986) provides the example of a two regressor model with $X_1^* \sim N(0,\sigma_1^2)$ and $X_2^* = (X_1^*)^2 - \sigma_1^2$. Then X_1^* is normal (Assumption 2.2(i) holds) but there does not exist a nonsingular matrix $(\boldsymbol{B},\boldsymbol{b})$ such that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ are independent (Assumption 2.2(ii) does not hold). Hence, the model is identified because the regressors cannot be decomposed into the sum of two independent components, one of which is normally distributed.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Now suppose that $\boldsymbol{X}^*$ is jointly normal and $(\boldsymbol{B},\boldsymbol{b})$ is nonsingular with $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ independent (Assumptions 2.2(i)–(iii) hold). Hong and Tamer (2003) show that the model is still identified if the errors are Laplace because they are not divisible by nondegenerate normal distributions (Assumptions 2.2(iv) and (v) fail).

To motivate the construction of an observationally equivalent model and to show how the covariances in Assumptions 2.2(iv) and (v) are obtained, the following lemma constructs a model with the same observable covariances as the underlying model in (1).

LEMMA 2.2. *Suppose that (1) holds and for simplicity assume that* $\boldsymbol{X}^* \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{X}^*})$. *Let* $\boldsymbol{b} \in \mathbb{R}^p$ *with* $||\boldsymbol{b}||$ *small and define* $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{b}$.

(i) *Suppose Assumptions 2.1(i), 2.2(iv), and 2.2(v) hold. Let* $Z_{1k} \sim N(0, |Cov(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k|)$, $Z'_{1k} \sim N(0, |Cov(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k|)$, *and* $Z_2 \sim N(0, |Cov(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)|)$. *If* $Cov(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k > 0$, *then define* $\widetilde{X}_k^* = X_k^* + Z_{1k}$ *and* $\eta_k = \widetilde{\eta}_k + Z'_{1k}$; *otherwise define* $X_k^* = \widetilde{X}_k^* + Z_{1k}$ *and* $\widetilde{\eta}_k = \eta_k + Z'_{1k}$. *If* $Cov(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*) < 0$, *then define* $\varepsilon = \widetilde{\varepsilon} + Z_2$; *otherwise define* $\widetilde{\varepsilon} = \varepsilon + Z_2$. *Then the observable covariances are,*

$$\begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}'_{X,Y} \\ \boldsymbol{\Sigma}_{X,Y} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{X}^*} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} & \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} + \sigma_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}}} & \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} \\ \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}} & \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}} + \sigma_{\widetilde{\varepsilon}}^2 \end{pmatrix}.$$

(ii) *Suppose Assumption 2.1(ii) holds. Define* $(\widetilde{\boldsymbol{\eta}}, \widetilde{\varepsilon}) \sim N(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}})$ *and* $\widetilde{\boldsymbol{X}}^* \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*})$, *where* $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Omega}$, $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} = \boldsymbol{\Sigma}_{\boldsymbol{X}^*} - \boldsymbol{\Omega}_{\boldsymbol{\eta}}$, *and* $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{\eta}} & \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon} & \Omega_\varepsilon \end{pmatrix}$ *satisfies* $\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}} = \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} - \boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}$ *and* $\Omega_\varepsilon = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} + \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}$. *Then the observable covariances are,*

$$\begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}'_{X,Y} \\ \boldsymbol{\Sigma}_{X,Y} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{X}^*} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\beta}\boldsymbol{\Sigma}_{\boldsymbol{X}^*} + \boldsymbol{\Sigma}'_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} & \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} + \sigma_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}}} & \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} + \boldsymbol{\Sigma}'_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}} \\ \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}} & \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}} + \sigma_{\widetilde{\varepsilon}}^2 \end{pmatrix}.$$

Lemma 2.2 constructs infinitely many observationally equivalent models, one for each choice of $\boldsymbol{b} \in \mathbb{R}^p$ as long as $\widetilde{\boldsymbol{\Sigma}} \succ 0$ and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} \succ 0$ (see Klepper and Leamer, 1984). If $||\boldsymbol{b}||$ is small enough (i.e., $\widetilde{\boldsymbol{\beta}}$ is close enough to $\boldsymbol{\beta}$) then $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} \succ 0$ and $\widetilde{\boldsymbol{\Sigma}} \succ 0$ because $\boldsymbol{\Sigma}_{\boldsymbol{X}^*} \succ 0$, $\boldsymbol{\Sigma} \succ 0$, and $||\boldsymbol{\Omega}|| \leq ||\boldsymbol{b}|| \cdot P_0$. For example, $\widetilde{\boldsymbol{\Sigma}} \succ 0$ implies that $\sigma_{\widetilde{\varepsilon}}^2 = \sigma_\varepsilon^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} + \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} > 0$. If σ_ε^2 is small then this can restrict $||\boldsymbol{b}||$ (and $||\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}||$) to be small.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

The construction in part (i) shifts the variance of ε by $|\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)|$ and the variance of η_k by $|\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k|$, and compensates for these shifts with opposing shifts in the variance of X_k^*. Depending on the direction of a shift, we either add a nondegenerate normal distribution or assume divisibility by a nondegenerate normal distribution by Assumptions 2.2(iv) and (v). The construction in part (ii) in the case $\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} = \mathbf{0}$ is the same as part (i) because $\Omega_\varepsilon = \mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$, $\Omega_{\eta_j,\eta_k} = -\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k$ if $k = j$, and $\Omega_{\eta_j,\eta_k} = 0$ if $k \neq j$. The errors are jointly normal by Assumption 2.1(ii) and hence automatically divisible by normal distributions.

Constructing models that are observationally equivalent to (1) will be analogous to the constructions in Lemma 2.2. The difference is that instead of adding and subtracting variances, we add and subtract quadratic terms to log characteristic functions.

The following theorem establishes identification of the errors-in-variables model with either mutually independent errors or with jointly normal errors, extending Reiersøl (1950) from the single regressor model to the multiple regressor model.

THEOREM 2.1.

(i) *Suppose that (1) and Assumption 2.1(i) hold. Assume that* $E[e^{\lambda|\varepsilon|}] < \infty$ *and* $E[e^{\lambda|\eta_k|}] < \infty$, $k = 1, \ldots, p$, *for some* $\lambda > 0$. *The vector* $\boldsymbol{\beta}$ *is not identified if and only if Assumptions 2.2(i)–(v) hold.*

(ii) *Suppose that (1) and Assumption 2.1(ii) hold. The vector* $\boldsymbol{\beta}$ *is not identified if and only if Assumptions 2.2(i)–(iii) hold.*

We provide a broad outline of the proof leaving the details to Section A.3 of the Appendix. The log characteristic function of (1) is,

$$\varphi_{\boldsymbol{X},Y}(\boldsymbol{s}) = \varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) + \varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{R}^{p+1}, \tag{2}$$

where $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = \varphi_\varepsilon(s_{p+1}) + \sum_{k=1}^p \varphi_{\eta_k}(s_k)$ is separable when the errors are mutually independent (Assumption 2.1(i) holds) or $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = -\frac{1}{2}\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}$ is quadratic when the errors are jointly normal (Assumption 2.1(ii) holds). Sufficient conditions for identification are then obtained by strategically choosing the arguments $\boldsymbol{s}$ (which represent linear combinations of $\boldsymbol{Y}$) and taking derivatives, thus eliminating $\varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s})$ from the equation. To prove necessity of these conditions, we construct an observationally equivalent model by using Assumptions 2.2(i)–(iii) to decompose $\boldsymbol{X}^*$ into the sum of two independent components, one of which is normal, and then transferring the normal component from the regressors to the errors by Assumptions 2.2(iv) and (v) or Assumption 2.1(ii).

The sufficiency part of the proof analyzes characteristic functions in neighborhoods of the origin (no moments need to exist). The necessity part of the proof, however, constructs an observationally equivalent model by defining log characteristic functions on their domains. The necessary and sufficient conditions

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

coincide because the errors (but not necessarily the regressors) have analytic characteristic functions so that relationships can be extended from neighborhoods of the origin to domains of functions by analytic continuation.[3,4]

Theorem 2.1(i) in the case of a single regressor does not need the assumption that the errors have analytic characteristic functions (Reiersøl, 1950).

Corollary 2.1. *Suppose that $X = X^* + \eta$ and $Y = \beta X^* + \varepsilon$. Assume that X^*, ε, and η are mutually independent. The coefficient β is not identified if and only if (i) X^* is nondegenerate normal and (ii) either η or ε is divisible by a nondegenerate normal distribution.*

The main difference between Theorem 2.1(ii) and Bekker (1986) is that he assumes $\boldsymbol{\eta}$ and ε are independent while we allow $\boldsymbol{\eta}$ and ε to be arbitrarily dependent. To show necessity, Bekker (1986) sets up a system of equations with more parameters than equations but does not explicitly provide multiple solutions to the system while we explicitly construct an observationally equivalent model.

3. ESTIMATION

In this section, we mimic the steps from our identification proof to estimate $\boldsymbol{\beta}$ in the errors-in-variables model (1). The estimator uses a continuum of moment conditions based on second derivatives of the log characteristic function of the observables, which has the potential to be more efficient than estimators based on third and higher moments because the characteristic function completely characterizes the distribution of the underlying random variables. The estimator is root-n consistent and asymptotically normal.

Consider the following second derivatives of the log characteristic function (2) evaluated at $\boldsymbol{s} = (\boldsymbol{b}, -1)v$, for $\boldsymbol{b} \in \mathcal{B} \subset \mathbb{R}^p$, less the derivatives evaluated at $\boldsymbol{s} = \boldsymbol{0}$,

$$\begin{aligned}
&\operatorname{Cov}(X_{k_1}, Y) + \frac{\partial^2 \varphi_{X,Y}(\boldsymbol{s})}{\partial s_{k_1} \partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v} \\
&\quad = \sum_{k=1}^{p} \beta_k \left(\operatorname{Cov}\left(X_{k_1}^*, X_k^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_k}\bigg|_{\boldsymbol{u}=(\boldsymbol{b}-\boldsymbol{\beta})v} \right), \quad 1 \le k_1 \le p, \\
&\operatorname{Cov}(X_{k_1}, X_{k_2}) + \frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_{k_1} \partial s_{k_2}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v} \\
&\quad = \operatorname{Cov}\left(X_{k_1}^*, X_{k_2}^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}}\bigg|_{\boldsymbol{u}=(\boldsymbol{b}-\boldsymbol{\beta})v}, \quad 1 \le k_1 < k_2 \le p.
\end{aligned}$$

[3] The assumption that the errors are analytic ($E[e^{\lambda|\varepsilon|}] < \infty$ and $E[e^{\lambda|\eta_k|}] < \infty$) can be replaced by the assumption that the errors are subgaussian or bounded.

[4] Székely and Rao (2000) provide examples where relationships in a neighborhood of the origin do not extend to domains of nonanalytic characteristic functions.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

The left sides of the equations are expressed in terms of covariances and second derivatives of the log characteristic function of the observables $(\boldsymbol{X}, Y)$ and the arguments $\boldsymbol{b}$. The right sides of the equations are expressed in terms of covariances and second derivatives of the log characteristic function of the unobservables $\boldsymbol{X}^*$, the unknown coefficients $\boldsymbol{\beta}$, and the arguments $\boldsymbol{b}$. We can interpret the above expressions as measures of nongaussianity since they are all identically 0 for any value of $\boldsymbol{b}$ for which $(\boldsymbol{b}-\boldsymbol{\beta})'\boldsymbol{X}^*$ is normal.

The estimator is now based on the following moments,

$$\boldsymbol{H}(\boldsymbol{b}, v) = \left(\operatorname{Cov}(X_1, X_2) + \frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_1 \partial s_2}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v}, \ldots, \operatorname{Cov}(X_{p-1}, X_p) + \frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_{p-1} \partial s_p}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\right.$$
$$\left.\operatorname{Cov}(X_1, Y) + \frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_1 \partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v}, \ldots, \operatorname{Cov}(X_p, Y) + \frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_p \partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v}\right)', \quad \textbf{(3)}$$

where $\boldsymbol{H}(\boldsymbol{b}, v)'$ is the conjugate transpose of $\boldsymbol{H}(\boldsymbol{b}, v) \in \mathbb{C}^{p(p+1)/2 \times 1}$.

We impose the identifying assumption that no linear combinations of $\boldsymbol{X}^*$ are normal (i.e., Assumption 2.2(i) does not hold).

Assumption 3.1. There does not exist a nonzero $\boldsymbol{b} \in \mathcal{B}$ such that $\boldsymbol{b}'\boldsymbol{X}^*$ is normal.

If Assumption 3.1 holds then $\boldsymbol{\beta}$ is identified from the moments $\boldsymbol{H}(\boldsymbol{b}, v) = \boldsymbol{0}$ (i.e., $\boldsymbol{H}(\boldsymbol{b}, v) = \boldsymbol{0}$ if and only if $\boldsymbol{b} = \boldsymbol{\beta}$). Variation in $\frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_{k_1} \partial s_{k_2}}$ implies that $(\boldsymbol{X}, Y)$ is not jointly normal and is a necessary condition for the assumption to hold.

Now to estimate the coefficients using the moments $\boldsymbol{H}(\boldsymbol{b}, v)$, we can fix v and choose how to weight each of the moments based on minimizing a norm. Then, we can choose how to weight $v \in \mathbb{R}$ based on minimizing another norm. Carrasco and Florens (2000) and Carrasco and Kotchoni (2017) analyze the latter of these choices taking the former as given. They show that the optimal weight on a continuum of moment conditions is the inverse of an asymptotic covariance operator in Hilbert space. This inverse is discontinuous and so its estimation requires some regularization. In this setting, only a few papers analyze which regularization technique to use and how to choose regularization parameters to estimate the optimal weights. Further, estimates are usually sensitive to the regularization technique and the choice of regularization parameters (which can be data intensive). We opt for the simpler IGMM estimator, which uses the optimal GMM weight matrix for each v and exponentially decaying weight on v,

$$\widehat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{b} \in \mathcal{B}} \int_{-\infty}^{\infty} \boldsymbol{H}_n(\boldsymbol{b}, v)' \boldsymbol{W}_n(v) \boldsymbol{H}_n(\boldsymbol{b}, v) \pi(v) dv, \quad \textbf{(4)}$$

where $\pi(v) = e^{-v^2/2}/\sqrt{2\pi}$ is the density of a standard normal distribution, $\boldsymbol{H}_n(\boldsymbol{b}, v)$ is the natural sample analog of $\boldsymbol{H}(\boldsymbol{b}, v)$, $\boldsymbol{W}_n(v)$ is a random positive definite matrix approaching the optimal GMM matrix weight $\boldsymbol{\Omega}(v, v)^{-1}$, where $\sqrt{n}\boldsymbol{H}_n(\boldsymbol{\beta}, v) \xrightarrow{p} N(\boldsymbol{0}, \boldsymbol{\Omega}(v, v))$ and $\boldsymbol{H}_n(\boldsymbol{b}, v)'\boldsymbol{W}_n(v)\boldsymbol{H}_n(\boldsymbol{b}, v)$ is the optimal GMM norm (see

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Section A.5 of the Appendix for further details). Exponentially decaying weights reflect that the characteristic function is more accurately estimated closer to the origin and contains more information about the underlying distribution closer to the origin. They are commonly used in estimation of parameters using characteristic functions (see Yu, 2004).

Next we assume random sampling and impose regularity conditions.

Assumption 3.2. We observe a sample $\{\boldsymbol{X}_i, Y_i\}_{i=1}^n$ of independently and identically distributed draws from the distribution of $(\boldsymbol{X}, Y)$.

Assumption 3.3.

(i) $\boldsymbol{\beta}$ belongs to the interior of the compact set $\mathcal{B}$,
(ii) $E[e^{\lambda|Y|}] < \infty$ and $E[e^{\lambda|X_k|}] < \infty$, $k = 1, \ldots, p$, for some $\lambda > 0$,
(iii) $E[e^{iv(\boldsymbol{b}'\boldsymbol{X}-Y)}] \neq 0$ for all $\boldsymbol{b} \in \mathcal{B}$ and v almost everywhere,
(iv) $\boldsymbol{W}_n(v) \xrightarrow{p} \boldsymbol{W}(v)$ almost everywhere for $\boldsymbol{W}(v)$ positive definite,
(v) $\boldsymbol{H}(\boldsymbol{b}, v)'\boldsymbol{W}(v)\boldsymbol{H}(\boldsymbol{b}, v) \in L^2(\pi)$ and all elements of $\frac{\partial \boldsymbol{H}(\boldsymbol{b}, v)}{\partial \boldsymbol{b}}\boldsymbol{W}(v)\boldsymbol{H}(\boldsymbol{b}, v) \in L^2(\pi)$, for all $\boldsymbol{b} \in \mathcal{B}$, where $L^2(\pi) = \{g : \mathbb{R} \to \mathbb{R} \mid \int g(v)g(v)'\pi(v)dv < \infty\}$, and
(vi) $\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta}, v)\boldsymbol{W}(v)\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta}, v)\pi(v)dv$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta}, v)\boldsymbol{W}(v)\boldsymbol{\Omega}(v, w)\boldsymbol{W}(w)\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta}, w)\pi(v)\pi(w)dvdw$ are nonsingular and finite.

Using Assumption 3.1 to show identification, a random sample by Assumption 3.2, and Assumption 3.3's regularity conditions, the following theorem establishes that $\widehat{\boldsymbol{\beta}}$ is root-n asymptotically normal.

THEOREM 3.1. *Suppose that (1) holds, either Assumption 2.1(i) holds or Assumption 2.1(ii) holds, and Assumptions 3.1, 3.2, and 3.3 with* $\boldsymbol{W}(v) = (\boldsymbol{\Omega}(v, v))^{-1}$ *hold. Then,*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{S}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta}, v)\boldsymbol{\Omega}^{-1}(v, v)\boldsymbol{\Omega}(v, w)\boldsymbol{\Omega}^{-1}(w, w) \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta}, w)\pi(v)\pi(w)dvdw\right)\boldsymbol{S}'\right),$$

where $\boldsymbol{S} = \left(\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta}, v)\boldsymbol{\Omega}^{-1}(v, v)\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta}, v)\pi(v)dv\right)^{-1}$ *and* $\boldsymbol{\Omega}(v, w)$ *is defined in Section A.5 of the appendix.*

4. MONTE CARLO SIMULATIONS

In this section, we compare the finite sample performance of the IGMM estimator to the OLS estimator and a GMM estimator based on third to fifth moments allowing for various amounts of measurement error and for some mismeasured and perfectly measured regressors. The IGMM and GMM estimators produce results that have small bias and decreasing standard deviation at root-n rate.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

These estimators are robust to measurement error and the number of mismeasured regressors. When there is measurement error, the OLS estimator produces results that are severely biased with zero coverage probability of the 95% confidence interval.

The data $\{\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i\}_{i=1}^n$, for $n \in \{500, 1,000, 2,000\}$, are simulated 100 times from the errors-in-variables model,

$$\begin{cases} \boldsymbol{X} &= \boldsymbol{X}^* + \boldsymbol{\eta}, \\ Y &= \boldsymbol{\beta}'\boldsymbol{X}^* + \boldsymbol{\gamma}'\boldsymbol{Z} + \varepsilon, \end{cases} \tag{5}$$

where $(\boldsymbol{X}^*, \boldsymbol{Z})$, $\boldsymbol{\eta}$, and ε are mutually independent. The vector $\boldsymbol{X}^*$ contains unobserved mismeasured regressors and the vector $\boldsymbol{Z}$ contains perfectly measured observed regressors. The regressors $(\boldsymbol{X}^*, \boldsymbol{Z})$ are multidimensional χ^2 distributed with 10 degrees of freedom and covariance matrix adjusted to have diagonal elements (variances) equal to 3 and off-diagonal elements (covariances) equal to 1, the structural error ε is standard normal, the measurement errors $\boldsymbol{\eta}$ are either identically zero (no measurement errors), independent normal with variance 1 (independent measurement errors), or jointly normal with covariance matrix having diagonal elements equal to 1 and off-diagonal elements equal to 1/2 (dependent measurement errors).

The parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are all equal to 1. Denote the dimension of $\boldsymbol{X}$ (and $\boldsymbol{\beta}$) by N_x and the dimension of $\boldsymbol{Z}$ (and $\boldsymbol{\gamma}$) by N_z, and let $N_x = 1, 2, 3$ and $N_x + N_z = 4$. The estimators all use the preliminary step of partialling out the perfectly measured regressors $\boldsymbol{Z}$. We compare three estimators that use the residuals of the regressions of Y on $\boldsymbol{Z}$ and $\boldsymbol{X}$ on $\boldsymbol{Z}$:

(a) the IGMM estimator in (4) with starting values for the optimization coming from the GMM estimates,[5]
(b) the GMM estimator based on third, fourth, and fifth moments (see Erickson, Parham, and Whited, 2017),[6] and
(c) the OLS estimator that ignores measurement error.

The integral for the IGMM estimator is estimated by $\sum_{j=1}^{50} \boldsymbol{H}_n(\boldsymbol{b}, v_j)'\boldsymbol{W}_n(v_j)\boldsymbol{H}_n(\boldsymbol{b}, v_j)$, where $v_1, \ldots, v_{50}$ are 50 independent draws from a standard normal distribution, $\boldsymbol{H}_n(\boldsymbol{b}, v_j)$ is the sample analog of $\boldsymbol{H}(\boldsymbol{b}, v_j)$ in (3), and $\boldsymbol{W}_n(v_j)$ is the sample analog of $(\boldsymbol{\Omega}(v_j, v_j))^{-1}$ in (A.21) with the unknown $\boldsymbol{\beta}$ replaced by the GMM estimator.[7] The weights $\pi(v) = e^{-v^2/2}/\sqrt{2\pi}$ decay rapidly and are not data-driven. The estimates are sensitive to these weights and we view the choice of the weights as similar to the choice of bandwidth in other nonparametric settings like deconvolution. By allowing the weights to depend on the data, the estimator could take into account how information decays as $|v|$ grows and how well the

[5]The estimator produces similar results when the starting values for the optimization are the GMM estimates, the OLS estimates or the underlying coefficient values.

[6]The estimates use the STATA function "xtewreg" available on the authors' websites.

[7]The results are similar when using the identity matrix as the weight matrix.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

characteristic function is estimated as $|v|$ grows. Incorporating these weights into the estimator as in Carrasco and Florens (2000) has the potential to improve the IGMM estimator even further (see also the discussion surrounding (4)).

We plot the sample analogs of $\frac{\partial^2 \varphi_{\boldsymbol{X},Y}(s)}{\partial s_{k_1} \partial s_{k_2}}$ to confirm variation in these second derivatives as suggestive evidence that no linear combinations of the unobserved regressors are normal.

Using 100 simulations each of sample size $n \in \{500, 1,000, 2,000\}$, Tables 1 and 2 report the bias, standard deviation (std), and coverage probability (cp) of the 95% confidence interval of the estimates for $\beta_1 = 1$ in Model (6). The confidence intervals use 50 bootstrapped samples to estimate standard errors. Table 1 displays the results when $N_x = N_z = 2$ and for various amounts of measurement errors and Table 2 displays the results for a varying number of mismeasured and perfectly measured regressors.

When there is no measurement error, the OLS estimator is unbiased, has smallest standard deviations, and has coverage probabilities that are close to the nominal level of 95%. However, when there is measurement error, then the OLS estimator has large bias that does not improve with sample size. The standard deviations are deceptively small and the coverage probabilities are zero. The OLS estimator has incorrect rejection rates and coverage probabilities and inference based on this estimator will not be trustworthy.

The IGMM and GMM estimators perform well in all the simulations. They have small bias and standard deviations that decrease at the root-n convergence rate in Theorem 3.1. The coverage probability is close to the nominal level for

Table 1. Performances of the IGMM, GMM, and OLS estimators with $N_x = N_z = 2$ and various amounts of measurement errors

		$(\sigma^2_\eta,\sigma_{\eta_1,\eta_2})$=(0,0)			$(\sigma^2_\eta,\sigma_{\eta_1,\eta_2})$=(1,0)			$(\sigma^2_\eta,\sigma_{\eta_1,\eta_2})$=(1,1/2)		
n		IGMM	GMM	OLS	IGMM	GMM	OLS	IGMM	GMM	OLS
500	bias	0.02	0.01	0.00	−0.06	−0.10	−0.25	−0.11	−0.15	−0.33
	std	0.10	0.09	0.03	0.15	0.16	0.03	0.16	0.19	0.04
	cp	0.90	0.72	0.92	0.94	0.64	0.00	0.92	0.59	0.00
1,000	bias	0.01	−0.01	0.00	−0.05	−0.05	−0.25	−0.08	−0.08	−0.33
	std	0.07	0.06	0.02	0.13	0.13	0.03	0.13	0.14	0.03
	cp	0.95	0.83	0.97	0.88	0.73	0.00	0.89	0.68	0.00
2,000	bias	0.00	0.00	0.00	−0.04	−0.03	−0.25	−0.06	−0.05	−0.33
	std	0.05	0.04	0.02	0.09	0.09	0.02	0.09	0.09	0.02
	cp	0.95	0.84	0.93	0.91	0.77	0.00	0.90	0.79	0.00

Note: Bias, standard deviation, and coverage probability (cp) of the 95% confidence interval are estimated from 100 simulations each with sample size n. The confidence intervals use 50 bootstrapped samples to estimate standard errors. The measurement errors are jointly normal with covariance matrix having diagonal elements σ^2_η and off-diagonal elements σ_{η_1,η_2}.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

TABLE 2. Performances of the IGMM, GMM, and OLS estimators with a varying number of mismeasured regressors N_x and perfectly measured regressors N_z

		$(N_x, N_z) = (1, 3)$			$(N_x, N_z) = (2, 2)$			$(N_x, N_z) = (3, 1)$		
n		IGMM	GMM	OLS	IGMM	GMM	OLS	IGMM	GMM	OLS
500	bias	−0.03	−0.06	−0.29	−0.06	−0.10	−0.25	−0.08	−0.11	−0.20
	std	0.18	0.15	0.03	0.15	0.16	0.03	0.14	0.16	0.04
	cp	0.95	0.77	0.00	0.94	0.64	0.00	0.90	0.43	0.00
1,000	bias	−0.03	−0.02	−0.29	−0.05	−0.05	−0.25	−0.07	−0.07	−0.20
	std	0.15	0.11	0.02	0.13	0.13	0.03	0.13	0.12	0.03
	cp	0.90	0.81	0.00	0.88	0.73	0.00	0.88	0.59	0.00
2,000	bias	−0.02	0.01	−0.29	−0.04	−0.03	−0.25	−0.05	−0.06	−0.20
	std	0.11	0.06	0.01	0.09	0.09	0.02	0.09	0.08	0.20
	cp	0.92	0.90	0.00	0.91	0.77	0.00	0.89	0.74	0.00

Note: Bias, standard deviation, and coverage probability (cp) of the 95% confidence interval are estimated from 100 simulations each with sample size n. The confidence intervals use 50 bootstrapped samples to estimate standard errors. The measurement errors are independent standard normals.

the IGMM estimator but sometimes far away for the GMM estimator, suggesting that inference using the GMM estimator might not be trustworthy. Table 1 shows that the bias and standard deviation increase with measurement error but remain relatively stable. Table 2 shows that the number of mismeasured regressor increases bias by a little but has a negligible effect on the standard deviation. We experimented with different distributional choices and covariance matrices for the regressors and errors, producing results that were qualitatively similar to the ones presented in Tables 1 and 2.

5. APPLICATION

Investment is a large component of a country's GDP (e.g., investment accounts for about 15% of U.S. GDP) and hence is central to our understanding of economic activity (see, e.g., Abel, 1990). Despite an elegant theory for firm investment decisions and extensive empirical research, some key variables and/or mechanisms that drive these investment decisions have not been identified.

The neoclassical theory of firm investment under convex adjustment costs implies that firm investment should be a function only of marginal returns to capital. The intuition is that a firm should increase capital when its value is greater than its cost and decrease capital when its value is less than its cost (Keynes, 1936).

Consider the linear regression model,

$$Y = \beta q^* + \varepsilon, \tag{6}$$

where Y is firm investment, q^* is marginal returns to capital, and ε is an error.

Empirical studies that regress investment on observed measures of marginal returns to capital explain little about investment decisions. This rejection of the

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

theory has been interpreted as a sign that financial constraints should be included in the model,

$$Y = \beta_1 q^* + \beta_2 FC^* + \varepsilon, \tag{7}$$

where FC^* is financial constraints.

Another common explanation for the theory's failure is that marginal returns to capital are measured with error,

$$q = q^* + \eta_q, \tag{8}$$

where q is the ratio of market to book value of equity plus liabilities, and η_q is measurement error.

We investigate the possibility that financial constraints are measured with error,

$$FC = FC^* + \eta_{FC}, \tag{9}$$

where FC is cash flows, often used as a measure for financial constraints (see, e.g., Cao, Lorenzi, and Walentin, 2019; Gilchrist and Himmelberg, 1995), and η_{FC} is measurement error.

We use data from Erickson et al. (2014) for the years 1992–1995.[8] There is a lot of variation in the second derivatives of the empirical log characteristic function. This is strong evidence that the observables are nongaussian and is a necessary condition for no linear combinations of the unobserved regressors to be normal (and hence identification of the coefficients in the models).[9]

Table 3 compares the IGMM, GMM, and OLS estimators in three models, with standard errors of the IGMM estimator calculated from 100 bootstrapped samples. Model 1 is based on (6) and (8) and excludes a variable for financial constraints, Model 2 is based on (7) and (8) and assumes that financial constraints are perfectly measured, and Model 3 is based on (7), (8), and (9) and allows both marginal returns to capital and financial constraints to have measurement errors.

Ignoring measurement errors, the OLS estimates of the effects of marginal returns to capital and financial constraints on investment decisions are always positive and significant at the 5% level. This supports the majority of the literature that both marginal returns to capital and financial constraints are important determinants of investment decisions.

In Model 1, which excludes financial constraints, the IGMM and GMM estimates are similar, and agree that the effects of marginal returns to capital on investment decisions are positive and significant. These estimates are larger than the OLS estimates, which suffers from attenuation bias.

In Model 2, which assumes that financial constraints are perfectly measured, the IGMM and GMM estimates of the coefficients on marginal returns to capital are all positive, significant, and similar. The estimates of the coefficients on financial

[8] Erickson et al. (2014) collect the data from the 2012 Compustat Industrial Files. There are 2,850, 3,031, 3,279, and 3,425 observations in 1992, 1993, 1994, and 1995, respectively.

[9] Erickson et al. (2014) provide evidence for nongaussianity by showing that the observed data are highly skewed.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Table 3. Estimates of firm investment decisions based on marginal returns to capital and financial constraints

	β_1 (returns to capital)			β_2 (financial constraints)		
Year	OLS	GMM	IGMM	OLS	GMM	IGMM
Model 1: $Y = \beta_1 q^* + \varepsilon$, $q = q^* + \eta_q$						
1992	0.014**	0.015**	0.029**		—	
	(0.001)	(0.002)	(0.008)			
1993	0.016**	0.029**	0.030**		—	
	(0.001)	(0.008)	(0.006)			
1994	0.017**	0.036**	0.032**		—	
	(0.002)	(0.012)	(0.009)			
1995	0.017**	0.075*	0.083**		—	
	(0.002)	(0.037)	(0.023)			
Model 2: $Y = \beta_1 q^* + \beta_2 FC^* + \varepsilon$, $q = q^* + \eta_q$						
1992	0.012**	0.017**	0.029**	0.077**	0.056	−0.001
	(0.002)	(0.007)	(0.005)	(0.019)	(0.043)	(0.032)
1993	0.013**	0.039**	0.034**	0.064**	−0.047	−0.026
	(0.002)	(0.008)	(0.008)	(0.012)	(0.037)	(0.031)
1994	0.012**	0.040**	0.051**	0.111**	−0.017	−0.070
	(0.002)	(0.017)	(0.015)	(0.016)	(0.078)	(0.077)
1995	0.014**	0.055**	0.120**	0.106**	−0.023	−0.215
	(0.002)	(0.025)	(0.032)	(0.016)	(0.082)	(0.118)
Model 3: $Y = \beta_1 q^* + \beta_2 FC^* + \varepsilon$, $q = q^* + \eta_q$, $FC = FC^* + \eta_{FC}$						
1992	0.012**	0.011**	0.017**	0.077**	−0.119**	0.082
	(0.002)	(0.003)	(0.006)	(0.019)	(0.060)	(0.147)
1993	0.013**	0.008	0.027**	0.064**	0.029	−0.003
	(0.002)	(0.009)	(0.010)	(0.012)	(0.141)	(0.167)
1994	0.012**	0.003	−0.008	0.111**	0.138	0.629**
	(0.002)	(0.007)	(0.012)	(0.016)	(0.142)	(0.200)
1995	0.014**	0.004	0.010	0.106**	0.061	0.386
	(0.002)	(0.008)	(0.025)	(0.016)	(0.129)	(0.236)

Note: Standard errors are in parentheses under the estimate. The IGMM standard errors use 100 bootstrapped samples. Double asterisks indicate significance at the 5% level.

constraints are all insignificant. Hence, allowing for measurement error in marginal returns to capital restores the validity of the neoclassical theory that returns to capital, and not financial constraints, are the main drivers of investment decisions (see Erickson and Whited, 2000).

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

In Model 3, which allows for measurement errors in both marginal returns to capital and financial constraints, the IGMM and GMM estimates of the coefficients are mostly insignificant. This suggests that the theory is missing an important component, or that the linear regression model allowing for measurement errors in both the regressors is inappropriate due to nonlinearities, or that the regressors and unobservables are statistically dependent. One possible direction to pursue is a nonlinear model that allows for measurement errors in both marginal returns to capital and financial constraints (see, e.g., Schennach and Hu, 2013).

6. CONCLUSION

We provide necessary and sufficient conditions for identification of the coefficients in the linear errors-in-variables model. This extends Reiersøl (1950) from a linear errors-in-variables model with a single mismeasured regressor to multiple mismeasured regressors. We mimic the identification procedure to produce a new estimator for the coefficients using a continuum of moments based on second derivatives of the log characteristic function of observables. Monte Carlo simulations provide evidence that the estimator has potential to perform well in practice and could be a competitive alternative to already available estimators. We apply the estimator to data on firm investment decisions and show that the relationship between investment and the regressors, marginal returns to capital and financial constraints, depends on whether or not the regressors contain measurement errors.

REFERENCES

Abel, A.B. (1983) Optimal investment under uncertainty. *The American Economic Review* 73(1), 228–233.

Abel, A.B. (1990) Chapter 14 consumption and investment. In *Handbook of Monetary Economics*, Vol. 2, pp. 725–778. Elsevier.

Abel, A.B. & J.C. Eberly (1994) A unified model of investment under uncertainty. *The American Economic Review* 84(5), 1369–1384.

Aigner, D.J., C. Hsiao, A. Kapteyn, & T. Wansbeek (1984) Latent variable models in econometrics. *Handbook of Econometrics* 2, 1321–1393.

Bekker, P.A. (1986) Comment on identification in the linear errors in variables model. *Econometrica* 54(1), 215–217.

Ben-Moshe, D., X. D'Haultfœuille, & A. Lewbel (2017) Identification of additive and polynomial models of mismeasured regressors without instruments. *Journal of Econometrics* 200(2), 207–222.

Bickel, P. & Y. Ritov (1987) Efficient estimation in the errors in variables model. *Annals of Statistics* 15, 513–540.

Blanchard, O.J., F. Lopez-de Silanes, & A. Shleifer (1994) What do firms do with cash windfalls? *Journal of Financial Economics* 36(3), 337–360.

Brainard, W.C. & J. Tobin (1968) Pitfalls in financial model building. *The American Economic Review* 58(2), 99–122.

Buonaccorsi, J.P. (2010) *Measurement Error: Models, Methods, and Applications*. Chapman & Hall/CRC Interdisciplinary Statistics.

Cao, D., G. Lorenzi, & K. Walentin (2019) Financial frictions, investment, and Tobin's *q*. *Journal of Monetary Economics* 103, 105–122.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Carrasco, M. & J.-P. Florens (2000) Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16(6), 797–834.
Carrasco, M. & R. Kotchoni (2017) Efficient estimation using the characteristic function. *Econometric Theory* 33(2), 479–526.
Carroll, R.J., D. Ruppert, L.A. Stefanski, & C.M. Crainiceanu (2006) *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press.
Chen, X., H. Hong, & D. Nekipelov (2011) Nonlinear models of measurement errors. *Journal of Economic Literature* 49(4), 901–937.
Cheng, C.-L. & J.W. Van Ness (1999) *Statistical Regression with Measurement Error*, Vol. 6. Arnold.
Chesher, A. (1991) The effect of measurement error. *Biometrika* 78(3), 451–462.
Cragg, J.G. (1997) Using higher moments to estimate the simple errors-in-variables model. *Rand Journal of Economics* 28, S71–S91.
Dagenais, M.G. & D.L. Dagenais (1997) Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics* 76, 193–221.
Erickson, T., C.H. Jiang, & T.M. Whited (2014) Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *Journal of Econometrics* 183(2), 211–221.
Erickson, T., R. Parham, & T.M. Whited (2017) Fitting the errors-in-variables model using high-order cumulants and moments. *Stata Journal* 17(1), 116–129.
Erickson, T. & T.M. Whited (2000) Measurement error and the relationship between investment and *q*. *Journal of Political Economy* 108(5), 1027–1057.
Erickson, T. & T.M. Whited (2002) Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory* 18(3), 776–799.
Fazzari, S., R.G. Hubbard, & B. Petersen (1988) Investment, financing decisions, and tax policy. *The American Economic Review* 78(2), 200–205.
Fuller, W.A. (2009) *Measurement Error Models*, Vol. 305. Wiley.
Geary, R.C. (1941) Inherent relations between random variables. In *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, pp. 63–76.
Gilchrist, S. & C.P. Himmelberg (1995) Evidence on the role of cash flow for investment. *Journal of Monetary Economics* 36(3), 541–572.
Gillard, J. (2010) An overview of linear structural models in errors in variables regression. *REVSTAT–Statistical Journal* 8(1), 57–80.
Hausman, J.A., W.K. Newey, H. Ichimura, & J.L. Powell (1991) Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics* 50, 273–295.
Hayashi, F. (1982) Tobin's marginal *q* and average *q*: A neoclassical interpretation. *Econometrica* 50, 213–224.
Hong, H. & E. Tamer (2003) A simple estimator for nonlinear error in variable models. *Journal of Econometrics* 117(1), 1–19.
Hu, Y. & Y. Sasaki (2015) Closed-form estimation of nonparametric models with non-classical measurement errors. *Journal of Econometrics* 185(2), 392–408.
Kaldor, N. (1966) Marginal productivity and the macro-economic theories of distribution: Comment on Samuelson and Modigliani. *The Review of Economic Studies* 33(4), 309–319.
Kapteyn, A. & T. Wansbeek (1983) Identification in the linear errors in variables model. *Econometrica* 51, 1847–1849.
Keynes, J.M. (1936) The general theory of money, interest and employment. *Reprinted in The Collected Writings of John Maynard Keynes*, p. 7.
Khatri, C. & C.R. Rao (1972) Functional equations and characterization of probability laws through linear functions of random variables. *Journal of Multivariate Analysis* 2(2), 162–173.
Klepper, S. & E.E. Leamer (1984) Consistent sets of estimates for regressions with errors in all variables. *Econometrica* 52, 163–183.
Koopmans, T.C. (1937) *Linear Regression Analysis of Economic Time Series*, Vol. 20. De erven F. Bohn nv.
Lewbel, A. (1997) Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65, 1201–1213.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Lewbel, A. (2012) Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics* 30(1), 67–80.
Li, T. (2002) Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics* 110(1), 1–26.
Marcinkiewicz, J. (1939) Sur une propriete de la loi de gauss. *Mathematische Zeitschrift* 44(1), 612–618.
Mussa, M. (1977) External and internal adjustment costs and the theory of aggregate and firm investment. *Economica* 44(174), 163–178.
Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4, 2111–2245.
Pal, M. (1980) Consistent moment estimators of regression coefficients in the presence of errors in variables. *Journal of Econometrics* 14(3), 349–364.
Reiersøl, O. (1945) *Confluence Analysis by Means of Instrumental Sets of Variables. PhD thesis*. Almqvist & Wiksell.
Reiersøl, O. (1950) Identifiability of a linear relation between variables which are subject to error. *Econometrica* 18, 375–389.
Schennach, S. (2007) Instrumental variables estimation of nonlinear errors-in-variables models. *Econometrica* 75, 201–239.
Schennach, S. & Y. Hu (2013) Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association* 108, 177–186.
Schennach, S.M. (2004) Estimation of nonlinear models with measurement error. *Econometrica* 72, 33–75.
Schennach, S.M. (2016) Recent advances in the measurement error literature. *Annual Review of Economics* 8(1), 341–377.
Spiegelman, C. (1979) On estimating the slope of a straight line when both variables are subject to error. *Annals of Statistics* 7(1), 201–206.
Székely, G. & C. Rao (2000) Identifiability of distributions of independent random variables by linear combinations and moments. *Sankhyā: The Indian Journal of Statistics Series A* 62, 193–202.
Tobin, J. (1969) A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking* 1(1), 15–29.
Willassen, Y. (1979) Extension of some results by reiersøl to multivariate models. *Scandinavian Journal of Statistics* 6, 89–91.
Yu, J. (2004) Empirical characteristic function estimation and its applications. *Econometric Reviews* 23(2), 93–123.

APPENDIX

A.1. Proof of Lemma 2.1

The matrix $(\boldsymbol{b},\boldsymbol{B})$ is nonsingular because,

$$\det(\boldsymbol{b},\boldsymbol{B}) = (-1)^{p-1}(\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*))^{p-2}\sum_{k=1}^{p} b_k \mathrm{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*) \neq 0,$$

where the inequality is by the assumptions that $\sum_{k=1}^{p} b_k \mathrm{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*) \neq 0$ and $\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*) \neq 0$. Next,

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

$$\mathrm{Cov}(\boldsymbol{b}'\boldsymbol{X}^*,\boldsymbol{B}'\boldsymbol{X}^*)=\begin{pmatrix}\mathrm{Cov}(X_2^*,\boldsymbol{b}'\boldsymbol{X}^*)\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*)-\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*)\mathrm{Cov}(X_2^*,\boldsymbol{b}'\boldsymbol{X}^*)\\ \vdots\\ \mathrm{Cov}(X_p^*,\boldsymbol{b}'\boldsymbol{X}^*)\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*)-\mathrm{Cov}(X_1^*,\boldsymbol{b}'\boldsymbol{X}^*)\mathrm{Cov}(X_p^*,\boldsymbol{b}'\boldsymbol{X}^*)\end{pmatrix}=\boldsymbol{0},$$

where the first equality follows by substituting in for $\boldsymbol{B}$.

A.2. Proof of Lemma 2.2

(i) Denote $d_k=\mathrm{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*)=\sum_{j=1}^p b_j\sigma_{X_k^*X_j^*}$. Then $\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)=\sum_{k=1}^p d_k\beta_k$, $\sigma_{\widetilde{\varepsilon}}^2=\sigma_\varepsilon^2+\sum_{k=1}^p d_k\beta_k$, $\sigma_{\widetilde{\eta}_k}^2=\sigma_{\eta_k}^2-\frac{d_k}{\widetilde{\beta}_k}$, $\sigma_{\widetilde{X}_k^*}^2=\sigma_{X_k^*}^2+\frac{d_k}{\widetilde{\beta}_k}$, and $\sigma_{\widetilde{X}_k^*\widetilde{X}_j^*}=\sigma_{X_k^*X_j^*}$. Then the observable variances and covariances are,

$$\sigma_{X_k}^2=\sigma_{\widetilde{X}_k^*}^2+\sigma_{\widetilde{\eta}_k}^2=(\sigma_{X_k^*}^2+\frac{d_k}{\widetilde{\beta}_k})+(\sigma_{\eta_k}^2-\frac{d_k}{\widetilde{\beta}_k})=\sigma_{X_k^*}^2+\sigma_{\eta_k}^2,$$

$$\sigma_{X_k,X_j}=\sigma_{\widetilde{X}_k^*\widetilde{X}_j^*}=\sigma_{X_k^*X_j^*},$$

$$\begin{aligned}\sigma_{X_j,Y}&=\sum_{k=1}^p\widetilde{\beta}_k\sigma_{\widetilde{X}_k^*\widetilde{X}_j^*}=\sum_{k\neq j}\widetilde{\beta}_k\sigma_{X_k^*X_j^*}+(\widetilde{\beta}_j\sigma_j^2+d_k)\\&=\sum_{k=1}^p\widetilde{\beta}_k\sigma_{X_k^*X_j^*}+\sum_{j=1}^p b_j\sigma_{X_k^*X_j^*}=\sum_{k=1}^p\beta_k\sigma_{X_k^*X_j^*},\end{aligned}$$

$$\begin{aligned}\sigma_Y^2&=\sum_{k=1}^p\sum_{j=1}^p\widetilde{\beta}_k\widetilde{\beta}_j\sigma_{\widetilde{X}_k^*\widetilde{X}_j^*}+\widetilde{\sigma}_\varepsilon^2=\sum_{k=1}^p\widetilde{\beta}_k^2(\sigma_{X_k^*}^2\\&\quad+\frac{d_k}{\widetilde{\beta}_k})+\sum_{k\neq j}\widetilde{\beta}_k\widetilde{\beta}_j\sigma_{X_k^*X_j^*}+(\sigma_\varepsilon^2+\sum_{k=1}^p d_k\beta_k)\\&=\sum_{k=1}^p\widetilde{\beta}_k^2\sigma_{X_k^*}^2+\sum_{k=1}^p\sum_{j=1}^p\widetilde{\beta}_k b_j\sigma_{X_k^*X_j^*}+\sum_{k\neq j}\widetilde{\beta}_k\widetilde{\beta}_j\sigma_{X_k^*X_j^*}+(\sigma_\varepsilon^2+\sum_{k=1}^p\sum_{j=1}^p b_j\beta_k\sigma_{X_k^*X_j^*})\\&=\sum_{k=1}^p(\widetilde{\beta}_k^2+b_k\beta_k+\widetilde{\beta}_k b_k)\sigma_{X_k^*}^2+\sum_{k\neq j}(\widetilde{\beta}_k b_j+\widetilde{\beta}_k\widetilde{\beta}_j+b_k\beta_j)\sigma_{X_k^*X_j^*}+\sigma_\varepsilon^2\\&=\sum_{k=1}^p\sum_{j=1}^p\beta_k\beta_j\sigma_{X_k^*X_j^*}+\sigma_\varepsilon^2.\end{aligned}$$

Also, $\sigma_{\widetilde{\varepsilon}}^2>0$, $\sigma_{\widetilde{\eta}_k}^2>0$, and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\succ 0$ for small enough $||\boldsymbol{b}||$ because $\sigma_\varepsilon^2>0$, $\sigma_{\eta_k}^2>0$, $\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\succ 0$, and $|d_k|\leq||\boldsymbol{b}||\cdot||\boldsymbol{\Sigma}_{\boldsymbol{X}^*}||$.

(ii) Using $\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}}=\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}-\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}$ and $\Omega_\varepsilon=\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}+\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}$, we obtain the observable variances and covariances,

$$\boldsymbol{\Sigma}_{\boldsymbol{X}}=\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}+\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}}}=(\boldsymbol{\Sigma}_{\boldsymbol{X}^*}-\boldsymbol{\Omega}_{\boldsymbol{\eta}})+(\boldsymbol{\Sigma}_{\boldsymbol{\eta}}+\boldsymbol{\Omega}_{\boldsymbol{\eta}})=\boldsymbol{\Sigma}_{\boldsymbol{X}^*}+\boldsymbol{\Sigma}_{\boldsymbol{\eta}},$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{X},Y}&=\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}}+\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}}=(\boldsymbol{\Sigma}_{\boldsymbol{X}^*}-\boldsymbol{\Omega}_{\boldsymbol{\eta}})\widetilde{\boldsymbol{\beta}}+(\boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon}+\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon})\\&=\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta}-\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}-\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}}+\boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon}+\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}=\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta}+\boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon},\end{aligned}$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

$$\sigma_Y^2 = \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\widetilde{\boldsymbol{\beta}} + \sigma_{\widetilde{\varepsilon}}^2 = \widetilde{\boldsymbol{\beta}}'(\boldsymbol{\Sigma}_{\boldsymbol{X}^*} - \boldsymbol{\Omega}_{\boldsymbol{\eta}})\widetilde{\boldsymbol{\beta}} + (\sigma_\varepsilon^2 + \Omega_\varepsilon)$$
$$= \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} - \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}} + \sigma_\varepsilon^2 + \Omega_\varepsilon = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta} + \sigma_\varepsilon^2.$$

Choose $\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} = \mathbf{0}$. Then, $\widetilde{\boldsymbol{\Sigma}} \succ 0$ and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} \succ 0$ for small enough $||\boldsymbol{b}||$ because $\boldsymbol{\Sigma} \succ 0$, $\boldsymbol{\Sigma}_{\boldsymbol{X}^*} \succ 0$, and $||\boldsymbol{\Omega}|| \leq ||\boldsymbol{b}|| \cdot P_0$.

A.3. Proof of Theorem 2.1

Assume, without loss of generality, that all random variables have zero means. This simply avoids the extra term $i\sum_{k=1}^{p+1} a_{0k}s_k$ in some of the expressions. Recall that we assume that $\beta_k \neq 0$, for $k = 1, \ldots, p$, and that there does not exist a nonzero $\boldsymbol{b}$ such that $\boldsymbol{b}'\boldsymbol{X}^*$ is degenerate. Otherwise, there is multicollinearity and $\boldsymbol{b}'\boldsymbol{X}^*$ can be added to $\boldsymbol{\beta}'\boldsymbol{X}^*$ and subtracted from ε to obtain an observationally equivalent model even in the linear regression model without measurement errors (i.e., $\boldsymbol{\eta} = \mathbf{0}$).

Let $(\boldsymbol{X}^*, \boldsymbol{\eta}, \varepsilon, \boldsymbol{\beta})$ and $(\widetilde{\boldsymbol{X}}^*, \widetilde{\boldsymbol{\eta}}, \widetilde{\varepsilon}, \widetilde{\boldsymbol{\beta}})$ be observationally equivalent. Then,

$$\begin{pmatrix} \boldsymbol{X} \\ Y \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{\beta}' \end{pmatrix}\boldsymbol{X}^* + \begin{pmatrix} \boldsymbol{\eta} \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_p \\ \widetilde{\boldsymbol{\beta}}' \end{pmatrix}\widetilde{\boldsymbol{X}}^* + \begin{pmatrix} \widetilde{\boldsymbol{\eta}} \\ \widetilde{\varepsilon} \end{pmatrix}.$$

The log characteristic function of $(\boldsymbol{X}, Y)$ is the functional equation,

$$\varphi_{\boldsymbol{X},Y}(\boldsymbol{s}) = \varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) + \varphi_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{s}) = \varphi_{\widetilde{\boldsymbol{X}}^*}((\boldsymbol{I}_p, \widetilde{\boldsymbol{\beta}})\boldsymbol{s}) + \varphi_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{R}^{p+1}, \tag{A.1}$$

where Assumption 2.1 that $\boldsymbol{X}^*$ is independent of $(\boldsymbol{\eta}, \varepsilon)$ is used to separate log characteristic functions. Hence,

$$\varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{X}}^*}((\boldsymbol{I}_p, \widetilde{\boldsymbol{\beta}})\boldsymbol{s}) = g(\boldsymbol{s}), \tag{A.2}$$

where either by Assumption 2.1(i) (independent errors), $g(\boldsymbol{s}) = \sum_{k=1}^{p+1} g_k(s_k)$, where $g_k(s_k) = \varphi_{\widetilde{\eta}_k}(s_k) - \varphi_{\eta_k}(s_k)$, for $k = 1, \ldots, p$, and $g_{p+1}(s_{p+1}) = \varphi_{\widetilde{\varepsilon}}(s_{p+1}) - \varphi_\varepsilon(s_{p+1})$, or by Assumption 2.1(ii) (joint normality of the errors), $g(\boldsymbol{s}) = -\frac{1}{2}\boldsymbol{s}'\boldsymbol{\Omega}\boldsymbol{s}$, where $\boldsymbol{\Omega} = \widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ is a $(p+1) \times (p+1)$ symmetric matrix.

(i) **Sufficiency:** Assume that $\boldsymbol{\beta}$ is not identified (i.e., $\widetilde{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$). We show that Assumptions 2.1(i)–(v) hold. The first derivatives of (A.2) evaluated at $\boldsymbol{s} = (\boldsymbol{I}_p, \mathbf{0})'\boldsymbol{u} + (-\widetilde{\boldsymbol{\beta}}', 1)'v = ((\boldsymbol{u} - \widetilde{\boldsymbol{\beta}}v)', v)'$ are,

$$\frac{\partial \varphi_{\boldsymbol{X}^*}(\boldsymbol{w})}{\partial w_k}\bigg|_{\boldsymbol{w}=(\boldsymbol{u}+(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})v)} - \frac{\partial \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{w})}{\partial w_k}\bigg|_{\boldsymbol{w}=\boldsymbol{u}} = g_k'(u_k - \widetilde{\beta}_k v), \qquad k = 1, \ldots, p,$$

$$\sum_{k=1}^{p} \beta_k \frac{\partial \varphi_{\boldsymbol{X}^*}(\boldsymbol{w})}{\partial w_k}\bigg|_{\boldsymbol{w}=(\boldsymbol{u}+(-\widetilde{\boldsymbol{\beta}})v)} - \sum_{k=1}^{p} \widetilde{\beta}_k \frac{\partial \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{w})}{\partial w_k}\bigg|_{\boldsymbol{w}=\boldsymbol{u}} = g_{p+1}'(v).$$

Next, the derivatives with respect to v evaluated at $\boldsymbol{u} = -(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})v$ are,

$$\sum_{k_2=1}^{p} (\beta_{k_2} - \widetilde{\beta}_{k_2}) \frac{\partial \varphi_{\boldsymbol{X}^*}^2(\boldsymbol{w})}{\partial w_{k_1}\partial w_{k_2}}\bigg|_{\boldsymbol{w}=\mathbf{0}} = -\widetilde{\beta}_{k_1} g_{k_1}''(-\beta_{k_1} v), \qquad k_1 = 1, \ldots, p,$$

$$\sum_{k_1,k_2} \beta_{k_1}(\beta_{k_2} - \widetilde{\beta}_{k_2}) \frac{\partial \varphi_{\boldsymbol{X}^*}^2(\boldsymbol{w})}{\partial w_{k_1}\partial w_{k_2}}\bigg|_{\boldsymbol{w}=\mathbf{0}} = g_{p+1}''(v).$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Hence, $g_k''(s_k) = 2\alpha_k$, where $\alpha_k = -\frac{1}{2\widetilde{\beta}_k}\sum_{k_2=1}^{p}(\beta_{k_2} - \widetilde{\beta}_{k_2})\mathrm{Cov}(X_k^*, X_{k_2}^*)$, for $k = 1, \ldots, p$, and $\alpha_{p+1} = \frac{1}{2}\sum_{k_1,k_2}\beta_{k_1}(\beta_{k_2} - \widetilde{\beta}_{k_2})\mathrm{Cov}(X_{k_1}^*, X_{k_2}^*)$. Integrating the above equations, we obtain $g_k(s_k) = \alpha_k s_k^2$ in a neighborhood of the origin.[10] By analytic continuation $g_k(s_k) = \alpha_k s_k^2$, $s_k \in \mathbb{R}$, for $k = 1, \ldots, p+1$. Substituting back into (A.2), we obtain,

$$\varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{X}}^*}((\boldsymbol{I}_p, \widetilde{\boldsymbol{\beta}})\boldsymbol{s}) = \sum_{k=1}^{p+1} g_k(s_k) = \sum_{k=1}^{p+1}\alpha_k s_k^2. \tag{A.3}$$

Next, substitute $\boldsymbol{s} = ((\boldsymbol{u} - \widetilde{\boldsymbol{\beta}}v)', v)'$, $\boldsymbol{s} = (-\widetilde{\boldsymbol{\beta}}', 1)'v$, and $\boldsymbol{s} = (\boldsymbol{u}, 0)'$ into (A.3), respectively,

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{u} + \boldsymbol{b}v) - \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{u}) = \sum_{k=1}^{p} g_k(u_k - \widetilde{\beta}_k v) + g_{p+1}(v) = \sum_{k=1}^{p}\alpha_k(u_k - \widetilde{\beta}_k v)^2 + \alpha_{p+1}v^2, \tag{A.4}$$

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v) = \sum_{k=1}^{p} g_k(\widetilde{\beta}_k v) + g_{p+1}(v) = \sum_{k=1}^{p}\alpha_k\widetilde{\beta}_k^2 v^2 + \alpha_{p+1}v^2, \tag{A.5}$$

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{u}) - \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{u}) = \sum_{k=1}^{p} g_k(u_k) = \sum_{k=1}^{p}\alpha_k u_k^2, \tag{A.6}$$

where $\boldsymbol{b} = \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}$. By (A.5) and Marcinkiewicz (1939), $\boldsymbol{b}'\boldsymbol{X}^*$ is normal.[11] If there does not exist a nonzero $\boldsymbol{b} \in \mathbb{R}^p$ such that $\boldsymbol{b}'\boldsymbol{X}^*$ is normal then $\boldsymbol{b} = \boldsymbol{0}$ and $\boldsymbol{\beta}$ is identified. Hence, nonidentifiability implies that $\boldsymbol{b}'\boldsymbol{X}^*$ is nondegenerate normal (Assumption 2.2(i) holds).Substitute (A.5) and (A.6) into (A.4) to obtain,

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{u} + \boldsymbol{b}v) = \varphi_{\boldsymbol{X}^*}(\boldsymbol{u}) + \varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v) - 2v\boldsymbol{d}'\boldsymbol{u}, \tag{A.7}$$

where $\boldsymbol{d} = \begin{pmatrix}\alpha_1\widetilde{\beta}_1 & \ldots & \alpha_p\widetilde{\beta}_p\end{pmatrix}'$. Next, let $\boldsymbol{B} \in \mathbb{R}^{p\times(p-1)}$ have full column rank and satisfy

$$\boldsymbol{d}'\boldsymbol{B} = \boldsymbol{0}, \tag{A.8}$$

and substitute $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{w}$, for $\boldsymbol{w} \in \mathbb{R}^{p-1}$, into (A.7), to obtain,

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w} + \boldsymbol{b}v) = \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) + \varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v). \tag{A.9}$$

Hence, nonidentifiability implies that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ are independent (Assumption 2.2(ii) holds). Now, if $(\boldsymbol{B}, \boldsymbol{b})$ is singular, then some nonzero linear combination of the columns of $\boldsymbol{B}$ is equal to $\boldsymbol{b}$, which implies that $\boldsymbol{b}'\boldsymbol{X}^*$ is independent of itself and $\boldsymbol{b}'\boldsymbol{X}^*$ is degenerate. Hence, nonidentifiability implies that $(\boldsymbol{B}, \boldsymbol{b})$ is nonsingular (Assumption 2.2(iii) holds).Next substitute $\boldsymbol{s} = ((\boldsymbol{B}\boldsymbol{w} + \boldsymbol{b}z - \widetilde{\boldsymbol{\beta}}v)', v)'$ into (A.2) and simplify

[10] Khatri and Rao (1972) show that $g_k(s_k) = \alpha_k s_k^2$ in a neighborhood of the origin using finite differences (and continuity of the log characteristic function) instead of derivatives so no moments need to exist up to this point in the proof.

[11] Even without analytic continuity, (A.5) holds in a neighborhood of the origin so $\boldsymbol{b}'\boldsymbol{X}^*$ is normal by Marcinkiewicz (1939).

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

to obtain,

$$0 = (\sum_{k=1}^{p} d_k\widetilde{\beta}_k + \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} + \alpha_{p+1})v^2 + (\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} - 2\sum_{k=1}^{p} d_k b_k)vz - 2\boldsymbol{d}'\boldsymbol{B}\boldsymbol{w}v.$$

Using (A.8) and equating coefficients, we obtain,

$$\frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} = \sum_{k=1}^{p} d_k b_k = -\sum_{k=1}^{p} d_k\widetilde{\beta}_k - \alpha_{p+1} > 0.$$

By the second equality $\alpha_{p+1} = -\sum_{k=1}^{p} d_k\beta_k$. By the first equality and (A.8), $\boldsymbol{d}$ solves the p linear equations,

$$\boldsymbol{d}'\begin{pmatrix}\boldsymbol{B} & \boldsymbol{b}\end{pmatrix} = \begin{pmatrix}\boldsymbol{0} & \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}\end{pmatrix}.$$

But by Assumption 2.2(ii), $\boldsymbol{0} = \mathrm{Cov}(\boldsymbol{B}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*) = \boldsymbol{B}'\mathrm{Cov}(\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$. Noting that $\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} = \boldsymbol{b}'\mathrm{Cov}(\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$, we obtain the solution $\boldsymbol{d} = \frac{1}{2}\mathrm{Cov}(\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$. By Assumption 2.2(iii), $(\boldsymbol{B}, \boldsymbol{b})$ is nonsingular so that this is the unique solution. Hence, $\alpha_k = d_k/\widetilde{\beta}_k = \frac{1}{2}\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k$, for $k = 1, \ldots, p$, and $\alpha_{p+1} = -\sum_{k=1}^{p} d_k\beta_k = -\frac{1}{2}\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$.

Recall that $g_k(s_k) = \varphi_{\widetilde{\eta}_k}(s_k) - \varphi_{\eta_k}(s_k) = \alpha_k s_k^2$, for $k = 1, \ldots, p+1$, so if $-2\alpha_{p+1} = 2\sum_{k=1}^{p} d_k\beta_k = \mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*) < 0$, then ε is divisible by a nondegenerate normal distribution (Assumption 2.2(iv) holds) and if $2\alpha_k = 2d_k/\widetilde{\beta}_k = \mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k > 0$, then η_k is divisible by a nondegenerate normal distribution (Assumption 2.2(v) holds).

Finally, $\sigma_{\widetilde{\varepsilon}}^2 = \sigma_\varepsilon^2 + \sum_{k=1}^{p} d_k\beta_k$, $\sigma_{\widetilde{\eta}_k}^2 = \sigma_{\eta_k}^2 - \frac{d_k}{\widetilde{\beta}_k}$, $\sigma_{\widetilde{X}_k^*}^2 = \sigma_{X_k^*}^2 + \frac{d_k}{\widetilde{\beta}_k}$, and $\sigma_{\widetilde{X}_k^*\widetilde{X}_j^*} = \sigma_{X_k^*X_j^*}$. Hence, $||\boldsymbol{b}||$ needs to be small enough so that $\sigma_{\widetilde{\varepsilon}}^2 > 0$, $\sigma_{\widetilde{\eta}_k}^2 > 0$, and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*} \succ 0$, which follows by $\sigma_\varepsilon^2 > 0$, $\sigma_{\eta_k}^2 > 0$, $\boldsymbol{\Sigma}_{\boldsymbol{X}^*} \succ 0$, and $|d_k| \leq ||\boldsymbol{b}|| \cdot ||\boldsymbol{\Sigma}_{\boldsymbol{X}^*}||$.

Necessity: Assume that Assumptions 2.2(i)–(v) hold. We show that $\boldsymbol{\beta}$ is not identified by constructing an observationally equivalent model that satisfies (A.1). By Assumptions 2.2(i)–(iii), there exists a nonsingular matrix $(\boldsymbol{B}, \boldsymbol{b})$ such that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^* \sim N(0, \boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})$ are independent with $\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b} > 0$ and $||\boldsymbol{b}||$ small. Let $Z_{1k} \sim N(0, |\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\beta_k|)$, $Z_{1k}' \sim N(0, |\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k|)$, and $Z_2 \sim N(0, |\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)|)$ be independent random variables. Define $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{b}$. If $\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k > 0$, then define $\widetilde{X}_k^* = X_k^* + Z_{1k}$ and $\eta_k = \widetilde{\eta}_k + Z_{1k}'$; otherwise define $X_k^* = \widetilde{X}_k^* + Z_{1k}$ and $\widetilde{\eta}_k = \eta_k + Z_{1k}'$, and if $\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*) < 0$, then define $\varepsilon = \widetilde{\varepsilon} + Z_2$; otherwise define $\widetilde{\varepsilon} = \varepsilon + Z_2$.

Then $\varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{u}) = \varphi_{\boldsymbol{X}^*}(\boldsymbol{u}) - \sum_{k=1}^{p} \alpha_k u_k^2$, $\varphi_{\widetilde{\eta}_k}(s_k) = \varphi_{\eta_k}(s_k) + \alpha_k s_k^2$, and $\varphi_{\widetilde{\varepsilon}}(s_{p+1}) = \varphi_\varepsilon(s_{p+1}) + \alpha_{p+1}s_{p+1}^2$, where $\alpha_k = \frac{1}{2}\mathrm{Cov}(X_k^*, \boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k$ and $\alpha_{p+1} = -\frac{1}{2}\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*, \boldsymbol{b}'\boldsymbol{X}^*)$.

Substitute $\boldsymbol{s} = ((\boldsymbol{I}_p, \boldsymbol{0})'(\boldsymbol{B}, \boldsymbol{b}), (-\widetilde{\boldsymbol{\beta}}', 1)')(\boldsymbol{w}', z, v)' = ((\boldsymbol{B}\boldsymbol{w} + \boldsymbol{b}z - \widetilde{\boldsymbol{\beta}}v)', v)' \in \mathbb{R}^{p+1}$ into (A.1) to obtain,

$$\begin{aligned}\varphi_{\boldsymbol{X},Y}(\boldsymbol{s}) &= \varphi_{\widetilde{\boldsymbol{X}}^*}((\boldsymbol{I}_p, \widetilde{\boldsymbol{\beta}})\boldsymbol{s}) + \sum_{k=1}^{p} \varphi_{\widetilde{\eta}_k}(s_k) + \varphi_{\widetilde{\varepsilon}}(s_{p+1}) \\ &= \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{B}\boldsymbol{w} + \boldsymbol{b}z) + \sum_{k=1}^{p} \varphi_{\widetilde{\eta}_k}(\boldsymbol{B}_k'\boldsymbol{w} + b_k z - \widetilde{\beta}_k v) + \varphi_{\widetilde{\varepsilon}}(v)\end{aligned}$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

$$
\begin{aligned}
&= \left[\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z) - \sum_{k=1}^{p}\alpha_k(B_k'\boldsymbol{w}+b_kz)^2\right] \\
&\quad + \sum_{k=1}^{p}\left[\varphi_{\eta_k}(\boldsymbol{B}_k'\boldsymbol{w}+b_kz-\widetilde{\beta}_kv)+\alpha_k(\boldsymbol{B}_k'\boldsymbol{w}+b_kz-\widetilde{\beta}_kv)^2\right]+\left[\varphi_{\varepsilon}(v)+\alpha_{p+1}v^2\right] \\
&= \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) + \sum_{k=1}^{p}\varphi_{\eta_k}(\boldsymbol{B}_k'\boldsymbol{w}+b_kz-\widetilde{\beta}_kv)+\varphi_{\varepsilon}(v) - \frac{1}{2}\boldsymbol{b}'\Sigma_{\boldsymbol{X}^*}\boldsymbol{b}z^2 \\
&\quad - 2\sum_{k=1}^{p}\alpha_k\widetilde{\beta}_k\boldsymbol{B}_k'\boldsymbol{w}v - 2\sum_{k=1}^{p}\alpha_k\widetilde{\beta}_kb_kzv + \sum_{k=1}^{p}\alpha_k\widetilde{\beta}_k^2v^2+\alpha_{p+1}v^2 \\
&= \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) + \sum_{k=1}^{p}\varphi_{\eta_k}(\boldsymbol{B}_k'\boldsymbol{w}+b_kz-\widetilde{\beta}_kv)+\varphi_{\varepsilon}(v) - \frac{1}{2}\boldsymbol{b}'\Sigma_{\boldsymbol{X}^*}\boldsymbol{b}z^2 \\
&\quad - \mathrm{Cov}(\boldsymbol{B}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)\boldsymbol{w}v - \mathrm{Cov}(\boldsymbol{b}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)zv + \frac{1}{2}\mathrm{Cov}(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)v^2 \\
&\quad - \frac{1}{2}\mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)v^2 \\
&= \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) + \sum_{k=1}^{p}\varphi_{\eta_k}(\boldsymbol{B}_k'\boldsymbol{w}+b_kz-\widetilde{\beta}_kv)+\varphi_{\varepsilon}(v) - \frac{1}{2}\boldsymbol{b}'\Sigma_{\boldsymbol{X}^*}\boldsymbol{b}(v+z)^2 \\
&= \varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p,\boldsymbol{\beta})\boldsymbol{s}) + \sum_{k=1}^{p}\varphi_{\eta_k}(s_p)+\varphi_{\varepsilon}(s_{p+1}),
\end{aligned}
$$

where $\boldsymbol{B}_k = (B_{k1},\ldots,B_{kp-1})'$ is the kth row of $\boldsymbol{B}$. The fourth, sixth, and last equalities follow by Assumptions 2.2(i) and (ii) that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ are independent (so $\mathrm{Cov}(\boldsymbol{B}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*) = 0$) and $\boldsymbol{b}'\boldsymbol{X}^*$ is normal.

(ii) **Sufficiency:** Assume that $\boldsymbol{\beta}$ is not identified. We show that Assumptions 2.2(i)–(iii) hold. Let

$$
\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{\eta}} & \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}' & \Omega_{\varepsilon} \end{pmatrix}, \quad \widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}}} & \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}} \\ \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}}' & \sigma_{\widetilde{\varepsilon}}^2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon}' & \sigma_{\varepsilon}^2, \end{pmatrix} \tag{A.10}
$$

and substitute $\boldsymbol{s} = ((\boldsymbol{u}-\widetilde{\boldsymbol{\beta}}v)',v)'$, $\boldsymbol{s} = (-\widetilde{\boldsymbol{\beta}}',1)'v$, and $\boldsymbol{s} = (\boldsymbol{u},0)'$ into (A.2), respectively,

$$
\varphi_{\boldsymbol{X}^*}(\boldsymbol{u}+\boldsymbol{b}v) = \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\boldsymbol{u} + v(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}} - \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}')\boldsymbol{u} - \frac{1}{2}(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}} - 2\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}'\widetilde{\boldsymbol{\beta}} + \Omega_{\varepsilon})v^2, \tag{A.11}
$$

$$
\varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v) = -\frac{1}{2}(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}} - 2\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}'\widetilde{\boldsymbol{\beta}} + \Omega_{\varepsilon})v^2, \tag{A.12}
$$

$$
\varphi_{\boldsymbol{X}^*}(\boldsymbol{u}) = \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\boldsymbol{u}, \tag{A.13}
$$

where $\boldsymbol{b} = \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}$. By (A.12) and Marcinkiewicz (1939), $\boldsymbol{b}'\boldsymbol{X}^*$ is normal. If there does not exist a nonzero $\boldsymbol{b} \in \mathbb{R}^p$ such that $\boldsymbol{b}'\boldsymbol{X}^*$ is normal then $\boldsymbol{b} = \boldsymbol{0}$ and $\boldsymbol{\beta}$ is identified. Hence, nonidentifiability implies that $\boldsymbol{b}'\boldsymbol{X}^*$ is nondegenerate normal (Assumption 2.2(i)

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

holds).Substitute (A.12) and (A.13) into (A.11) to obtain,

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{u}+\boldsymbol{b}v)=\varphi_{\boldsymbol{X}^*}(\boldsymbol{u})+\varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v)+v(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}-\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon})\boldsymbol{u}. \tag{A.14}$$

Next, let $\boldsymbol{B}\in\mathbb{R}^{p\times(p-1)}$ have full column rank and satisfy

$$(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}-\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon})\boldsymbol{B}=\boldsymbol{0}, \tag{A.15}$$

and substitute $\boldsymbol{u}=\boldsymbol{B}\boldsymbol{w}$, for $\boldsymbol{w}\in\mathbb{R}^{p-1}$, into (A.14) to obtain,

$$\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}v)=\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w})+\varphi_{\boldsymbol{X}^*}(\boldsymbol{b}v). \tag{A.16}$$

Hence, nonidentifiability implies that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ are independent (Assumption 2.2(ii) holds). Now, if $(\boldsymbol{B},\boldsymbol{b})$ is singular then some nonzero linear combination of the columns of $\boldsymbol{B}$ is equal to $\boldsymbol{b}$, which implies that $\boldsymbol{b}'\boldsymbol{X}^*$ is independent of itself and $\boldsymbol{b}'\boldsymbol{X}^*$ is degenerate. Hence, nonidentifiability implies that $(\boldsymbol{B},\boldsymbol{b})$ is nonsingular (Assumption 2.2(iii) holds).

Next substitute $\boldsymbol{s}=((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)',v)'$ into (A.2) and simplify to obtain,

$$0=(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}-\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon})\boldsymbol{B}\boldsymbol{w}v-(\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}\boldsymbol{b}-\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\boldsymbol{b}-\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})zv-\frac{1}{2}(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}}-2\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}\widetilde{\boldsymbol{\beta}}$$
$$+\Omega_{\varepsilon}-\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})v^2.$$

Using (A.15) and equating coefficients, we obtain,

$$\left(\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}-\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}},\Omega_{\varepsilon}-\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}\widetilde{\boldsymbol{\beta}}\right)=\left((\boldsymbol{0},\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})(\boldsymbol{B},\boldsymbol{b})^{-1},(\boldsymbol{0},\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})(\boldsymbol{B},\boldsymbol{b})^{-1}\widetilde{\boldsymbol{\beta}}+\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}\right)$$
$$=\left(\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*},\quad \boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta}\right),$$

where the last equality follows because

$$(\boldsymbol{0},\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})(\boldsymbol{B},\boldsymbol{b})^{-1}=(\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{B},\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})(\boldsymbol{B},\boldsymbol{b})^{-1}=\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}(\boldsymbol{B},\boldsymbol{b})(\boldsymbol{B},\boldsymbol{b})^{-1}=\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*},$$

where $\boldsymbol{0}=\mathrm{Cov}(\boldsymbol{b}'\boldsymbol{X}^*,\boldsymbol{B}'\boldsymbol{X}^*)=\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{B}$ by Assumption 2.2(ii) and $(\boldsymbol{B},\boldsymbol{b})$ is nonsingular by Assumption 2.2(iii). Hence, $\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}=\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}-\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}$ and $\Omega_{\varepsilon}=\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{\beta}+\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}\widetilde{\boldsymbol{\beta}}$. Let $\boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon}=\boldsymbol{0}$ and solve for $\boldsymbol{\Omega}$ to obtain,

$$\boldsymbol{\Omega}=\begin{pmatrix}-\mathrm{Cov}(\frac{X_1^*}{\widetilde{\beta}_1},\boldsymbol{b}'\boldsymbol{X}^*) & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & \ddots & -\mathrm{Cov}(\frac{X_p^*}{\widetilde{\beta}_p},\boldsymbol{b}'\boldsymbol{X}^*) & 0\\ 0 & \cdots & 0 & \mathrm{Cov}(\boldsymbol{\beta}'\boldsymbol{X}^*,\boldsymbol{b}'\boldsymbol{X}^*)\end{pmatrix}. \tag{A.17}$$

Finally, $||\boldsymbol{b}||$ needs to be small enough so that $\widetilde{\boldsymbol{\Sigma}}\succ 0$ and $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{X}}^*}\succ 0$, which follows by $\boldsymbol{\Sigma}\succ 0$, $\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\succ 0$, and $||\boldsymbol{\Omega}||\le||\boldsymbol{b}||\cdot P_0$.

Necessity: Assume that Assumptions 2.2(i)–(iii) hold. We construct an observationally equivalent model that satisfies (A.1). By Assumptions 2.2(i)–(iii), there exists a nonsingular matrix $(\boldsymbol{B},\boldsymbol{b})$ such that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*\sim N(0,\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b})$ are independent with $\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}>0$ and $||\boldsymbol{b}||$ small. Let $Z_{1k}\sim N(0,|\mathrm{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k|)$ be an independent random variable. Define $\widetilde{\boldsymbol{\beta}}=\boldsymbol{\beta}-\boldsymbol{b}$, $(\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon})\sim N(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}})$, where $\widetilde{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}+\boldsymbol{\Omega}$ with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ defined in (A.10) and (A.17), and if $\mathrm{Cov}(X_k^*,\boldsymbol{b}'\boldsymbol{X}^*)/\widetilde{\beta}_k>0$, then $\widetilde{X}_k^*=X_k^*+Z_{1k}$; otherwise $X_k^*=\widetilde{X}_k^*+Z_{1k}$.

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Then $\varphi_{\widetilde{X}^*}(\boldsymbol{u}) = \varphi_{X^*}(\boldsymbol{u}) + \frac{1}{2}\boldsymbol{u}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\boldsymbol{u}$, and $\varphi_{\widetilde{\eta},\widetilde{\varepsilon}}(\boldsymbol{s}) = -\frac{1}{2}\boldsymbol{s}'\widetilde{\boldsymbol{\Sigma}}\boldsymbol{s}$. Substitute $\boldsymbol{s} = ((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)' \in \mathbb{R}^{p+1}$ into (A.1) to obtain,

$$
\begin{aligned}
\varphi_{\boldsymbol{X},Y}(\boldsymbol{s}) &= \varphi_{\widetilde{\boldsymbol{X}}^*}((\boldsymbol{I}_p, \widetilde{\boldsymbol{\beta}})\boldsymbol{s}) - \frac{1}{2}\boldsymbol{s}'\widetilde{\boldsymbol{\Sigma}}\boldsymbol{s} \\
&= \varphi_{\widetilde{\boldsymbol{X}}^*}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z) - \frac{1}{2}((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)\begin{pmatrix}\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}}} & \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}} \\ \boldsymbol{\Sigma}'_{\widetilde{\boldsymbol{\eta}},\widetilde{\varepsilon}} & \sigma^2_{\widetilde{\varepsilon}}\end{pmatrix}\begin{pmatrix}\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v \\ v\end{pmatrix} \\
&= \left[\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z) + \frac{1}{2}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z)'\boldsymbol{\Omega}_{\boldsymbol{\eta}}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z)\right] \\
&\quad - \frac{1}{2}((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)\left[\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}'_{\boldsymbol{\eta},\varepsilon} & \sigma^2_{\varepsilon}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\Omega}_{\boldsymbol{\eta}} & \boldsymbol{\Omega}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon} & \Omega_{\varepsilon}\end{pmatrix}\right]\begin{pmatrix}\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v \\ v\end{pmatrix} \\
&= \left[\varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) - \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}z^2\right] \\
&\quad - \frac{1}{2}((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}'_{\boldsymbol{\eta},\varepsilon} & \sigma^2_{\varepsilon}\end{pmatrix}\begin{pmatrix}\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v \\ v\end{pmatrix} \\
&\quad + \frac{1}{2}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z)'\boldsymbol{\Omega}_{\boldsymbol{\eta}}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z) - \frac{1}{2}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)'\boldsymbol{\Omega}_{\boldsymbol{\eta}}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v) \\
&\quad - \boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}(\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)v - \frac{1}{2}\Omega_{\varepsilon}v^2 \\
&= \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) - \frac{1}{2}((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}'_{\boldsymbol{\eta},\varepsilon} & \sigma^2_{\varepsilon}\end{pmatrix}\begin{pmatrix}\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v \\ v\end{pmatrix} \\
&\quad - \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}z^2 - (\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon} - \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}})\boldsymbol{B}\boldsymbol{w}v - (\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon} - \widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}})\boldsymbol{b}zv \\
&\quad - \frac{1}{2}(\widetilde{\boldsymbol{\beta}}'\boldsymbol{\Omega}_{\boldsymbol{\eta}}\widetilde{\boldsymbol{\beta}} - 2\boldsymbol{\Omega}'_{\boldsymbol{\eta},\varepsilon}\widetilde{\boldsymbol{\beta}} + \Omega_{\varepsilon})v^2 \\
&= \varphi_{\boldsymbol{X}^*}(\boldsymbol{B}\boldsymbol{w}) - \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{b}(v+z)^2 \\
&\quad - \frac{1}{2}((\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v)', v)\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta},\varepsilon} \\ \boldsymbol{\Sigma}'_{\boldsymbol{\eta},\varepsilon} & \sigma^2_{\varepsilon}\end{pmatrix}\begin{pmatrix}\boldsymbol{B}\boldsymbol{w}+\boldsymbol{b}z-\widetilde{\boldsymbol{\beta}}v \\ v\end{pmatrix} \\
&= \varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) - \frac{1}{2}\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s},
\end{aligned}
$$

where the fourth, sixth, and last equalities follow by Assumptions 2.2(i) and (ii) that $\boldsymbol{B}'\boldsymbol{X}^*$ and $\boldsymbol{b}'\boldsymbol{X}^*$ are independent (so $\boldsymbol{b}'\boldsymbol{\Sigma}_{\boldsymbol{X}^*}\boldsymbol{B} = Cov(\boldsymbol{b}'\boldsymbol{X}^*, \boldsymbol{B}'\boldsymbol{X}^*) = 0$) and $\boldsymbol{b}'\boldsymbol{X}^*$ is normal.

A.4. Proof of Corollary 2.1

In the case $p = 1$, (A.2) is,

$$
\varphi_{X^*}(s_1 + \beta s_2) - \varphi_{\widetilde{X}^*}(s_1 + \widetilde{\beta}s_2) = g_1(s_1) + g_2(s_2), \tag{A.18}
$$

where $g_1(s_1) = \varphi_{\widetilde{\eta}}(s_1) - \varphi_{\eta}(s_1)$ and $g_2(s_2) = \varphi_{\widetilde{\varepsilon}}(s_2) - \varphi_{\varepsilon}(s_2)$. By Khatri and Rao (1972), $g_1(s_1) = \alpha_1 s_1^2$ and $g_2(s_2) = \alpha_2 s_2^2$ in some neighborhood of the origin.

Sufficiency: Assume that β is not identified. We show that X^* is normal and either ε or η is divisible by a nondegenerate normal distribution. Substitute $\boldsymbol{s} = (-\widetilde{\beta}v, v)'$ and $\boldsymbol{s} =$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

$(-\beta v, v)'$ respectively into (A.18) to obtain,

$$\varphi_{X^*}(bv) = g_1(-\widetilde{\beta}v) + g_2(v) = \alpha_1\widetilde{\beta}^2v^2 + \alpha_2v^2, \tag{A.19}$$

$$-\varphi_{\widetilde{X}^*}(-bv) = g_1(-\beta v) + g_2(v) = \alpha_1\beta^2v^2 + \alpha_2v^2, \tag{A.20}$$

where $b = \beta - \widetilde{\beta}$. By Marcinkiewicz (1939), bX^* and $b\widetilde{X}^*$ are normal. If X^* is nongaussian, then $b = 0$ and β is identified. Hence, nonidentification implies that X^* and $\widetilde{X}^*$ are nondegenerate normal.

Using (A.18) and that X^* and $\widetilde{X}^*$ are nondegenerate normal,

$$\frac{1}{2}(\sigma^2_{\widetilde{X}^*} - \sigma^2_{X^*})s_1^2 + (\sigma^2_{\widetilde{X}^*}\widetilde{\beta} - \sigma^2_{X^*}\beta)s_1s_2 + \frac{1}{2}(\sigma^2_{\widetilde{X}^*}\widetilde{\beta}^2 - \sigma^2_{X^*}\beta^2)s_2^2 = g_1(s_1) + g_2(s_2),$$
$$(s_1, s_2) \in \mathbb{R}^2.$$

Equating coefficients, $\sigma^2_{\widetilde{X}^*}\widetilde{\beta} = \sigma^2_{X^*}\beta$ ($|b| = |\beta - \widetilde{\beta}|$ needs to be small enough so that β and $\widetilde{\beta}$ have the same sign), $g_1(s_1) = \alpha_1 s_1^2$, and $g_2(s_2) = \alpha_2 s_2^2$, where $\alpha_1 = \frac{1}{2}\sigma^2_{X^*}b/\widetilde{\beta}$ and $\alpha_2 = -\frac{1}{2}\sigma^2_{X^*}\beta b$. If $\alpha_1 = \alpha_2 = 0$, then $b = 0$ and β is identified. Hence, nonidentifiability implies that either ε or η is divisible by a nondegenerate normal distribution.

Necessity: Assume that $X^* \sim N(0, \sigma^2_{X^*})$ with $\sigma^2_{X^*} > 0$ and either ε or η is divisible by a nondegenerate normal distribution. We construct an observationally equivalent model that satisfies (A.1). Define $\widetilde{\beta} = \beta - b$ for some small nonzero $b \in \mathbb{R}$ so that β and $\widetilde{\beta}$ have the same sign, and let $Z_1 \sim N(0, |b/\widetilde{\beta}|\sigma^2_{X^*})$, $Z_1' \sim N(0, |b/\widetilde{\beta}|\sigma^2_{X^*})$, and $Z_2 \sim N(0, |b\beta|\sigma^2_{X^*})$ be independent random variables. If $b\beta > 0$, then define $\widetilde{X}^* = X^* + Z_1$, $\eta = \widetilde{\eta} + Z_1'$, and $\widetilde{\varepsilon} = \varepsilon + Z_2$ otherwise define $X^* = \widetilde{X}^* + Z_1$, $\widetilde{\eta} = \eta + Z_1'$, and $\varepsilon = \widetilde{\varepsilon} + Z_2$.

Then $\varphi_{\widetilde{X}^*}(u) = \varphi_{X^*}(u) - \alpha_1u^2 = -(\frac{1}{2}\sigma^2_{\widetilde{X}^*} + \alpha_1)u^2$, $\varphi_{\widetilde{\eta}}(s_1) = \varphi_\eta(s_1) + \alpha_1 s_1^2$, and $\varphi_{\widetilde{\varepsilon}}(s_2) = \varphi_\varepsilon(s_2) + \alpha_2 s_2^2$, where $\alpha_1 = \frac{1}{2}\sigma^2_{X^*}b/\widetilde{\beta}$ and $\alpha_2 = -\frac{1}{2}b\beta\sigma^2_{X^*}$. Now substituting into (A.1) we obtain,

$$\begin{aligned}
\varphi_{X,Y}(s_1, s_2) &= \varphi_{X^*}(s_1 + s_2\beta) + \varphi_\eta(s_1) + \varphi_\varepsilon(s_2) \\
&= -\frac{1}{2}\sigma^2_{X^*}(s_1 + s_2\beta)^2 + \left(\varphi_{\widetilde{\eta}}(s_1) - \frac{1}{2}\frac{b\sigma^2_{X^*}}{\widetilde{\beta}}s_1^2\right) + \left(\varphi_{\widetilde{\varepsilon}}(s_2) + \frac{1}{2}b\beta\sigma^2_{X^*}s_2^2\right) \\
&= -\frac{1}{2}\sigma^2_{X^*}s_1^2 - \sigma^2_{X^*}\beta s_1s_2 - \frac{1}{2}\sigma^2_{X^*}s_2^2\beta^2 \\
&\quad - \frac{1}{2}\frac{b\sigma^2_{X^*}}{\widetilde{\beta}}s_1^2 + \frac{1}{2}b\beta\sigma^2_{X^*}s_2^2 + \varphi_{\widetilde{\eta}}(s_1) + \varphi_{\widetilde{\varepsilon}}(s_2) \\
&= -\frac{1}{2}\sigma^2_{X^*}\frac{\beta}{\widetilde{\beta}}s_1^2 - \sigma^2_{X^*}\beta s_1s_2 - \frac{1}{2}\beta\widetilde{\beta}\sigma^2_{X^*}s_2^2 + \varphi_{\widetilde{\eta}}(s_1) + \varphi_{\widetilde{\varepsilon}}(s_2) \\
&= -\frac{1}{2}\sigma^2_{X^*}\frac{\beta}{\widetilde{\beta}}(s_1 + \widetilde{\beta}s_2)^2 + \varphi_{\widetilde{\eta}}(s_1) + \varphi_{\widetilde{\varepsilon}}(s_2) \\
&= \varphi_{\widetilde{X}^*}(s_1 + s_2\widetilde{\beta}) + \varphi_{\widetilde{\eta}}(s_1) + \varphi_{\widetilde{\varepsilon}}(s_2).
\end{aligned}$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

A.5. Proof of Theorem 3.1

The estimator in (4) minimizes the integral of a GMM norm. The proof of root-n asymptotic normality is basically the same as the proof for any extremum estimator (see, e.g., Newey and McFadden, 1994). See Carrasco and Florens (2000) for a proof in a general setting with a continuum of moment conditions and with higher level assumptions.

Define

$$\begin{aligned}
&\boldsymbol{H}(\boldsymbol{b},v)\\
&= \left(\operatorname{Cov}(X_1,X_2)+\frac{\partial^2\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_1\partial s_2}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\ldots,\operatorname{Cov}(X_{p-1},X_p)+\frac{\partial^2\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_{p-1}\partial s_p}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\right.\\
&\quad\left.\operatorname{Cov}(X_1,Y)+\frac{\partial^2\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_1\partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\ldots,\operatorname{Cov}(X_p,Y)+\frac{\partial^2\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_p\partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v}\right)',\\
&\boldsymbol{H}_n(\boldsymbol{b},v)\\
&= \left(\widehat{\operatorname{Cov}}(X_1,X_2)+\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_1\partial s_2}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\ldots,\widehat{\operatorname{Cov}}(X_{p-1},X_p)+\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_{p-1}\partial s_p}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\right.\\
&\quad\left.\widehat{\operatorname{Cov}}(X_1,Y)+\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_1\partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v},\ldots,\widehat{\operatorname{Cov}}(X_p,Y)+\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_p\partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v}\right)',
\end{aligned}$$

where $\boldsymbol{b}\in\mathcal{B}\subset\mathbb{R}^p$, $\boldsymbol{H}(\boldsymbol{b},v)'$ is the conjugate transpose of $H(\boldsymbol{b},v)\in\mathbb{C}^{p(p+1)/2\times 1}$, which is finite for all $\boldsymbol{b}\in\mathcal{B}$ and v almost everywhere by Assumptions 3.3(ii) and (iii), $\boldsymbol{H}_n(\boldsymbol{b},v)'$ is the conjugate transpose of $\boldsymbol{H}_n(\boldsymbol{b},v)\in\mathbb{C}^{p(p+1)/2\times 1}$, and

$$\begin{aligned}
\widehat{\operatorname{Cov}}(X_{k_1},X_{k_2}) &= \frac{1}{n}\sum_{i=1}^n X_{ik_1}X_{ik_2}-\left(\frac{1}{n}\sum_{i=1}^n X_{ik_1}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{ik_2}\right),\\
\widehat{\operatorname{Cov}}(X_k,Y) &= \frac{1}{n}\sum_{i=1}^n X_{ik}Y_i-\left(\frac{1}{n}\sum_{i=1}^n X_{ik}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_i\right),\\
\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_{k_1}\partial s_{k_2}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v} &= \frac{\left(\frac{1}{n}\sum_{i=1}^n X_{ik_1}e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)\left(\frac{1}{n}\sum_{i=1}^n X_{ik_2}e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)^2}\\
&\quad-\frac{\frac{1}{n}\sum_{i=1}^n X_{ik_1}X_{ik_2}e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}}{\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}},\\
\frac{\partial^2\widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_k\partial s_{p+1}}\bigg|_{\boldsymbol{s}=(\boldsymbol{b},-1)v} &= \frac{\left(\frac{1}{n}\sum_{i=1}^n X_{ik}e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_ie^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}\right)^2}\\
&\quad-\frac{\frac{1}{n}\sum_{i=1}^n X_{ik}Y_ie^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}}{\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v(\boldsymbol{b}'\boldsymbol{X}_i-Y_i)}}.
\end{aligned}$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Define

$$Q(\boldsymbol{b}) = -\int_{-\infty}^{\infty} \boldsymbol{H}(\boldsymbol{b}, v)' \boldsymbol{W}(v) \boldsymbol{H}(\boldsymbol{b}, v) \pi(v) dv,$$

$$Q_n(\boldsymbol{b}) = -\int_{-\infty}^{\infty} \boldsymbol{H}_n(\boldsymbol{b}, v)' \boldsymbol{W}_n(v) \boldsymbol{H}_n(\boldsymbol{b}, v) \pi(v) dv,$$

where $\boldsymbol{W}_n(v) \xrightarrow{p} \boldsymbol{W}(v)$ almost everywhere for some positive definite matrix $\boldsymbol{W}(v)$ by Assumption 3.3(iv) and $\pi(v) > 0$ is a probability density function. Then the IGMM estimator is,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{b} \in \mathcal{B}}{\operatorname{argmax}} Q_n(\boldsymbol{b}).$$

We show that under Assumption 3.1, Q is maximized if and only if $\boldsymbol{b} = \boldsymbol{\beta}$ or equivalently that $\boldsymbol{H}(\boldsymbol{b}, v) = 0$ if and only if $\boldsymbol{b} = \boldsymbol{\beta}$. The log characteristic function of $(\boldsymbol{X}, Y)$ is

$$\varphi_{\boldsymbol{X}, Y}(\boldsymbol{s}) = \varphi_{\boldsymbol{X}^*}((\boldsymbol{I}_p, \boldsymbol{\beta})\boldsymbol{s}) + \varphi_{\boldsymbol{\eta}, \varepsilon}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{R}^{p+1},$$

where $\varphi_{\boldsymbol{\eta}, \varepsilon}(\boldsymbol{s}) = \varphi_{\varepsilon}(s_{p+1}) + \sum_{k=1}^{p} \varphi_{\eta_k}(s_k)$ when Assumption 2.1(i) holds and $\varphi_{\boldsymbol{\eta}, \varepsilon}(\boldsymbol{s}) = -\frac{1}{2}\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}$ when Assumption 2.1(ii) holds. The second derivatives evaluated at $\boldsymbol{s} = (\boldsymbol{b}, -1)v$ less the derivatives evaluated at $\boldsymbol{s} = \boldsymbol{0}$ are,

$$\operatorname{Cov}(X_{k_1}, Y) + \frac{\partial^2 \varphi_{\boldsymbol{X}, Y}(\boldsymbol{s})}{\partial s_{k_1} \partial s_{p+1}}\bigg|_{\boldsymbol{s} = (\boldsymbol{b}, -1)v}$$

$$= \sum_{k=1}^{p} \beta_k \left(\operatorname{Cov}\left(X_{k_1}^*, X_k^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_k}\bigg|_{\boldsymbol{u} = (\boldsymbol{b} - \boldsymbol{\beta})v} \right), \quad 1 \le k_1 \le p,$$

$$\operatorname{Cov}(X_{k_1}, X_{k_2}) + \frac{\partial^2 \varphi_{\boldsymbol{X}, Y}(\boldsymbol{s})}{\partial s_{k_1} \partial s_{k_2}}\bigg|_{\boldsymbol{s} = (\boldsymbol{b}, -1)v}$$

$$= \operatorname{Cov}\left(X_{k_1}^*, X_{k_2}^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}}\bigg|_{\boldsymbol{u} = (\boldsymbol{b} - \boldsymbol{\beta})v}, \quad 1 \le k_1 < k_2 \le p,$$

where all derivatives are finite because of Assumptions 3.3(ii) and (iii). The above expressions are equal to zero when $\boldsymbol{b} = \boldsymbol{\beta}$ so Q is maximized at $\boldsymbol{b} = \boldsymbol{\beta}$ because $Q(\boldsymbol{\beta}) = 0$. Now assume that $\widetilde{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$ and $Q(\widetilde{\boldsymbol{\beta}}) = 0$. Then,

$$0 = \sum_{k=1}^{p} \beta_k \left(\operatorname{Cov}\left(X_{k_1}^*, X_k^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_k}\bigg|_{\boldsymbol{u} = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})v} \right), \quad 1 \le k_1 \le p,$$

$$0 = \operatorname{Cov}\left(X_{k_1}^*, X_{k_2}^*\right) + \frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}}\bigg|_{\boldsymbol{u} = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})v}, \quad 1 \le k_1 < k_2 \le p,$$

and so

$$\frac{\partial^2 \varphi_{\boldsymbol{X}^*}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}}\bigg|_{\boldsymbol{u} = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})v} = -\operatorname{Cov}\left(X_{k_1}^*, X_{k_2}^*\right), \quad 1 \le k_1 \le k_2 \le p.$$

We can now twice apply the fundamental theorem of calculus, because all the functions are continuous by Assumptions 3.3(ii) and (iii), to obtain $\varphi_{\boldsymbol{X}^*}((\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})v) = P_2(v)$. Hence, by

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

Marcinkiewicz (1939), $(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}^*$ is normal. This is a contradiction by Assumption 3.1. Hence Q is uniquely maximized at $\boldsymbol{b}=\boldsymbol{\beta}$.

By Assumptions 3.2 and 3.3(i)–(iii), we have $\frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}'}(\boldsymbol{b},v) \xrightarrow{p} \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{b},v)$ and $\boldsymbol{H}_n(\boldsymbol{b},v) \xrightarrow{p} \boldsymbol{H}(\boldsymbol{b},v)$ almost everywhere and by the standard argument $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. We now use Assumption 3.3, consistency of $\widehat{\boldsymbol{\beta}}$, and identification of $\boldsymbol{\beta}$ to show asymptotic normality.

Let n be large enough. The first-order conditions of (4) are,

$$\begin{aligned}\mathbf{0} &= \int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}}(\widehat{\boldsymbol{\beta}},v)\boldsymbol{W}_n(v)\boldsymbol{H}_n(\widehat{\boldsymbol{\beta}},v)\pi(v)dv \\ &= \int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}}(\widehat{\boldsymbol{\beta}},v)\boldsymbol{W}_n(v)\left(\boldsymbol{H}_n(\boldsymbol{\beta},v)+\frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}'}(\widetilde{\boldsymbol{\beta}},v)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right)\pi(v)dv,\end{aligned}$$

where the first equality follows by Assumption 3.3(v), and the second equality by the mean value theorem that uses Assumptions 3.3(ii) and (iii) ($\widetilde{\boldsymbol{\beta}}$ is on the line segment joining $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$). Rearranging and multiplying by $\sqrt{n}$,

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = &-\left(\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}}(\widehat{\boldsymbol{\beta}},v)\boldsymbol{W}_n(v)\frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}'}(\widetilde{\boldsymbol{\beta}},v)\pi(v)dv\right)^{-1} \\ &\times\left(\int_{-\infty}^{\infty} \frac{\partial \boldsymbol{H}_n}{\partial \boldsymbol{b}}(\widehat{\boldsymbol{\beta}},v)\boldsymbol{W}_n(v)\sqrt{n}\boldsymbol{H}_n(\boldsymbol{\beta},v)\pi(v)dv\right),\end{aligned}$$

where the matrix is invertible by Assumption 3.3(vi). Now we linearize $\boldsymbol{H}_n(\boldsymbol{\beta},v)$ to obtain,

$$\begin{aligned}&\sqrt{n}\boldsymbol{H}_n(\boldsymbol{\beta},v) = \Big(\cdots \\ &\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\Big\{a_{k_2}(v)\left(X_{ik_1}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}_i-Y_i)}-E[X_{k_1}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]\right) \\ &+a_{k_1}(v)\left(X_{ik_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}_i-Y_i)}-E[X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]\right) \\ &+a_0(v)\left(X_{ik_1}X_{ik_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}_i-Y_i)}-E[X_{k_1}X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]\right) \\ &+a_{0k_1k_2}(v)\left(e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}_i-Y_i)}-E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]\right)\Big\}+o_p(1), \\ &\qquad\cdots\Big),\end{aligned}$$

where

$$\begin{aligned}a_{k_1}(v) &= \frac{E[X_{k_1}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]}{(E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}])^2}, \quad a_{k_2}(v) = \frac{E[X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]}{(E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}])^2}, \quad a_0(v) = \frac{1}{(E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}])^2}, \\ a_{0k_1k_2}(v) &= \frac{E[X_{k_1}X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]-2E[X_{k_1}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]E[X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}]}{(E[e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}])^2},\end{aligned}$$

are finite by Assumptions 3.3(ii) and (iii). Using Assumptions 3.3(v) and (vi) to apply Slutsky's Theorem and Fatou's Lemma, and also Assumption 3.2 to apply the Lindeberg–Feller central limit theorem,

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$
$$\xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{S}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta},v)\boldsymbol{W}(v)\boldsymbol{\Omega}(v,w)\boldsymbol{W}(w)\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta},w)\pi(v)\pi(w)dvdw\right)\boldsymbol{S}'\right),$$

where $\boldsymbol{S}=\left(\int_{-\infty}^{\infty}\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}}(\boldsymbol{\beta},v)\boldsymbol{W}(v)\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{b}'}(\boldsymbol{\beta},v)\pi(v)dv\right)^{-1}$ and the terms in $\boldsymbol{\Omega}(v,w)$ are of the form,

$$\begin{aligned}\mathrm{Cov}\Big(&a_{k_2}(v)X_{k_1}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}+a_{k_1}(v)X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}\\&+a_0(v)X_{k_1}X_{k_2}e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)}+a_{0k_1k_2}(v)e^{\mathrm{i}v(\boldsymbol{\beta}'\boldsymbol{X}-Y)},\\&a_{k_4}(w)X_{k_3}e^{\mathrm{i}w(\boldsymbol{\beta}'\boldsymbol{X}-Y)}+a_{k_3}(w)X_{k_4}e^{\mathrm{i}w(\boldsymbol{\beta}'\boldsymbol{X}-Y)}\\&+a_0(w)X_{k_3}X_{k_4}e^{\mathrm{i}w(\boldsymbol{\beta}'\boldsymbol{X}-Y)}+a_{0k_3k_4}(w)e^{\mathrm{i}w(\boldsymbol{\beta}'\boldsymbol{X}-Y)}\Big).\end{aligned}\tag{A.21}$$

https://doi.org/10.1017/S0266466620000250 Published online by Cambridge University Press