

המכון למחקר כלכלי בישראל  
על-שם מוריס פאלק בע"מ (חל"צ)  
The Maurice Falk Institute for  
Economic Research in Israel Ltd.



## **Identification of Additive and Polynomial Models of Mismeasured Regressors Without Instruments**

by

Dan Ben-Moshe, Xavier D'Haultfoeuille,  
and Arthur Lewbel

**Discussion Paper No. 15.03**

**May 2015**

בנין פרץ נפתלי, קמפוס האוניברסיטה העברית, הר הצופים, ירושלים 91905  
The Hebrew University Campus, MT. Scopus, 91905 Jerusalem, Israel  
[www.falk.huji.ac.il](http://www.falk.huji.ac.il)

# Identification of Additive and Polynomial Models of Mismeasured Regressors Without Instruments\*

By

Dan Ben-Moshe  
The Hebrew University of Jerusalem

Xavier D'Haultfoeuille  
Centre de Recherche en Économie et Statistique

and

Arthur Lewbel  
Boston College

## Abstract

Suppose we observe only a dependent variable  $Y$ , a mismeasured  $X$  (with unobserved true value  $X^*$ ), and another covariate  $Z$ . We show nonparametric point identification of the model  $E[Y \setminus X^*, Z] = g(X^*) + h(Z)$ , for unknown functions  $g$  and  $h$ . If  $h(Z)$  were identically zero, then  $g(X^*)$  could be identified by Schennach (2007), using  $Z$  as an instrument. Our results replace her exclusion restriction for  $Z$  with the much weaker assumption of additivity in  $Z$ . Additive models are popular and measurement errors are ubiquitous, so our results combining the two should have widespread potential application. We also identify the model where  $g(X^*)$  is replaced with a polynomial in both  $X^*$  and  $Z$ . Our identification proofs are constructive, and so can be used to form estimators.

JEL codes: C14, C26

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

**The Maurice Falk Institute for Economic Research in Israel Ltd.**

Jerusalem, May 2015 • Discussion Paper No. 15.03

---

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

# 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 error 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  $Y$ , a correctly measured covariate  $Z$ , and the mismeasured  $X$ . The main model we consider is

$$E[Y | X^*, Z] = g(X^*) + h(Z), \quad X = X^* + U, \quad (1)$$

where  $g$  and  $h$  are unknown functions, the true  $X^*$  is unobserved, and  $U$  is the unobserved measurement error. We also assume low order moments of  $U$  are independent of  $X^*$  and  $Z$ , which is weaker than the standard assumption of independent 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 & 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 type of exclusion restriction that is ordinarily needed for nonparametric identification with measurement error. If the function  $h$  were not present in equation (1), then  $Z$  would be excluded from the model and could therefore be used as an instrument. Identification in this case could be based on, e.g., Schennach (2007). At the other extreme, if no restrictions were placed on  $E[Y | X^*, Z]$ , 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  $Y = g(X^*) + \varepsilon^*$  where we replace the usual exclusion assumption that  $E[\varepsilon^* | X^*, Z] = 0$  with the much weaker assumption that  $E[\varepsilon^* | X^*, Z] = h(Z)$ . Essentially, we still interpret  $Z$  as an instrument, but instead of the usual exclusion restriction that  $Z$  drops out entirely from the model, we allow  $Z$  to appear in the model, but only additively.

In our main result, we place restrictions on how  $X^*$  covaries with  $Z$ , and show nonparametric identification of both  $g$  and  $h$ . In an extension section, we consider alternatives in which these restrictions are weakened, at the price of a higher level assumption. We also extend to the case where  $g(X^*)$  is replaced with a polynomial in both  $X^*$  and  $Z$ , still keeping  $h(Z)$  nonparametric, in which case the model is  $E[Y | X^*, Z] = \sum_{j=0}^J \sum_{k=1}^K \alpha_{jk} Z^j X^{*k} + h(Z)$ . Despite the considerable interaction between  $X^*$  and the only available instrument  $Z$  in this model, we still can obtain point identification.

Consider the general class of models  $Y = M(X^*, Z) + \varepsilon$  with restrictions placed on  $M$  and  $\varepsilon$ . 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 moments of  $\varepsilon$  (in addition to placing restrictions on  $M$ ). For example, Chen et al. (2008, 2009) and Schennach & Hu (2013) assume  $\varepsilon$  is independent of  $X^*$ , Erickson & Whited (2002) and Lewbel (1997) assume  $\varepsilon$  has a conditional third moment of zero, while Klein & Vella (2010) and Lewbel (2012) impose constraints on how the variance of  $\varepsilon$  depends on  $X^*$  and  $Z$ . In contrast, the only constraint the present paper imposes on  $\varepsilon$  is the standard conditional mean (nonparametric regression) assumption  $E[\varepsilon | Z, X^*] = 0$ . This should be useful in practice because many if not most behavioral models do not provide higher moment or alternative additional restrictions on  $\varepsilon$ .

In the model  $E[Y | X^*, Z] = g(X^*)$ , Schennach (2007) gains identification by exploiting moments like  $E[YX | Z]$ , which in her model equals  $E[g(X^*)X | Z]$ . We cannot use this same method because in our model  $E[YX | Z] = E[g(X^*)X + h(Z)X | Z]$ , and we would not be able to separate the effect of  $Z$  on  $h$  from the effect of  $Z$  on  $g$ . So we instead consider moments like  $\text{Cov}(Y, X | Z)$ . The function  $h(Z)$  drops out of this covariance, that is,  $\text{Cov}(Y, X | Z) = \text{Cov}(g(X^*), X | Z)$ . Essentially, we construct conditional moments for identifying  $g$  by using covariances that project off the unknown function  $h$ , analogous to the first step of Robinson's (1988) estimator for partially linear regression models. These conditional covariances contain less information than conditional means, so we need to work with more moments and more restrictions on how  $X^*$  covaries with  $Z$  than Schennach (2007) does to gain identification.

When  $g$  is nonparametric, we use covariances like  $\text{Cov}(Y, X^k | Z)$  for small integers  $k$  to obtain expressions for convolutions between  $g(X^*)$  and  $X^k - E[X^k | Z]$ . We manipulate these expressions to solve for the characteristic function of the latter, which then lets us back out the function  $g$ . The analysis is simplified when  $g$  is a polynomial. In that case we can directly obtain the coefficients of the polynomial by regressing  $\text{Cov}(Y, X | Z)$  on moments of  $X$  conditional on  $Z$ . Unlike related results in, e.g., Hausman et al. (1991), we obtain closed form expressions for these coefficients.

The next section provides our main model and its identification, and we then provide extensions in Section 3. One set of alternative results in this section replaces a nonparametric  $g(X^*)$  with a polynomial in both  $X^*$  and  $Z$ . Other results in Section 3 weaken our main assumptions regarding the relationship between  $X^*$  and  $Z$ . Our identification strategies are constructive, so estimators could be based on them. Although our main focus is on identification rather than estimation, in Section 4 we provide some informal discussion on how corresponding nonparametric or semiparametric estimators might be obtained. We then present some simulation studies that compare these estimators to an estimator that ignores measurement error. Section 6 concludes.

## 2 The model and main result

We consider in this section the nonparametric additive model,

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

We impose the following normalization and moment conditions.

**Assumption 2.1.** (i)  $g(x_0^*) = 0$  for some  $x_0^* \in \text{Support}(X^*)$ ; (ii)  $E[\varepsilon|X^*, Z] = 0$ ; and (iii)  $E[U^k|X^*, Z] = E[U^k] = \mu_k$  for  $k \in \{1, 2, 3\}$  and  $\mu_1 = 0$ .

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 the estimand  $g(X^*) + h(Z)$  is the conditional mean function  $E[Y | X^*, Z]$ . Importantly, this allows for heteroscedasticity of unknown form in  $\varepsilon$ , as well as not restricting dependence in any higher moments of  $\varepsilon$ . 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  $\text{Support}(m(Z)) = \mathbb{R}$ ,  $V \perp\!\!\!\perp Z$ ,  $E[V] = 0$  and  $E[V^2] > 0$ .

Here  $m(Z)$  can be defined 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 very 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 section 3 we provide additional results that weaken this assumption. The large support condition on  $m(Z)$ , which implies that  $\text{Support}(X^*)$  also equals  $\mathbb{R}$ , is also made by Schennach (2007) and is required for our approach based on Fourier transforms. We do not impose this assumption directly on  $\text{Support}(X^*)$  because we provide other results later that are not based on Fourier transforms and so do not require this restriction.

The condition  $E[V] = 0$  is a free location normalization, while the condition  $E[V^2] > 0$  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  $h$ .

As noted in the introduction, when the function  $h(Z)$  is known to equal zero for all  $Z$ , Schennach (2007) shows that identification of  $g$  can be achieved using  $E[Y|Z]$  and  $E[XY|Z]$ . Our idea here is to obtain similar identifying equations on  $g$  that are conditional covariances rather than conditional means, and thereby do not depend on the  $h$  function. For that purpose, note first that  $m(Z)$  is

identified by  $m(Z) = E[X | Z]$ . Then the functions  $\text{Cov}(Y, (X - m(Z))^k | Z = z)$  satisfy

$$\text{Cov}(Y, X - m(Z) | Z = z) = E[Vg(m(z) + V)]$$

$$\text{Cov}(Y, (X - m(Z))^2 | Z = z) = E[(V^2 - \nu_2)g(m(z) + V)]$$

$$\text{Cov}(Y, (X - m(Z))^3 | Z = z) = E[(V^3 - \nu_3)g(m(z) + V)] + 3(m_2 - \nu_2)\text{Cov}(Y, X - m(Z) | Z = z),$$

where  $\nu_k = E[V^k]$  and  $m_2 = E[(U + V)^2]$ . The constant  $m_2$  is identified by  $m_2 = E[(X - E[X|Z])^2]$ . These covariances depend on  $z$  only through  $m(z)$ , so let  $q_k(m) = \text{Cov}(Y, (X - m(Z))^k | m(Z) = m)$ . These expressions show that each  $q_k$  can be written as a convolution between  $g$  and the distribution of  $-V$ . Following similar reasoning as Schennach (2007) and Zinde-Walsh (2014) we work with the Fourier transforms  $F_k$  of  $q_k$  because Fourier transforms of convolutions are products of Fourier transforms. Let  $\Psi$  denote the characteristic function of  $-V$  and let  $\Gamma$  be the Fourier transform of  $g$ . Ignoring details for now, taking Fourier transforms of the above covariances gives the following equations that we use for identification:

$$F_1 = \Gamma \times (i\Psi') \tag{3}$$

$$F_2 = -\Gamma \times (\Psi'' + \nu_2\Psi) \tag{4}$$

$$F_3 = -\Gamma \times (i\Psi''' - 3i(m_2 - \nu_2)\Psi' + \nu_3\Psi). \tag{5}$$

We then have  $\lambda_0 F_1 + iF_2 = 0$  and  $\mu_0 F_1 + iF_3 = 0$ , with

$$\lambda_0 = \frac{\Psi'' + \nu_2\Psi}{\Psi'} \tag{6}$$

$$\mu_0 = \frac{i\Psi'''(t) + \nu_3\Psi}{\Psi'} - 3i(m_2 - \nu_2). \tag{7}$$

Suppose that the functional equation  $\lambda \times F_1 = 0$  has the unique solution  $\lambda = 0$ . Then  $\lambda_0$  and  $\mu_0$  are the unique solutions to  $\lambda F_1 + iF_2 = 0$  and  $\mu F_1 + iF_3 = 0$ , respectively. Because  $F_1, F_2$  and  $F_3$  are identified, this implies that  $\lambda_0$  and  $\mu_0$  are identified as well. Then by L'Hôpital's rule and Taylor expansions, we can prove that  $\nu_2 = (3m_2 - (\lambda_0(0))^2 - 2\lambda_0'(0) - i\mu_0(0))/2$  and  $\nu_3 = i\nu_2\lambda_0(0)$ . Hence,  $\nu_2$  and  $\nu_3$  are identified. Solving the differential equation (6), we then recover  $\Psi$ , and given  $\Psi$  we obtain  $\Gamma$  by (3). Because the Fourier transform is invertible, we then obtain  $g$  from  $\Gamma$ . Finally, because  $g$  and the distribution of  $V$  are identified, so is  $E[g(X^*)|Z]$ , and then  $h(Z)$  is identified by  $E[Y - g(X^*)|Z]$ .

These identification steps rely on three claims that need to be satisfied. The first is that Equations (3)-(5) hold. Because  $q_1, q_2$  and  $q_3$  may not be integrable on the real line, the Fourier transforms above are not defined in the usual sense. But if  $q_k$  is bounded by a polynomial,  $F_k$  can still be defined as its Fourier transform, seen as a tempered distribution (see Appendix A.1 for details). The second claim is that the functional equation in  $\lambda$ ,  $\lambda \times F_1 = 0$ , has the unique solution  $\lambda = 0$ . Because we are dealing with tempered distributions, this is not immediate and must be verified. The third is that there is a unique tempered distribution  $G$ , namely  $\Gamma$ , that solves  $F_1 - G \times (i\Psi') = 0$ . To ensure that these claims hold, we make the following assumptions.

**Assumption 2.3.** (i)  $E[\exp(|V|^\beta)] < +\infty$  for some  $\beta > 0$  and (ii)  $\Psi'$ , the derivative of the characteristic function of  $V$ , only vanishes at 0.

**Assumption 2.4.** We have either: (i)  $g$  is bounded by a polynomial and the interior of the support of  $\Gamma$  is not empty or (ii)  $g$  is a polynomial of some order  $K > 1$ .

The condition on  $\Psi'$  in Assumption 2.3 is a strengthening of the standard condition in measurement error problems that  $\Psi$  does not vanish. Assumption 2.4 puts some restrictions on  $g$ . Assumption 2.4(i) by itself would rule out polynomials, since for polynomial  $g$  the support of  $\Gamma$  is  $\{0\}$  (see Appendix A.1 for a definition of the support of a tempered distribution). This assumption also rules out having  $g$  be finite combinations of sine and cosine functions. Note that this assumption is weaker than those made by Schennach (2007) and Zinde-Walsh (2014). Still, polynomials are popular models and we do not want to rule them out, so we add Assumption 2.4(ii) that allows for any polynomial, except a linear,  $g$ . Failure of identification in the case of  $g$  linear is not a limitation of our identification proof, but rather is a fundamental feature of the model. We prove this below by giving a specific example of a data generating process with a linear  $g$  that is not identified. However, we also show (in Section 3) that nonidentification of the linear model is due to the restriction in Assumption 2.2 that  $V$  is homoscedastic, or more precisely that  $E[X^2|Z]$  is linear in  $(E[X|Z])^2$ . In our extensions we consider cases where  $V$  can be heteroscedastic, which permits identification of a linear  $g$  can be identified. Finally, we note that Assumption 2.4 has observable implications and is partly testable. We employ these properties in our proof, by using observables both to tell if  $g$  is a polynomial or not, and to uncover information about  $\Gamma$  when  $g$  is not a polynomial.

**Theorem 2.1.** *Suppose that Equation (2) and Assumptions 2.1-2.3 hold. Then:*

- (i) *Assumption 2.4(ii) holds if and only if  $q_1$  is a polynomial;*
- (ii)  *$g$  and  $h$  are identified when Assumption 2.4(i) or 2.4(ii) holds.*

Proofs are provided in the appendix. Theorem 2.1(i), which allows us to tell if  $g$  is a polynomial or not, is important because our proof of identification is different under Assumption 2.4(i) than under 2.4(ii). Under Assumption 2.4(i), we first identify the full distribution of  $V$ , before identifying  $g$  and  $h$ . Under Assumption 2.4(ii), on the other hand, we identify simultaneously  $g$  and the low order moments of  $V$ . One might therefore use different estimators in the two cases. Theorem 2.1(i) shows that Assumption 2.4(ii) holds if and only if  $q_1$  is polynomial, which is convenient because  $q_1$  can be estimated in a straightforward way. Later in Section 4 we discuss possible estimators and a possible procedure for testing whether  $q_1$  is a polynomial or not.

We end this section by proving, using an example, that the model can fail to be identified when  $g$  is linear. Suppose that

$$- g(x^*) = \alpha(x^* - x_0^*) \text{ with } \alpha \neq 0;$$

- $U = U_1 + U_2$  with  $U_1, U_2$  and  $(X^*, Z, \varepsilon)$  mutually independent,  $E[U_1] = E[U_2] = 0$  and  $E[\exp(|U_1|^\gamma)] < \infty$  for some  $\gamma > 0$ ;
- $E[V|U_1 + V = x] = \rho x$  with  $\rho \neq 0$  and for all  $t \neq 0$ ,  $\Psi'_{-U_1}(t)\Psi(t) + \Psi_{-U_1}\Psi'(t) \neq 0$ , with  $\Psi_{-U_1}$  the characteristic function of  $-U_1$ .

The first two restrictions are compatible with Assumptions 2.1. The last condition is satisfied if, e.g.,  $(U_1, V)$  has a nondegenerate gaussian distribution. Let  $\tilde{X}^* = X^* + U_1$ ,  $\tilde{g}(x^*) = \alpha\rho(x^* - x_0^*)$ ,  $\tilde{h}(z) = h(z) + \alpha(1 - \rho)(m(z) - x_0^*)$  and  $\tilde{\varepsilon} = \varepsilon + \alpha(1 - \rho)V - \alpha\rho U_1$ . We then have, by construction,

$$Y = \tilde{g}(\tilde{X}^*) + \tilde{h}(Z) + \tilde{\varepsilon},$$

with  $\tilde{g}(x_0^*) = 0$  and  $E[\tilde{\varepsilon}|\tilde{X}^*, Z] = 0$ . Moreover,  $X = \tilde{X}^* + U_2$  with  $U_2 \perp\!\!\!\perp (X^*, Z)$  and  $E[U_2] = 0$ . Therefore, this alternative model still satisfies Assumption 2.1. Also, since  $\tilde{X}^* = m(Z) + \tilde{V}$  with  $\tilde{V} = V + U_1$  independent of  $Z$  and  $E[\tilde{V}] = 0$ , we have that  $\tilde{X}^*$  satisfies Assumption 2.2. Finally, the conditions above ensure that  $\tilde{V}$  satisfies Assumption 2.3. As a result, we can rationalize  $(\tilde{g}, \tilde{h})$  with the data and the model. This means that  $(g, h)$  and  $(\tilde{g}, \tilde{h})$  are observationally equivalent, and so  $(g, h)$  cannot be identified in this case. We show later that heteroscedasticity in  $V$  can overcome this nonidentification.

### 3 Extensions

As discussed earlier, Assumption 2.2 is particularly strong. We now investigate ways that it can be weakened. First, we replace  $g$  with a function that is restricted to be a polynomial, though we allow it to be a polynomial in both  $X^*$  and  $Z$ . In this model we can completely replace Assumption 2.1 with a testable rank condition on moments of  $X$  conditional on  $Z$ . Second, we return to the original model where  $g$  is a nonparametric function of  $X^*$ , but now we allow for multiplicative heteroscedasticity in the expression for  $X^*$ , that is, we replace Assumption 2.2 with  $X^* = m(Z) + \sigma(Z)V$ . We show this alternative identifies the model, (including the case where  $g$  is linear), up to an unknown scalar  $\sigma(z_0)$ , and that  $\sigma(z_0)$  is identified by an assumption that it is the unique solution to one of an infinite number of equations. The advantage of this result over our original theorem is the greater generality of our expression for  $X^*$ , but it has the disadvantage of requiring a high level assumption on the identification of the value of the function  $\sigma(Z)$  at a single point  $Z = z_0$ .

#### 3.1 A polynomial restriction

For our next result we replace the function  $g$  with a polynomial in both  $X^*$  and  $Z$ , so

$$\begin{cases} Y &= g(X^*, Z) + h(Z) + \varepsilon = \sum_{j=0}^J \sum_{k=1}^K \alpha_{jk} Z^j X^{*k} + h(Z) + \varepsilon \\ X &= X^* + U \end{cases} \quad (8)$$



where  $\alpha_{jK} \neq 0$  for some  $j$ . This model is more general than our original one in relaxing additivity by allowing interactions between  $X^*$  and  $Z$ , but it is less general in constraining  $g$  to be a polynomial. Our identification strategy now is to express  $\text{Cov}(Y, X|Z)$  in terms of the coefficients  $\alpha_{jk}$ , moments of  $X$  conditional on  $Z$ , and moments of  $U$ .

**Assumption 3.1.** (i)  $E[\varepsilon|X^*, Z] = 0$  and (ii)  $E[U^k|X^*, Z] = E[U^k] = \mu_k$  for  $k \in \{1, 2, \dots, K+1\}$  and  $\mu_1 = 0$ .

This assumption replaces Assumption 2.1. When  $K > 2$  this is stronger than Assumption 2.1 in that 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  $K$ . We do not need to include an explicit location normalization on  $g$  now, because Equation (8) already satisfies the location normalization  $g(0, z) = \sum_{j=0}^K \sum_{k=1}^K \alpha_{jk} z^j 0^k = 0$ .

Equation 8 and Assumption 3.1 imply that

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

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

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

to end up expressing  $\text{Cov}(X, Y|Z)$  as a linear combination of terms of the form  $Z^j E[X^k|Z]$  and  $Z^j E[X|Z]E[X^k|Z]$ . The coefficients of these observable conditional moments are then complicated but known functions of  $\alpha_{j1}, \dots, \alpha_{jK}$  and  $\mu_1, \dots, \mu_{K+1}$ , which can be solved for  $\alpha_{j1}, \dots, \alpha_{jK}$ , thereby identifying  $g$ . Replacing Assumption 2.2, all we need now for identification is a rank condition that allows us to obtain these conditional moment coefficients.

**Assumption 3.2.** *Define*

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

*Then  $E[R(Z)R(Z)']$  is finite and nonsingular.*

Assumption 3.2 gives us identification by ensuring that variation in  $Z$  induces sufficient relative variation in the moments  $E[X^k | Z]$  for  $k = 1, \dots, K+1$ . The vector  $R(Z)$  includes, for example, the functions  $E[X^2 | Z]$  and  $(E[X | Z])^2$  so the required nonsingularity of  $E[R(Z)R(Z)']$  requires relative variation in  $E[X | Z]$  and  $\text{Var}(X | Z)$ . Relative variation is also required for higher conditional moments of  $X$ . Assumption 3.2 may therefore conflict with, and so should be considered as an alternative to Assumption 2.2, where  $\text{Var}(X | Z)$  is constant. Note that the case where  $g$  is a

nonlinear polynomial in  $X^*$  is covered by both theorems, and so can be identified under either set of conditions.

**Theorem 3.1.** *Suppose that Equation (8) and Assumptions 2.1 and 3.2 hold. Then the functions  $h$  and  $g$  and the moments  $\mu_1, \mu_2, \dots, \mu_{K+1}$  are identified.*

The proof is based directly on the above covariance expansions. It is similar to the proof of Theorem 2.1 under Assumption 2.4(ii) but instead of first jointly identifying  $g$  and moments of  $V$ , Theorem 3.1 first identifies  $g$  and moments of  $U$ . Based on the discussion above, the proof uses Assumption 3.2 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  $\alpha_{j1}, \dots, \alpha_{jK}$  and  $\mu_1, \dots, \mu_{K+1}$ , which are then manipulated to recover these parameters and moments.

Our main theorem assumed  $m(Z)$  had support  $\mathbb{R}$ . In contrast, Theorem 3.1 allows  $Z$  to be limited or even discrete, even if  $X$  is continuous. However, the rank condition in Assumption 3.2 does require that  $Z$  have at least  $K + 2$  points of support.

Theorem 3.1 extends readily to the case where  $Z$  or  $X^*$  is a vector of covariates. As shown in Appendix B, this result can also be adapted quite easily to the case of multiplicative instead of additive measurement errors.

Unlike our main theorem, Theorem 3.1 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 3.1.** *Suppose that Equation (8) holds with  $Y = h(Z) + \alpha_1 X^* + \varepsilon$ , i.e.  $g$  is linear, and Assumptions 3.1 and 3.2 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 3.1 shows this model can be identified even when the model and measurement errors are normal. The key for identification here, in contrast to Theorem 2.1, is that Assumption 3.1 requires  $\text{var}(X|Z)$  to vary with  $Z$ , thereby requiring heteroscedasticity in the relationship between  $X^*$  and  $Z$ . Klein & Vella (2010) and Lewbel (2012) are other examples of models that require and exploit heteroscedasticity to obtain identification.

### 3.2 Additive model between $X^*$ and $Z$ with heteroscedasticity

Theorem 3.1 shows that the homoscedasticity of  $V$  imposed by Assumption 2.2, while helpful for constructing moments for nonparametric identification in our main theorem, will in some cases actually prevent identification. In this section we revisit the model of our main theorem, but now we weaken Assumption 2.2 by replacing it with the following assumption that allows for heteroscedasticity, albeit in multiplicative form.

**Assumption 3.3.**  $X^* = m(Z) + \sigma(Z)V$  with  $\text{Support}(m(Z)) = \mathbb{R}$ ,  $V \perp\!\!\!\perp Z$ ,  $E[V] = 0$  and  $E[V^2] = 1$ .  $\sigma(\cdot)$  is differentiable and there exists  $z_0 \in \text{Support}(Z)$  such that  $\sigma(z_0) > 0$  and  $\sigma'(z_0) \neq 0$ .

Under this assumption  $\text{Var}(\sigma(Z)V) = \sigma^2(Z)$  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 that  $E[V^2] = 1$  here is a free normalization because we can always divide  $\sigma(Z)$  and multiply  $V$  by a constant. In this heteroscedastic setting, we first separate  $U$  from  $V$  using  $X$  and  $Z$  only. To operationalize this idea, we need to slightly strengthen Assumption 2.1(iii) as follows. We also place the free location normalization on  $h$  instead of  $g$  because it turns out to be more convenient in this setting.

**Assumption 3.4.** (i)  $h(z_0) = 0$  for some  $z_0 \in \text{Support}(Z)$ ; (ii)  $E[\varepsilon|X^*, Z] = 0$ ; and (iii)  $U$  is independent of  $(V, Z)$ .

Assumption 3.4 is stronger than Assumption 2.1 only in that it assumes full independence of the measurement error  $U$  from the components of  $X^*$ . Identification now 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. Finally, we need to impose regularity conditions similar to those of Assumption 2.3.

**Assumption 3.5.** (i)  $E[U^2] < +\infty$  and (ii) The characteristic functions of  $U$  and  $V$  do not vanish.  $V$  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 (details are in Appendix A.4). First, using  $X - m(Z) = \sigma(Z)V + U$ , independence of  $U$ , and nonconstant  $\sigma(\cdot)$ , we show in the appendix 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  $g$ , up to the unknown  $\sigma_0$ , using the equation for  $Y$ ,

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

where the equality follows by Assumptions 3.4(i) and (ii) and  $\Psi_U$  denotes the characteristic function of  $U$ . Hence,

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

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

$$g(m(z_0) + x) = \frac{1}{f_{\sigma_0 V}(x)} \mathcal{F}^{-1} \left( \frac{E [Y \exp(it(X - m(Z))(\cdot))|Z = z_0]}{\Psi_U(\cdot)} \right) (x), \quad (9)$$

where  $\mathcal{F}^{-1}$  denotes the inverse Fourier transform. All terms on the right-hand side are either identified or depend on  $\sigma_0$ , so  $g$  and thus also  $h$  are identified up to the scalar constant  $\sigma_0$ . Note that  $\sigma(z)$  satisfies  $\sigma(z) = \sqrt{\text{Var}(X|Z = z) - \text{Var}(X|Z = z_0) + \sigma_0^2}$ , so given what has been identified, the function  $\sigma(z)$  is also identified up to the single constant  $\sigma_0$ .

Finally, to identify  $\sigma_0$ , we can rely as previously on  $\text{Cov}(Y, X|Z = z)$ ,

$$\text{Cov}(Y, X|Z = z) = E [g(m(z) + \sigma(z)V)\sigma(z)V]. \quad (10)$$

The left-hand side is identified, while the right-hand side consists only of functions that are identified up to  $\sigma_0$ . Since equation (10) holds for all  $z$ , this provides an infinite number of equations (through variation in  $z$ ) in the single scalar unknown  $\sigma_0$ . We therefore expect  $\sigma_0$  to be greatly overidentified. These equations are, however, extremely complicated functions of  $\sigma_0$ , and so we cannot produce low-level assumptions that guarantee that these equations identify  $\sigma_0$ . We therefore impose the following high-level condition. Henceforth, we add the superscript  $\sigma$  to indicate the dependence in  $\sigma_0$ , e.g.,  $g^\sigma$  is the  $g$  function obtained by (9) when  $\sigma_0$  is set equal to  $\sigma$ .

**Assumption 3.6.** *The mapping  $\sigma \mapsto [z \mapsto \sigma^\sigma(z) \int g^\sigma(m(z) + \sigma^\sigma(z)v) v f_V^\sigma(v) dv]$  is injective.*

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

**Theorem 3.2.** *Suppose that Equation (2) and Assumptions 3.3-3.6 hold. Then the functions  $g$  and  $h$  are identified.*

This result relies on Assumption 3.6, 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 3.1.** *Suppose that Equation (2) and Assumptions 3.3-3.5 hold. Suppose also that  $g$  is linear, not constant, and  $U$  and  $V$  are normally distributed. Then Assumption 3.6 holds and thus  $g$  and  $h$  are identified.*

This special case is of interest because we showed earlier that under the assumptions of our main theorem, the model was not identified if  $g$  is linear and  $U$  and  $V$  are both normal. Now, using Theorem 3.2 instead of Theorem 2.1, and in particular replacing Assumption 2.2 with Assumption 3.3, the model with linear  $g$  and normal  $U$  and  $V$  is identified, since we have verified in this case that Assumption 3.6 holds. We conjecture that this holds more generally. Basically, the heteroscedasticity function  $\sigma(z)$  provides additional variation to help in the identification of  $g$ , and is itself identified up to  $\sigma_0$  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  $\sigma_0$ .

## 4 Discussion on inference

The focus of our paper is identification rather than estimation. However, our identification proofs are largely constructive, and so could be used to motivate corresponding estimators. Though we do not provide formal results, we explain in this section how inference might be conducted with a sample  $\{Y_i, X_i, Z_i\}_{i=1, \dots, n}$  of independent and identically distribution variables under Assumptions 2.1-2.4.

As emphasized before, Theorem 2.1 establishes identification of  $g$  and  $h$  both with polynomial or nonpolynomial  $g$ . However, the steps involved in identification are different in the two cases, and so they lead to different corresponding estimators. We therefore discuss estimation of the two cases separately below. We then indicate how one might test which of these two cases holds. A thorough limiting distribution theory for these estimators, which we do not attempt to provide, would ideally take the impact of this pretesting into account.

## 4.1 Polynomial case

We first consider the case where  $g$  is a nonlinear polynomial, so  $g(x) = \sum_{k=0}^K \alpha_k x^k$  where  $K > 1$  and an upper bound on  $K$  is assumed to be fixed and known. We only need to estimate  $(\alpha_1, \dots, \alpha_K)$ , since  $\alpha_0$  can then be obtained directly from the normalization  $g(x_0^*) = 0$ . As shown in the appendix,  $q_1, q_2$  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  $(\alpha_1, \dots, \alpha_K)$  and  $(\nu_2, \dots, \nu_{K+3})$  (with  $\nu_k = E[V^k]$ ) from the estimates of  $\beta_{kj}$ . Finally, we estimate  $h$ .

First consider the estimation of the  $\beta_{kj}$ , for  $k \in \{1, 2, 3\}$ . We have

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

and, under Assumptions 2.1-2.2,

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

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

$$q_k(m) = E [Y ((X - m(Z))^k - m_k) | m(Z) = m]. \quad (11)$$

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  $E[X|Z]$ , such as a standard kernel or local polynomial estimator. While  $m_1 = 0$  by construction, we can estimate  $m_k$  for  $k \in \{2, 3\}$  by simply taking the average of  $(X_i - \hat{m}(Z_i))^k$ . Then we can estimate the  $\{\beta_{kj}\}_{j=0}^{K-1}$  by regressing  $\hat{Q}_k = Y [(X - \hat{m}(Z))^k - \hat{m}_k]$  on  $(1, \hat{m}(Z), \dots, \hat{m}(Z)^{K-1})$ .

In a second step, we can then use a classical minimum distance estimator (see, e.g., Wooldridge 2002, Section 14.6) to estimate the  $2K+1$  unknown coefficients and moments  $(\alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+2})$  from

the  $3K$  equations (see the appendix for more details),

$$\begin{aligned}\beta_{1j+1} &= \sum_{k=j+1}^K \alpha_k \binom{k}{j} \nu_{k-j+1}, \\ \beta_{2j+1} &= \sum_{k=j+1}^K \alpha_k \binom{k}{j} (\nu_{k-j+2} - \nu_2 \nu_{k-j}), \\ \beta_{3j+1} &= \sum_{k=j+1}^K \binom{k}{j} \alpha_k (\nu_{k-j+3} - 3(m_2 - \nu_2) \nu_{k-j+1} - \nu_3 \nu_{k-j}).\end{aligned}$$

where  $j \in \{0, \dots, K-1\}$ . In the proof of Theorem 2.1 identification comes from closed-form expressions of  $(\alpha_1, \dots, \alpha_K)$  and  $(\nu_2, \dots, \nu_{K+2})$  in terms of  $\beta_{kj}$ , which in turn could be used for estimation. However,  $(\alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+2})$  is generally overidentified by these moments, making minimum distance generally more efficient than using the closed-form expressions for  $(\alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+2})$  (though the latter could be used to obtain consistent starting values for the minimum distance estimation).

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

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

Thus,  $E[g(X^*)|Z = z]$  can be simply estimated by

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

and we can then estimate  $h$  by  $\widehat{h}(z) = \widehat{E}[Y|Z = z] - \widehat{E}[g(X^*)|Z = z]$ .

Note that estimation based on Theorem 3.1 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  $\beta$  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  $(\alpha_{01}, \dots, \alpha_{KK})$ ,  $(\mu_2, \dots, \mu_{K+1})$ , and  $(E[X^*|Z], \dots, E[X^{*K}|Z])$  are estimated by replacing  $\beta$  with  $\widehat{\beta}$ . Finally, we can estimate  $h(Z)$  by  $\widehat{h}(Z) = E[Y|Z] - \sum_{j=0}^J Z^j \sum_{k=1}^K \widehat{\alpha}_{jk} \widehat{E}[X^{*k}|Z]$ .

## 4.2 Non-polynomial case

In the non-polynomial case, identification is based on Equations (3)-(5) using Fourier transforms of tempered distributions. Such equations are far more delicate than the previous moment equations to turn into an estimator, so we only sketch how a consistent estimator might be obtained

here, following Zinde-Walsh (2014, Section 5.2). The idea is first to estimate  $q_k(\cdot)$  by kernel estimators, using (11), and then take their Fourier transforms. To ensure convergence of these Fourier transforms in the space of tempered distributions, the growth at infinity of the estimator has to be controlled. One possibility is to trim an initial estimator  $\tilde{q}_k$ , and consider  $\hat{q}_k(m) = \min(\max(\tilde{q}_k(m), -C(1+m^2)^K), C(1+m^2)^K)$ , for some tuning constants  $C$  and  $K$ . In a similar context, Zinde-Walsh (2014) proves that  $\hat{F}_k = \mathcal{F}(\hat{q})$  then converges to  $F_k$  under appropriate conditions. If the kernel has compact support,  $\hat{q}_k$  also has compact support, implying that  $\hat{F}_k$  is infinitely differentiable. We then estimate  $\lambda_0 = -iF_2/F_1$  and  $\mu_0 = -iF_3/F_1$  by  $\hat{\lambda} = -i\hat{F}_2/\hat{F}_1$  and  $\hat{\mu} = -i\hat{F}_3/\hat{F}_1$ .

The next step is to observe that Equations (14), (15) and (17) in the Appendix imply

$$\begin{aligned}\nu_2 &= \frac{1}{2}(3m_2 - (\lambda_0(0))^2 - 2\lambda_0'(0) - i\mu_0(0)), \\ \nu_3 &= i\nu_2\lambda_0(0), \\ \Psi(t) &= \exp\left(\int_0^t \frac{\lambda_0(s)\nu_2 + i\nu_3}{\lambda_0(s)^2 + \lambda_0'(s) + i\mu_0(s) - 3m_2 + 2\nu_2} ds\right).\end{aligned}$$

We can then estimate these quantities by plug-in estimators, replacing  $\lambda_0$  and  $\mu_0$  by  $\hat{\lambda}$  and  $\hat{\mu}$ . In turn, we can estimate  $\Gamma$  using (3). As in Zinde-Walsh (2014), one may have to trim the estimator of  $\Psi'$  to consistently estimate  $\Gamma$  and then  $g = \mathcal{F}^{-1}(\Gamma)$ . 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  $f_{-V} = \mathcal{F}^{-1}(\Psi)$ .

### 4.3 Testing the polynomial restriction

To distinguish between the polynomial and nonpolynomial cases above, note that under the maintained Assumptions 2.1-2.3, Theorem 2.1(i) ensures that  $g$  is polynomial if and only if  $E[Y(X - m(Z))|m(Z) = m]$  is a polynomial in  $m$ . As discussed above, we can estimate  $Q_1 = Y(X - m(Z))$  by  $\hat{Q}_1 = Y(X - \hat{m}(Z))$ . We can then test whether the nonparametric regression of  $\hat{Q}_1$  on  $\hat{m}(Z)$  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.

## 5 Simulations

The main point of our identification theorems is to account for regressor measurement error in an additive nonparametric regression model. To illustrate the importance of accounting for measure-

ment error, we present results from a few Monte-Carlo simulations. The data  $\{Y_i, X_i, Z_i\}_{i=1}^n$  is generated from the following model that has quadratic  $g(X^*) = \alpha_1 X^* + \alpha_2 X^{*2}$ ,

$$Y = Z + Z^2 + \alpha_1 X^* + \alpha_2 X^{*2} + \varepsilon, \quad X^* = Z + V, \quad X = X^* + U$$

where  $U$ ,  $V$ , and  $\varepsilon$  are i.i.d. standard normal,  $Z$  follows the mixture of normals  $0.5N(-2, 3) + 0.5N(3, 2)$  and  $\alpha_1 = \alpha_2 = 1$ . We consider samples of size 1,000, 5,000 and 10,000 and draw 100 simulations for each.

Table 1 reports the mean and standard deviations of the estimates of  $\alpha_1$  and  $\alpha_2$  using (a) Robinson’s (1988) estimator which ignores measurement error and (b) Theorem 2.1 as described in Section 4.1, which takes measurement error into account.

Table 1: Mean and standard deviation of estimators of  $\alpha_1$  and  $\alpha_2$

$N$	$\alpha_1$			$\alpha_2$		
	1,000	5,000	10,000	1,000	5,000	10,000
Ignoring error	0.52	0.50	0.48	0.49	0.50	0.50
	(0.14)	(0.08)	(0.08)	(0.04)	(0.02)	(0.01)
Taking error into account	1.16	1.12	1.11	1.04	1.06	0.93
	(2.40)	(0.43)	(0.31)	(2.52)	(0.32)	(0.17)

Notes: results from 100 simulations of sample size  $N$ . Standard deviations are in parentheses.

The estimators that ignore measurement error are severely biased. In this case, we observe an attenuation bias, but this need not be the case in our framework because of the nonlinearity of the model. As the sample size increases, the means of the estimators based on Theorem 2.1 get closer to their true value and the standard deviations become smaller.

Next we generate data from a model with linear  $g(X^*) = \alpha_1 X^*$  and with heteroscedasticity,

$$Y = Z + \alpha_1 X^* + \varepsilon, \quad X^* = Z + ZV, \quad X = X^* + U$$

where again  $U$ ,  $V$ , and  $\varepsilon$  are i.i.d. standard normal,  $Z$  follows the same mixture of normals, and  $\alpha_1 = 1$ . Table 2, like Table 1, reports the mean and standard deviations of the estimates of  $\alpha_1$  using (a) Robinson’s (1988) estimator which ignores measurement error and (b) Theorem 3.1 as described in Section 4.1 which takes measurement error into account.

Table 2: Mean and standard deviations of estimates of  $\alpha_1$

$N$	1,000	5,000	10,000
Ignoring error	0.52	0.50	0.50
	(0.03)	(0.02)	(0.01)
Taking error into account	0.95	0.98	0.99
	(0.06)	(0.04)	(0.03)



The estimator that ignores measurement error is again severely biased. The estimator based on Theorem 3.1, on the other hand, is very close to the true value of  $\alpha_1$ . Its bias decreases quickly to zero, while its standard deviation is only twice as large as Robinson’s (1988) estimator.

## 6 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  $\varepsilon$  other than having conditional mean zero.

Measurement error is a common source of endogeneity in models, and two of the classic ways to obtain identification in structural econometric models is either by exclusion restrictions or by 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 of 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. One potential application of these results would be to estimate  $h(Z)$ , and thereby test whether the standard exclusion assumption for identification holds, by testing whether the estimated function  $h$  is identically zero. Additive models are popular in statistics, and measurement errors are ubiquitous, so our results that combine the two should have widespread potential application. 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. & Lewbel, A. (2008), ‘Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments’, *Economics Letters* **100**, 381–384.
- Chen, X., Hu, Y. & Lewbel, A. (2009), ‘Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information’, *Statistica Sinica* **19**, 949–968.
- D’Haultfœuille, X. (2011), ‘On the completeness condition in nonparametric instrumental regression’, *Econometric Theory* **27**, 460–471.
- Erickson, T. & Whited, T. M. (2002), ‘Two-step gmm estimation of the errors-in-variables model using high-order moments’, *Econometric Theory* **18**, 776–799.
- Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman and Hall.
- Hausman, J. A., Newey, W. K., Ichimura, H. & Powell, J. L. (1991), ‘Identification and estimation of polynomial errors-in-variables models’, *Journal of Econometrics* **50**, 273–295.
- Klein, R. & Vella, F. (2010), ‘Estimating a class of triangular simultaneous equations models without exclusion restrictions’, *Journal of Econometrics* **154**(2), 154–164.
- 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., Powell, J. L. & 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. & 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.
- 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 Proof of Theorem 2.1

Before proving the results, we recall some definitions related to the theory of distributions (see, e.g., Schwartz 1973). The Schwartz space  $\mathcal{S}$  is the subspace of  $C^\infty$  functions  $s$  such that for any  $(k, j) \in \mathbb{N}^2$ ,  $\lim_{x \rightarrow \pm\infty} |x|^j s^{(k)}(x) = 0$ . 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  $\phi$ , seen as a function in  $L^1(\mathbb{R})$ . When  $f$  is a function bounded by a polynomial, the linear form  $\tilde{f} : s \mapsto \int f(u)s(u)du$  defined on  $\mathcal{S}$  is a tempered distribution. In the absence of ambiguity, we assimilate  $f$  with  $\tilde{f}$  hereafter.

We proceed in several steps. We first show that Equations (3)-(5) hold, as these equations are needed for both Theorem 2.1(i) and (ii). We then prove Theorem 2.1(i). Then we show that  $g$  and  $h$  are identified under Assumption 2.4(i). Finally, we prove the same result under Assumption 2.4(ii).

### 1. Equations (3)-(5) hold.

We only prove that Equation (3) holds, as the exact same reasoning applies to Equations (4) and (5). 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

$$|q_1(m)| \leq E[|V||g(m + V)|] \leq C [E[|V|] + 2^{k-1}(E[|V|^{k+1}] + E[|V|]m^k)], \quad (12)$$

with  $E[|V|^{k+1}] < \infty$  by Assumption 2.3(i). 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'Haultfoeuille (2011, pp.469-470), we also have  $g_n \rightarrow g$  in  $\mathcal{S}'$ . Finally, let us show that  $q_{1n} = f \star g_n \rightarrow q_1$  in  $\mathcal{S}'$ . Let  $\Phi$  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  $|x| \rightarrow \infty$  and for every  $j$ , such that  $|\phi(x)| \leq b(x)$  for every  $x \in \mathbb{R}$  and every  $\phi \in \Phi$ . Then (12) implies that

$b \times q_1$  is integrable. The same inequality (12) applies to  $q_{1n}$ , implying that  $b \times (q_{1n} - q_1)$  is also integrable. Besides, by dominated convergence,

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

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

$$F_{1|\mathcal{U}} = \Gamma_{|\mathcal{U}} \times \mathcal{F}(f)_{|\mathcal{U}},$$

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(i),  $\mathcal{F}(f)$  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,  $\mathcal{F}(f) = i\Psi'$ . As a result, Equation (3) holds.

## 2. Assumption 2.4(ii) holds if and only if $q_1$ is a polynomial.

First suppose that Assumption 2.4(ii) holds. Let  $g(x) = \sum_{k=0}^K \alpha_k x^k$  and  $\nu_k = E[V^k]$  for  $k \geq 0$ . Then,

$$\begin{aligned} q_1(m) &= \sum_{k=1}^K \alpha_k E[V(m+V)^k] = \sum_{k=1}^K \alpha_k \sum_{j=0}^k \binom{k}{j} m^j \nu_{k-j+1} \\ &= \sum_{j=0}^{K-1} \left[ \sum_{k=j+1}^K \alpha_k \binom{k}{j} \nu_{k-j+1} \right] m^j. \end{aligned} \tag{13}$$

Hence,  $q_1$  is a polynomial of order at most  $K - 1$ .

Let us show that conversely, if  $q_1$  is a polynomial then Assumption 2.4(ii) is satisfied. If  $q_1$  is a polynomial, then  $F_1 = \sum_{k=0}^K a_k \delta_0^{(k)}$ , where  $\delta_0^{(k)}$  denotes the  $k$ -th derivative of the dirac distribution at 0. Hence, the support of  $F_1$  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,  $\Psi'$  does not vanish on  $\mathcal{O}$ . Because it is continuous,  $1/\Psi'$  is bounded and  $\phi/\Psi' \in \mathcal{S}$  with support included in  $\mathcal{O}$ . Then, by Equation (3),

$$0 = F_1(\phi/\Psi') = [\Gamma \times \Psi'](\phi/\Psi') = \Gamma(\phi),$$

and  $\Gamma$  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  $\Gamma$  is then  $\{0\}$ . This and the fact that  $\Gamma$  is tempered implies (Schwartz 1973, Théorème XXXV) that  $\Gamma$  is a finite linear combination of derivatives (in the sense of distributions) of Dirac measures, located at 0. By inverting the Fourier transform, this implies that  $g$  is a polynomial.

### 3. $g$ and $h$ are identified if Assumption 2.4(i) holds.

To show the identification of  $g$  and  $h$ , we prove first that (3)-(5) admit a unique solution in  $\Psi$  and  $\Gamma$ , up to a parameter. By taking the inverse Fourier transform of  $\Gamma$  and using the normalization  $g(x_0^*) = 0$ , we then recover  $g$ , and finally  $h$ . We decompose the proof into several steps.

(a) *The equation  $\lambda F_1 = 0$  for  $\lambda$  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  $\lambda$  such that  $\lambda F_1 = 0$ . Similarly to analytic functions, nonzero meromorphic functions have isolated zeros (see, e.g., Rudin 1987, p.209) and thus  $\lambda$  does not vanish on a bounded open set  $\mathcal{O} \subset \text{Support}(\Gamma) \setminus \{0\}$ . By Assumption 2.3,  $\Psi'$  does not vanish on  $\mathcal{O}$  either. Hence, for any  $\phi \in \mathcal{S}$  with support included in  $\mathcal{O}$ ,  $\phi/(\lambda\Psi')$  belongs to  $\mathcal{S}$  and has a support included in  $\mathcal{O}$ . Further, by Equation (3),

$$\Gamma(\phi) = \lambda \times [\Gamma \times \Psi'](\phi/\lambda\Psi') = (\lambda F_1)(\phi/\lambda\Psi') = 0.$$

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

(b)  *$\Psi$  is identified.*

Because  $E[\exp(|\beta|V)] < \infty$ ,  $\Psi$  is analytical on the strip  $\{z \in \mathbb{C} : |\text{Im}(z)| < \beta\}$ . Thus, the functions  $\lambda_0$  and  $\mu_0$  defined by (6) and (7) respectively are meromorphic on that strip, as ratios of analytic functions. By step (a), the equations  $\lambda F_1 + iF_2 = 0$  and  $\mu F_1 + iF_3 = 0$  in  $\lambda$  and  $\mu$  respectively and restricted to meromorphic functions admit unique solutions,  $\lambda_0$  and  $\mu_0$ . As a result,  $\lambda_0$  and  $\mu_0$  are identified. Then some algebra show that

$$\frac{\Psi'(t)}{\Psi(t)} = \frac{\lambda_0(t)\nu_2 + i\nu_3}{\lambda_0(t)^2 + \lambda_0'(t) + i\mu_0(t) - 3m_2 + 2\nu_2}. \quad (14)$$

Thus,  $\Psi'/\Psi$  is identified. Besides, by L'Hôpital's rule and a Taylor expansion of  $\Psi'(t)/\Psi(t)$  around 0,

$$\lambda_0(0) = \lim_{t \rightarrow 0} \lambda_0(t) = \lim_{t \rightarrow 0} \frac{\Psi''(t) + \nu_2\Psi(t)}{\Psi'(t)} = \lim_{t \rightarrow 0} \frac{\Psi'''(t) + \nu_2\Psi'(t)}{\Psi''(t)} = \frac{\Psi'''(0) + \nu_2\Psi'(0)}{\Psi''(0)} = -\frac{i\nu_3}{\nu_2}, \quad (15)$$

$$\frac{\Psi'(t)}{\Psi(t)} = \frac{\Psi'(0)}{\Psi(0)} + \left( \frac{\Psi''(0)\Psi(0) - \Psi'^2(0)}{(\Psi(0))^2} \right) t + o(t) = -\nu_2 t + o(t). \quad (16)$$

Substituting (15) and (16) into (14) and letting  $t \rightarrow 0$ , we obtain

$$\nu_2 = \frac{1}{2}(3m_2 - (\lambda_0(0))^2 - 2\lambda_0'(0) - i\mu_0(0)). \quad (17)$$

This implies that  $\nu_2$  and  $\nu_3 = i\nu_2\lambda_0(0)$  are identified. In turn  $\Psi$  is identified as the unique solution of the differential equation (14) satisfying  $\Psi(0) = 1$  and  $\Psi'(0) = 0$ .

(c)  $g$  and  $h$  are identified.

By Assumption 2.3,  $\Psi'$  vanishes only at 0. Moreover,  $\Psi''(0) = -\nu_2 \neq 0$ . Then, any other solution  $\tilde{\Gamma}$  of (3) satisfies  $\tilde{\Gamma} - \Gamma = c\delta_0$  for some real  $c$ . Because the Fourier transform is an automorphism on the space of tempered distributions, any  $\tilde{g}$  whose Fourier transform  $\tilde{\Gamma}$  satisfies (3) is such that  $\tilde{g} = g + c$ . The normalization  $g(x_0^*) = 0$  then implies that  $\tilde{g} = g$ . 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.

#### 4. $g$ and $h$ are identified if Assumption 2.4(ii) holds.

As before, 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  $(\alpha_1, \dots, \alpha_K)$ , since  $\alpha_0$  is then identified by the normalization  $g(x_0^*) = 0$ . First, Equation (13) 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 (13) also shows that for  $j \in \{0, \dots, K - 1\}$ , the quantities

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

are identified. Further, some algebra shows that

$$\begin{aligned} q_2(m) &= \sum_{j=0}^K \left[ \sum_{k=j+1}^K \alpha_k \binom{k}{j} (\nu_{k-j+2} - \nu_2 \nu_{k-j}) \right] m^j, \\ q_3(m) &= \sum_{j=0}^K \left[ \sum_{k=j+1}^K \binom{k}{j} \alpha_k (\nu_{k-j+3} + 3(m_2 - \nu_2) \nu_{k-j+1} - \nu_3 \nu_{k-j}) \right] m^j. \end{aligned}$$

Hence, we also identify, for  $j \in \{0, \dots, K - 1\}$ , the quantities

$$\begin{aligned} \beta_{2j+1} &= \sum_{k=j+1}^K \alpha_k \binom{k}{j} (\nu_{k-j+2} - \nu_2 \nu_{k-j}), \\ \beta_{3j+1} &= \sum_{k=j+1}^K \alpha_k \binom{k}{j} (\nu_{k-j+3} + 3(m_2 - \nu_2) \nu_{k-j+1} - \nu_3 \nu_{k-j}). \end{aligned}$$

We now show that this information allows us to identify  $(\alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+2})$ . For that purpose, let us first show that  $\nu_2$  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

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

which ensures that  $\nu_2$  is identified.

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

By induction, this shows that  $\alpha_1, \dots, \alpha_K, \nu_2, \dots, \nu_{K+2}$  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  $\beta_{31}, \dots, \beta_{3K-1}$  here.

Finally,

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

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

## A.2 Proof of Theorem 3.1

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,

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



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  $\text{Cov}(X, Y|Z)$  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  $\alpha_1, \dots, \alpha_K$  and  $\mu_1, \dots, \mu_{K+1}$ ,

$$\begin{aligned}
& \text{Cov}(X, Y|Z) \\
&= \sum_{j=0}^J Z^j \sum_{k=1}^K \alpha_{jk} \text{Cov}(X, X^{*k}|Z) \\
&= \sum_{j=0}^J Z^j \sum_{k=1}^K \alpha_{jk} (E[XX^{*k}|Z] - E[X|Z]E[X^{*k}|Z]) \\
&= \sum_{j=0}^J Z^j \sum_{k=1}^K \alpha_{jk} (E[X^{*k+1}|Z] - E[X|Z]E[X^{*k}|Z]) \\
&= \sum_{j=0}^J Z^j \sum_{k=1}^K \left[ \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) - \dots \right. \\
&\quad - \sum_{k_1=2l}^K \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} + \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} - \dots \right. \\
&\quad \left. \left. (-1)^{l-1} \sum_{k_2=2l-2}^{k_1-1} \binom{k_1}{k_2} \mu_{k_1-k_2} \dots \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) \right] E[X^{k+1}|Z] \\
&\quad - \sum_{j=0}^J Z^j \sum_{k=0}^K \left[ \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) - \dots \right. \\
&\quad - \sum_{k_1=2l}^K \alpha_{jk+k_1} \binom{k+k_1}{k} \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} - \dots \right. \\
&\quad \left. \left. (-1)^{l-1} \sum_{k_2=2l-2}^{k_1-2} \binom{k_1}{k_2} \mu_{k_1-k_2} \dots \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) \right] 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

$$\begin{aligned}
Q(Z) &= (E[X^{K+1}|Z], -E[X^K|Z]E[X|Z], \dots, E[X^{k+1}|Z], -E[X^k|Z]E[X|Z], \dots, E[X^2|Z], -E[X|Z]E[X|Z], E[X|Z], 1)', \\
R(Z) &= (Z^0 Q(Z)', Z^1 Q(Z)', \dots, Z^J Q(Z)')', \quad \beta_j = (\beta_{j1}, \dots, \beta_{j2K+2})', \quad \beta = (\beta'_0, \dots, \beta'_J)'.
\end{aligned}$$

By Assumption 3.2,  $E[R(Z)R(Z)']$  is finite and nonsingular so  $\beta = E[R(Z)R(Z)']^{-1}E[R(Z)\text{Cov}(X, Y|Z)]$  is identified. Further,  $\mu_k$  and  $\alpha_{jk}$  are recursively identified by

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

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

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

### A.3 Proof of Example 3.1

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

$$\begin{aligned}
\text{Cov}(X, Y|Z) &= \alpha_1 \text{Cov}(X, X^*|Z) = \alpha_1 (E[XX^*|Z] - E[X|Z]E[X^*|Z]) \\
&= \alpha_1 E[X^2|Z] - \alpha_1 (E[X|Z])^2 - \alpha_1 \mu_2.
\end{aligned}$$

By Assumption 2.1,  $\text{Cov}(X, Y|Z) = R(Z)' \beta$  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 3.2,  $\beta = E[R(Z)R(Z)']^{-1} E[R(Z)\text{Cov}(X, Y|Z)]$ . 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.4 Proof of Theorem 3.2

The function  $m(Z) = E[X|Z]$  is identified. Let  $\Psi_{X-m(Z)|Z}$ ,  $\Psi_V$  and  $\Psi_U$  be the characteristic functions of  $X - m(Z)|Z = z$ ,  $V$ , and  $U$  respectively.

Using Assumption 3.4 and  $X - m(Z) = \sigma(Z)V + U$ , we obtain

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

Let

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

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

$$\begin{aligned}\Psi'_V(t) &= 2\sigma(z_0)r\left(\frac{t}{\sigma(z_0)}\right)\Psi_V(t), \\ \Psi_V(t) &= \exp\left(2\sigma(z_0)\int_0^t r\left(\frac{u}{\sigma(z_0)}\right)du\right).\end{aligned}\tag{18}$$

Hence,  $\Psi_V$  is identified up to  $\sigma(z_0)$ .

Now  $\sigma_U^2 = \text{Var}(X - m(Z)|Z = z_0) - \sigma^2(z_0)$ ,  $\sigma^2(z) = \text{Var}(X - m(Z)|Z = z) - \text{Var}(X - m(Z)|Z = z_0) + \sigma^2(z_0)$  and  $\Psi_U(t) = \frac{\Psi_{X-m(Z)|Z}(t|z_0)}{\Psi_V(\sigma(z_0)t)}$ . Equation (9) shows that  $g$  is identified up to  $\sigma(z_0)$ . Then, using (10) and Assumption 3.6,  $\sigma(z_0)$  is identified. This implies that,  $g$ ,  $\sigma(\cdot)$  and the distribution of  $V$  are identified. Finally, the equality

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

shows that  $h$  is identified as well.

## A.5 Proof of Proposition 3.1

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 of a nonconstant zero is similar but slightly more cumbersome. We first compute  $\Psi_U^\sigma$  and  $\Psi_V^\sigma$  in this context. Consider the function  $r(\cdot)$  defined in the proof of Theorem 3.2. 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 (18),  $\Psi_V^\sigma(t) = \exp(-t^2/2)$ . In other words, the distribution of  $V$  is identified in this case, as it does not depend on  $\sigma(z_0)$ . Then

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

Next, we compute  $g^\sigma$  using (9). First,

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

Second,

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

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

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

where  $\phi$  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 (9) yields

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

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

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

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

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

is injective. Hence, Assumption 3.6 holds and  $g$  and  $h$  are identified.

## B A polynomial restriction on $g$ with multiplicative errors

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

$$\begin{cases} Y &= \sum_{j=0}^J \sum_{k=1}^K \alpha_{jk} Z^j X^{*k} + h(Z) + \varepsilon \\ X &= X^* U \end{cases} \quad (19)$$

and assume  $\alpha_{jk} \neq 0$  for all  $j$  and  $k$ . The following assumptions replace Assumptions 3.1 and 3.2 from the main text.

**Assumption B.1.** (i)  $E[\varepsilon|X^*, Z] = 0$  and (ii)  $E[U^k|X^*, Z] = \mu_k$  for  $k \in \{1, 2, \dots, K\}$  and  $\mu_1 = 1$ .

**Assumption B.2.** Define

$$\begin{aligned} Q(Z) &= (-E[X|Z]E[X|Z], E[X^2|Z], -E[X|Z]E[X^2|Z], E[X^3|Z], \dots, -E[X|Z]E[X^K|Z], E[X^{K+1}|Z])' \\ R(Z) &= (Z^0 Q(Z)', Z^1 Q(Z)', \dots, Z^J Q(Z))'. \end{aligned}$$

$R(Z)$  is finite and nonsingular.

**Theorem B.1.** Suppose Equation (19) and Assumptions B.1 and B.2 hold. Then the functions  $g$  and  $h$  and the moments  $\mu_1, \dots, \mu_{K+1}$  are identified.

**Proof:** First, we have

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

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

$$\begin{aligned} \beta_{j2k-1} &= \frac{\alpha_{jk}}{\mu_k}, \quad \beta_{j2k} = \frac{\alpha_{jk}}{\mu_{k+1}}, \quad \beta_j = (\beta_{j1}, \dots, \beta_{jK}), \quad (j, k) \in \{1, \dots, K\}^2, \\ \beta &= (\beta'_0, \dots, \beta'_K)'. \end{aligned}$$

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

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

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,  $h(Z) = E[Y|Z] - \sum_{j=0}^J \sum_{k=1}^K \alpha_{jk} Z^j E[X^{*k}|Z] = E[Y|Z] - \sum_{j=0}^J \sum_{k=1}^K \beta_{j2k-1} Z^j E[X^k|Z]$ .