Identification of Additive and Polynomial Models of Mismeasured Regressors Without Instruments[*]

Dan Ben-Moshe[†], Xavier D'Haultfœuille[‡], and Arthur Lewbel[§]

The Hebrew University of Jerusalem, Centre de Recherche en Économie et Statistique, and Boston College

Original: March 2015. Revised: July 2016

Abstract

We show nonparametric point identification of a measurement error model with covariates that can be interpreted as invalid instruments. Our main contribution is to replace standard exclusion restrictions with the weaker assumption of additivity in the covariates. Measurement errors are ubiquitous and additive models are popular, so our results combining the two should have widespread potential application. We also identify a model that replaces the nonparametric function of the mismeasured regressor with a polynomial in that regressor and other covariates. This allows for rich interactions between the variables, at the expense of introducing a parametric restriction. Our identification proofs are constructive, and so can be used to form estimators. We establish root-n asymptotic normality for one of our estimators.

JEL codes: C14, C26

Keywords: Nonparametric, semiparametric, measurement error, additive regression, polynomial regression, identification.

[*] Ben-Moshe gratefully acknowledges the support of a grant from the Maurice Falk Institute.

[†] Department of Economics, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel.

[‡] Centre de Recherche en Économie et Statistique (CREST), 15 Boulevard Gabriel Peri 92245 Malakoff Cedex, Paris, France.

[§] Corresponding address: Arthur Lewbel, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA, (617)-552-3678, lewbel@bc.edu, http://www2.bc.edu/arthur-lewbel.

1 Introduction

This paper provides point identification for additive nonparametric and semiparametric models in which some continuously distributed regressor X^* is measured with error, and none of the additional information that is usually used to deal with measurement errors is available. In particular, there are no excluded regressors, no repeated measures, and no validation samples or other outside sources of error distribution information. All we are assumed to observe is a dependent variable Y, correctly measured covariates Z, and the mismeasured X. The main model we consider is

$$E\left[Y \mid X^*, Z\right] = g\left(X^*\right) + h\left(Z\right), \qquad X = X^* + U, \tag{1}$$

where g and h are unknown functions, the true X^* is unobserved, and U is the unobserved measurement error. Our goal is point identification of the functions g and h.

These results extend the literature on nonparametric additive models, widely used in statistics and econometrics (see, e.g., Hastie and Tibshirani, 1990; Linton, 2000; Wood, 2006, and many references therein), to allow for measurement error. A common motivation for additivity (relative to a general nonparametric regression) is to overcome the curse of dimensionality, since additive models typically converge at faster rates than ordinary nonparametric regressions. However, our motivation is different. In our case, we are looking to relax the exclusion restrictions that are ordinarily needed for nonparametric identification with measurement error. If the function h was identically zero, then Z would be excluded from the model and could be used as instruments. Identification could then be based on, e.g., Schennach (2007). See also Hu (2008) and Hu and Schennach (2008) for general approaches based on exclusion restrictions, with respectively discrete and continuous mismeasured regressors. At the other extreme, if no restrictions are placed on $E\left[Y \mid X^*, Z\right]$, then identification would not be possible at all. Additivity substantially relaxes the usual instrumental variables exclusion assumption, while, as we show in this paper, still allowing for identification.

Another way to think of this same framework is to consider a nonparametric structural model of the form $Y = g\left(X^*\right) + \varepsilon^*$ where we replace the usual exogeneity assumption that $E\left[\varepsilon^* \mid X^*, Z\right] = 0$ with the much weaker assumption that $E\left[\varepsilon^* \mid X^*, Z\right] = h\left(Z\right)$. Essentially, we still interpret Z as a vector of instruments, but instead of the usual exclusion restrictions that Z drops out entirely from the model, we allow Z to appear in the model, but only additively. So Z are invalid instruments, but we nevertheless are able to use them like instruments to obtain point identification. Note that we allow $h\left(Z\right)$ to be identically zero, but unlike existing results, we do not require it.

It is often difficult to be certain that candidate instruments satisfy exclusion restrictions, yet consistency of almost all instrumental variable estimators critically depend on instruments satisfying these exclusions. Allowing the instruments to directly affect outcomes, thereby allowing for violations of the exclusion restrictions, should therefore be a valuable contribution for empirical researchers. For example, Chetty et al. (2011) investigate the effects of unobserved early childhood achievement, X^*, on later life outcomes, Y, using observed kindergarten performance, X, and random assignment to kindergarten teachers, Z. However, Kolesár et al. (2014) argue that the assignment to kindergarten teachers may have a direct effect on outcomes, so Z may not satisfy the exclusion assumption, despite having been randomly assigned. They deal with the issue by assuming a linear model (which is a special case of our additive model) and with some uncorrelatedness assumptions. In contrast, we allow for measurement error in X^* and for Y to be a nonparametric function of X^*.

In our main result, we place restrictions on how X^* covaries with Z, and show nonparametric identification of both g and h. We then consider an alternative model where $g\left(X^*\right)$ is replaced with a polynomial in both X^* and Z, still keeping $h\left(Z\right)$ nonparametric. So this alternative model is $E\left[Y \mid X^*, Z\right] = \sum_{j=0}^{J} \sum_{k=1}^{K} \alpha_{jk} Z^j X^{*k} + h\left(Z\right)$. To give an empirical example where such interactions matter, consider the returns to education literature, where it is important to include interactions between education and measures of cognitive ability in wage models. See, e.g., Heckman et al. (2006), who measure cognitive ability using the Armed Forces Qualifying Test, and include in their analysis the covariates schooling, parental income, family background variables and interaction terms.

Consider the general class of models $Y = M\left(X^*, Z\right) + \varepsilon$ with restrictions placed on M and ε. There exists a small literature on point identifying such models, where no additional information like excluded instruments, multiple measures, or known error distributions are available to deal with the measurement error problem. The existing results in this literature impose restrictions on higher order moments of ε (in addition to placing restrictions on M). For example, Chen et al. (2008, 2009) and Schennach and Hu (2013) assume ε is independent of X^*, Erickson and Whited (2002) and Lewbel (1997) assume ε has a conditional third moment of zero, while Klein and Vella (2010) and Lewbel (2012) impose constraints on how the variance of ε depends on X^* and Z. In contrast, the only constraint the present paper imposes on ε is the standard conditional mean (nonparametric regression) assumption $E\left[\varepsilon \mid Z, X^*\right] = 0$. This should be useful in practice because many if not most behavioral models do not provide higher order moment or alternative additional restrictions on ε.

Our model conceptually combines features of Robinson (1988) and Schennach (2007), however, neither of their approaches can be used to establish identification of our model. In the measurement error model $E\left[Y \mid X^*, Z\right] = g\left(X^*\right)$, Schennach (2007) gains identification by exploiting moments like $E\left[YX \mid Z\right]$, which in her model equals $E\left[g\left(X^*\right) X \mid Z\right]$. We cannot use this same method because in our model $E\left[YX \mid Z\right] = E\left[g\left(X^*\right) X + h\left(Z\right) X \mid Z\right]$, and we would not be able to separate the effect of Z on h from the effect of Z on g. Robinson's (1988) estimator of a partially linear model can be interpreted as using the conditional covariance $\operatorname{Cov}\left(Y, X \mid Z\right)$ to project off the unknown function h. This can suffice to identify g when it is linear and when X^* is not mismeasured as in Robinson's model, but not in our case.

Our identification strategy starts by extending the moments from Robinson (1988) and Schennach (2007) to the set of conditional covariances $\operatorname{Cov}\big(Y, (X - E[X \mid Z])^k \mid Z\big)$. These moments are related to convolutions of $g(X^*)$ with the density of $(X^* - E[X^* \mid Z])$. We find that the function $g(X^*)$ can be identified from these moments either when g equals a polynomial of unknown degree, or when it is bounded by a polynomial of unknown degree. When g is bounded by a polynomial, we convert the problem from a system of equations of convolutions to a system of equations that involve products of Fourier transforms. We then use methods from Mattner (1992), D'Haultfœuille (2011) and Zinde-Walsh (2014) to show that these equations exist and have a unique solution, and then solve the resulting system of equations to recover g. When g is a polynomial, we first identify reduced form coefficients by regressing the above conditional covariances on moments of X conditional on Z. We then identify the structural coefficients by solving for them using the reduced form coefficients, analogous to the indirect least squares estimator of linear models. Unlike related polynomial model results in, e.g., Hausman et al. (1991), we can obtain closed-form expressions for these coefficients. Finally we show that we can distinguish between the non-polynomial and polynomial cases.

The next section provides our main model and its identification, and then some alternative ways to achieve identification. One such way is to replace a nonparametric $g(X^*)$ with a polynomial in both X^* and Z. Another variant we consider weakens our main assumptions regarding the relationship between X^* and Z. Our identification strategies are constructive, so estimators can be based on them. Although our main focus is on identification rather than estimation, in Section 3 we show root-n convergence and asymptotic normality in the semiparametric model when g is a polynomial and h is nonparametric. We then present some Monte Carlo simulation studies of the corresponding estimators. Section 5 concludes.

2 The model and its identification

2.1 Main result

We consider the nonparametric additive model

$$\begin{cases} Y &= g(X^*) + h(Z) + \varepsilon, \\ X &= X^* + U, \end{cases} \tag{2}$$

where $Y \in \mathbb{R}$ and $X \in \mathbb{R}$ are observed random variables, $Z \in \mathbb{R}^r$ is an observed random vector, $X^* \in \mathbb{R}$, $\varepsilon \in \mathbb{R}$ and $U \in \mathbb{R}$ are unobserved random variables and $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R}^r \to \mathbb{R}$ are unknown functions to be identified. We impose the following normalization and moment conditions.

Assumption 2.1. *(i) $g(x_0^*) = 0$ for some $x_0^* \in$ Support(X^*); (ii) $E[\varepsilon|X^*, Z, U] = 0$ and (iii) $E[U^k|X^*, Z] = E[U^k]$ for $k \in \{1, 2, 3\}$ and $E[U] = 0$.*

Condition (i) is a harmless location normalization because we can always add a constant to g and subtract it from h. Condition (ii) says that X^* and Z are exogenous regressors, or equivalently that $E\left[Y \mid X^*, Z\right] = g\left(X^*\right) + h\left(Z\right)$. Importantly, this allows for heteroscedasticity of unknown form in ε, as well as not restricting dependence in any higher order moments of ε. As noted in the introduction, this is in sharp contrast to previously existing results that obtain identification without outside information, and may be of considerable importance in practice. Condition (iii) is similar to, but strictly weaker than, the classical measurement error assumption of full independence between U and (X^*, Z).

Assumption 2.2. *$X^* = m(Z) + V$ with $V \perp\!\!\!\perp Z$, $\nu_1 = 0$ and $\nu_2 > 0$, where $\nu_k = E[V^k]$.*

The function $m(Z)$ is defined by $m(Z) = E\left[X^* \mid Z\right]$ and throughout the rest of the paper is identified by $m(Z) = E[X|Z]$. The assumption that V is independent of Z is a strong restriction on how X^* covaries with Z, but it is also a common assumption both in the measurement error literature (see, e.g., Hausman et al., 1991; Schennach, 2007), and in control function type estimators of endogeneity (see, e.g., Newey et al., 1999). In the next subsections, we provide additional results without such an assumption. The condition $\nu_1 = 0$ is a free location normalization, while the condition $\nu_2 > 0$ simply rules out the degenerate case where X^* is a deterministic function of Z, in which case g could obviously not be separately identified from h.

When the function h is known to be identically zero, Schennach (2007) shows that identification of g can be achieved using $E[Y|Z]$ and $E[XY|Z]$. We instead obtain identifying equations on g using conditional covariances rather than conditional means, which are equations that do not depend on the h function. The functions $\mathrm{Cov}(Y, (X - m(Z))^k|Z = z)$ depend on z only through $m(z)$, so we let $q_k(m) = \mathrm{Cov}(Y, (X - m(Z))^k|m(Z) = m)$. These conditional covariances satisfy

$$q_1(m) = E[Vg(m + V)], \tag{3}$$

$$q_2(m) = E[(V^2 - \nu_2)g(m + V)], \tag{4}$$

$$q_3(m) = E[(V^3 - \nu_3)g(m + V)] + 3(m_2 - \nu_2)q_1(m), \tag{5}$$

where $m_k = E[(X - m(Z))^k]$, for $k \geq 1$. These equations, which we use to identify the function g, are functionals of the unknown density of V and of the function g. The identification strategies are different for polynomial and non-polynomial g so we divide identification into two theorems, one for each case, and then present a proposition that shows how to distinguish between the cases.

First, we consider the case with non-polynomial g. For this case, as in Schennach (2007) and Zinde-Walsh (2014), we work with Fourier transforms because Fourier transforms of convolutions are products of Fourier transforms. Despite this fundamental similarity, the details of our proof differ substantially from theirs due primarily to the greater complexity of our identifying equations above. Denote the characteristic function of a random variable A by $\Psi_A(t) = E[e^{itA}]$ and the Fourier transform of a real function f by $\mathcal{F}(f)$. A technical issue here is that f may not be integrable, so $\mathcal{F}(f)$ cannot be defined in the usual sense. In such a case, $\mathcal{F}(f)$ is the Fourier transform of f seen as a tempered distribution. Formal definitions related to the theory of distributions are provided in Appendix A.

Assumption 2.3. *(i) g is bounded by a polynomial and the interior of the support of $\mathcal{F}(g)$ is not empty; (ii) Support$(m(Z)) = \mathbb{R}$; (iii) $E[\exp(|V|\beta)] < \infty$ for some $\beta > 0$ and (iv) Ψ'_{-V}, the derivative of the characteristic function of $-V$, only vanishes at 0.*

Assumption 2.3(i) is weaker than those made by Schennach (2007) and Zinde-Walsh (2014). In particular, we do not require 0 to belong to the support of $\mathcal{F}(g)$. Nevertheless, it rules out polynomials and finite combinations of sine and cosine functions, since the support of their Fourier transforms is discrete. The large support condition on $m(Z)$, which implies that Support$(X^*) = \mathbb{R}$, is also made by Schennach (2007) and is required for our approach based on Fourier transforms. We will not require this large support assumption for the case of a polynomial g below. Assumption 2.3(iii) ensures that Ψ_{-V} is analytic on a strip of the complex plane including the real line, so that Ψ'_{-V} and all higher order derivatives are well defined. Assumption 2.3(iv) is similar to the standard condition in measurement error problems that Ψ_{-V} does not vanish.

Theorem 2.1. *Suppose that Equation* (2) *and Assumptions 2.1, 2.2 and 2.3 hold. Then g and h are identified.*

We provide some intuition for the proof, leaving the details to Appendix A. The expressions for q_k in Equations (3), (4) and (5) can be written as a convolution between g and the density of $-V$. Hence, taking Fourier transforms of these covariances gives

$$\mathcal{F}(q_1) = \mathcal{F}(g) \times (i\Psi'_{-V}), \tag{6}$$

$$\mathcal{F}(q_2) = -\mathcal{F}(g) \times (\Psi''_{-V} + \nu_2\Psi_{-V}), \tag{7}$$

$$\mathcal{F}(q_3) = -\mathcal{F}(g) \times (i\Psi'''_{-V} - 3i(m_2 - \nu_2)\Psi'_{-V} + \nu_3\Psi_{-V}). \tag{8}$$

Intuitively, identification comes from showing that a unique $\mathcal{F}(g)$ solves the above equations. Despite the simplicity of these equations, however, they are functionals of tempered distributions and standard algebraic manipulations may not be valid. Even without this complication, and unlike Schennach (2007), it is not obvious that we do not need more than three equations to solve for the unknown Fourier transform $\mathcal{F}(g)$ and the characteristic function Ψ_{-V} (the moments ν_2 and ν_3 being given by Ψ_{-V}). Further, adding additional equations q_k, for $k > 3$, leads to additional unknown moments and more complicated differential equations. In Appendix A we in fact show that there is a unique g and distribution of V that solve this system.

Now consider the case with polynomial g. Note that we do not impose hereafter that the degree of g, nor an upper bound on it, is known by the econometrician. The assumptions made by Schennach (2007) rule out polynomials, so we adopt a completely different proof strategy for this case that loosely resembles Hausman et al. (1991). Some of the conditions required for the nonparametric g case can be substantially weakened here where g is a polynomial. In particular, we do not require a large support condition on $m(Z)$, but only that it takes a sufficient number of distinct values, so Z can even be discrete.

Assumption 2.4. *(i) g is a polynomial of unknown degree $K > 1$; (ii) $E[|V|^{K+3}] < \infty$ and (iii) the support of $m(Z)$ contains at least $K+1$ elements.*

Theorem 2.2. *Suppose that Equation* (2) *and Assumptions 2.1, 2.2 and 2.4 hold. Then g and h are identified.*

The proof is quite different from the non-polynomial case. Instead of using Fourier transforms we substitute $g(x) = \sum_{k=1}^{K} \alpha_k x^k$ into Equations (3), (4) and (5) to obtain, after some algebra, the regression equations

$$q_1(m) = \sum_{j=0}^{K-1} \left[\sum_{k=j+1}^{K} \alpha_k \binom{k}{j} \nu_{k-j+1} \right] m^j, \tag{9}$$

$$q_2(m) = \sum_{j=0}^{K-1} \left[\sum_{k=j+1}^{K} \binom{k}{j} \alpha_k (\nu_{k-j+2} - \nu_2 \nu_{k-j}) \right] m^j, \tag{10}$$

$$q_3(m) = \sum_{j=0}^{K-1} \left[\sum_{k=j+1}^{K} \binom{k}{j} \alpha_k (\nu_{k-j+3} + 3(m_2 - \nu_2)\nu_{k-j+1} - \nu_3 \nu_{k-j}) \right] m^j. \tag{11}$$

In Appendix A we show that these polynomials in m can be solved to identify the unknown coefficients and moments $(\alpha_1, \ldots, \alpha_K, \nu_2, \ldots, \nu_{K+3})$.

Because the identification proofs in the polynomial and non-polynomial cases are distinct, one may be worried that in practice, we cannot tell which case holds. Fortunately, it is possible to identify a priori whether g is a polynomial or not, using the following proposition.

Proposition 2.1. *Suppose that Equation* (2)*, Assumptions 2.1, 2.2 and either Assumption 2.3 or 2.4 hold. Then g is a polynomial if and only if q_1 is a polynomial.*

Taken together, Proposition 2.1 and Theorems 2.1 and 2.2 imply that under the conditions of Proposition 2.1, g and h are identified.

Finally, our results do not cover the case of a linear function g. The same is true of related identification theorems including Schennach (2007). In our case, this is not a limitation of our proofs, but rather a fundamental feature of the model, as the lemma below shows. In the next subsections, we show how identification of a linear g can be restored with heteroscedasticity.

Lemma 2.1. *Suppose that Equation* (2) *and Assumptions 2.1 and 2.2 hold. Assume that g is linear, $E[|V|] < \infty$ and the support of $m(Z)$ contains at least two elements. Then g is not identified in general.*

We prove this lemma by providing a specific example of a data generating process with a linear g that is not identified. The intuition for this non-identification is that if $g(x^*) = \alpha x^*$, then $q_k(m) = \alpha E[V(V+U)^k]$ does not depend on Z. Hence variation in Z cannot distinguish between α and $E[V(V+U)^k]$.

2.2 Identification through heteroscedasticity of the first stage

Homoscedasticity of V imposed by Assumption 2.2, while helpful for constructing and simplifying moments for nonparametric identification in Theorems 2.1 and 2.2, actually prevents identification in some cases, as emphasized in Lemma 2.1. In this section we achieve identification by allowing for heteroscedasticity.

We note that identification based on assumptions regarding higher order moments can sometimes be fragile. Still, there are many examples of such methods being used successfully in empirical applications. For example, Erickson and Whited (2012) use estimates of higher order moments to deal with measurement error in Tobin's Q model. In the literature on market risk assessment and asset allocation, several papers (e.g. Fama, 1965 and Jondeau and Rockinger, 2006) have used the higher order moments of non-normal asset returns to analyze optimal portfolio choices. Lewbel (2012) exploits heteroscedasticity to identify Engel curves with measurement errors without exclusion restrictions.

Consider again the model in Equation (2) with the following assumption replacing Assumption 2.1.

Assumption 2.5. *(i) $h(z_0) = 0$ for some $z_0 \in Support(Z) \subset \mathbb{R}$; (ii) $E[\varepsilon|X^*, Z, U] = 0$ and (iii) U is independent of (X^*, Z).*

Assumption 2.5(i) places the free location normalization on h instead of g because it will be more convenient in this setting. Assumption 2.5(iii) is stronger than Assumption 2.1(iii) because identification will use characteristic functions that require full independence instead of just having independent low order moments. As discussed earlier, this independence is a standard assumption of measurement error models.

The next assumption replaces Assumption 2.2 by allowing for heteroscedasticity, albeit in multiplicative form.

Assumption 2.6. *$X^* = m(Z) + \sigma(Z)V$ with $Support(m(Z)) = \mathbb{R}$, $V \perp\!\!\!\perp Z$, $E[V] = 0$ and $E[V^2] = 1$. The function $\sigma(.)$ is differentiable and there exists $z_0 \in Support(Z)$ such that $\sigma(z_0) > 0$ and $\sigma'(z_0) \neq 0$.*

Under this assumption $\text{Var}(\sigma(Z)V|Z) = \sigma^2(Z)$ and so the variance of the relationship between X^* and Z is now permitted to depend nonparametrically on Z by equaling the unknown function $\sigma^2(Z)$. The condition $E[V^2] = 1$ is a free normalization because we can always divide V by $E[V^2]$ and multiply $\sigma(Z)$ by the same constant.

Finally, we need to impose regularity conditions similar to Assumption 2.3(iv).

Assumption 2.7. *(i) $E[U^2] < \infty$; (ii) the characteristic functions of U and V do not vanish and (iii) V admits a density with respect to the Lebesgue measure with support equal to the real line.*

Under these conditions, identification proceeds by the following steps (with details in Appendix A). First, using $X - m(Z) = \sigma(Z)V + U$, independence of U, and nonconstant $\sigma(.)$, we show that the distributions of U and V are identified up to the scalar $\sigma_0 = \sigma(z_0)$, the value of the function $\sigma(Z)$ at one point $Z = z_0$. Next we identify g and h, up to the unknown σ_0, using the moment $E[Y \exp(it(X - m(Z)))|Z = z_0]$. Finally, to identify σ_0, we use $\mathrm{Cov}(Y, X|Z = z)$, which holds for all z, and thus provides an infinite number of equations (through variation in z) in the single scalar unknown σ_0. We therefore expect σ_0 to be greatly overidentified. These equations are, however, extremely complicated functions of σ_0, and so we cannot produce low-level assumptions that guarantee that these equations identify σ_0. We therefore impose the following high-level condition.

Assumption 2.8. *The mapping $\sigma \mapsto \left[z \mapsto \sigma^\sigma(z) \int g^\sigma \left(m(z) + \sigma^\sigma(z)v\right) v f_V^\sigma(v) dv\right]$ is injective; where the superscript σ indicates the dependence in σ_0, e.g., g^σ is the g function obtained when σ_0 is set equal to σ.*

Under this condition and the previous ones, the model is identified.

Theorem 2.3. *Suppose that Equation* (2) *and Assumptions 2.6, 2.7 and 2.8 hold. Then g and h are identified.*

This result relies on Assumption 2.8, which despite being high-level, can be verified in some particular settings. For example, the following proposition shows that this assumption holds when U and V are normal and g is linear.

Proposition 2.2. *Suppose that Equation* (2) *and Assumptions 2.5 and 2.6 hold. Suppose also that g is linear, not constant, and U and V are normally distributed. Then Assumption 2.8 holds and thus g and h are identified.*

This proposition is of interest in view of Lemma 2.1, where we showed that the model is not identified under homoscedasticity in the equation $X^* = m(Z) + V$ when g is linear and U and V are normal. Now, replacing Assumption 2.2 with Assumption 2.6, the model with linear g and normal U and V is identified, since we have verified in this case that Assumption 2.8 holds. We conjecture that heteroscedasticity identifies the model more generally. Basically, the heteroscedasticity function $\sigma(z)$ provides additional variation to help in the identification of g, and is identified up to σ_0 using only X and Z. Then using Y, we have an infinite number of additional equations that should generally suffice to identify the single scalar σ_0.

2.3 A polynomial restriction with interaction terms

In this subsection, we replace the function g with a polynomial in both $X^* \in \mathbb{R}$ and $Z \in \mathbb{R}$, so

$$\left\{ \begin{array}{lcl} Y & = & g\left(X^*, Z\right) + h\left(Z\right) + \varepsilon = \sum_{j=0}^{J} \sum_{k=1}^{K} \alpha_{jk} Z^j X^{*k} + h\left(Z\right) + \varepsilon, \\ X & = & X^* + U. \end{array} \right. \tag{12}$$

Here, $h : \mathbb{R} \to \mathbb{R}$ is an unknown function and $\{\alpha_{jk}\}_{j,k}$ are unknown parameters with $\alpha_{jK} \neq 0$ for some j. This model is more general than Model (2) in relaxing additivity by allowing interactions between X^* and Z, but it is less general in constraining g to be a polynomial. The model extends readily to the cases where Z or X^* are vectors, and in Appendix B we also show that identification can be adapted quite easily to the case of multiplicative instead of additive measurement errors.

The following assumption replaces Assumption 2.1.

Assumption 2.9. *(i) $E[\varepsilon|X^*, Z, U] = 0$ and (ii) $E[U^k|X^*, Z] = E[U^k]$ for $k \in \{1, 2, \ldots, K+1\}$ and $E[U] = 0$.*

When $K > 2$ Assumption 2.9(ii) is stronger than Assumption 2.1(iii) because higher order moments of U are assumed to not depend on X^* or Z, though in practice, one would typically assume that the measurement error is independent of the true covariates, which would then satisfy either assumption regardless of K. Finally, we do not need to include an explicit location normalization on g here, because Equation (12) already satisfies the location normalization $g(0, z) = \sum_{j=0}^{K} \sum_{k=1}^{K} \alpha_{jk} z^j 0^k = 0$.

Equation (12) and Assumption 2.9 imply that

$$\mathrm{Cov}(X, Y|Z) = \sum_{j=0}^{J} Z^j \sum_{k=1}^{K} \alpha_{jk} \left(E[X^{*k+1}|Z] - E[X|Z]\alpha_{jk} E[X^{*k}|Z] \right). \tag{13}$$

To identify this model, we recursively substitute $X^k = (X^* + U)^{*k}$ into the binomial expansion

$$X^{*k} = X^k - \sum_{l=0}^{k-1} \binom{k}{l} X^{*l} U^{k-l}$$

and end up expressing $\mathrm{Cov}(X, Y|Z)$ as a linear combination of terms of the form $Z^j E[X^k|Z]$ and $Z^j E[X|Z]E[X^k|Z]$. All we need now for identification is to replace Assumptions 2.2 and 2.4 with a rank condition that allows us to obtain the coefficients on these conditional moments.

Assumption 2.10. *Define*

$$Q(Z) = (E[X^{K+1}|Z], -E[X^K|Z]E[X|Z], \ldots, E[X^2|Z], -E[X|Z]E[X|Z], E[X|Z], 1)',$$
$$R(Z) = (Z^0 Q(Z)', Z^1 Q(Z)', \ldots, Z^J Q(Z)')'.$$

Let $E\left[R(Z)\, R(Z)'\right]$ be finite and nonsingular.

Assumption 2.10, like Assumption 2.4(iii), allows the support of Z to be limited or discrete, even if X is continuous. However, the rank condition requires that Z have at least $K+2$ points of support. The assumption gives us identification by ensuring that variation in Z induces sufficient relative variation in the moments $E[X^k|Z]$ for $k = 1, \ldots, K+1$. The vector $R(Z)$ includes, for example, $E[X^2|Z]$ and $(E[X|Z])^2$ so that nonsingularity of $E\left[R(Z)\, R(Z)'\right]$ requires relative variation in $E[X|Z]$ and $\mathrm{Var}(X|Z)$. Relative variation is also required for higher conditional moments of X. Assumption 2.10 therefore conflicts with Assumption 2.2, where $\mathrm{Var}(X|Z)$ is constant, and so should be considered as an alternative to it.

Theorem 2.4. *Suppose that Equation* (12) *and Assumptions 2.9 and 2.10 hold. Then the functions g and h and the moments $E[U], E[U^2], \ldots, E[U^{K+1}]$ are identified.*

The proof is based directly on the above covariance expansion in Equation (13). It is similar to the proof of Theorem 2.2 but instead of first identifying g and moments of V, Theorem 2.4 first identifies g and moments of U. The proof uses Assumption 2.10 to identify the reduced form coefficients on $R(Z)$ by projecting $\text{Cov}(X, Y|Z)$ on $R(Z)$. The coefficients on $R(Z)$ are known but complicated functions of $\alpha_{j1}, \ldots, \alpha_{jK}$ and $E[U], \ldots, E[U^{K+1}]$, which are then manipulated to recover these parameters and moments and hence g.

Both Theorems 2.2 and 2.4 identify g when it is a nonlinear polynomial in X^* only. However, unlike the theorems in Section 2.1, Theorem 2.4 can identify a linear g. This is shown by the following example, which is the classical linear errors-in-variables model but with an additional nonparametric term that is a function of a correctly measured Z.

Example 2.1. *Suppose that Equation* (12) *holds with $Y = \alpha_1 X^* + h(Z) + \varepsilon$, i.e., g is linear, and Assumptions 2.9 and 2.10 hold. Then g and h are identified.*

Results like Reiersøl (1950) show that without Z this model would not be identified under normality. In contrast, by projecting off Z and using it as an instrument for X^*, Theorem 2.4 shows that this model can be identified even when the model and measurement errors are normal. As in the previous subsection, the key for identification here is that Assumption 2.9 requires $\text{Var}\,(X|Z)$ to vary with Z, thereby requiring heteroscedasticity in the relationship between X^* and Z. The main tradeoff between this result and that of the previous subsection is that here we do not require a location-scale model, but now g is restricted to be parametric.

3 Semiparametric estimation

In this section, we show how the steps in our identification proof can be used to construct estimators of g and h in our main Model (2) when g is a polynomial while h is left unspecified. The estimator is simple in this case, as it only involves linear ordinary least squares estimators and a minimum distance step that can be achieved without numerical optimization. We show that the corresponding estimators are asymptotically normal with the parametric part achieving a root-n convergence rate. Nonparametric estimation of g based on Theorem 2.1 and testing between polynomial and non-polynomial g are considered in Appendix C.

Let $g(x) = \sum_{k=1}^{K} \alpha_k x^k$ where $K > 1$ is assumed to be known. We impose the normalization $g(0) = 0$ so that $\alpha_0 = 0$. As proved in Proposition 2.1 and shown in Equations (9), (10) and (11), q_1, q_2 and q_3 are polynomials in this setting. The idea is then to estimate first the coefficients $\{\beta_{kj}\}_{j=0}^{K-1}$, for $k \in \{1, 2, 3\}$, of these polynomials, and then recover $(\alpha_1, \ldots, \alpha_K)$ and $(\nu_2, \ldots, \nu_{K+3})$ (with $\nu_k = E[V^k]$) from the estimates of β_{kj}. Finally, we estimate h.

First consider the estimation of $\{\beta_{kj}\}_{j=0}^{K-1}$, for $k \in \{1,2,3\}$. We have

$$q_k(m) = E\left[Y\left((X-m(Z))^k - E[(X-m(Z))^k|m(Z)]\right)|m(Z)=m\right]$$

and, under Assumptions 2.1 and 2.2,

$$E[(X-m(Z))^k|m(Z)] = E[(U+V)^k|m(Z)] = E[(U+V)^k] = E[(X-m(Z))^k].$$

Letting $m_k = E[(X-m(Z))^k]$, we obtain

$$q_k(m) = E\left[Y\left((X-m(Z))^k - m_k\right)|m(Z)=m\right]. \tag{14}$$

This equality is convenient because it shows that q_k corresponds to a simple conditional expectation. Here, $m(Z) = E[X|Z]$ can be estimated by any uniformly consistent nonparametric regression estimator of $E[X|Z]$. The case where Z is discrete is straightforward because $m(.)$ can be estimated at the root-n rate by simple averages, so we focus hereafter on the continuous Z case. We consider a series estimator for $m(.)$. For any positive integer L, let $p^L(z) = (p_{1L}(z), \dots, p_{LL}(z))'$ be a vector of basis functions and $P^L = (p^L(Z_1), \dots, p^L(Z_n))$. We then estimate $m(z)$ by

$$\widehat{m}(z) = p^{L_n}(z)'\left(P^{L_n}P^{L_n\prime}\right)^{-} P^{L_n}\left(X_1, \dots, X_n\right)',$$

where $(.)^-$ denotes a generalized inverse and $(L_n)_{n\in\mathbb{N}}$ is a sequence of integers tending to infinity at a rate specified below. Then we estimate $m_k = E[(X-m(Z))^k]$ for $k \in \{2,3\}$ by simply taking the average of $(X_i - \widehat{m}(Z_i))^k$. Note that $m_1 = 0$ need not be estimated. Next, we estimate $\{\beta_{kj}\}_{j=0}^{K-1}$ by regressing $\widehat{Q}_k = Y\left[(X-\widehat{m}(Z))^k - \widehat{m}_k\right]$ on $(1, \widehat{m}(Z), \dots, \widehat{m}(Z)^{K-1})$. We denote hereafter by $\widehat{\theta}$ the estimator of $\theta_0 = (m_2, m_3, \beta_{10}, \dots, \beta_{3K-1})'$.

In a second step, we can then use a classical minimum distance estimator (see, e.g., Wooldridge, 2002, Section 14.6) to estimate the $2K+2$ unknown parameters $\eta_0 = (\tau_2, \tau_3, \alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+3})$ from $\theta_0 = \Pi(\eta_0)$, where $\Pi(\eta_0) = (\Pi_1(\eta_0), \dots, \Pi_{3K+2}(\eta_0))'$ and

$$\Pi_j(\eta) = \left|\begin{array}{ll} \tau_{j+1} & \text{if } j \in \{1,2\}, \\ \sum_{k=1}^{K-j+3}\binom{k+j-3}{j-3}\alpha_{k+j-3}\nu_{k+1} & \text{if } j \in \{3,\dots,K+2\}, \\ \sum_{k=1}^{2K+3-j}\binom{k+j-K-3}{j-K-3}\alpha_{k+j-K-3}(\nu_{k+2}-\nu_2\nu_k) & \text{if } j \in \{K+3,\dots,2K+2\}, \\ \sum_{k=1}^{3K+3-j}\binom{k+j-2K-3}{j-2K-3}\alpha_{k+j-2K-3}(\nu_{k+3}-3(\tau_2-\nu_2)\nu_{k+1}-\nu_3\nu_k) & \text{if } j \in \{2K+3,\dots,3K+2\}. \end{array}\right.$$

We refer the reader to the proof of Theorem 2.2 in Appendix A on how we obtain these equations and note that $\tau_k = m_k$ is introduced here to ensure that $\Pi(.)$ is a function of η only. We also recall that $\nu_1 = 0$ need not be estimated. We therefore estimate η_0 by

$$\widehat{\eta} = \arg\min_{\eta\in\mathcal{H}}\left(\widehat{\theta} - \Pi(\eta)\right)' W_n\left(\widehat{\theta} - \Pi(\eta)\right), \tag{15}$$

where W_n is a random, symmetric positive definite matrix and $\mathcal{H}$ is a compact set. Using $\widehat{\eta}$, we estimate $g(.)$ by

$$\widehat{g}(x) = \sum_{k=1}^{K}\widehat{\alpha}_k x^k.$$

Note that in the proof of Theorem 2.2, identification comes from a closed-form expression of η_0 in terms of β_0, which can also be used for estimation. The corresponding estimator is not efficient in general, though. On the other hand, we can use it to compute a one-step estimator (see, e.g., van der Vaart, 2000, Section 5.7) based on (15). Such an estimator is asymptotically efficient and does not require numerical optimization.

Our asymptotic results on $\widehat{\eta}$ and $\widehat{g}(.)$ are based on the following conditions.

Assumption 3.1. *We observe a sample $(X_i, Y_i, Z_i)_{i=1,\dots,n}$ of i.i.d. variables with the same distribution as (X, Y, Z).*

Assumption 3.2. *(i) θ_0 and η_0 belong to the interior of two compact sets, Θ and $\mathcal{H}$ respectively;*

(ii) $E[X^6] < \infty$, $E[Y^2] < \infty$ and $E[Y^2X^6] < \infty$;

(iii) $z \mapsto E[X|Z = z]$ is not constant and is s times continuously differentiable on Support$(Z) \subseteq \mathbb{R}^r$, with $s > 3r$;

(iv) Support(Z) is a Cartesian product of compact intervals on which Z has a probability density function that is bounded away from zero;

(v) The series terms $p_{\ell L_n}$, $1 \leq \ell \leq L_n$, are products of polynomials orthonormal with respect to the uniform weight. Moreover, $L_n^{4(s/r-1)}/n \to \infty$ and $L_n^7/n \to 0$;

(vi) $W_n \xrightarrow{p} W$, which is nonstochastic and positive definite and

(vii) The matrix $J = \partial\Pi/\partial\eta_{|\eta=\eta_0}$ has full rank.

These are standard regularity conditions (e.g., Frölich, 2007). In Condition (iii), the fact that $z \mapsto E[X|Z = z]$ is not constant may be seen as our rank condition. In the polynomial model, in contrast to the nonparametric case, we do not require large variation on $z \mapsto E[X|Z = z]$ to achieve identification but just that Z has some effect on the conditional expectation of X.

Theorem 3.1. *Suppose that Equation (2) and Assumptions 2.1, 2.2, 2.4, 3.1 and 3.2 hold. Then*

$$\sqrt{n}\left(\widehat{\eta} - \eta_0\right) \xrightarrow{d} \mathcal{N}\left(0, (J'WJ)^{-1}J'WG^{-1}HG^{-1\prime}WJ(J'WJ)^{-1}\right),$$

where G and H are defined respectively in Equations (31) and (35) in Appendix A. The optimal weighting matrix is $W^* = (G^{-1}HG^{-1\prime})^{-1}$. Finally, $\widehat{g}(x)$ is also asymptotically normal for any $x \in \mathbb{R}$.

Finally, we estimate h using $h(Z) = E[Y|Z] - E[g(X^*)|Z]$. The term $E[Y|Z]$ can be estimated by standard nonparametric regression, while

$$E[g(X^*)|Z = z] = \sum_{k=0}^{K} \alpha_k E[(m(z) + V)^k] = \sum_{j=0}^{K} \left[\sum_{k=j}^{K} \binom{k}{j} \alpha_k \nu_{k-j}\right] m(z)^j$$

can be estimated by

$$\widehat{E}[g(X^*)|Z=z] = \sum_{j=0}^{K}\left[\sum_{k=j}^{K}\binom{k}{j}\widehat{\alpha}_k\widehat{\nu}_{k-j}\right]\widehat{m}(z)^j$$

and we can then estimate h by $\widehat{h}(z) = \widehat{E}[Y|Z=z] - \widehat{E}[g(X^*)|Z=z]$. Because $\widehat{\eta}$ is root-n consistent by Theorem 3.1 above, it will have no effect on the asymptotic distribution of $\widehat{h}(z)$: only the nonparametric estimation of $E[X|Z=z]$ and $E[Y|Z=z]$ will matter. Under standard regularity conditions, $\widehat{h}(z)$ will be asymptotically normal (with nonparametric rate of convergence) by the joint asymptotic normality of $\widehat{m}(z)$ and $\widehat{E}[Y|Z=z]$ and the delta method.

Note that estimation based on Theorem 2.4 can be done in a similar way as the polynomial case above. Specifically, $E[X^k|Z]$ is estimated using a standard nonparametric estimator and β is then estimated by regressing $\widehat{\text{Cov}}(X,Y|Z)$ on $\widehat{R}(Z)$:

$$\widehat{\beta} = \widehat{E}[\widehat{R}(Z)\widehat{R}(Z)'^{-1}]\widehat{E}[\widehat{R}(Z)\widehat{\text{Cov}}(X,Y|Z)].$$

Then $(\alpha_{01},\ldots,\alpha_{KK})$, $(E[U^2],\ldots,E[U^{K+1}])$, and $(E[X^*|Z],\ldots,E[X^{*K}|Z])$ are estimated by replacing β with $\widehat{\beta}$. Finally, we can estimate $h(Z)$ by $\widehat{h}(Z) = \widehat{E}[Y|Z] - \sum_{j=0}^{J} Z^j \sum_{k=1}^{K}\widehat{\alpha}_{jk}\widehat{E}[X^{*k}|Z]$.

4 Monte Carlo Simulations

In this section, we explore the finite sample properties of our measurement error-corrected estimators developed above. Our simulation designs are chosen to illustrate the stability of the estimator for different amounts of measurement error.

Data $\{Y_i, X_i, Z_i\}_{i=1}^n$ are generated from one of the following two model designs:

$$Y = \alpha_1 X^* + \alpha_2 X^{*2} + \ln(|Z|) + \varepsilon, \qquad X^* = Z + V, \qquad X = X^* + U \qquad \text{(Model 1)}$$
$$Y = \alpha_1 X^* + \ln(|Z|) + \varepsilon, \qquad X^* = Z + ZV, \qquad X = X^* + U \qquad \text{(Model 2)}$$

where $\alpha_1 = \alpha_2 = 1$ and $V \sim \mathcal{N}(0,1)$, $\varepsilon \sim \mathcal{N}(0,1)$, $Z \sim \mathcal{N}(0,2)$ and $U \sim \mathcal{N}(0,\sigma_U^2)$, are independent. The various choices of $\sigma_U^2 \in \{0, 1/4, 1\}$ allow us to consider different amounts of measurement error. Model 1 is identified by Theorem 2.2 while Model 2, for which $g(.)$ is linear, is identified because of the heteroscedasticity by Proposition 2.2 or Theorems 2.4. We consider 100 simulations with sample size $n \in \{500, 1000, 2000\}$.

Even though we have parameterized g as a polynomial, for estimation we do not assume that the $h\,(Z)$ function (given by $\ln(|Z|)$ in the simulations) is known or parameterized, and therefore the regression is not parametrically specified. If σ_U^2 and hence the measurement errors were zero, then the model would be equivalent to a partially linear specification. We therefore compare our estimator to the partially linear model estimator of Robinson (1988).

Table 1 compares biases, standard deviations (SD) and root mean squared errors (RMSE) in Model 1 and Model 2, respectively, using (a) Robinson's (1988) estimator, which first nonparametrically estimates conditional moments and then estimates α by regressing $Y - \widehat{E}[Y|Z]$ on $X - \widehat{E}[X|Z]$ and (b) the measurement error-corrected estimator (MEC) we propose in Section 3.

Table 1: Performances of the measurement error corrected (MEC) and Robinson's estimators with $V \sim \mathcal{N}(0,1)$

				α_1			α_2	
Model	Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
Model 1	MEC	0	-0.019	0.102	0.103	0.009	0.069	0.069
		1/4	-0.017	0.115	0.116	-0.015	0.065	0.066
		1	-0.028	0.132	0.132	0.001	0.072	0.072
	Robinson's	0	0.004	0.032	0.033	0.002	0.012	0.012
		1/4	-0.201	0.098	0.223	-0.200	0.033	0.203
		1	-0.492	0.113	0.505	-0.505	0.038	0.506
Model 2	MEC	0	-0.020	0.043	0.047			
		1/4	-0.003	0.052	0.052		—	
		1	0.005	0.057	0.057			
	Robinson's	0	0.015	0.033	0.036			
		1/4	-0.197	0.028	0.199		—	
		1	-0.495	0.026	0.496			

Notes: results from 100 simulations of sample size $1,000$.

The MEC estimator for Model 1 nonparametrically estimates the conditional moments $\widehat{\text{Cov}}(X, Y|Z)$, $\widehat{\text{Cov}}(X - E[X|Z]^2, Y|Z)$ and $\widehat{\text{Cov}}(X - E[X|Z]^3, Y|Z)$, regresses these estimated covariances on $\widehat{E}[X|Z]$ and an intercept to obtain $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$ and finally uses the minimum distance estimator to estimate α_1 and α_2. The quadratic function g in Model 1 is exactly identified because the number of unknowns is exactly equal to the number of equations $(\dim(\tau_2, \tau_3, \alpha_1, \alpha_2, \nu_2, \nu_3, \nu_4, \nu_5) = 8 = 3K + 2 = \dim(\theta_0))$, so we have closed-form solutions for the parameters that make the minimum distance exactly zero. Thus, the minimum distance estimator is in this case equivalent to using sample analogs of the formulas for the parameters in Equation (24) and the paragraph that follows it. For example $\widehat{\alpha}_1 = \widehat{\beta}_{11} - \widehat{\alpha}_2\widehat{\nu}_3/\nu_2$ and $\widehat{\alpha}_2 = \widehat{\beta}_{12}/\widehat{\nu}_2$. The MEC estimator for Model 2 nonparametrically estimates the conditional moment $\widehat{\text{Cov}}(X, Y|Z)$, regresses it on $(\widehat{E}[X|Z])^2$, $\widehat{E}[X^2|Z]$ and an intercept to obtain $\widehat{\beta}$. Then, by Equation (28) and the paragraph that follows it, we use $\widehat{\alpha}_1 = \widehat{\beta}_1$.

Two observations can be made. First, the MEC estimators are quite insensitive to the amount of measurement error, with very little increase of their RMSE as σ_U^2 grows. Second, Robinson's (1988) estimator has the smallest RMSE when there is no measurement error but large RMSE even with small amounts of measurement error. For example, the average of the estimates using Robinson

(1988) in Model 1 with $\sigma_U^2 = 1/4$, which represents about 6% of the total variance of X, is over 2 standard deviations away from α_1 and 6 standard deviations away from α_2.

Table 2 shows the biases, SDs and RMSEs of the MEC estimators in Models 1 and 2 respectively with $\sigma_U^2 = 1$ and sample sizes 500, $1,000$ and $2,000$. Even when $n = 500$, the MEC estimators have small biases and the standard deviations decline with sample size.

Table 2: Performances of the MEC estimator as a function of n

			α_1			α_2	
Model	n	bias	SD	RMSE	bias	SD	RMSE
Model 1	500	-0.027	0.196	0.197	-0.016	0.109	0.109
	$1,000$	-0.018	0.132	0.132	0.001	0.072	0.072
	$2,000$	0.013	0.093	0.093	0.015	0.044	0.046
Model 2	500	0.014	0.079	0.079			
	$1,000$	-0.003	0.067	0.067		—	
	$2,000$	-0.004	0.054	0.054			

Notes: results from 100 simulations.

To test the robustness of the results to fat-tailed, bimodal or discontinuous densities we conducted Monte-Carlo simulations where the measurement error had a t, bimodal or uniform distribution. We found that although fat tails have a slight adverse affect on all the estimators the results are qualitatively unchanged. Specifically, the MEC estimators had relatively small RMSEs and were almost the same for all choices of σ_U^2. Robinson's estimator had the lowest RMSE when there was no measurement error but high RMSEs with even small amounts of measurement error. We refer to Appendix D for tables with these additional simulations.

5 Conclusions

Observing only Y, X, and Z, we have provided conditions for point identification of the models $Y = g(X^*) + h(Z) + \varepsilon$ and $Y = \sum_{j=0}^{K} \sum_{k=1}^{K} \alpha_{jk} Z^j X^{*k} + h(Z) + \varepsilon$, where g and h are unknown functions, and X is a mismeasured version of X^*. Unlike previous results in the literature that identify measurement error models without exclusion restrictions or other outside information, we place no assumptions on ε other than having conditional mean zero.

Measurement error is a common source of endogeneity in economic models, and two of the classic ways to obtain identification in structural econometric models is either by exclusion restrictions or by imposing parametric functional forms. This paper's results can be interpreted as a middle ground between these cases. The potential instrument Z is not excluded, and can affect the outcome through the unknown function h, but the model either rules out interactions between X^* and Z, or only allows parametric (polynomial) interactions. These types of restrictions on interaction terms are much weaker than imposing exclusion restrictions, but as we show, still suffice for model identification.

Our identification proofs are constructive, and so can be used to form estimators. In the polynomial case, we have provided a two-step estimator that consists only of linear ordinary least squares to obtain reduced form parameters, followed by a minimum distance estimator. No numerical optimization is required for this estimator because we have a closed-form expression for the structural model in terms of these reduced form parameters. This closed form can be used to obtain a consistent estimator. Then a one-step estimator, as in van der Vaart (2000, Section 5.7), based on the minimum distance program in Equation (15), can be applied for efficiency.

Estimation based on our identification constructions in the nonparametric g case is more challenging. A possible route, detailed in Appendix C, is to consider plug-in estimators. While the statistical properties of such estimators may be hard to establish, they are also computationally tractable. F_k are tempered distributions but the estimators $\widehat{F_k}$ are proper functions, so they can be handled more easily. Also, Fourier transforms and their inverses can be computed numerically at low computation cost using Fast Fourier Transforms. Further, the nonparametric g case does not require any numerical searches or optimization procedures (apart from bandwidth selection). Still, one might want to consider alternative estimators such as sieve maximum likelihood for the nonparametric g case, or to provide estimators that are the same in both polynomial and non-polynomial cases. Analysis of such alternative estimators is left for future research.

One general application of our results would be to test instrument validity, i.e., testing whether the standard exclusion assumption for identification holds. This could be done by estimating $h\left(Z\right)$ and testing whether this estimated function is identically zero. Our model also nests standard linear and partially linear models, and so could be applied in some of those contexts as well.

References

Chen, X., Hu, Y. and Lewbel, A. (2008), 'Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments', *Economics Letters* **100**, 381–384.

Chen, X., Hu, Y. and Lewbel, A. (2009), 'Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information', *Statistica Sinica* **19**, 949–968.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011), 'How does your kindergarten classroom affect your earnings? evidence from project star', *The Quarterly Journal of Economics* **126**(4), 1593–1660.

D'Haultfœuille, X. (2011), 'On the completeness condition in nonparametric instrumental regression', *Econometric Theory* **27**, 460–471.

Erickson, T. and Whited, T. M. (2002), 'Two-step gmm estimation of the errors-in-variables model using high-order moments', *Econometric Theory* **18**, 776–799.

Erickson, T. and Whited, T. M. (2012), 'Treating measurement error in tobin's q', *Review of Financial Studies* **25**(4), 1286–1329.

Fama, E. F. (1965), 'The behavior of stock-market prices', *Journal of business* pp. 34–105.

Frölich, M. (2007), 'Nonparametric iv estimation of local average treatment effects with covariates', *Journal of Econometrics* **139**(1), 35–75.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman and Hall.

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

Heckman, J. J., Urzua, S. and Vytlacil, E. (2006), 'Understanding instrumental variables in models with essential heterogeneity', *The Review of Economics and Statistics* **88**(3), 389–432.

Hu, Y. (2008), 'Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution', *Journal of Econometrics* **144**(1), 27–61.

Hu, Y. and Schennach, S. M. (2008), 'Instrumental variable treatment of nonclassical measurement error models', *Econometrica* **76**(1), 195–216.

Jondeau, E. and Rockinger, M. (2006), 'Optimal portfolio allocation under higher moments', *European Financial Management* **12**(1), 29–55.

Klein, R. and Vella, F. (2010), 'Estimating a class of triangular simultaneous equations models without exclusion restrictions', *Journal of Econometrics* **154**(2), 154–164.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E. and Imbens, G. W. (2014), 'Identification and inference with many invalid instruments', *Journal of Business & Economic Statistics* (just-accepted), 00–00.

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* pp. 1201–1213.

Lewbel, A. (2012), 'Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models', *Journal of Business and Economic Statistics* **30**, 67–80.

Linton, O. B. (2000), 'Efficient estimation of generalized additive nonparametric regression models', *Econometric Theory* **16**, 502–523.

Mattner, L. (1992), 'Completeness of location families, translated moments, and uniqueness of charges', *Probability Theory and Related Fields* **92**, 137–149.

Newey, W. K. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica* **62**(6), pp. 1349–1382.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**(1), 147 – 168.

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

Newey, W. K., Powell, J. L. and Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

Reiersøl, O. (1950), 'Identifiability of a linear relation between variables which are subject to error', *Econometrica* pp. 375–389.

Robinson, P. M. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**, 931–954.

Rudin, W. (1987), *Real and Complex Analysis*, McGraw-Hill.

Schennach, S. (2007), 'Instrumental variables estimation of nonlinear errors-in-variables models', *Econometrica* **75**, 201–239.

Schennach, S. and Hu, Y. (2013), 'Nonparametric identification and semiparametric estimation of classical measurement error models without side information', *Journal of the American Statistical Association* **108**, 177–186.

Schwartz, L. (1973), *Thorie des distributions, deuxime dition*, Hemann.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Wood, S. (2006), *Generalized additive models: an introduction with R*, CRC press.

Wooldridge, J. M. (2002), *The Econometrics of Cross Section and Panel Data*, MIT Press.

Zheng, J. X. (1996), ‘A consistent test of functional form via nonparametric estimation techniques’, *Journal of Econometrics* **75**(2), 263 – 289.

Zinde-Walsh, V. (2014), ‘Measurement error and deconvolution in spaces of generalized functions’, *Econometric Theory* **30**, 1207–1246.

A Proofs

A.1 Definitions related to the theory of distributions

We recall here some definitions related to the theory of distributions (see, e.g., Schwartz, 1973). The Schwartz space $\mathcal{S}$ is the subspace of C^∞ functions s such that for any $(k,j) \in \mathbb{N}^2$, $\lim_{x\to\pm\infty} |x|^j s^{(k)}(x) = 0$. Tempered distributions are then linear forms defined on $\mathcal{S}$. We say that f in $\mathcal{S}'$, the space of tempered distributions, is zero on an open set $\mathcal{O}$ if for any $\phi \in \mathcal{S}$ with support included in $\mathcal{O}$, $f(\phi) = 0$. Then the support of f is the complement of the largest open set on which f is zero. For any $f \in \mathcal{S}'$, its Fourier transform is the unique $F \in \mathcal{S}'$ satisfying, for any $\phi \in \mathcal{S}$, $F(\phi) = f(\mathcal{F}(\phi))$, where $\mathcal{F}(\phi) = \int_{\mathbb{R}} \exp(itu)\phi(u)du$ denotes the Fourier transform of ϕ, seen as a function in $L^1(\mathbb{R})$. When f is a function bounded by a polynomial, the linear form $\widetilde{f} : s \mapsto \int f(u)s(u)du$ defined on $\mathcal{S}$ is a tempered distribution. In the absence of ambiguity, we assimilate f with $\widetilde{f}$ hereafter.

A.2 Proof of Theorem 2.1

We proceed in two main steps. First, we show that Equations (6), (7) and (8) hold. Then we show that g and h are identified from these equations.

1. Equations (6), (7) and (8) hold.

We only prove that Equation (6) holds, as the exact same reasoning applies to Equations (7) and (8). We use a similar approach as Mattner (1992) in the beginning of the proof of his Theorem 1.3. We check in particular that the conditions of his Lemma 2.1 apply. For that purpose, let $g_n = g \times \mathbb{1}_{[-n,n]}$ and f be the linear form defined by $f(\phi) = E[\phi(-V)V]$. Mattner's h function is q_1 in our context. First, because g is bounded by a polynomial, it is tempered. Second, because $E[|V|] < \infty$, the total variation measure associated with f is finite, which implies that f is tempered (see Schwartz, 1973, Théorème VII p. 242). Third, by assumption, there exists $C > 0, k \geq 1$ such that for all x, $|g(x)| \leq C(1 + |x|^k)$. Then the inequality $(x+y)^k \leq 2^{k-1}(x^k + y^k)$ yields

$$|q_1(m)| \leq E[|V||g(m+V)|] \leq C\left[E[|V|] + 2^{k-1}(E[|V|^{k+1}] + E[|V|]m^k)\right], \tag{16}$$

with $E[|V|^{k+1}] < \infty$ by Assumption 2.3(ii). Thus q_1 is bounded by a polynomial and as such, is tempered. Fourth, because g_n is a tempered distribution with compact support, it belongs to the space of quickly decreasing distributions $\mathcal{O}'_C$ (see Schwartz, 1973, p. 244). Reasoning exactly as in D'Haultfœuille (2011, pp. 469-470), we also have $g_n \to g$ in $\mathcal{S}'$. Finally, let us show that $q_{1n} = f \star g_n \to q_1$ in $\mathcal{S}'$. Let Φ be any bounded set in $\mathcal{S}$, the space of rapidly decreasing functions. There exists (see Schwartz, 1973, p. 235) a continuous function b with $b(x) = o(|x|^{-j})$ as $|x| \to \infty$ and for every j, such that $|\phi(x)| \leq b(x)$ for every $x \in \mathbb{R}$ and every $\phi \in \Phi$. Then (16) implies that

$b \times q_1$ is integrable. The same inequality (16) applies to q_{1n}, implying that $b \times (q_{1n} - q_1)$ is also integrable. Further, by dominated convergence,

$$\sup_{\phi \in \Phi} \left| \int \phi(m)(q_{1n}(m) - q_1(m))dm \right| \leq \int \int b(m) \mathbb{1}_{^c[-n,n]}(m-v)|vg(m-v)|dmdP^{-V}(v) \longrightarrow 0,$$

where P^{-V} denotes the probability measure of $-V$. Hence, all conditions of Mattner's Lemma 2.1 are fulfilled. As a result, for any open set $\mathcal{U} \subset \mathbb{R}$ such that $\mathcal{F}(f)$, the Fourier transform of f, is infinitely differentiable, we have

$$\mathcal{F}(F_{1|\mathcal{U}}) = \mathcal{F}(g_{|\mathcal{U}}) \times \mathcal{F}(f_{|\mathcal{U}}),$$

where $q_{|\mathcal{U}}$ denotes the restriction of the distribution q to $\mathcal{U}$. Given the definition of f, its Fourier transform satisfies $\mathcal{F}(f)(t) = E[\exp(-itV)V]$. By Assumption 2.3(iii), $\mathcal{F}(f)$ is analytic on the strip $\{z \in \mathbb{C} : |\text{Im}(z)| < \beta\}$ and therefore infinitely differentiable on $\mathbb{R}$. Thus, we can choose $\mathcal{U} = \mathbb{R}$. Moreover, by dominated convergence, $\mathcal{F}(f) = i\Psi'_{-V}$. As a result, Equation (6) holds.

2. g and h are identified from Equations (6), (7) and (8).

To show the identification of g and h, we prove first that Equations(6), (7) and (8) admit a unique solution in Ψ_{-V} and $\mathcal{F}(g)$, up to a parameter. By taking the inverse Fourier transform of $\mathcal{F}(g)$ and using the normalization $g(x_0^*) = 0$, we then recover g, and finally h. We decompose the proof into several substeps.

(a) The equation $\lambda\mathcal{F}(q_1) = 0$ for λ meromorphic admits a unique solution, $\lambda = 0$.

Recall that a meromorphic function is the ratio between two analytic functions. We use a similar reasoning as Zinde-Walsh (2014, p. 1224). Let us reason by contradiction and suppose that there exists a nonzero meromorphic function λ such that $\lambda\mathcal{F}(q_1) = 0$. Similarly to analytic functions, non-zero meromorphic functions have isolated zeros (see, e.g., Rudin, 1987, p. 209) and thus λ does not vanish on a bounded open set $\mathcal{O} \subset \text{Support}(\mathcal{F}(g))\backslash\{0\}$. By Assumption 2.3(iv), Ψ'_{-V} does not vanish on $\mathcal{O}$ either. Hence, for any $\phi \in \mathcal{S}$ with support included in $\mathcal{O}$, $\phi/(\lambda\Psi'_{-V})$ belongs to $\mathcal{S}$ and has a support included in $\mathcal{O}$. Further, by Equation (6),

$$\mathcal{F}(g)(\phi) = \lambda \times \left[\mathcal{F}(g) \times \Psi'_{-V}\right](\phi/\lambda\Psi'_{-V}) = (\lambda\mathcal{F}(q_1))(\phi/\lambda\Psi'_{-V}) = 0.$$

This implies that $\mathcal{F}(g)$ is zero on $\mathcal{O}$, a contradiction. Hence, $\lambda = 0$.

(b) Ψ_{-V} is identified.

From Equations (6), (7) and (8) we get $\lambda_0\mathcal{F}(q_1) + i\mathcal{F}(q_2) = 0$ and $\mu_0\mathcal{F}(q_1) + i\mathcal{F}(q_3) = 0$, with

$$\lambda_0 = \frac{\Psi''_{-V} + \nu_2\Psi_{-V}}{\Psi'_{-V}}, \tag{17}$$

$$\mu_0 = \frac{i\Psi'''_{-V} + \nu_3\Psi_{-V}}{\Psi'_{-V}} - 3i(m_2 - \nu_2). \tag{18}$$

Because $E[\exp(|V|\beta)] < \infty$, Ψ_{-V} is analytical on the strip $\{z \in \mathbb{C} : |\text{Im}(z)| < \beta\}$. Thus, the functions λ_0 and μ_0 defined by Equations (17) and (18) respectively are meromorphic on that strip, as ratios of analytic functions. By step (a), the equations $\lambda\mathcal{F}(q_1)+i\mathcal{F}(q_2) = 0$ and $\mu\mathcal{F}(q_1)+i\mathcal{F}(q_3) = 0$ in λ and μ respectively and restricted to meromorphic functions admit unique solutions λ_0 and μ_0. As a result, λ_0 and μ_0 are identified. Then some algebra shows that

$$\left[\lambda_0(t)^2 + \lambda_0'(t) + i\mu_0(t) - 3m_2 + 2\nu_2\right] \Psi'_{-V}(t) - \left[\lambda_0(t)\nu_2 + i\nu_3\right] \Psi_{-V}(t) = 0. \tag{19}$$

Now, suppose that $\Psi_{-V}(t) = 0$ for some $t > 0$. Then, because $\lim_{t\to+\infty} \Psi_{-V}(t) = 0$, we would have, by Rolle's theorem, $\Psi'_{-V}(t') = 0$ for some $t' > t$, a contradiction by Assumption 2.3(iv). The same argument for $t < 0$ and $\Psi_{-V}(0) = 1$ then implies that Ψ_{-V} does not vanish on the real line. Further, because $\lambda_0^2 + \lambda_0' + i\mu_0 - 3m_2 + 2\nu_2$ is meromorphic, it has isolated zeros on the real line. Let $\mathcal{Z}$ denote this set of zeros. Equation (19) then implies that for all $t \notin \mathcal{Z}$,

$$\frac{\Psi'_{-V}(t)}{\Psi_{-V}(t)} = \frac{\lambda_0(t)\nu_2 + i\nu_3}{\lambda_0(t)^2 + \lambda_0'(t) + i\mu_0(t) - 3m_2 + 2\nu_2}. \tag{20}$$

Thus, Ψ'_{-V}/Ψ_{-V} is identified on $\mathbb{R}\backslash\mathcal{Z}$ and, by continuity, on the whole real line. Next, by L'Hôpital's rule, a Taylor expansion of $\Psi'_{-V}(t)/\Psi_{-V}(t)$ around 0 and $\Psi_{-V}(0) = 1$ and $\Psi''_{-V}(0) = -\nu_2 \neq 0$, we obtain

$$\begin{aligned}\lambda_0(0) &= \lim_{t\to 0} \lambda_0(t) = \lim_{t\to 0} \frac{\Psi''_{-V}(t) + \nu_2\Psi_{-V}(t)}{\Psi'_{-V}(t)} = \lim_{t\to 0} \frac{\Psi'''_{-V}(t) + \nu_2\Psi'_{-V}(t)}{\Psi''_{-V}(t)} \\ &= \frac{\Psi'''_{-V}(0) + \nu_2\Psi'_{-V}(0)}{\Psi''_{-V}(0)} = -\frac{i\nu_3}{\nu_2}, \end{aligned} \tag{21}$$

$$\frac{\Psi'_{-V}(t)}{\Psi_{-V}(t)} = \frac{\Psi'_{-V}(0)}{\Psi_{-V}(0)} + \left(\frac{\Psi''_{-V}(0)\Psi_{-V}(0) - \Psi'^2_{-V}(0)}{(\Psi_{-V}(0))^2}\right) t + o(t) = -\nu_2 t + o(t). \tag{22}$$

Substituting (21) and (22) into (20) and letting $t \to 0$, we obtain

$$\nu_2 = \frac{1}{2}(3m_2 - (\lambda_0(0))^2 - 2\lambda_0'(0) - i\mu_0(0)). \tag{23}$$

This implies that ν_2 and $\nu_3 = i\nu_2\lambda_0(0)$ are identified. In turn Ψ_{-V} is identified as the unique solution of the differential equation (20) satisfying $\Psi_{-V}(0) = 1$ and $\Psi'_{-V}(0) = 0$.

(c) g and h are identified.

By Assumption 2.3(iv), Ψ'_{-V} vanishes only at 0. Moreover, $\Psi''_{-V}(0) = -\nu_2 \neq 0$. Then, any other solution $\widetilde{F}_g$ of (6) satisfies $\widetilde{F}_g - \mathcal{F}(g) = c\delta_0$ for some real c. Because the Fourier transform is an automorphism on the space of tempered distributions, any $\widetilde{g}$ whose Fourier transform $\widetilde{F}_g$ satisfies (6) is such that $\widetilde{g} = g + c$. The normalization $g(x_0^*) = 0$ then implies that $\widetilde{g} = g$. Hence g is identified. Finally, because g and the distribution of V are identified, so is $E[g(X^*)|Z = z] = E[g(m(z) + V)]$. Hence $h(Z) = E[Y - g(X^*)|Z]$ is also identified.

A.3 Proof of Theorem 2.2.

Let $g(x) = \sum_{k=0}^{K} \alpha_k x^k$ with $\alpha_K \neq 0$ and $\nu_k = E[V^k]$ for $k \geq 0$. Note first that we just have to identify K and $(\alpha_1, \ldots, \alpha_K)$, since α_0 is then identified by the normalization $g(x_0^*) = 0$. First, Equation (9) shows that q_1 is a polynomial of order at most $K-1$. The coefficient corresponding to m^{K-1} is $\alpha_K \nu_2 \neq 0$, so its degree is actually equal to $K-1$. Thus K is identified. Equation (9) also shows that for $j \in \{0, .., K-1\}$, the quantities

$$\beta_{1j+1} = \sum_{k=j+1}^{K} \binom{k}{j} \alpha_k \nu_{k-j+1}$$

are identified. Next Equations (10) and (11) identify, for $j \in \{0, \ldots, K-1\}$, the quantities

$$\beta_{2j+1} = \sum_{k=j+1}^{K} \binom{k}{j} \alpha_k (\nu_{k-j+2} - \nu_2 \nu_{k-j}),$$

$$\beta_{3j+1} = \sum_{k=j+1}^{K} \binom{k}{j} \alpha_k (\nu_{k-j+3} + 3(m_2 - \nu_2)\nu_{k-j+1} - \nu_3 \nu_{k-j}).$$

We now show that this information allows us to identify $(\alpha_1, \ldots, \alpha_K, \nu_2, \ldots, \nu_{K+3})$. For that purpose, let us first show that ν_2 is identified. From above, we identify $\beta_{1K} = K\alpha_K \nu_2$, $\beta_{1K-1} = (K-1)\alpha_{K-1}\nu_2 + K(K-1)\alpha_K \nu_3/2$, $\beta_{2K} = K\alpha_K \nu_3$, $\beta_{2K-1} = (K-1)\alpha_{K-1}\nu_3 + K(K-1)\alpha_K(\nu_4 - \nu_2^2)/2$ and $\beta_{3K} = K\alpha_K(\nu_4 + 3(m_2 - \nu_2)\nu_2)$. Note also that $\beta_{1K} \neq 0$. Then, after some tedious but straightforward algebra, we obtain

$$\nu_2 = \frac{3m_2\beta_{1K} - \beta_{3K} + \frac{2\beta_{2K-1}}{K-1} - \frac{2\beta_{1K-1}\beta_{2K}}{(K-1)\beta_{1K}} + \frac{\beta_{2K}^2}{\beta_{1K}}}{2\beta_{1K}}, \tag{24}$$

which ensures that ν_2 is identified.

Now, let us prove that if we know $\alpha_{k+1}, \ldots, \alpha_K$ and $\nu_2, \ldots, \nu_{K-k+2}$, with $1 \leq k \leq K$ (in the case $k = K$, this amounts to supposing that we only know ν_2), then we identify α_k and ν_{K-k+3}. Taking $j = k-1$, we know $\sum_{\ell=k}^{K} \binom{\ell}{k-1} \alpha_\ell \nu_{\ell-k+2}$. If we know $\alpha_{k+1}, \ldots, \alpha_K$ and $\nu_2, \ldots, \nu_{K-k+2}$, we know each term of this sum except the first, that is to say $\alpha_k \nu_2$. Hence, we identify α_k. Similarly, we know $\sum_{\ell=k}^{K} \binom{\ell}{k-1} \alpha_\ell (\nu_{\ell-k+3} - \nu_2 \nu_{\ell-k+1})$. Each term of this sum is known except the last, that is to say $\alpha_K(\nu_{K-k+3} - \nu_2 \nu_{K-k+1})$. This implies that ν_{K-k+3} is identified.

By induction, this shows that $\alpha_1, \ldots, \alpha_K, \nu_2, \ldots, \nu_{K+3}$ are identified and hence g is identified. In fact, there are $3K$ equations and only $2K+2$ unknowns so, in general the model is overidentified. This is not surprising because we have not used $\beta_{31}, \ldots, \beta_{3K-1}$ here.

Finally,

$$E[g(X^*)|Z = z] = \sum_{k=0}^{K} \alpha_k E\left[(m(z) + V)^k\right],$$

and the right-hand side is identified by what precedes. Hence, $h(Z) = E[Y|Z] - E[g(X^*)|Z]$ is also identified.

A.4 Proof of Proposition 2.1.

First suppose that g is a polynomial, with $g(x) = \sum_{k=0}^{K} \alpha_k x^k$ for some $K > 1$. Then Assumption 2.4 holds, and Equation (9) implies that q_1 is a polynomial of order at most $K - 1$.

Let us show conversely, that if q_1 is a polynomial, then g is a polynomial. We reason by contradiction by supposing that Assumption 2.3 holds. Because q_1 is a polynomial, its Fourier transform satisfies $\mathcal{F}(q_1) = \sum_{k=0}^{K} a_k \delta_0^{(k)}$, where $\delta_0^{(k)}$ denotes the k-th derivative of the Dirac distribution at 0. Hence, the support of $\mathcal{F}(q_1)$ is zero. Let $\mathcal{O}$ denote a bounded open set that does not include 0. Let $\phi \in \mathcal{S}$ with support included in $\mathcal{O}$. By Assumption 2.3(iv), Ψ'_{-V} does not vanish on $\mathcal{O}$. Because it is continuous, $1/\Psi'_{-V}$ is bounded and $\phi/\Psi'_{-V} \in \mathcal{S}$ with support included in $\mathcal{O}$. Then, by Equation (6),

$$0 = \mathcal{F}(q_1)(\phi/\Psi'_{-V}) = \left[\mathcal{F}(g) \times \Psi'_{-V}\right](\phi/\Psi'_{-V}) = \mathcal{F}(g)(\phi),$$

and $\mathcal{F}(g)$ is zero on $\mathcal{O}$. Because the support of a distribution is the complement of the union on all open sets where the distribution is zero, the support of $\mathcal{F}(g)$ is then $\{0\}$. This contradicts the support condition in Assumption 2.3(i), which concludes the proof.

A.5 Proof of Lemma 2.1.

We prove the lemma by exhibiting a specific data generating process where we can construct $\widetilde{g} \neq g$ that is observationally equivalent to g.

Let $\varepsilon, (U_1, V), U_2$ and Z be mutually independent random variables with $E[\varepsilon] = E[U_1] = E[U_2] = E[V] = 0$. Let (U_1, V) satisfy $E[V|U_1 + V] = \rho(U_1 + V)$ with $\rho \neq 1$. This is the case for instance if (U_1, V) is normal with $\text{Cov}(U_1, V) \neq -\text{Var}(U_1)$. Then let $U = U_1 + U_2$, $X^* = m(Z) + V$ for some function m taking at least two elements and $g(x) = \alpha(x - x_0^*)$ with $\alpha \neq 0$. h is left unspecified. Finally, define

$$\begin{cases} Y &= g(X^*) + h(Z) + \varepsilon, \\ X &= X^* + U. \end{cases}$$

By construction, Equation (2) and Assumptions 2.1 and 2.2 are satisfied, g is linear, $E[|V|] < \infty$ and $m(Z)$ contains at least two elements.

Now, we show that this model is observationally equivalent to one with $\widetilde{\varepsilon} = \varepsilon + \alpha(1-\rho)V - \alpha\rho U_1$, $\widetilde{U} = U_2$, $\widetilde{V} = V + U_1$, $\widetilde{X}^* = X^* + U_1 = m(Z) + \widetilde{V}$, $\widetilde{g}(x) = \alpha\rho(x - x_0^*)$ and $\widetilde{h}(z) = h(z) + \alpha(1-\rho)(m(z) - x_0^*)$. For that purpose, we check that Equation (2) and Assumptions 2.1 and 2.2 are satisfied with these new objects. This is sufficient to show observational equivalence because we already have that $\widetilde{g}$ is linear, $E[|\widetilde{V}|] < \infty$ and $m(Z)$ contains at least two elements.

First, by construction,

$$\begin{cases} Y &= \widetilde{g}(\widetilde{X}^*) + \widetilde{h}(Z) + \widetilde{\varepsilon}, \\ X &= \widetilde{X}^* + \widetilde{U}. \end{cases}$$

Second, mutual independence between $\varepsilon, (U_1, V), U_2$ and (X^*, Z) and $E[V|U_1 + V] = \rho(U_1 + V)$ imply that

$$\begin{aligned} E[\widetilde{\varepsilon}|\widetilde{X}^*, Z, \widetilde{U}] &= E[\varepsilon|\widetilde{X}^*, Z, \widetilde{U}] + \alpha(1-\rho)E[V|\widetilde{X}^*, Z, \widetilde{U}] - \alpha\rho E[U_1|\widetilde{X}^*, Z, \widetilde{U}] \\ &= \alpha(1-\rho)E[V|m(Z) + V + U_1, Z] - \alpha\rho(E[m(Z) + U_1 + V - (V + m(Z))|m(Z) + U_1 + V, Z]) \\ &= \alpha(1-\rho)\rho(U_1 + V) - \alpha\rho(1-\rho)(U_1 + V) \\ &= 0. \end{aligned}$$

Moreover, $E[\widetilde{U}^k|X^*, Z] = E[U_2^k]$. Thus, Assumption 2.1 is also satisfied. Finally, $\widetilde{X}^* = m(Z) + \widetilde{V}$ and $\widetilde{V}$ is independent of Z, with $E[\widetilde{V}] = 0$. Hence, Assumption 2.2 is satisfied, which implies that $\widetilde{g} \neq g$ is observationally equivalent to g.

A.6 Proof of Theorem 2.3

Using Assumption 2.5 and $X - m(Z) = \sigma(Z)V + U$, we obtain

$$\begin{aligned} \Psi_{X-m(Z)|Z}(t|z) &= \Psi_V(t\sigma(z))\Psi_U(t), \\ \frac{\frac{\partial \Psi_{X-m(Z)|Z}}{\partial z}(t|z)}{\Psi_{X-m(Z)|Z}(t|z)} &= t\sigma'(z)\frac{\Psi'_V(t\sigma(z))}{\Psi_V(t\sigma(z))}. \end{aligned}$$

Let

$$r(t) = \frac{\frac{\partial \Psi_{X-m(Z)|Z}}{\partial z}(t|z_0)}{2\sigma(z_0)\sigma'(z_0)t\Psi_{X-m(Z)|Z}(t|z_0)}.$$

The function $2\sigma(z_0)\sigma'(z_0) = \frac{\partial \mathrm{Var}(X|Z=z_0)}{\partial z_0}$ is identified, so $r(t)$ is identified as well. Moreover,

$$\begin{aligned} \Psi'_V(t) &= 2\sigma(z_0) r\left(\frac{t}{\sigma(z_0)}\right)\Psi_V(t), \\ \Psi_V(t) &= \exp\left(2\sigma(z_0)\int_0^t r\left(\frac{u}{\sigma(z_0)}\right) du\right), \end{aligned} \tag{25}$$

$\sigma_U^2 = \mathrm{Var}(X - m(Z)|Z = z_0) - \sigma^2(z_0)$, $\sigma^2(z) = \mathrm{Var}(X - m(Z)|Z = z) - \mathrm{Var}(X - m(Z)|Z = z_0) + \sigma^2(z_0)$ and $\Psi_U(t) = \frac{\Psi_{X-m(Z)|Z}(t|z_0)}{\Psi_V(\sigma(z_0)t)}$ are all identified up to $\sigma_0 = \sigma(z_0)$, the value of the function $\sigma(Z)$ at one point $Z = z_0$.

Next we identify g, up to the unknown σ_0, using the equation

$$E\left[Y\exp(it(X - m(Z)))|Z = z_0\right] = E\left[g(m(z_0) + \sigma_0 V)\exp(it\sigma_0 V)\right]\Psi_U(t),$$

where the equality follows by Assumptions 2.5(i) and 2.5(ii). Hence,

$$\mathcal{F}\left[g(m(z_0) + .) \times f_{\sigma_0 V}(.)\right](t) = \frac{E\left[Y\exp(it(X - m(Z)))|Z = z_0\right]}{\Psi_U(t)},$$

where $f_{\sigma_0 V}$ denotes the density of $\sigma_0 V$. This implies in turn that

$$g(m(z_0)+x) = \frac{1}{f_{\sigma_0 V}(x)}\mathcal{F}^{-1}\left(\frac{E\left[Y\exp(i(X-m(Z))(.))|Z=z_0\right]}{\Psi_U(.)}\right)(x), \tag{26}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. All terms on the right-hand side are either identified or depend on σ_0, so g is identified up to the scalar constant σ_0.

To identify σ_0, we use

$$\text{Cov}(Y,X|Z=z) = E\left[g(m(z)+\sigma(z)V)\sigma(z)V\right]. \tag{27}$$

The left-hand side is identified, while the right-hand side consists only of functions that are identified up to σ_0. By Assumption 2.8, Equation (27) identifies σ_0. This implies that g, $\sigma(.)$ and the distribution of V are identified. Finally, $h(z)$ is identified by

$$h(z) = E[Y|Z=z] - E\left[g(m(z)+\sigma(z)V)\right].$$

A.7 Proof of Proposition 2.2

By assumption, $V\sim\mathcal{N}(0,1)$ and $g(x^*)=\alpha+\beta x^*$, $\beta\neq 0$. For simplicity, we consider here the case where $\alpha=0$. The case $\alpha\neq 0$ is similar but more cumbersome. We first compute Ψ_U^σ and Ψ_V^σ in this context. Consider the function $r(.)$ defined in the proof of Theorem 2.3. By this proof,

$$r(t/\sigma(z_0)) = \frac{\Psi_V'(t)}{2\sigma(z_0)\Psi_V(t)}.$$

Hence, $r(u)=-u/2$ here. By (25), $\Psi_V^\sigma(t)=\exp(-t^2/2)$. In other words, the distribution of V is identified in this case, as it does not depend on $\sigma(z_0)$. Then

$$\Psi_U^\sigma(t) = \exp\left(-\frac{1}{2}\left(\text{Var}(X|Z=z_0)-\sigma^2\right)t^2\right).$$

Next, we compute g^σ using Equation (26). First,

$$\begin{aligned} E\left[Y\exp(it(X-m(Z)))|Z=z_0\right] &= \beta E[\exp(itU)]\left\{m(z_0)E[\exp(it\sigma(z_0)V)]+\sigma(z_0)E[V\exp(it\sigma(z_0)V)]\right\} \\ &= \beta\exp\left(-\frac{1}{2}\text{Var}(X|Z=z_0)t^2\right)\left[m(z_0)+i\sigma(z_0)^2t\right]. \end{aligned}$$

Second,

$$\frac{E\left[Y\exp(it(X-m(Z)))|Z=z_0\right]}{\Psi_U^\sigma(t)} = \beta\exp\left(-\frac{1}{2}\sigma^2t^2\right)\left[m(z_0)+i\sigma(z_0)^2t\right].$$

Third, recall that $\exp\left(-\frac{1}{2}\sigma^2t^2\right)$ is the Fourier transform of the density of a $\mathcal{N}(0,\sigma^2)$ variable. Hence,

$$\mathcal{F}^{-1}\left[\exp\left(-\frac{1}{2}\sigma^2t^2\right)\right](x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right),$$

where ϕ is the density of a standard normal variable. Using the fact that $\mathcal{F}^{-1}(q) = \mathcal{F}(q \circ s)/2\pi$, with $s(x) = -x$, we also obtain, after some algebra,

$$\mathcal{F}^{-1}\left[t \exp\left(-\frac{1}{2}\sigma^2 t^2\right)\right](x) = -\frac{ix}{\sigma^3}\phi\left(\frac{x}{\sigma}\right).$$

Combining the previous equations with Equation (26) yields

$$g^\sigma(m(z_0)+x) = \frac{\sigma}{f_V\left(\frac{x}{\sigma}\right)}\left\{\frac{\beta}{\sigma}\phi\left(\frac{x}{\sigma}\right)\left[m(z_0) + \frac{\sigma(z_0)^2}{\sigma^2}x\right]\right\} = \beta\left[m(z_0) + \frac{\sigma(z_0)^2}{\sigma^2}x\right].$$

Finally, let us consider the mapping $\sigma_0 \mapsto \left[z \mapsto \sigma^{\sigma_0}(z) \int g^{\sigma_0}\left(m(z) + \sigma^{\sigma_0}(z)v\right) v f_V^{\sigma_0}(v)dv\right]$. By what preceded,

$$\sigma^\sigma(z)\int g^\sigma\left(m(z)+\sigma^\sigma(z)v\right) v f_V^\sigma(v)dv = \beta\frac{\sigma(z_0)^2}{\sigma^2}(\sigma^\sigma(z))^2 = \beta\frac{\sigma(z_0)^2}{\sigma^2}\left(\sigma(z)^2 + \sigma^2 - \sigma(z_0)^2\right).$$

This and the fact that $\sigma(.)$ is not constant shows that the mapping

$$\sigma \mapsto \left[z \mapsto \sigma^\sigma(z)\int g^\sigma\left(m(z)+\sigma^\sigma(z)v\right) v f_V^\sigma(v)dv\right]$$

is injective. Hence, Assumption 2.8 holds and g and h are identified.

A.8 Proof of Theorem 2.4

We let hereafter $\mu_k = E(U^k)$ for $k = 1 \dots K+1$. First, we find an expression for $E[X^{*k}|Z]$ in terms of moments of U and moments of X conditional on Z. By the binomial expansion $X^k = (X^*+U)^k = X^{*k} + UX^{*k-1} + \sum_{l=0}^{k-2}\binom{k}{l}U^{k-l}X^{*l}$ and using $\mu_1 = 0$,

$$E[X^{*k}|Z] = E[X^k|Z] - \sum_{l=0}^{k-2}\binom{k}{l}\mu_{k-l}E[X^{*l}|Z].$$

After recursively substituting in for $E[X^{*l}|Z]$, for $l = 1, \dots, k-2$, and tedious algebraic manipulation,

$$\begin{aligned}
E[X^{*k}|Z] &= E[X^k|Z] - \sum_{k_1=2}^{3}\binom{k}{k-k_1}\mu_{k_1}E[X^{k-k_1}|Z] \\
&\quad - \sum_{k_1=4}^{5}\binom{k}{k-k_1}\left(\mu_{k_1} - \sum_{k_2=2}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\mu_{k_2}\right)E[X^{k-k_1}|Z] \\
&\quad - \sum_{k_1=6}^{7}\binom{k}{k-k_1}\left(\mu_{k_1} - \sum_{k_2=2}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\mu_{k_2} + \sum_{k_2=4}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\sum_{k_3=2}^{k_2-2}\binom{k_2}{k_3}\mu_{k_2-k_3}\mu_{k_3}\right)E[X^{k-k_1}|Z] - \cdots \\
&\quad - \sum_{k_1=2l}^{k}\binom{k}{k-k_1}\left(\mu_{k_1} - \sum_{k_2=2}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\mu_{k_2} + \sum_{k_2=4}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\sum_{k_3=2}^{k_2-2}\binom{k_2}{k_3}\mu_{k_2-k_3}\mu_{k_3} - \cdots\right. \\
&\quad \left. + (-1)^{l-1}\sum_{k_2=2l-2}^{k_1-2}\binom{k_1}{k_2}\mu_{k_1-k_2}\cdots\sum_{k_{l-1}=4}^{k_{l-2}-2}\binom{k_{l-2}}{k_{l-1}}\mu_{k_{l-2}-k_{l-1}}\sum_{k_l=2}^{k_{l-1}-2}\binom{k_{l-1}}{k_l}\mu_{k_{l-1}-k_l}\mu_{k_l}\right)E[X^{k-k_1}|Z],
\end{aligned}$$

where $l = \lfloor \frac{k}{2} \rfloor$ and $\lfloor x \rfloor$ denotes the largest integer less than or equal to $\frac{k}{2}$.

Now adopt the notation $\alpha_{jk} = 0$ when $K < k$ and substitute the above binomial expansion into $\mathrm{Cov}(X, Y|Z)$ to get an expression that is a linear combination of moments of X conditional on Z with coefficients that are complicated (but known) linear combinations of $\alpha_1, \ldots, \alpha_K$ and $\mu_1, \ldots, \mu_{K+1}$,

$$
\begin{aligned}
&\mathrm{Cov}(X, Y|Z) \\
&= \sum_{j=0}^{J} Z^j \sum_{k=1}^{K} \alpha_{jk} \mathrm{Cov}(X, X^{*k}|Z) \\
&= \sum_{j=0}^{J} Z^j \sum_{k=1}^{K} \alpha_{jk} \left(E[XX^{*k}|Z] - E[X|Z]E[X^{*k}|Z] \right) \\
&= \sum_{j=0}^{J} Z^j \sum_{k=1}^{K} \alpha_{jk} \left(E[X^{*k+1}|Z] - E[X|Z]E[X^{*k}|Z] \right) \\
&= \sum_{j=0}^{J} Z^j \sum_{k=-1}^{K} \Bigg[\alpha_{jk} - \sum_{k_1=2}^{3} \alpha_{jk+k_1} \binom{k+k_1+1}{k+1} \mu_{k_1} - \sum_{k_1=4}^{5} \alpha_{jk+k_1} \binom{k+k_1+1}{k+1} \left(\mu_{k_1} - \sum_{k_2=2}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \mu_{k_2} \right) - \cdots \\
&\qquad - \sum_{k_1=2l}^{K} \alpha_{jk+k_1} \binom{k+k_1+1}{k+1} \Bigg(\mu_{k_1} - \sum_{k_2=2}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \mu_{k_2} + \sum_{k_2=4}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \sum_{k_3=2}^{k_2-2} \binom{k_2}{k_3} \mu_{k_2-k_3} \mu_{k_3} - \cdots \\
&\qquad (-1)^{l-1} \sum_{k_2=2l-2}^{k_1-1} \binom{k_1}{k_2} \mu_{k_1-k_2} \cdots \sum_{k_{l-1}=4}^{k_{l-2}-2} \binom{k_{l-2}}{k_{l-1}} \mu_{k_{l-2}-k_{l-1}} \sum_{k_l=2}^{k_{l-1}-2} \binom{k_{l-1}}{k_l} \mu_{k_{l-1}-k_l} \mu_{k_l} \Bigg) \Bigg] E[X^{k+1}|Z] \\
&- \sum_{j=0}^{J} Z^j \sum_{k=0}^{K} \Bigg[\alpha_{jk} - \sum_{k_1=2}^{3} \alpha_{jk+k_1} \binom{k+k_1}{k} \mu_{k_1} - \sum_{k_1=4}^{5} \binom{k+k_1}{k} \left(\mu_{k_1} - \sum_{k_2=2}^{k_1-2} \alpha_{jk+k_1} \binom{k_1}{k_2} \mu_{k_1-k_2} \mu_{k_2} \right) - \ldots \\
&\qquad - \sum_{k_1=2l}^{K} \alpha_{jk+k_1} \binom{k+k_1}{k} \Bigg(\mu_{k_1} - \sum_{k_2=2}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \mu_{k_2} + \sum_{k_2=4}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \sum_{k_3=2}^{k_2-2} \binom{k_2}{k_3} \mu_{k_2-k_3} \mu_{k_3} - \cdots \\
&\qquad (-1)^{l-1} \sum_{k_2=2l-2}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \cdots \sum_{k_{l-1}=4}^{k_{l-2}-2} \binom{k_{l-2}}{k_{l-1}} \mu_{k_{l-2}-k_{l-1}} \sum_{k_l=2}^{k_{l-1}-2} \binom{k_{l-1}}{k_l} \mu_{k_{l-1}-k_l} \mu_{k_l} \Bigg) \Bigg] E[X|Z]E[X^k|Z] \\
&= \sum_{j=0}^{J} Z^j Q(Z)' \beta_j = R(Z)' \beta,
\end{aligned}
$$

where the fourth equality follows by substituting in the binomial expansion and

$$
\begin{aligned}
&Q(Z) = (E[X^{K+1}|Z], -E[X^K|Z]E[X|Z], \ldots, E[X^{k+1}|Z], -E[X^k|Z]E[X|Z], \ldots, E[X^2|Z], -E[X|Z]E[X|Z], E[X|Z], 1)', \\
&R(Z) = (Z^0 Q(Z)', Z^1 Q(Z)', \ldots, Z^J Q(Z)')', \qquad \beta_j = (\beta_{j1}, \ldots, \beta_{j2K+2})', \qquad \beta = (\beta_0', \ldots, \beta_J')'.
\end{aligned}
$$

By Assumption 2.10, $E[R(Z)R(Z)']$ is finite and nonsingular so $\beta = E[R(Z)R(Z)'^{-1}]E[R(Z)\mathrm{Cov}(X, Y|Z)]$ is identified. Further, μ_k and α_{jk} are recursively identified by

$$
\begin{aligned}
\alpha_{jK} &= \beta_{j1} = \beta_{j2}, \quad \alpha_{jK-1} = \beta_{j3} = \beta_{j4} \\
\mu_2 &= \frac{\beta_{j6} - \beta_{j5}}{\alpha_{jK} \left(\binom{K+1}{K-1} - \binom{K}{K-2} \right)}, \quad \alpha_{jK-2} = \beta_{j5} + \alpha_{jK} \binom{K+1}{K-1} \mu_2, \\
&\vdots \\
\mu_k &= \frac{\beta_{j2k+2} - \beta_{j2k+1} - \sum_{k_1=2}^{3} \alpha_{jK-k+k_1} \left(\binom{K-k+k_1+1}{K-k+1} - \binom{K-k+k_1}{K-k} \right) \mu_{k_1} - \ldots}{\alpha_{jK} \left(\binom{K+1}{K-k+1} - \binom{K}{K-k} \right)},
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{jK-k} &= \beta_{j2k+1} + \sum_{k_1=2}^{3} \alpha_{jK-k+k_1} \binom{K-k+k_1+1}{K-k+1} \mu_{k_1} + \dots, \\
&\vdots \\
\mu_{K-1} &= \frac{\beta_{j2K} - \beta_{j2K-1} - \sum_{k_1=2}^{3} \alpha_{jk_1+1} \left(\binom{k_1+2}{2} - \binom{k_1+1}{1}\right) \mu_{k_1} - \dots}{\alpha_{jK} \left(\binom{K+1}{2} - \binom{K}{1}\right)}, \\
\alpha_{j1} &= \beta_{j2K-1} + \sum_{k_1=2}^{3} \alpha_{jk_1+1} \binom{k_1+2}{2} \mu_{k_1} + \dots, \\
\mu_K &= \frac{-\beta_{j2K+1} - \sum_{k_1=2}^{3} \alpha_{jk_1} \left(\binom{k_1+1}{1} - \binom{k_1}{0}\right) \mu_{k_1} - \dots}{K\alpha_{jK}} \\
\mu_{K+1} &= \frac{-\beta_{j2K+2} - \sum_{k_1=2}^{3} \alpha_{jk_1-1} \binom{k_1}{0} \mu_{k_1} - \dots}{\alpha_{jK}}.
\end{aligned}
$$

This identifies $g(X^*, Z)$. Identification of $h(Z)$ follows by

$$
h(Z) = E[Y|Z] - \sum_{j=0}^{J} Z^j \sum_{k=1}^{K} \alpha_{jk} E[X^{*k}|Z],
$$

where α_{jk} and $E[X^{*k}|Z]$ are identified above.

A.9 Proof of Example 2.1

When $Y = h(Z) + \alpha_1 X^* + \varepsilon$, we have

$$
\begin{aligned}
\mathrm{Cov}(X, Y|Z) &= \alpha_1 \mathrm{Cov}(X, X^*|Z) = \alpha_1 \left(E[XX^*|Z] - E[X|Z]E[X^*|Z]\right) \\
&= \alpha_1 E[X^2|Z] - \alpha_1 (E[X|Z])^2 - \alpha_1 \mu_2 \qquad (28)
\end{aligned}
$$

By Assumption 2.1, $\mathrm{Cov}(X, Y|Z) = R(Z)'\beta$ where $R(Z) = (E[X^2|Z], -E[X|Z]E[X|Z], E[X|Z], 1)'$ and $\beta = (\alpha_1, \alpha_1, 0, -\alpha_1\mu_2)'$. By Assumption 2.10, $\beta = E[R(Z)R(Z)'^{-1}]E[R(Z)\mathrm{Cov}(X, Y|Z)]$. Hence, $\alpha_1 = \beta_1 = \beta_2$, $\mu_1 = -\beta_3$, $\mu_2 = -\beta_4/\alpha_1$ and $h(Z) = E[Y|Z] - \alpha_1 E[X|Z]$ are identified.

A.10 Proof of Theorem 3.1

The proof is divided into two main steps: we first show that $\widehat{\theta}$ is asymptotically normal. Then we prove asymptotic normality of $\widehat{\eta}$ and $\widehat{g}$.

1. Asymptotic normality of $\widehat{\theta}$

First, we note that $\widehat{\theta}$ is a two-step GMM estimator, with a nonparametric first step. $\widehat{\theta}$ satisfies

$$
\frac{1}{n} \sum_{i=1}^{n} M(U_i, \widehat{m}; \widehat{\theta}) = 0,
$$

with $U = (X, Y, Z)$, $M(U, m; \theta) = (M_1(U, m; \theta), \dots, M_{3K+2}(U, m; \theta))'$ and

$$\begin{aligned}
& M_j(U, m; \theta) \\
= \;& (X - m(Z))^{j+1} - m_{j+1} && \text{if } j \in \{1, 2\}, \\
= \;& m(Z)^{j-2} \left[Y \left(X - m(Z) \right) - \textstyle\sum_{k=0}^{K-1} \beta_{1k} m(Z)^k \right] && \text{if } j \in \{3, \dots, K+2\}, \\
= \;& m(Z)^{j-(K+3)} \left[Y \left((X - m(Z))^2 - m_2 \right) - \textstyle\sum_{k=0}^{K-1} \beta_{2k} m(Z)^k \right] && \text{if } j \in \{K+3, \dots, 2K+2\}, \\
= \;& m(Z)^{j-(2K+3)} \left[Y \left((X - m(Z))^3 - m_3 \right) - \textstyle\sum_{k=0}^{K-1} \beta_{3k} m(Z)^k \right] && \text{if } j \in \{2K+3, \dots, 3K+2\}.
\end{aligned}$$

To prove the asymptotic normality of $\widehat{\theta}$, we check the conditions of Theorem 6.1 of Newey (1994). We first introduce some notation. Denote by $m_0(.)$ the true function $E[X|Z = .]$. Let $\|u\| = \max_{j=1,\dots,J} |u_j|$ for $u = (u_1, \dots, u_J)' \in \mathbb{R}^J$ and $J \in \mathbb{N}^+$. For any real function q that is d times differentiable on Support$(m_0(Z))$, let

$$|q|_d = \max_{\lambda \in \mathbb{N}^r : \sum_{k=1}^r \lambda_k \leq d} \; \sup_{z \in \text{Support}(m_0(Z))} \left| \frac{\partial^\lambda q(z)}{\partial z_1^{\lambda_1} \dots \partial z_r^{\lambda_r}} \right|.$$

For any $\xi > 0$, define $\mathcal{S}_\xi = \{x \in \mathbb{R} : \exists m \in \text{Support}(m_0(Z)) : |x - m| < \xi\}$ and $\|q\|_{\xi,\infty} = \sup_{x \in \mathcal{S}_\xi} \|q(x)\|$, for any vector function $q(.)$ defined on $\mathcal{S}_\xi$. Let $\widetilde{M}_j(U, x, \theta_0)$ be defined as $M_j(U, m, \theta_0)$ except that the function $m(Z)$ is replaced by the number x,

$$\begin{aligned}
M_j(U, x; \theta) = \;& (X - x)^{j+1} - m_{j+1} && \text{if } j \in \{1, 2\}, \\
= \;& x^{j-2} \left[Y \left(X - x \right) - \textstyle\sum_{k=0}^{K-1} \beta_{1k} x^k \right] && \text{if } j \in \{3, \dots, K+2\}, \\
= \;& x^{j-(K+3)} \left[Y \left((X - x)^2 - m_2 \right) - \textstyle\sum_{k=0}^{K-1} \beta_{2k} x^k \right] && \text{if } j \in \{K+3, \dots, 2K+2\}, \\
= \;& x^{j-(2K+3)} \left[Y \left((X - x)^3 - m_3 \right) - \textstyle\sum_{k=0}^{K-1} \beta_{3k} x^k \right] && \text{if } j \in \{2K+3, \dots, 3K+2\}.
\end{aligned}$$

To bound $M_j(U, m, \theta_0)$, for all $x \in \mathcal{S}_\xi$, we use repeatedly the triangle inequality and that there exists $C_0 > 0$ such that for all $\theta \in \Theta$ and all $k \in \{1, \dots, 3K+2\}$, $|\theta_k| \leq C_0$ (because Θ is compact),

$$\begin{aligned}
|\widetilde{M}_j(U, x; \theta)| \leq \;& 2^j \left(|X|^{j+1} + \overline{m}^{j+1} \right) + C_0 && \text{if } j \in \{1, 2\}, \\
\leq \;& \overline{m}^{j-2} \left[|Y| \left(|X| + |\overline{m}| \right) + C_0 K \overline{m}^{K-1} \right] && \text{if } j \in \{3, \dots, K+2\}, \\
\leq \;& \overline{m}^{j-(K+3)} \left[|Y| \left(2|X|^2 + 2|\overline{m}|^2 + C_0 \right) + C_0 K \overline{m}^{K-1} \right] && \text{if } j \in \{K+3, \dots, 2K+2\}, \\
\leq \;& \overline{m}^{j-(2K+3)} \left[|Y| \left(4|X|^3 + 4|\overline{m}|^3 + C_0 \right) + C_0 K \overline{m}^{K-1} \right] && \text{if } j \in \{2K+3, \dots, 3K+2\},
\end{aligned}$$

where $\overline{m} = \max\left(1, \sup_{x \in \mathcal{S}_\xi} \|x\|\right)$. Further, because $x \mapsto \widetilde{M}_j(U, x; \theta)$ is a polynomial, we obtain similar inequalities for its derivatives. Hence, for $\ell \in \{0, 1, 2\}$, there exist constants $(C_{\ell 1}, C_{\ell 2}, C_{\ell 3})$ such that for all $\theta \in \Theta$,

$$\left\| \widetilde{M}^{(\ell)}(U, .; \theta) \right\|_{\xi,\infty} \leq b_\ell(U) \equiv \left(4|X|^3 + 2|X|^2 + |X| + C_{\ell 1} \right) \left(C_{\ell 2} |Y| + C_{\ell 3} \right), \tag{29}$$

where $\widetilde{M}^{(\ell)}(U, .; \theta)$ denotes the derivative of order ℓ of $\widetilde{M}(U, .; \theta)$.

We now check that Conditions 5.4-5.6 and 6.1-6.6 of Newey (1994) are satisfied, so that we can apply his Theorem 6.1. Instead of checking Condition 6.4(i), we verify his weaker Condition 5.1(i), since 6.4(i) is only needed for the consistency of the asymptotic variance estimator.

First, $\theta \mapsto M(U, \theta, m_0)$ is continuous. It is then bounded on the compact set Θ. Hence, Condition 5.4(i) holds. Moreover, for all m such that $|m - m_0|_0 < \xi$,

$$\begin{aligned}\|M(U, m, \theta_0) - M(U, m_0, \theta_0)\| &\leq \sup_{x \in \mathcal{S}_\xi} \left| \widetilde{M}_j^{(1)}(U, x, \theta_0) \right| |m(Z) - m_0(Z)| \\ &\leq b_1(U) \, |m - m_0|_0 \, ,\end{aligned}$$

which implies that Condition 5.4(ii) holds.

Second, in our framework the weighting matrix of the GMM is the identity matrix. Further, suppose that $E[M(U, m_0; \theta)] = 0$ for some $\theta = (\widetilde{m}_2, \widetilde{m}_3, \widetilde{\beta}_{10}, \dots, \widetilde{\beta}_{3K-1})'$. Then from the first two equations $\widetilde{m}_k = m_k$ for $k \in \{2, 3\}$. Let $P_m(Z) = (1, m(Z), \dots, m(Z)^{K-1})'$ and $H_m(Z) = P_m(Z) P_m(Z)'$. $E[M(U, m_0; \theta)] = 0$ then implies that $E[H_{m_0}(Z)](\widetilde{\beta} - \beta_0) = 0$. Thus, to prove that $\widetilde{\beta} = \beta_0$, we have to show that $E[H_{m_0}(Z)]$ is full rank, which is equivalent to

$$\sum_{j=0}^{K-1} \gamma_j m_0(Z)^j = 0 \text{ almost surely } \implies \gamma_0 = \ldots = \gamma_{K-1} = 0. \tag{30}$$

By Assumptions 3.2(iii) and (iv), m_0 is differentiable and not constant on a Cartesian product of intervals. Therefore, the support of $m_0(Z)$ contains an interval, which ensures that (30) holds. Hence, $\widetilde{\beta} = \beta_0$, and $\theta = \theta_0$. Finally, Θ is compact. Thus Condition 5.5 holds.

Condition 5.6(i) follows by Assumption 3.2(i). $\theta \mapsto M(u, m; \theta)$ is linear and therefore differentiable for any (m, u), so (ii) holds as well. Now, let I_2 be the 2×2 identity matrix, 0_{IJ} the $I \times J$ zero matrix, $G_{1m} = (Y P_m(Z), 0_{K1})$, $G_{2m} = (0_{K1}, Y P_m(Z))$. Then,

$$\frac{\partial M(U, m; \theta)}{\partial \theta} = - \begin{pmatrix} I_2 & 0_{2K} & 0_{2K} & 0_{2K} \\ 0_{K2} & H_m(Z) & 0_{KK} & 0_{KK} \\ G_{1m} & 0_{KK} & H_m(Z) & 0_{KK} \\ G_{2m} & 0_{KK} & 0_{KK} & H_m(Z) \\ \end{pmatrix}.$$

Because this matrix is triangular and $E[H_{m_0}(Z)]$ is full rank as shown above,

$$G = E\left[\left. \frac{\partial M(U, m_0; \theta)}{\partial \theta} \right|_{\theta = \theta_0} \right] \tag{31}$$

is nonsingular and Condition 5.6(iii) holds. Next, by Equation (29),

$$E\left[\|M(U, m, \theta_0)\|^2 \right] \leq E\left[b_0(U)^2 \right] < \infty,$$

where the last equality holds by Assumption 3.2(ii). Thus, Condition 5.6(iv) holds. Condition 5.6(v), amounts to verifying Condition 5.4 on $\partial M_j(U, m_0; \theta)/\partial\theta$ for all $j \in \{1, \dots, 3K+2\}$. First, because this function does not depend on θ, it satisfies Condition 5.4(i). Now, note that for all $i \in \{1, \dots, 3K+2\}$, $\partial M_j(U, m; \theta)/\partial\theta_i$ is either constant, equal to $-m(Z)^{j'}$ or equal to $-Ym(Z)^{j'}$ for some j'. Hence, for all m such that $|m - m_0|_0 < \xi$, there exists C_j such that

$$\left\|\frac{\partial M_j(U, m; \theta)}{\partial\theta} - \frac{\partial M_j(U, m_0; \theta)}{\partial\theta}\right\| \leq C_j \left|m - m_0\right|_0.$$

This ensures that $\partial M_j(U, m; \theta)/\partial\theta$ satisfies Condition 5.4(ii). Thus, Condition 5.6(v) holds.

Condition 6.1 holds since $z \mapsto \text{Var}(X|Z = z) = \text{Var}(U) + \text{Var}(V)$ is bounded. Conditions 6.2 and 6.3 are satisfied here by Assumption 3.2(iv) and (v), as shown in Section 5 of Newey (1997).

We now check Condition 5.1(i), instead of Condition 6.4(i) as explained above. Define

$$D(U, m) = \widetilde{M}^{(1)}(U, m_0(Z); \theta_0) \times m(Z). \tag{32}$$

Then, by a second order Taylor expansion, we have, for all m such that $|m - m_0|_0 < \xi$,

$$\|M(U, m; \theta_0) - M(U, m_0; \theta_0) - D(U, m; \theta)\| \leq \frac{1}{2}\left\|\widetilde{M}^{(2)}(U, .; \theta_0)\right\|_{\xi,\infty} |m - m_0|_0^2. \tag{33}$$

So Condition 5.1(i) holds. Turning to Condition 6.4(ii), note that in our case, and using Newey's (1994) notation, $d = 0$, $\alpha = s/r$ (see (Newey, 1994), p. 1370) and $b(U) = b_0(U)/2$, the latter in view of (29) and (33). Then Assumption 3.2(ii) ensures that $E[|b(U)|] < \infty$. Second, $\zeta_0(L_n) \leq C_1 L_n$ for some constant C_1 (see (Newey, 1994), p. 1371). Therefore, the two statements of Condition 6.4(ii) hold because $L_n\left[\sqrt{L_n/n} + L_n^{-s/r}\right] = o(n^{-1/4})$ by Assumptions 3.2(v) and (vi).

We check Condition 6.5 with $d = 1$. Given Equations (29) and (32), we have

$$\|D(U, m; \theta, m_0)\| \leq b_1(U)\left|m\right|_0 \leq b_1(U)\left|m\right|_1.$$

Moreover, Assumption 3.2(ii) implies that $E[b_1(U)^2] < \infty$, and the first statement follows. The second and third statement are satisfied by exactly the reasoning of Frölich (2007), p. 68, and the conditions $s > 3r$ and $L_n^7/n \to 0$.

Finally, let $\delta(U) = \widetilde{M}^{(1)}(U, m_0(Z); \theta_0)$. Then Condition 6.6(i) holds by (32). Condition 6.6(ii) holds by applying once more Frölich (2007), pp. 68-69, and because both m_0 and δ are s times differentiable.

Hence, we can apply the first part of Theorem 6.1 in Newey (1994), which implies that

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, G^{-1}HG^{-1\prime}\right), \tag{34}$$

where G is defined in (31) and

$$H = V\left[M(U, m_0; \theta_0) + \delta(U)(X - m_0(Z))\right]. \tag{35}$$

Asymptotic normality of $\widehat{\eta}$ and $\widehat{g}(x)$

We first show that $\widehat{\eta}$ is consistent. For that purpose, we check that the conditions of Theorem 2.1 of Newey and McFadden (1994) apply with, taking the same notation as Newey and McFadden, $Q_n(\eta) = -(\widehat{\theta} - \Pi(\eta))'W_n(\widehat{\theta} - \Pi(\eta))$ and $Q_0(\eta) = -(\theta_0 - \Pi(\eta))'W(\theta_0 - \Pi(\eta))$. We have $Q_0(\eta) \leq 0$ and because W is nonsingular, $Q_0(\eta) = 0$ if and only if $\theta_0 = \Pi(\eta)$. We showed in the proof of Theorem 2.2 that this implies that $\eta = \eta_0$. Hence, $Q_0(.)$ is uniquely maximized at η_0, and their Condition 2.1(i) holds. By Assumption 3.2(i), $\mathcal{H}$ is compact so their Condition 2.1(ii) holds. Next, $\Pi(.)$ is continuous so $Q_0(.)$ is continuous as well, and their Condition 2.1(iii) holds. Finally,

$$Q_n(\eta) - Q_0(\eta) = \theta_0'(W - W_n)(\theta_0 - 2\Pi(\eta)) + [\theta_0 + \widehat{\theta} - 2\Pi(\eta)]'W_n(\theta_0 - \widehat{\theta}) + \Pi(\eta)'(W - W_n)\Pi(\eta).$$

Hence, because $\widehat{\theta}$ and $\Pi(\eta)$ belong to a compact set by Assumption 3.2(i), $\widehat{\theta}$ is consistent and $W_n \xrightarrow{\text{p}} W$, we obtain, by the triangular and Cauchy-Schwartz inequalities,

$$\sup_{\eta \in \mathcal{H}} |Q_n(\eta) - Q_0(\eta)| \xrightarrow{\text{p}} 0,$$

and their Condition 2.1(iv) holds. Therefore, $\widehat{\eta}$ is consistent by Theorem 2.1 of Newey and McFadden (1994).

Now, we check that the conditions of Theorem 3.2 of Newey and McFadden (1994) apply. First, by Assumption 3.2(vi), $W_n \xrightarrow{\text{p}} W$ where W is positive definite and $\widehat{\eta}$ is consistent by the paragraph above. Second by Assumption 3.2(i), η is in the interior of the compact set $\mathcal{H}$, so their Condition 3.2(i) holds. Third, $\Pi(.)$ is continuously differentiable so their Condition 3.2(ii) is satisfied as well. Fourth, by Equation (34),

$$\sqrt{n}\left(\widehat{\theta} - \Pi(\eta_0)\right) \xrightarrow{d} \mathcal{N}\left(0, G^{-1}HG^{-1\prime}\right),$$

so their Condition 3.2(iii) holds. Fifth, their Condition 3.2(iv) is automatically satisfied since $\Pi(.)$ is nonstochastic. Finally, J and W are full rank by Assumption 3.2(vii), therefore $J'WJ$ is full rank as well. Then, by Theorem 3.2 of Newey and McFadden (1994),

$$\sqrt{n}\left(\widehat{\eta} - \eta_0\right) \xrightarrow{d} \mathcal{N}\left(0, (J'WJ)^{-1}J'WG^{-1}HG^{-1\prime}WJ(J'WJ)^{-1}\right).$$

By standard results (see, e.g., Wooldridge, 2002, p. 424), the optimal weighting matrix is $W^* = \left[G^{-1}HG^{-1\prime}\right]^{-1}$. Finally, the asymptotic normality of $\widehat{\eta}$ implies that any linear combination of this vector is also asymptotically normal, which in particular implies that $\widehat{g}(x)$ is asymptotically normal.

B A polynomial restriction on g with multiplicative errors

We briefly consider the model with polynomial g and multiplicative errors,

$$\begin{cases} Y &= \sum_{j=0}^{J}\sum_{k=1}^{K}\alpha_{jk}Z^{j}X^{*k}+h\left(Z\right)+\varepsilon \\ X &= X^{*}U \end{cases} \tag{36}$$

and assume $\alpha_{jk}\neq 0$ for all j and k. The following assumptions replace Assumptions 2.9 and 2.10 from the main text.

Assumption B.1. *(i) $E[\varepsilon|X^{*},Z]=0$ and (ii) $E[U^{k}|X^{*},Z]=E[U^{k}]$ for $k\in\{1,2,\ldots,K\}$ and $E[U]=1$.*

Assumption B.2. *Define*

$$Q(Z)=\left(-E[X|Z]E[X|Z],E[X^{2}|Z],-E[X|Z]E[X^{2}|Z],E[X^{3}|Z],\ldots,-E[X|Z]E[X^{K}|Z],E[X^{K+1}|Z]\right)'$$
$$R(Z)=(Z^{0}Q(Z)',Z^{1}Q(Z)',\ldots,Z^{J}Q(Z)')'$$

$R(Z)$ is finite and nonsingular.

Theorem B.1. *Suppose Equation (36) and Assumptions B.1 and B.2 hold. Then g and h and the moments $E[U^{1}],\ldots,E[U^{K+1}]$ are identified.*

Proof: we let $\mu_k=E[U^k]$ for $k=1\ldots K+1$. First, we have

$$\begin{aligned}\operatorname{Cov}(X,Y|Z)&=\sum_{j=0}^{J}\sum_{k=1}^{K}\alpha_{jk}Z^{j}\operatorname{Cov}(X,X^{*k}|Z)\\&=\sum_{j=0}^{J}\sum_{k=1}^{K}\alpha_{jk}Z^{j}\left(E[XX^{*k}|Z]-E[X|Z]E[X^{*k}|Z]\right)\\&=\sum_{j=0}^{J}\sum_{k=1}^{K}\alpha_{jk}\left(\frac{Z^{j}E[X^{k+1}|Z]}{\mu_{k+1}}-\frac{Z^{j}E[X|Z]E[X^{k}|Z]}{\mu_{k}}\right)\\&=\sum_{j=0}^{J}Z^{j}Q(Z)'\beta_{j}=R(Z)'\beta,\end{aligned}$$

where the third equality follows by $E[X^{*k}|Z]=\frac{E[X^{k}|Z]}{\mu_k}$, $R(Z)$ is defined in Assumption B.2 and

$$\begin{aligned}&\beta_{j2k-1}=\frac{\alpha_{jk}}{\mu_k},\quad \beta_{j2k}=\frac{\alpha_{jk}}{\mu_{k+1}},\quad \beta_j=(\beta_{j1},\ldots,\beta_{jK}),\quad (j,k)\in\{1,\ldots,K\}^2,\\&\beta=(\beta_0',\ldots,\beta_K')'.\end{aligned}$$

By Assumption B.2, $E[R(Z)R(Z)']$ is finite and nonsingular. Thus,

$$\beta=E[R(Z)R(Z)'^{-1}]E[R(Z)\operatorname{Cov}(X,Y|Z)].$$

Then $\alpha_{j1}=\beta_{j1}$ and for $k>1$, $\mu_k=\prod_{i=1}^{k-1}\beta_{j2i-1}/\prod_{i=1}^{k-1}\beta_{j2i}$ and $\alpha_{jk}=\prod_{i=1}^{k}\beta_{j2i-1}/\prod_{i=1}^{k-1}\beta_{j2i}$. Further, $h(Z)=E[Y|Z]-\sum_{j=0}^{J}\sum_{k=1}^{K}\alpha_{jk}Z^{j}E[X^{*k}|Z]=E[Y|Z]-\sum_{j=0}^{J}\sum_{k=1}^{K}\beta_{j2k-1}Z^{j}E[X^{k}|Z]$.

C Further discussion on inference

C.1 Non-polynomial case

In the non-polynomial case, identification is based on Equations (6)-(8) using Fourier transforms of tempered distributions. An idea to develop nonparametric estimation is to consider nonparametric estimators $\widehat{q}_k$ of q_k, for $k \in \{1, 2, 3\}$, and then several plug-in estimators based on the same equalities as those we use to prove identification. Specifically, we compute first $\mathcal{F}(\widehat{q}_k)$ for $k \in \{1, 2, 3\}$, $\widehat{\lambda} = -i\mathcal{F}(\widehat{q}_2)/\mathcal{F}(\widehat{q}_1)$ and $\widehat{\mu} = -i\mathcal{F}(\widehat{q}_3)/\mathcal{F}(\widehat{q}_1)$. Using Equations (20), (21) and (23), we then consider

$$\begin{aligned}
\widehat{\nu}_2 &= \frac{1}{2}(3\widehat{m}_2 - (\widehat{\lambda}(0))^2 - 2\widehat{\lambda}'(0) - i\widehat{\mu}(0)), \\
\widehat{\nu}_3 &= i\widehat{\nu}_2\widehat{\lambda}(0), \\
\widehat{\Psi}_{-V}(t) &= \exp\left(\int_0^t \frac{\widehat{\lambda}(s)\widehat{\nu}_2 + i\widehat{\nu}_3}{\widehat{\lambda}(s)^2 + \widehat{\lambda}'(s) + i\widehat{\mu}(s) - 3\widehat{m}_2 + 2\widehat{\nu}_2} ds\right).
\end{aligned}$$

We can then compute $\mathcal{F}(\widehat{g})$ using (6), and in turn $\widehat{g}$. Finally, h can be estimated using $h(Z) = E[Y|Z] - E[g(m(z) + V)]$. The second term involves the density of $-V$, which can be estimated using $f_{-V} = \mathcal{F}^{-1}(\Psi_{-V})$.

Let us now sketch how we could achieve consistency, following Zinde-Walsh (2014). Because we deal with tempered distributions here, it is convenient to rely on the corresponding notion of convergence. A sequence (T_n) of tempered distributions is said to converge to $T \in \mathcal{S}'$ (and we denote $T_n \rightharpoonup T$) if for all $\varphi \in \mathcal{S}$, $T_n(\varphi) \to T(\varphi)$. Similarly, a sequence of random tempered distributions T_n converges in probability to $T \in \mathcal{S}'$ ($T_n \overset{p}{\rightharpoonup} T$) if for all $\varphi \in \mathcal{S}$, $T_n(\varphi) \xrightarrow{\text{p}} T(\varphi)$. Such a notion of convergence is useful here because the Fourier transform preserves it, namely $T_n \rightharpoonup T$ implies that $\mathcal{F}(T_n) \rightharpoonup \mathcal{F}(T)$. Convergence in probability of $\widehat{g}$ can then be achieved if (i) the estimators $\widehat{q}_k$ of q_k satisfy $\widehat{q}_k \overset{p}{\rightharpoonup} q_k$ for $k \in \{1, 2, 3\}$ and (ii) we can prove that the problem is well-posed, namely $q_{kn} \rightharpoonup q_k$ for $k \in \{1, 2, 3\}$ implies that the corresponding g_n satisfies $g_n \rightharpoonup g$. (i) can be obtained by using, e.g., a trimmed kernel estimator of q_k,

$$\widehat{q}_k(m) = \min(\max(\widetilde{q}_k(m), -C(1 + m^2)^K), C(1 + m^2)^K),$$

where $\widetilde{q}_k$ is the usual kernel estimator of q_k and C and K are two tuning parameters. We refer to Zinde-Walsh (2014) for a proof of (i) in such a context. (ii) is more challenging. It has been established by Zinde-Walsh (2014, see Theorem 5) in a similar but simpler context. Note that $q_{kn} \rightharpoonup q_k$ implies $\mathcal{F}(q_{kn}) \rightharpoonup \mathcal{F}(q_{kn})$. Similarly, $\mathcal{F}(g_n) \rightharpoonup \mathcal{F}(g)$ implies $g_n \rightharpoonup g$. But the intermediate step establishing that $\mathcal{F}(q_{kn}) \rightharpoonup q_{kn}$ implies $\mathcal{F}(g_n) \rightharpoonup \mathcal{F}(g)$ is difficult, as it involves nonlinear operations on the tempered distributions at hand. We leave this issue for future research.

C.2 Testing the polynomial restriction

To distinguish between the polynomial and non-polynomial cases, note that under Assumptions 2.1, 2.2 and either Assumption 2.3 or 2.4, Proposition 2.1 ensures that g is a polynomial if and only if $E[Y(X - m(Z))|m(Z) = m]$ is a polynomial in m. A statistical test can be developed based on this proposition. First, we estimate $Q_1 = Y(X - m(Z))$ by $\widehat{Q}_1 = Y(X - \widehat{m}(Z))$. Second, we test whether the nonparametric regression of $\widehat{Q}_1$ on $\widehat{m}(Z)$ is a polynomial (of degree at most K, say) or not. There are several such specification tests in the literature, see e.g., Zheng (1996). However, one would need to take into account the fact that both the dependent and independent variables are generated here. This is likely to modify the asymptotic distribution of the test statistic, so some procedure like a bootstrap may be convenient for proper inference.

D Additional simulation results

We use the same models and data generating processes as in the simulations section but check the robustness of the estimators to measurement error that follows a t distribution with 12 degrees of freedom, a uniform distribution and a bimodal distribution. The tables below show results that are qualitatively similar to those in the main section: the RMSEs for the MEC estimators are stable for different amounts of measurement error while Robinson's estimator has lowest RMSE when there is no measurement error and quickly increases with small amounts of measurement error.

Table 3: Performances of MEC and Robinson's estimators with $V \sim U[-2, 2]$

				α_1			α_2	
Model	Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
Model 1	MEC	0	-0.020	0.098	0.100	0.006	0.035	0.035
		1/4	-0.017	0.113	0.114	0.004	0.050	0.050
		1	-0.013	0.141	0.141	0.015	0.067	0.068
	Robinson's	0	0.001	0.034	0.034	0.000	0.009	0.009
		1/4	-0.155	0.081	0.175	-0.160	0.025	0.162
		1	-0.435	0.101	0.446	-0.439	0.036	0.441
Model 2	MEC	0	0.004	0.041	0.041			
		1/4	-0.003	0.048	0.048		—	
		1	-0.004	0.059	0.059			
	Robinson's	0	0.002	0.036	0.036			
		1/4	-0.156	0.034	0.160		—	
		1	-0.426	0.029	0.427			

Notes: results from 100 simulations of sample size $1,000$.

Table 4: Performances of the MEC and Robinson's estimators with $V \sim \frac{1}{2}\mathcal{N}(-2,1) + \frac{1}{2}\mathcal{N}(2,1)$

				α_1			α_2	
Model	Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
Model 1	MEC	0	-0.001	0.085	0.085	-0.000	0.069	0.069
		1/4	0.062	0.110	0.126	0.032	0.069	0.076
		1	0.062	0.116	0.131	0.076	0.078	0.108
	Robinson's	0	0.001	0.018	0.018	0.000	0.004	0.004
		1/4	-0.045	0.073	0.085	-0.052	0.023	0.057
		1	-0.171	0.122	0.209	-0.181	0.040	0.186
Model 2	MEC	0	0.000	0.026	0.026		—	
		1/4	-0.003	0.043	0.043			
		1	0.001	0.045	0.045			
	Robinson's	0	0.001	0.018	0.018		—	
		1/4	-0.047	0.019	0.051			
		1	-0.168	0.021	0.170			

Notes: results from 100 simulations of sample size $1,000$.

Table 5: Performances of the MEC and Robinson's estimators with $V \sim t(12)$

				α_1			α_2	
Model	Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
Model 1	MEC	0	-0.026	0.129	0.130	-0.005	0.070	0.070
		1/4	-0.025	0.134	0.136	0.000	0.082	0.082
		1	-0.023	0.139	0.141	-0.009	0.078	0.078
	Robinson's	0	-0.001	0.037	0.037	0.001	0.011	0.011
		1/4	-0.177	0.132	0.221	-0.169	0.041	0.174
		1	-0.461	0.172	0.492	-0.451	0.053	0.454
Model 2	MEC	0	-0.001	0.052	0.052		—	
		1/4	-0.003	0.055	0.055			
		1	-0.004	0.069	0.069			
	Robinson's	0	0.001	0.039	0.039		—	
		1/4	-0.169	0.041	0.173			
		1	-0.447	0.038	0.448			

Notes: results from 100 simulations of sample size $1,000$.