

Biørn, Erik

Working Paper

Identification and method of moments estimation in polynomial measurement error models

Memorandum, No. 01/2017

Provided in Cooperation with:

Department of Economics, University of Oslo

Suggested Citation: Biørn, Erik (2017) : Identification and method of moments estimation in polynomial measurement error models, Memorandum, No. 01/2017, University of Oslo, Department of Economics, Oslo

This Version is available at:

<https://hdl.handle.net/10419/165963>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

MEMORANDUM

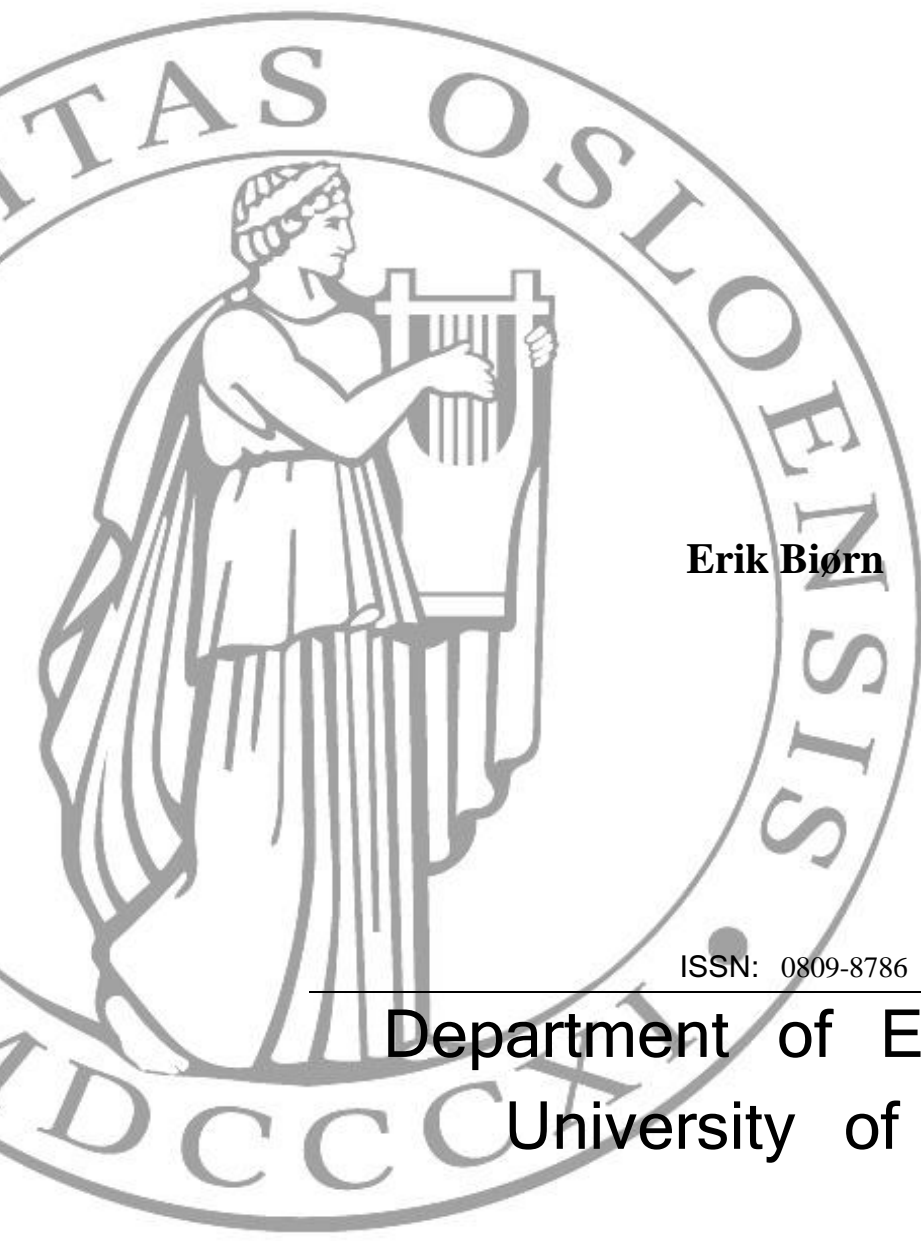
No 01/2017

Identification and Method of Moments Estimation in Polynomial Measurement Error Models

Erik Biørn

ISSN: 0809-8786

Department of Economics
University of Oslo



This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: <http://www.sv.uio.no/econ>
e-mail: econdep@econ.uio.no

In co-operation with
**The Frisch Centre for Economic
Research**

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: <http://www.frisch.uio.no>
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 19/16	Erik Biørn <i>Panel data estimators and aggregation</i>
No 18/16	Olav Bjerkholt <i>Wassily Leontief and the discovery of the input-output approach</i>
No 17/16	Øystein Kravdal <i>New Evidence about effects of reproductive variables on child mortality in sub-Saharan Africa</i>
No 16/16	Moti Michaeli and Daniel Spiro <i>The dynamics of revolutions</i>
No 15/16	Geir B. Asheim, Mark Voorneveld and Jörgen W. Weibull <i>Epistemically robust strategy subsets</i>
No 14/16	Torbjørn Hanson <i>Estimating output mix effectiveness: A scenario approach</i>
No 13/16	Halvor Mehlum and Kalle Moene <i>Unequal power and the dynamics of rivalry</i>
No 12/16	Halvor Mehlum <i>Another model of sales. Price discrimination in a horizontally differentiated duopoly market</i>
No 11/16	Vladimir W. Krivonozhko, Finn R. Førsund and Andrey V. Lychev <i>Smoothing the frontier in the DEA models</i>
No 10/16	Finn R. Førsund <i>Pollution Modelling and Multiple-Output Production Theory*</i>

Previous issues of the memo-series are available in a PDF® format at:
<http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/>

IDENTIFICATION AND METHOD OF MOMENTS ESTIMATION
IN POLYNOMIAL MEASUREMENT ERROR MODELS

ERIK BIØRN

Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway

E-mail: erik.biorn@econ.uio.no

ABSTRACT: Estimation of polynomial regression equations in one error-ridden variable and a number of error-free regressors, as well as an instrument set for the former is considered. Procedures for identification, operating on moments up to a certain order, are elaborated for single- and multi-equation models. Weak distributional assumptions are made for the error and the latent regressor. Simple order-conditions are derived, and procedures involving recursive identification of the moments of the regressor and its measurement errors together with the coefficients of the polynomials are considered. A Generalized Method of Moments (GMM) algorithm involving the instruments and proceeding stepwise from the identification procedures, is presented. An illustration for systems of linear, quadratic and cubic Engel functions, with household consumption and income data is given.

KEYWORDS: Errors in variables. Polynomial regression. Error distribution. Identification. Instrumental variables. Method of Moments. Engel functions.

JEL CLASSIFICATION: C21, C23, C31, C33, C51, E21

ACKNOWLEDGEMENTS: Preliminary versions have been presented at, *inter alia*, the 59th Econometric Society European Meeting (ESEM), Madrid, August 2004, and the 22nd Latin American Meeting of the Econometric Society (LAMES), Mexico City, November 2006. I thank Jørgen Aasness, Arthur Lewbel, Terje Skjerpen, and Yngve Willassen for helpful comments.

1 INTRODUCTION

Estimation of linear regression models with measurement errors (errors in variables, EIV) by using instrumental variables (IV) is discussed in almost any econometric textbook. However, several applications, especially in micro-econometrics, call for more flexible functional forms. Polynomial models with additive errors is one answer. Examples are quadratic or cubic Engel functions with error in the income measure, and quadratic or translog factor cost functions, where output and some input prices may be error-contaminated. Imposing, for example, linearity of Engel functions, as in Aasness *et al.* (1993, 2003) and Satorra (2002), may inadequately represent the data. Polynomial specifications can be used for testing linearity against low-order polynomials.

Polynomial, usually quadratic, EIV models, often under restrictive distributional assumptions, are discussed in Wolter and Fuller (1982), Montfort *et al.* (1987), Hausman *et al.* (1991, 1995), Carroll *et al.* (2006, sections 4.7.1, 4.9.2 and Appendix B.3.2), Moon and Gunst (1995), Cheng and Schneeweiss (1998), Wansbeek and Meijer (2000, section 11.2), and Kuha and Temple (2003). Depending on the strictness of the distributional assumptions, identification may be possible without access to valid IVs for the error-ridden variable, as exemplified by Cheng and Schneeweiss (1998) and Kuha and Temple (2003). Identification and estimation of linear EIV models exploiting moments of order higher than two are discussed in Lewbel (1997), Dagenais and Dagenais (1997), and Erickson and Whited (2002). Inference problems related to general non-linear specifications are considered in Newey (2001), Schennach (2004, 2007) and Hu and Schennach (2008). Of relevance is also the literature on non-linear (in variables) two-stage least squares; see, *e.g.*, Amemiya (1974; 1983, Section 5.1). For estimation of quadratic equations, moments of the observable variables of order at least four are needed. A more general question concerns the relationship between the polynomial order and the required moment order of the observable variables.

In this paper we reconsider identification and estimation of polynomial EIV models from moments of observable variables when imposing and not imposing strong distributional assumptions on the latent structural variables and the errors, but including IVs, whose relation to the latent regressor is formalized by linear equations. This resembles the setup in classical factor analysis. The structure of the identification problem motivates procedures for moment-based estimation of the coefficients of the polynomial and the measurement equations along with the moments of the latent regressor and the measurement errors. It is shown how under exact identification the identifying conditions can be established by recursions, from which there is a short way to estimation procedures, while in cases of overidentification simple methods of moments estimators can be extended to Generalized Method of Moments (GMM) procedures.

The identification conditions and the estimation procedures they motivate will involve *origo* moments of the latent structural variables, the errors, and the observable variables. High-order moments will often be involved. Since it is well known that, in particular in the absence of further restrictions on the variables' distribution, *e.g.*, symmetry or normality, such moments tend to be numerically unstable for realistic sample sizes, high-

order polynomial models have been argued to have limited practical interest.¹

The approach in this paper has similarities with that of Hausman *et al.* (1991, 1995). A notable discrepancy is that the measurement equations are specified with multiple indicators, resembling the format in classical factor analysis. We give order-conditions that ensure identification of the coefficients together with the moments of the latent regressor and of the measurement errors, and show how the identification status of the polynomial's coefficients depends on its order, the order of the moments of the observable variables exploited, and the number of IVs for the latent regressor. An extension to handle polynomial equation systems is considered.

The paper proceeds as follows. Section 2 formulates a one-equation polynomial model in one latent variable and a recursive procedure for identification is described. Section 3 modifies the procedure when including error-free regressors. Section 4 further extends to a model of a system of such equations. Estimation procedures utilizing a block-recursive organization of the moment equations are discussed in Section 5, with reference to a five-equation system of cubic equations. The latter is supplemented by an application to Engel functions. Concluding remarks follow in Section 6.

2 BASELINE MODEL

2.1 Basic assumptions

Consider an I th-order polynomial in the latent variable ξ :

$$(1) \quad Y = \alpha + \sum_{i=1}^I \xi^i \beta_i + u,$$

$$(2) \quad X = \xi + v,$$

where $\alpha, \beta_1, \dots, \beta_I$ are unknown constants, Y and X are observable variables, u is a disturbance and v is the measurement error, specifying the model for one generic observation, suppressing the observation subscript. All moments of ξ up to a suitable order are assumed to exist, all observations are independent. In addition J indicators of ξ , Q_1, \dots, Q_J , exist, related to ξ by J measurement equations:

$$(3) \quad Q_j = a_j + \xi b_j + w_j, \quad j = 1, \dots, J,$$

where a_j, b_j are constants and (w_1, \dots, w_J) are disturbances.

This resembles the setup in factor analysis; see Harman (1976). While (1)–(3) take ξ as a common latent factor in Y , X , Q_1, \dots, Q_J , treating ξ as exogenous and X , Y and Q_1, \dots, Q_J as jointly endogenous,² Hausman *et al.* (1991), see their Equation (3.3), adopt the *single equation* formulation, whose counterpart to (3) has the form $\xi = a^* + \sum_{j=1}^J b_j^* Q_j + w$, where a^* and b_j^* are constants, which, if v is stochastically independent of all w_j , implies a block-recursive structure from (Q_1, \dots, Q_J) , via ξ to X

¹Wansbeek and Meijer (2000, pp. 342-343) for example assert that since “the order of the moments that have to be fitted increases with the degree of the polynomial, and huge samples are needed to obtain relatively stable estimators of the parameters based on fitting higher-order moments. In most situations occurring in practice, it will not be possible to obtain useful estimators of polynomials higher than second or third degree”.

²If the relationship between ξ and its IVs has unknown form, (3) may be interpreted as *projections*, rather than as representing a structure with some permanence.

and Y . To fix ideas, (1)–(2) may represent a polynomial Engel function model in latent income ξ , X being an indicator of income and (Q_1, \dots, Q_J) variables related to ξ , *e.g.*, taxable income, net wealth and its components, indicators of human capital, ability, etc. We assume that u, v and ξ have moments up to a sufficiently high order and that³

$$(4) \quad \mathbf{E}(u) = \mathbf{E}(v) = \mathbf{E}(w_1) = \dots = \mathbf{E}(w_J) = 0, \quad \xi \perp u \perp v \perp (w_1, \dots, w_J),$$

with \perp denoting stochastic independence. Assumptions (2)–(4) imply

$$(5) \quad Q_j = a_j + Xb_j + w_j - vb_j, \quad j = 1, \dots, J.$$

Utilizing $(\xi + v)^g \equiv \sum_{r=0}^g \binom{g}{r} \xi^r v^{g-r}$, the following notation for moments is convenient

$$(6) \quad C_g \equiv \mathbf{E}(\xi^g), \quad \lambda_g \equiv \mathbf{E}(v^g),$$

$$(7) \quad F_g \equiv \mathbf{E}(Y\xi^g) = C_g\alpha + \sum_{i=1}^I C_{i+g}\beta_i,$$

$$(8) \quad H_{gh} \equiv \mathbf{E}(X^g\xi^h) = \sum_{r=0}^g \binom{g}{r} C_{h+r}\lambda_{g-r}.$$

In particular, $C_0 = H_{00} = \lambda_0 = 1$, $\lambda_1 = 0$, and

$$\begin{aligned} F_0 &= \mathbf{E}(Y) = \alpha + \sum_{i=1}^I C_i\beta_i, \\ F_1 &= \mathbf{E}(YX) = C_1\alpha + \sum_{i=1}^I C_{i+1}\beta_i. \end{aligned}$$

2.2 Moments of observable variables

We first express moments of the observable variables by the moments of ξ and v , and the coefficients. From (1)–(4) and (6)–(8) it follows that

$$(9) \quad \begin{aligned} \mathbf{E}(Q_j) &= a_j + C_1b_j = a_j + \mathbf{E}(X)b_j, \\ \mathbf{E}(YQ_j) &= F_0a_j + F_1b_j = \mathbf{E}(Y)a_j + \mathbf{E}(YX)b_j, \end{aligned} \quad j = 1, \dots, J.$$

Hence, $b_j = \text{cov}(Y, Q_j) / \text{cov}(Y, X)$, $a_i = \mathbf{E}(Q_j) - \mathbf{E}(X)b_j$, which together with (4) imply that Y is a valid IV for X relative to (5).⁴ From (1), (4), (7) and (8) we obtain

$$(10) \quad \mathbf{E}(YX^k) = \sum_{r=0}^k \binom{k}{r} F_r \lambda_{k-r},$$

$$(11) \quad \mathbf{E}(Q_j X^g) = \mathbf{E}(X^g)a_j + H_{g1}b_j.$$

Let the maximal power of X considered be K when it interacts with Y , and G when it interacts with any Q_j or stands alone. From (2), (6)–(8), (10), and (11) we have the moment equations:

$$(12) \quad \mathbf{E}(YX^k) = \sum_{r=0}^k \binom{k}{r} [\alpha C_r + \sum_{i=1}^I \beta_i C_{i+r}] \lambda_{k-r}, \quad k = 0, \dots, K,$$

$$(13) \quad \mathbf{E}(YQ_j) = (\alpha + \sum_{i=1}^I C_i\beta_i)a_j + (C_1\alpha + \sum_{i=1}^I C_{i+1}\beta_i)b_j, \quad j = 1, \dots, J,$$

$$(14) \quad \mathbf{E}(X^g) = \sum_{r=0}^g \binom{g}{r} C_r \lambda_{g-r}, \quad g = 1, \dots, G,$$

$$(15) \quad \begin{aligned} \mathbf{E}(Q_j X^g) &= \sum_{r=0}^g \binom{g}{r} [C_r a_j + C_{r+1} b_j] \lambda_{g-r}, \\ &= \mathbf{E}(X^g)a_j + [\sum_{r=0}^g \binom{g}{r} C_{r+1} \lambda_{g-r}] b_j, \quad j = 1, \dots, J; \quad g = 0, \dots, G. \end{aligned}$$

³Any functions $g(\xi)$, $h(u)$, $k(v)$, and $l(w_j)$ will then be uncorrelated. Normality would have imposed strong restrictions on the moments of these variables; see Evans *et al.* (1993, Chapter 29) and would therefore have had implications on the models' identification status.

⁴This exemplifies a case where an *endogenous* variable serves as an IV for an error-ridden *exogenous* variable.

They have two convenient properties: linearity in $(\alpha, \beta_1, \dots, \beta_I)$ for given (a_j, b_j) & (C_r, λ_k) , and linearity in (a_j, b_j) for given (C_r, λ_k) & $(\alpha, \beta_1, \dots, \beta_I)$.⁵

If the sample is sufficiently large and the convergence conditions for sample moments hold, see Davidson (1994, Chapter 18), then $E(YX^k)$, $E(YQ_j)$, $E(X^g)$ and $E(Q_jX^g)$ can, under weak conditions involving central limit theorems, be estimated consistently from their sample analogues.⁶ In exploring identification, we therefore proceed as if these moments are known.

2.3 Identification

Equations (9)⁷ and (12)–(15) motivate the following stepwise procedure:

- [1] Identify (a_j, b_j) from (9).
- [2] Identify C_g ($g=1, \dots, G+1$) and λ_g ($g=2, \dots, G$) from (14)–(15).
- [3] Identify $\alpha, \beta_1, \dots, \beta_I$ from (12).

It follows that:

- (i) The number of restrictions imposed by (15) on the (C_g, λ_g) s increases with J .
- (ii) To obtain from [2] a number of (C_g, λ_g) s sufficiently large to solve [3], the number of C_g s must be at least $K+I$. This requires $G+1 \geq K+I$.
- (iii) To determine $\alpha, \beta_1, \dots, \beta_I$ from (12), given the (C_g, λ_g) s, requires $K+1 \geq I+1$.

Altogether, (i)–(iii) imply that *necessary order-conditions* for identification are

$$(16) \quad J \geq 1, \quad K \geq I, \quad G \geq K+I-1 \implies G \geq 2I-1.$$

The boundary case $J=1, K=I, G=K+I-1 \implies G=2I-1$ gives exact identification.

To summarize: If $J=1$, knowledge of the origo moments $E(YQ_1)$, $E(YX^k)$ ($k=0, 1, \dots, I$), $E(X^g)$ ($g=1, \dots, 2I-1$) and $E(Q_1X^g)$ ($g=0, \dots, 2I-1$) ensures exact identification of $(\alpha, \beta_1, \dots, \beta_I)$ together with (a_1, b_1) , (C_1, \dots, C_{2I}) , $(\lambda_1, \dots, \lambda_{2I-1})$.

2.4 Examples

We will consider in some detail examples for the linear, quadratic and cubic cases, knowing from (16) that exact identification is ensured

- for $I=1$, when $J=1, G=1, K=1$,
- for $I=2$, when $J=1, G=3, K=2$,
- for $I=3$, when $J=1, G=5, K=3$.

How is the moment equation system changed when J, G and K are increased?⁸

⁵Inclusion also of powers of Y, Q_1, \dots, Q_J and interactions between Q_1, \dots, Q_J would have created more messy moment conditions, containing $\phi_r = E(u^r)$, $\mu_{jr} = E(w_j^r)$; $r=2, 3, \dots$ and second- and higher-order terms in $(\alpha, \beta_i, a_j, b_j)$, since for example, $E(Q_jQ_k) = a_ja_k + (a_jb_k + a_kb_j)C_1 + b_jb_kC_2 + E(w_jw_k)$. Also in this respect our approach differs from the Hausman *et al.* (1991) setup, which exploits moment conditions involving powers of the *indicators* rather than powers of the error-ridden regressor; see their equations (2.6)–(2.8).

⁶For high-order moments, the speed of convergence may, however, be slow if certain higher-order moments required by the standard central limit theorems do not exist. Then a more general central limit theorem is appropriate, see Embrechts, Klüppelberg, and Mikosch (1997, pp. 71–81).

⁷Note that (9) follows by combining (13), (12) for $k=1$, (14) for $g=1$ and (15) for $g=0$.

⁸The following expressions utilize $\lambda_1 = 0$ and the binomial coefficients $(1, 2, 1)$, $(1, 3, 3, 1)$, $(1, 4, 6, 4, 1)$, etc.

First, (13) in combination with (12) for $k = 0$ and 1, give

$$(a.0) \quad \mathbb{E}(YQ_j) = a_j\mathbb{E}(Y) + b_j\mathbb{E}(YX).$$

Next, (14) with $g=1, \dots, 6$ and (15) with $g=0, 1, \dots, 6$ give, respectively.

$$\begin{aligned} (b.1) \quad & \mathbb{E}(X) = C_1, \\ (b.2) \quad & \mathbb{E}(X^2) = C_2 + \lambda_2, \\ (b.3) \quad & \mathbb{E}(X^3) = C_3 + 3C_1\lambda_2 + \lambda_3, \\ (b.4) \quad & \mathbb{E}(X^4) = C_4 + 6C_2\lambda_2 + 4C_1\lambda_3 + \lambda_4, \\ (b.5) \quad & \mathbb{E}(X^5) = C_5 + 10C_3\lambda_2 + 10C_2\lambda_3 + 5C_1\lambda_4 + \lambda_5, \\ (b.6) \quad & \mathbb{E}(X^6) = C_6 + 15C_4\lambda_2 + 20C_3\lambda_3 + 15C_2\lambda_4 + 6C_1\lambda_5 + \lambda_6, \\ (c.0) \quad & \mathbb{E}(Q_j) = a_j + C_1b_j, \\ (c.1) \quad & \mathbb{E}(Q_jX) = \mathbb{E}(X)a_j + C_2b_j, \\ (c.2) \quad & \mathbb{E}(Q_jX^2) = \mathbb{E}(X^2)a_j + (C_3 + C_1\lambda_2)b_j, \\ (c.3) \quad & \mathbb{E}(Q_jX^3) = \mathbb{E}(X^3)a_j + (C_4 + 3C_2\lambda_2 + C_1\lambda_3)b_j, \\ (c.4) \quad & \mathbb{E}(Q_jX^4) = \mathbb{E}(X^4)a_j + (C_5 + 6C_3\lambda_2 + 4C_2\lambda_3 + C_1\lambda_4)b_j, \\ (c.5) \quad & \mathbb{E}(Q_jX^5) = \mathbb{E}(X^5)a_j + (C_6 + 10C_4\lambda_2 + 10C_3\lambda_3 + 5C_2\lambda_4 + C_1\lambda_5)b_j, \\ (c.6) \quad & \mathbb{E}(Q_jX^6) = \mathbb{E}(X^6)a_j + (C_7 + 15C_5\lambda_2 + 20C_4\lambda_3 + 15C_3\lambda_4 + 6C_2\lambda_5 + C_1\lambda_6)b_j. \end{aligned}$$

Further, from (12) with $(I, K) = (1, 3), (2, 4), (3, 5)$, we obtain, respectively:

for $I = 1; k = 0, 1, 2, 3$:

$$\begin{aligned} (d1.0) \quad & \mathbb{E}(Y) = \alpha + C_1\beta_1, \\ (d1.1) \quad & \mathbb{E}(YX) = C_1\alpha + C_2\beta_1, \\ (d1.2) \quad & \mathbb{E}(YX^2) = (C_2\alpha + C_3\beta_1) + \mathbb{E}(Y)\lambda_2, \\ (d1.3) \quad & \mathbb{E}(YX^3) = (C_3\alpha + C_4\beta_1) + 3\mathbb{E}(YX)\lambda_2 + \mathbb{E}(Y)\lambda_3. \end{aligned}$$

for $I = 2; k = 0, 1, \dots, 4$:

$$\begin{aligned} (d2.0) \quad & \mathbb{E}(Y) = \alpha + C_1\beta_1 + C_2\beta_2, \\ (d2.1) \quad & \mathbb{E}(YX) = C_1\alpha + C_2\beta_1 + C_3\beta_2, \\ (d2.2) \quad & \mathbb{E}(YX^2) = (C_2\alpha + C_3\beta_1 + C_4\beta_2) + \mathbb{E}(Y)\lambda_2, \\ (d2.3) \quad & \mathbb{E}(YX^3) = (C_3\alpha + C_4\beta_1 + C_5\beta_2) + 3\mathbb{E}(YX)\lambda_2 + \mathbb{E}(Y)\lambda_3, \\ (d2.4) \quad & \mathbb{E}(YX^4) = (C_4\alpha + C_5\beta_1 + C_6\beta_2) + 6(C_2\alpha + C_3\beta_1 + C_4\beta_2)\lambda_2 + 4\mathbb{E}(YX)\lambda_3 + \mathbb{E}(Y)\lambda_4. \end{aligned}$$

for $I = 3; k = 0, 1, \dots, 5$:

$$\begin{aligned} (d3.0) \quad & \mathbb{E}(Y) = \alpha + C_1\beta_1 + C_2\beta_2 + C_3\beta_3, \\ (d3.1) \quad & \mathbb{E}(YX) = C_1\alpha + C_2\beta_1 + C_3\beta_2 + C_4\beta_3, \\ (d3.2) \quad & \mathbb{E}(YX^2) = (C_2\alpha + C_3\beta_1 + C_4\beta_2 + C_5\beta_3) + \mathbb{E}(Y)\lambda_2, \\ (d3.3) \quad & \mathbb{E}(YX^3) = (C_3\alpha + C_4\beta_1 + C_5\beta_2 + C_6\beta_3) + 3\mathbb{E}(YX)\lambda_2 + \mathbb{E}(Y)\lambda_3, \\ (d3.4) \quad & \mathbb{E}(YX^4) = (C_4\alpha + C_5\beta_1 + C_6\beta_2 + C_7\beta_3) \\ & \quad + 6(C_2\alpha + C_3\beta_1 + C_4\beta_2 + C_5\beta_3)\lambda_2 + 4\mathbb{E}(YX)\lambda_3 + \mathbb{E}(Y)\lambda_4, \\ (d3.5) \quad & \mathbb{E}(YX^5) = (C_5\alpha + C_6\beta_1 + C_7\beta_2 + C_8\beta_3) \\ & \quad + 10(C_3\alpha + C_4\beta_1 + C_5\beta_2 + C_6\beta_3)\lambda_2 \\ & \quad + 10(C_2\alpha + C_3\beta_1 + C_4\beta_2 + C_5\beta_3)\lambda_3 + 5\mathbb{E}(YX)\lambda_4 + \mathbb{E}(Y)\lambda_5. \end{aligned}$$

We can proceed by picking moment equations sequentially: first (a.0), next equations in the (b) and (c) sets and finally equations in the (d) sets.

One indicator ($J=1$): Tables 1, 2 and 3 exemplify the cases $I=1, 2$ and 3, respectively. Their interpretation is explained by some examples. For the *linear* equation, going from case 1.1 to case 1.2, G increases from 1 to 2, while the 6 equations, which ensure (exact) identification of (α, β_1) and four additional parameters, are supplemented with (b.2) and (c.2). The latter involve the additional parameters (C_3, λ_2) and do not influence coefficient identification. Going to case 1.3, G increases from 2 to 3, while (b.3) and (c.3), which contains (C_4, λ_3) , are added, so their inclusion does not affect the identification status. Cases 1.4, 1.5 and 1.6, which by increasing K from 1 to 2 or 3 ($\implies K > I=1$),

introduce more additional equations than parameters, leads to overidentification. For example, cases 1.5 ($K=2$) and 1.6 ($K=3$), have the same parameters as case 1.3 and one and two overidentifying equations, respectively, (d1.2) and (d1.2)–(d1.3), in the latter. Cases 1.4, 1.5 and 1.6 concern identification of a linear equation by exploiting moments of higher order than two, a problem also considered by Lewbel (1997) and Erickson and Whited (2002). For the *quadratic* equation (Table 2), the moment equation system in the exactly identified case 2.3 ($K=I=2$) has the same number of parameters as case 2.5 ($K=3$) and case 2.6 ($K=4$), which hence gives overidentification, the *cubic* equation (Table 3) gives exact identification in case 3.2 ($K=I=3$) and overidentification in case 3.3, etc.

J indicators: The only change when the number of indicators for ξ is increased to J is that (a.0) and (c.1)–(c.6) occur J times. The systems of moment equations, after rearrangement and elimination of C_1 , in cases 1.1, 2.1 and 3.1, are:

Case 1.1: Linear model ($I=1, G=1, K=1$):

$$\begin{aligned}
[1] \quad & \mathbf{E}(Q_j) = a_j + \mathbf{E}(X)b_j, \\
& \mathbf{E}(YQ_j) = \mathbf{E}(Y)a_j + \mathbf{E}(YX)b_j, \\
[2] \quad & \mathbf{E}(Q_jX) = \mathbf{E}(X)a_j + C_2b_j, \\
[3] \quad & \mathbf{E}(Y) = \alpha + \mathbf{E}(X)\beta_1, \\
& \mathbf{E}(YX) = \mathbf{E}(X)\alpha + C_2\beta_1.
\end{aligned}$$

Case 2.1: Quadratic model ($I=2, G=3, K=2$):

$$\begin{aligned}
[1] \quad & \mathbf{E}(Q_j) = a_j + \mathbf{E}(X)b_j, \\
& \mathbf{E}(YQ_j) = \mathbf{E}(Y)a_j + \mathbf{E}(YX)b_j, \\
[2] \quad & \mathbf{E}(X^2) = C_2 + \lambda_2, \\
& \mathbf{E}(X^3) = C_3 + 3\mathbf{E}(X)\lambda_2 + \lambda_3, \\
& \mathbf{E}(Q_jX) = \mathbf{E}(X)a_j + C_2b_j, \\
& \mathbf{E}(Q_jX^2) = \mathbf{E}(X^2)a_j + [C_3 + \mathbf{E}(X)\lambda_2]b_j, \\
& \mathbf{E}(Q_jX^3) = \mathbf{E}(X^3)a_j + [C_4 + 3C_2\lambda_2 + \mathbf{E}(X)\lambda_3]b_j, \\
[3] \quad & \mathbf{E}(Y) = \alpha + \mathbf{E}(X)\beta_1 + C_2\beta_2, \\
& \mathbf{E}(YX) = \mathbf{E}(X)\alpha + C_2\beta_1 + C_3\beta_2, \\
& \mathbf{E}(YX^2) = C_2\alpha + C_3\beta_1 + C_4\beta_2 + \mathbf{E}(Y)\lambda_2.
\end{aligned}$$

Case 3.1: Cubic model ($I=3, G=5, K=3$):

$$\begin{aligned}
[1] \quad & \mathbf{E}(Q_j) = a_j + \mathbf{E}(X)b_j, \\
& \mathbf{E}(YQ_j) = \mathbf{E}(Y)a_j + \mathbf{E}(YX)b_j, \\
[2] \quad & \mathbf{E}(X^2) = C_2 + \lambda_2, \\
& \mathbf{E}(X^3) = C_3 + 3\mathbf{E}(X)\lambda_2 + \lambda_3, \\
& \mathbf{E}(X^4) = C_4 + 6C_2\lambda_2 + 4\mathbf{E}(X)\lambda_3 + \lambda_4, \\
& \mathbf{E}(X^5) = C_5 + 10C_3\lambda_2 + 10C_2\lambda_3 + 5\mathbf{E}(X)\lambda_4 + \lambda_5, \\
& \mathbf{E}(Q_jX) = \mathbf{E}(X)a_j + C_2b_j, \\
& \mathbf{E}(Q_jX^2) = \mathbf{E}(X^2)a_j + [C_3 + \mathbf{E}(X)\lambda_2]b_j, \\
& \mathbf{E}(Q_jX^3) = \mathbf{E}(X^3)a_j + [C_4 + 3C_2\lambda_2 + \mathbf{E}(X)\lambda_3]b_j, \\
& \mathbf{E}(Q_jX^4) = \mathbf{E}(X^4)a_j + [C_5 + 6C_3\lambda_2 + 4C_2\lambda_3 + \mathbf{E}(X)\lambda_4]b_j, \\
& \mathbf{E}(Q_jX^5) = \mathbf{E}(X^5)a_j + [C_6 + 10C_4\lambda_2 + 10C_3\lambda_3 + 5C_2\lambda_4 + \mathbf{E}(X)\lambda_5]b_j, \\
[3] \quad & \mathbf{E}(Y) = \alpha + \mathbf{E}(X)\beta_1 + C_2\beta_2 + C_3\beta_3, \\
& \mathbf{E}(YX) = \mathbf{E}(X)\alpha + C_2\beta_1 + C_3\beta_2 + C_4\beta_3, \\
& \mathbf{E}(YX^2) = C_2\alpha + C_3\beta_1 + C_4\beta_2 + C_5\beta_3 + \mathbf{E}(Y)\lambda_2, \\
& \mathbf{E}(YX^3) = C_3\alpha + C_4\beta_1 + C_5\beta_2 + C_6\beta_3 + 3\mathbf{E}(YX)\lambda_2 + \mathbf{E}(Y)\lambda_3.
\end{aligned}$$

In all cases, Block [1] identifies a_j, b_j . In case 1.1, Block [2] identifies C_2 (overidentifies for $J > 1$), Block [3] identifies α, β_1 . In case 2.1, Block [2] identifies $C_2, C_3, C_4, \lambda_2, \lambda_3$ (overidentifies for $J > 1$), Block [3] identifies α, β_1, β_2 . In case 3.1, Block [2] identifies $C_2, \dots, C_6, \lambda_2, \dots, \lambda_5$ (overidentifies for $J > 1$), Block [3] identifies $\alpha, \beta_1, \beta_2, \beta_3$. The three cases have systems with, respectively, $(3J+2)$ equations and $(3J+2)-(2J+3) = J-1$ overidentifying restrictions, $(5J+5)$ equations and $(5J+5)-(2J+8) = 3(J-1)$ overidentifying restrictions, and $(7J+8)$ equations and $(7J+8) - (2J+13) = 5(J-1)$ overidentifying restrictions.

Table 1: MOMENT EQUATIONS FOR $I = 1, J = 1$

	Baseline		G increased				G and K increased					
	1.1: $(G, K) = (1, 1)$		1.2: $(G, K) = (2, 1)$		1.3: $(G, K) = (3, 1)$		1.4: $(G, K) = (2, 2)$		1.5: $(G, K) = (3, 2)$		1.6: $(G, K) = (3, 3)$	
	Eqs	Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param
(a.0)	C_1		(b.2)	C_3	(b.2)	C_3	(b.2)	C_3	(b.2)	C_3	(b.2)	C_3
(b.1)	C_2		(c.2)	λ_2	(c.2)	λ_2	(c.2)	λ_2	(c.2)	λ_2	(c.2)	λ_2
(c.0)	a_1				(b.3)	C_4	(d1.2)		(b.3)	C_4	(b.3)	C_4
(c.1)	b_1				(c.3)	λ_3			(c.3)	λ_3	(c.3)	λ_3
(d1.0)	α								(d1.2)		(d1.2)	
(d1.1)	β_1										(d1.3)	
#	6	6 (6)	8	8 (8)	10	10 (9)	9	8 (8)	11	10 (9)	12	10 (9)

Bottom line: In parenthesis, no. of parameters under symmetry of error distribution

Table 2: MOMENT EQUATIONS FOR $I = 2, J = 1$

	Baseline		G increased				G and K increased					
	2.1: $(G, K) = (3, 2)$		2.2: $(G, K) = (4, 2)$		2.3: $(G, K) = (5, 2)$		2.4: $(G, K) = (4, 3)$		2.5: $(G, K) = (5, 3)$		2.6: $(G, K) = (5, 4)$	
	Eqs	Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param
(a.0)	C_1		(b.4)	C_5	(b.4)	C_5	(b.4)	C_5	(b.4)	C_5	(b.4)	C_5
(b.1)	C_2		(c.4)	λ_4	(c.4)	λ_4	(c.4)	λ_4	(c.4)	λ_4	(c.4)	λ_4
(b.2)	C_3				(b.5)	C_6	(d2.3)		(b.5)	C_6	(b.5)	C_6
(b.3)	C_4				(c.5)	λ_5			(c.5)	λ_5	(c.5)	λ_5
(c.0)	λ_2								(d2.3)		(d2.3)	
(c.1)	λ_3										(d2.4)	
(c.2)	a_1											
(c.3)	b_1											
(d2.0)	α											
(d2.1)	β_1											
(d2.2)	β_2											
#	11	11 (10)	13	13 (11)	15	15 (13)	14	13 (12)	16	15 (15)	17	15 (13)

Bottom line: In parenthesis, no. of parameters under symmetry of error distribution

Table 3: MOMENT EQUATIONS FOR $I = 3, J = 1$

	Baseline		G increased		G and K increased			
	3.1: $(G, K) = (5, 3)$		3.2: $(G, K) = (6, 3)$		3.3: $(G, K) = (6, 4)$		3.4: $(G, K) = (6, 5)$	
	Eqs	Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param	Δ Eqs	Δ Param
(a.0)	C_1		(b.6)	C_7	(b.6)	C_7	(b.6)	C_7
(b.1)	C_2		(c.6)	λ_6	(c.6)	λ_6	(c.6)	λ_6
(b.2)	C_3				(d3.4)		(d3.4)	C_8
(b.3)	C_4						(d3.5)	
(b.4)	C_5							
(b.5)	C_6							
(c.0)	λ_2							
(c.1)	λ_3							
(c.2)	λ_4							
(c.3)	λ_5							
(c.4)	a_1							
(c.5)	b_1							
(d3.0)	α							
(d3.1)	β_1							
(d3.2)	β_2							
(d3.3)	β_3							
#	16	16 (14)	18	18 (16)	19	18 (16)	20	19 (17)

Bottom line: In parenthesis, no. of parameters under symmetry of error distribution

Remark on the solution under exact identification, $J=1$: In case 2.1, the first two equations in [2] can be solved for (λ_2, λ_3) , giving $\lambda_2 = \mathbf{E}(X^2) - C_2$, $\lambda_3 = \mathbf{E}(X^3) - C_3 - 3\mathbf{E}(X)\lambda_2$. This gives a non-linear system in $C_2, C_3, C_4, \alpha, \beta_1, \beta_2$ when combined with the last three equations in [2] and [3]. In case 3.1, the first four equations in [2] can be solved for $(\lambda_2, \lambda_3, \lambda_4, \lambda_5)$, giving

$$\lambda_2 = \mathbf{E}(X^2) - C_2,$$

$$\lambda_3 = \mathbf{E}(X^3) - C_3 - 3\mathbf{E}(X)\lambda_2,$$

$$\lambda_4 = \mathbf{E}(X^4) - C_4 - 6C_2\lambda_2 - 4\mathbf{E}(X)\lambda_3,$$

$$\lambda_5 = \mathbf{E}(X^5) - C_5 - 10C_3\lambda_2 - 10C_2\lambda_3 - 5\mathbf{E}(X)\lambda_4.$$

The result is a non-linear system in $C_2, C_3, C_4, C_5, C_6, \alpha, \beta_1, \beta_2, \beta_3$ when combined with the last five equations in [2] and [3].

Symmetry of the error distribution restricts all odd-numbered moments of v to be zero ($\lambda_3 = \lambda_5 = \dots = 0$). The bottom parts of Tables 1–3 show the change in the identification status. For example, cases 1.3, 2.1 and 3.1 change from being exactly identified to becoming overidentified.⁹

3 MODEL WITH ERROR-FREE EXOGENOUS VARIABLES

3.1 Basic assumptions

We include in the polynomial equation (1) a $(1 \times L)$ -vector of error-free exogenous variables \mathbf{Z} , which also enter measurement equations (4). It is not difficult to demonstrate that this extension does not affect the model's identification status, as the added number of parameters equal the added number of moment equations. The model is

$$(17) \quad Y = \alpha + \sum_{i=1}^I \xi^i \beta_i + \mathbf{Z}\boldsymbol{\gamma} + u,$$

$$(18) \quad X = \xi + v,$$

$$(19) \quad Q_j = a_j + \xi b_j + \mathbf{Z}\mathbf{c}_j + w_j, \quad j = 1, \dots, J,$$

$$(20) \quad \begin{aligned} \mathbf{E}(u) = \mathbf{E}(v) = \mathbf{E}(w_1) = \dots = \mathbf{E}(w_J) = 0, \\ (\xi, \mathbf{Z}) \perp u \perp v \perp (w_1, \dots, w_J), \end{aligned}$$

where $\boldsymbol{\gamma}$ and \mathbf{c}_j are $(L \times 1)$ coefficient vectors. Boldface letters denote vectors or matrices. In the Engel function example \mathbf{Z} may contain demographic and household characteristics, assumed to be relevant also for the way measured income is related to true income (some elements of \mathbf{c}_j may well be set to zero). From (17) and (18) it follows that

$$(21) \quad Q_j = a_j + Xb_j + \mathbf{Z}\mathbf{c}_j + w_j - b_j v, \quad j = 1, \dots, J.$$

We accordingly extend (6) and (7) to

$$(22) \quad C_g \equiv \mathbf{E}(\xi^g), \quad \mathbf{D}_g \equiv \mathbf{E}(\xi^g \mathbf{Z}), \quad \boldsymbol{\Sigma} \equiv \mathbf{E}(\mathbf{Z}' \mathbf{Z}), \quad \lambda_g \equiv \mathbf{E}(v^g),$$

$$(23) \quad F_g \equiv \mathbf{E}(Y \xi^g) = C_g \alpha + \sum_{i=1}^I C_{i+g} \beta_i + \mathbf{D}_g \boldsymbol{\gamma}.$$

⁹If the stronger assumption of *error normality* is imposed, all even-numbered moments of v of fourth and higher order are simple functions of its variance λ_2 . Then $\lambda_{2k} \equiv \lambda_2^k \prod_{i=1}^k (2i-1)$, which implies $\lambda_{2k}/\lambda_{2k-2} = \lambda_2^2(2k-1)$, $k=2, 3, \dots$; see Balakrishnan and Nevzorov (2003, Section 23.7). Imposing error symmetry will give overidentification, as in, e.g., cases 1.3 for $I=1$, cases 2.1, 2.2 and 2.3 for $I=2$, and cases 3.1 and 3.2 for $I=3$. On the other hand, symmetry of the distribution of ξ will often be quite unrealistic (as in the Engel function example) and will not be considered. Although it may further reduce the number of parameters to be estimated it may complicate identification and estimation; see Reiersøl (1950) for discussion of the normality case.

In particular,

$$\begin{aligned} F_0 &= \mathbf{E}(Y) = \alpha + \sum_{i=1}^I C_i \beta_i + \mathbf{D}_0 \boldsymbol{\gamma}, \\ F_1 &= \mathbf{E}(YX) = C_1 \alpha + \sum_{i=1}^I C_{i+1} \beta_i + \mathbf{D}_1 \boldsymbol{\gamma}. \end{aligned}$$

3.2 Moments of observable variables

From (18)–(20) and (22)–(23) we find, since $\mathbf{D}_0 = \mathbf{E}(\mathbf{Z})$, $\mathbf{D}_1 = \mathbf{E}(X\mathbf{Z})$, and

$$\mathbf{S} \equiv \mathbf{E}(\mathbf{Z}'Y) = \mathbf{D}'_0 \alpha + \sum_{i=1}^I \mathbf{D}'_i \beta_i + \boldsymbol{\Sigma} \boldsymbol{\gamma},$$

that (9)–(11) are generalized to

$$\begin{aligned} \mathbf{E}(Q_j) &= a_j + C_1 b_j + \mathbf{D}_0 \mathbf{c}_j = a_j + \mathbf{E}(X) b_j + \mathbf{E}(\mathbf{Z}) \mathbf{c}_j, \\ (24) \quad \mathbf{E}(YQ_j) &= F_0 a_j + F_1 b_j + \mathbf{S} \mathbf{c}_j = \mathbf{E}(Y) a_j + \mathbf{E}(YX) b_j + \mathbf{E}(Y\mathbf{Z}) \mathbf{c}_j, \\ \mathbf{E}(\mathbf{Z}'Q_j) &= \mathbf{D}'_0 a_j + \mathbf{D}'_1 b_j + \boldsymbol{\Sigma} \mathbf{c}_j = \mathbf{E}(\mathbf{Z}') a_j + \mathbf{E}(\mathbf{Z}'X) b_j + \mathbf{E}(\mathbf{Z}'\mathbf{Z}) \mathbf{c}_j, \end{aligned}$$

$$(25) \quad \mathbf{E}(YX^k) = \sum_{r=0}^k \binom{k}{r} F_r \lambda_{k-r},$$

$$(26) \quad \mathbf{E}(Q_j X^g) = \mathbf{E}(X^g) a_j + H_{g1} b_j + \mathbf{D}_g \mathbf{c}_j.$$

The system (24), together with (20), implies that (Y, \mathbf{Z}) are valid IVs for (X, \mathbf{Z}) in (21).¹⁰ The generalizations of the moment equations (12)–(15) implied by (18), (22), (23), (25), and (26) are:¹¹

$$(27) \quad \mathbf{E}(YX^k) = \sum_{r=0}^k \binom{k}{r} [C_r \alpha + \sum_{i=1}^I C_{i+r} \beta_i + \mathbf{D}_r \boldsymbol{\gamma}] \lambda_{k-r}, \quad k=0, \dots, K,$$

$$(28) \quad \mathbf{E}(\mathbf{Z}'Y) = \mathbf{D}'_0 \alpha + \sum_{i=1}^I \mathbf{D}'_i \beta_i + \boldsymbol{\Sigma} \boldsymbol{\gamma},$$

$$(29) \quad \mathbf{E}(YQ_j) = (\alpha + \sum_{i=1}^I C_i \beta_i + \mathbf{D}_0 \boldsymbol{\gamma}) a_j + (C_1 \alpha + \sum_{i=1}^I C_{i+1} \beta_i + \mathbf{D}_1 \boldsymbol{\gamma}) b_j + \mathbf{S} \mathbf{c}_j, \quad j=1, \dots, J,$$

$$(30) \quad \mathbf{E}(X^g) = \sum_{r=0}^g \binom{g}{r} C_r \lambda_{g-r}, \quad g=1, \dots, G,$$

$$(31) \quad \mathbf{E}(X^g \mathbf{Z}) = \sum_{r=1}^g \binom{g}{r} \mathbf{D}_r \lambda_{g-r}, \quad g=1, \dots, G,$$

$$(32) \quad \mathbf{E}(Q_j X^g) = \sum_{r=0}^g \binom{g}{r} [C_r a_j + C_{r+1} b_j + \mathbf{D}_r \mathbf{c}_j] \lambda_{g-r}, \quad j=1, \dots, J; \quad g=0, \dots, G.$$

3.3 Identification

Equations (24) and (27)–(32) motivate the stepwise procedure:¹²

[1] Identify (a_j, b_j, \mathbf{c}_j) from (24).

[2] Identify $C_g (g=1, \dots, G+1), \mathbf{D}_g (g=1, \dots, G)$ and $\lambda_g (g=2, \dots, G)$ from (30)–(32).

[3] Identify $\alpha, \beta_1, \dots, \beta_I, \boldsymbol{\gamma}$ from (27)–(29).

It follows that:

- (i) The number of restrictions imposed by (32) on the $(C_g, \mathbf{D}_g, \lambda_g)$ s increases with J .
- (ii) To obtain from [2] a number of $(C_g, \mathbf{D}_g, \lambda_g)$ s sufficiently large to solve [3], the number of C_g s and \mathbf{D}_g s must be at least $K+I$. This requires $G+1 \geq K+I$.
- (iii) To determine $\alpha, \beta_1, \dots, \beta_I, \boldsymbol{\gamma}$ from (27), given the $(C_g, \mathbf{D}_g, \lambda_g)$ s, requires $K+1 \geq I+1$.

Therefore, inclusion of *observable* regressors does not alter the order-conditions (16).

¹⁰This again exemplifies an endogenous variable serving as an IV for an error-ridden exogenous variable.

¹¹These equations are linear in $(\alpha, \beta_1, \dots, \beta_I, \boldsymbol{\gamma})$ for given (a_j, b_j, \mathbf{c}_j) & $(C_r, \mathbf{D}_r, \lambda_k)$, and linear in (a_j, b_j, \mathbf{c}_j) for given $(\alpha, \beta_1, \dots, \beta_I, \boldsymbol{\gamma})$ & $(C_r, \mathbf{D}_r, \lambda_k)$.

¹²Note that (24) follows from (29), (27) for $k=1$, (30) for $g=1$, (31) for $g=1$, and (32) for $g=0$.

4 MULTI-EQUATION MODEL

4.1 Basic assumptions

We further extend (17) to a system of N polynomial equations:

$$(33) \quad Y_n = \alpha_n + \sum_{i=1}^I \xi^i \beta_{ni} + \mathbf{Z} \boldsymbol{\gamma}_n + u_n, \quad n = 1, \dots, N,$$

still assuming (18)–(19), and extend (20) to

$$(34) \quad \begin{aligned} \mathbf{E}(u_1) = \dots = \mathbf{E}(u_N) = \mathbf{E}(v) = \mathbf{E}(w_1) = \dots = \mathbf{E}(w_J) = 0, \\ (\boldsymbol{\xi}, \mathbf{Z}) \perp (u_1, \dots, u_N) \perp v \perp (w_1, \dots, w_J). \end{aligned}$$

The model then obtained may represent an N -commodity system of I 'th order polynomial Engel functions with common latent income and correctly measured demographic and household characteristics included for all commodities. The problem now is to identify $\{\alpha_n, \beta_{n1}, \dots, \beta_{nI}, \boldsymbol{\gamma}_n\}_{n=1}^N$ from observations on $(Y_1, \dots, Y_N, X, \mathbf{Z}, Q_1, \dots, Q_J)$. To explore this, we extend (23) to

$$(35) \quad F_{gn} \equiv \mathbf{E}(Y_n \xi^g) = C_g \alpha_n + \sum_{i=1}^I C_{i+g} \beta_{ni} + \mathbf{D}_g \boldsymbol{\gamma}_n.$$

In particular,

$$\begin{aligned} F_{0n} &= \mathbf{E}(Y_n) = \alpha_n + \sum_{i=1}^I C_i \beta_{ni} + \mathbf{D}_0 \boldsymbol{\gamma}_n, \\ F_{1n} &= \mathbf{E}(Y_n X) = C_1 \alpha_n + \sum_{i=1}^I C_{i+1} \beta_{ni} + \mathbf{D}_1 \boldsymbol{\gamma}_n. \end{aligned}$$

4.2 Moments of observable variables

Using (18), (19), (22) and (33)–(35), we find since

$$\mathbf{S}_n = \mathbf{E}(\mathbf{Z}' Y_n) = \mathbf{D}'_0 \alpha_n + \sum_{i=1}^I \mathbf{D}'_i \beta_{ni} + \boldsymbol{\Sigma} \boldsymbol{\gamma}_n,$$

that (24) and (25) are generalized to

$$(36) \quad \begin{aligned} \mathbf{E}(Q_j) &= a_j + C_1 b_j + \mathbf{D}_0 \mathbf{c}_j = a_j + \mathbf{E}(X) b_j + \mathbf{E}(\mathbf{Z}) \mathbf{c}_j, \\ \mathbf{E}(Y_n Q_j) &= F_{0n} a_j + F_{1n} b_j + \mathbf{S}_n \mathbf{c}_j = \mathbf{E}(Y_n) a_j + \mathbf{E}(Y_n X) b_j + \mathbf{E}(Y_n \mathbf{Z}) \mathbf{c}_j, \\ \mathbf{E}(\mathbf{Z}' Q_j) &= \mathbf{D}'_0 a_j + \mathbf{D}'_1 b_j + \boldsymbol{\Sigma} \mathbf{c}_j = \mathbf{E}(\mathbf{Z}') a_j + \mathbf{E}(\mathbf{Z}' X) b_j + \mathbf{E}(\mathbf{Z}' \mathbf{Z}) \mathbf{c}_j, \end{aligned}$$

$$(37) \quad \mathbf{E}(Y_n X^k) = \sum_{r=0}^k \binom{k}{r} F_{rn} \lambda_{k-r},$$

while (26) is unchanged. The system (36), together with (34), implies that $(Y_1, \dots, Y_N, \mathbf{Z})$ are valid IVs for (X, \mathbf{Z}) in (21).¹³ Combining (19) and (33)–(35) we obtain the following generalization of (27)–(29):¹⁴

$$(38) \quad \mathbf{E}(Y_n X^k) = \sum_{r=0}^k \binom{k}{r} [C_r \alpha_n + \sum_{i=1}^I C_{i+r} \beta_{ni} + \mathbf{D}_r \boldsymbol{\gamma}_n] \lambda_{k-r}, \quad n = 1, \dots, N; k = 0, \dots, K,$$

$$(39) \quad \mathbf{E}(\mathbf{Z}' Y_n) = \mathbf{D}'_0 \alpha_n + \sum_{i=1}^I \mathbf{D}'_i \beta_{ni} + \boldsymbol{\Sigma} \boldsymbol{\gamma}_n, \quad n = 1, \dots, N;$$

$$(40) \quad \begin{aligned} \mathbf{E}(Y_n Q_j) &= (\alpha_n + \sum_{i=1}^I C_i \beta_{ni} + \mathbf{D}_0 \boldsymbol{\gamma}_n) a_j \\ &\quad + (C_1 \alpha_n + \sum_{i=1}^I C_{i+1} \beta_{ni} + \mathbf{D}_1 \boldsymbol{\gamma}_n) b_j, \quad n = 1, \dots, N; j = 1, \dots, J, \end{aligned}$$

while (30)–(32) are unchanged.

¹³When $N > 1$, there are more equations than needed, which exemplifies a case where a set of formally endogenous variables lead to overidentification.

¹⁴These equations are linear in $(\alpha_n, \beta_{n1}, \dots, \beta_{nI}, \boldsymbol{\gamma}_n)$ for given (a_j, b_j, \mathbf{c}_j) & $(C_r, \mathbf{D}_r, \lambda_k)$, and linear in (a_j, b_j, \mathbf{c}_j) for given $(\alpha_n, \beta_{n1}, \dots, \beta_{nI}, \boldsymbol{\gamma}_n)$ & $(C_r, \mathbf{D}_r, \lambda_k)$.

4.3 Identification

The stepwise procedure for identification this motivates is:

- [1] Identify (a_j, b_j, \mathbf{c}_j) from (36).
- [2] Identify $C_g(g=1, \dots, G+1), \mathbf{D}_g(g=1, \dots, G), \lambda_g(g=2, \dots, G)$ from (30)–(32).
- [3] Identify $\alpha_n, \beta_{n1}, \dots, \beta_{nI}, \gamma_n$ ($n = 1, \dots, N$) from (38)–(40).

Still, (16) is a necessary condition for identification. For $N > 1$ there is overidentification. This follows because (i) extending the number of equations from one to N increases the number of moment equations from $K + J + 1$ in (27)–(29) to $N(K + J + 1)$ in (38)–(40), while (ii) the number of coefficients in the polynomials $(\alpha_n, \beta_{n1}, \dots, \beta_{nI}, \gamma_n)$ increases from $1 + I + L$ to $N(1 + I + L)$. We summarize the identification conditions in:

Proposition: *Let for an N -equation model, each equation containing an I -th-order polynomial, the maximal power of the error-ridden regressor X be K in moments where X is combined with the regressand Y , and G in moments where X is alone or is combined with the error-free regressor vector \mathbf{Z} or the indicators Q_1, \dots, Q_J . Necessary conditions for identification are then:*

$$N \geq 1, \quad J \geq 1, \quad K \geq I, \quad G \geq K + I - 1.$$

When $N = 1, J = 1, K = I, G = K + I - 1 \implies G = 2I - 1 = 2K - 1$ there is exact identification. Overidentification occurs if there is at least one strict inequality.

5 ESTIMATION

In this section, we consider procedures for estimation of an N -equation, J -indicator cubic example model with error-free exogenous variables, *corresponding to case 3.1 in Section 2.4*; see Table 3. At the end we give an empirical illustration. The example model has, in its three blocks, $NJ + 7J + 5N + 8$ moment equations.

5.1 The example

Block [1] has $(N + 2)J$ equations identifying a_j, b_j, \mathbf{c}_j :

$$\begin{aligned} \mathbf{E}(Q_j) &= a_j + \mathbf{E}(X)b_j + \mathbf{E}(\mathbf{Z})\mathbf{c}_j, \\ \mathbf{E}(Y_n Q_j) &= \mathbf{E}(Y_n)a_j + \mathbf{E}(Y_n X)b_j + \mathbf{E}(Y_n \mathbf{Z})\mathbf{c}_j, \\ \mathbf{E}(\mathbf{Z}' Q_j) &= \mathbf{E}(\mathbf{Z}')a_j + \mathbf{E}(\mathbf{Z}' X)b_j + \mathbf{E}(\mathbf{Z}' \mathbf{Z})\mathbf{c}_j. \end{aligned}$$

Block [2] has $5J + 8$ equations identifying $C_2, \dots, C_6, \mathbf{D}_2, \dots, \mathbf{D}_5, \lambda_2, \dots, \lambda_5$:

$$\begin{aligned} \mathbf{E}(X^2) &= C_2 + \lambda_2, \\ \mathbf{E}(X^3) &= C_3 + 3\mathbf{E}(X)\lambda_2 + \lambda_3, \\ \mathbf{E}(X^4) &= C_4 + 6C_2\lambda_2 + 4\mathbf{E}(X)\lambda_3 + \lambda_4, \\ \mathbf{E}(X^5) &= C_5 + 10C_3\lambda_2 + 10C_2\lambda_3 + 5\mathbf{E}(X)\lambda_4 + \lambda_5, \\ \mathbf{E}(Q_j X) &= \mathbf{E}(X)a_j + C_2 b_j + \mathbf{E}(X \mathbf{Z})\mathbf{c}_j, \\ \mathbf{E}(Q_j X^2) &= \mathbf{E}(X^2)a_j + [C_3 + \mathbf{E}(X)\lambda_2]b_j + [\mathbf{D}_2 + \mathbf{E}(\mathbf{Z})\lambda_2]\mathbf{c}_j, \\ \mathbf{E}(Q_j X^3) &= \mathbf{E}(X^3)a_j + [C_4 + 3C_2\lambda_2 + \mathbf{E}(X)\lambda_3]b_j \\ &\quad + [\mathbf{D}_3 + 3\mathbf{E}(X \mathbf{Z})\lambda_2 + \mathbf{E}(\mathbf{Z})\lambda_3]\mathbf{c}_j, \\ \mathbf{E}(Q_j X^4) &= \mathbf{E}(X^4)a_j + [C_5 + 6C_3\lambda_2 + 4C_2\lambda_3 + \mathbf{E}(X)\lambda_4]b_j \\ &\quad + [\mathbf{D}_4 + 6\mathbf{D}_2\lambda_2 + 4\mathbf{E}(X \mathbf{Z})\lambda_3 + \mathbf{E}(\mathbf{Z})\lambda_4]\mathbf{c}_j, \\ \mathbf{E}(Q_j X^5) &= \mathbf{E}(X^5)a_j + [C_6 + 10C_4\lambda_2 + 10C_3\lambda_3 + 5C_2\lambda_4 + \mathbf{E}(X)\lambda_5]b_j \\ &\quad + [\mathbf{D}_5 + 10\mathbf{D}_3\lambda_2 + 10\mathbf{D}_2\lambda_3 + 5\mathbf{E}(X \mathbf{Z})\lambda_4 + \mathbf{E}(\mathbf{Z})\lambda_5]\mathbf{c}_j, \end{aligned}$$

$$\begin{aligned}
\mathbf{E}(X^2\mathbf{Z}) &= \mathbf{D}_2 + \mathbf{E}(\mathbf{Z})\lambda_2, \\
\mathbf{E}(X^3\mathbf{Z}) &= \mathbf{D}_3 + 3\mathbf{E}(X\mathbf{Z})\lambda_2 + \mathbf{E}(\mathbf{Z})\lambda_3, \\
\mathbf{E}(X^4\mathbf{Z}) &= \mathbf{D}_4 + 6\mathbf{D}_2\lambda_2 + 4\mathbf{E}(X\mathbf{Z})\lambda_3 + \mathbf{E}(\mathbf{Z})\lambda_4, \\
\mathbf{E}(X^5\mathbf{Z}) &= \mathbf{D}_5 + 10\mathbf{D}_3\lambda_2 + 10\mathbf{D}_2\lambda_3 + 5\mathbf{E}(X\mathbf{Z})\lambda_4 + \mathbf{E}(\mathbf{Z})\lambda_5.
\end{aligned}$$

Block [3] has $5N$ equations identifying $\alpha_n, \beta_{n1}, \beta_{n2}, \beta_{n3}, \gamma_n$:

$$\begin{aligned}
\mathbf{E}(Y_n) &= \alpha_n + \mathbf{E}(X)\beta_{n1} + C_2\beta_{n2} + C_3\beta_{n3} + \mathbf{E}(\mathbf{Z})\gamma_n, \\
\mathbf{E}(Y_n X) &= \mathbf{E}(X)\alpha_n + C_2\beta_{n1} + C_3\beta_{n2} + C_4\beta_{n3} + \mathbf{E}(X\mathbf{Z})\gamma_n, \\
\mathbf{E}(Y_n X^2) &= C_2\alpha_n + C_3\beta_{n1} + C_4\beta_{n2} + C_5\beta_{n3} + \mathbf{D}_2\gamma_n + \mathbf{E}(Y_n)\lambda_2, \\
\mathbf{E}(Y_n X^3) &= C_3\alpha_n + C_4\beta_{n1} + C_5\beta_{n2} + C_6\beta_{n3} + \mathbf{D}_3\gamma_n + 3\mathbf{E}(Y_n X)\lambda_2 + \mathbf{E}(Y_n)\lambda_3, \\
\mathbf{E}(\mathbf{Z}'Y_n) &= \mathbf{E}(\mathbf{Z}')\alpha_n + \mathbf{E}(X\mathbf{Z}')\beta_{n1} + \mathbf{D}'_2\beta_{n2} + \mathbf{D}'_3\beta_{n3} + \mathbf{E}(\mathbf{Z}'\mathbf{Z})\gamma_n.
\end{aligned}$$

There are $(NJ+7J+5N+8) - (3J+5N+13) = (N+4)J - 5$ overidentifying restrictions.

5.2 The block-structure and the GMM problem in general

Let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$ denote the row vectors containing the elements of, respectively,

$$\begin{aligned}
a_j, b_j, \mathbf{c}'_j, \quad j = 1, \dots, J & \quad (\text{in Blocks [1] and [2]}), \\
C_2, \dots, C_6, \lambda_2, \dots, \lambda_5, \mathbf{D}_2, \dots, \mathbf{D}_5 & \quad (\text{in Blocks [2] and [3]}), \\
\alpha_n, \beta_{n1}, \beta_{n2}, \beta_{n3}, \gamma'_n, \quad n = 1, \dots, N & \quad (\text{in Block [3]}).
\end{aligned}$$

The observable variables be ordered in 5 column vectors: $\mathbf{w}_{[1]}, \mathbf{w}_{[2]}, \mathbf{w}_{[3]}$, containing the variables whose expectations occur only at the LHS of the equations:

$$\begin{aligned}
Q_j, Q_j\mathbf{Z}, Y_n Q_j, \quad j = 1, \dots, J; \quad n = 1, \dots, N & \quad (\text{Block [1]}), \\
X^g, X^g\mathbf{Z}, Q_j X, Q_j X^g, \quad g = 2, \dots, 5 & \quad (\text{Block [2]}), \\
Y_n X^2, Y_n X^3 & \quad (\text{Block [3]}),
\end{aligned}$$

and \mathbf{v} and \mathbf{q} containing the variables whose expectations occur either at the LHS or the RHS of the equations or enter more than one block, respectively,

$$Y_n, Y_n X, Y_n \mathbf{Z} \quad \text{and} \quad X, \mathbf{Z}, X\mathbf{Z}, \mathbf{Z}'\mathbf{Z}.$$

Finally, let $\phi_1(\cdot), \phi_2(\cdot), \phi_3(\cdot)$ be matrix functions and write the equations blockwise as

$$\begin{aligned}
\mathbf{E}(\mathbf{w}_{[1]}) &= \mathbf{E}(\mathbf{q}, \mathbf{v})\phi_1(\boldsymbol{\theta}_1), \\
\mathbf{E}(\mathbf{w}_{[2]}) &= \mathbf{E}(\mathbf{q})\phi_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \\
\mathbf{E}\begin{bmatrix} \mathbf{w}_{[3]} \\ \mathbf{v} \end{bmatrix} &= \mathbf{E}(\mathbf{q}, \mathbf{v})\phi_3(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3),
\end{aligned}$$

compactly,

$$\mathbf{E}\begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} = \mathbf{E}[\mathbf{q}\mathbf{v}]\phi(\boldsymbol{\theta}) \iff \mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{with} \quad \boldsymbol{\epsilon} \equiv \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} - [\mathbf{q}\mathbf{v}]\phi(\boldsymbol{\theta}).$$

The sample counterpart to $\mathbf{E}(\boldsymbol{\epsilon})$, with M observations, indexed by m , becomes

$$\frac{1}{M}\sum_{m=1}^M \boldsymbol{\epsilon}_m = \frac{1}{M}\sum_{m=1}^M \begin{bmatrix} \mathbf{w}_m \\ \mathbf{v}_m \end{bmatrix} - \frac{1}{M}\sum_{m=1}^M [\mathbf{q}_m \mathbf{v}_m]\phi(\boldsymbol{\theta}) \equiv \mathbf{g}_M(\mathbf{w}, \mathbf{v}, \mathbf{q}; \boldsymbol{\theta}),$$

where $\mathbf{w}_m, \mathbf{v}_m, \mathbf{q}_m, \boldsymbol{\epsilon}_m$ denote the values of $\mathbf{w}, \mathbf{v}, \mathbf{q}, \boldsymbol{\epsilon}$ for observation m . GMM estimation of $\boldsymbol{\theta}$ solves, for a positive definite weighting matrix $\mathbf{Q}(\boldsymbol{\theta})$,¹⁵

¹⁵This description is quite general and can be applied not only to this specific model ($I=3, K=3, G=5$), but to any model with a (N, I, J, K, G) -constellation that ensures identification.

$$\widehat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} [\mathbf{g}_M(\mathbf{w}, \mathbf{v}, \mathbf{q}; \boldsymbol{\theta})' \mathbf{Q}(\boldsymbol{\theta}) \mathbf{g}_M(\mathbf{w}, \mathbf{v}, \mathbf{q}; \boldsymbol{\theta})].$$

Asymptotically efficient estimation requires $\mathbf{Q}(\boldsymbol{\theta})$ to be proportional to the inverse asymptotic covariance matrix of $\bar{\boldsymbol{\epsilon}} = \frac{1}{M} \sum_m \boldsymbol{\epsilon}_m$; see Davidson and MacKinnon (1993, Theorem 17.3) and Harris and Mátyás (1999, section 1.3.3). The minimand depends on $\boldsymbol{\theta}$ in a complex way, however, which makes operationalization of this claim difficult.

5.3 A simplified, stepwise procedure

Instead of attempting to solve the GMM problem we proceed by estimating $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$ in three steps, utilizing that (i) the parameters in Block [1] enter linearly, (ii) the equations in Block [2] are linear in $C_i, \mathbf{D}_i, C_i \lambda_r, \mathbf{D}_i \lambda_r$, given the parameters in Block [1], and (iii) the parameters in Block [3] enter linearly, given the parameters in Blocks [1] and [2].

Step 1: Estimation of $\boldsymbol{\theta}_1$: Estimate $(a_j, b_j, \mathbf{c}'_j)$ from the equations in Block [1] by Two-Stage Least Squares (2SLS), using $(Y_1, \dots, Y_n, \mathbf{Z})$ as IV set for (X, \mathbf{Z}) , *i.e.*, regressing (Q_1, \dots, Q_J) on $(\widehat{X}, \mathbf{Z})$, where \widehat{X} is the fitted value in the first-stage regression of X on $(Y_1, \dots, Y_n, \mathbf{Z})$. This gives $(\widehat{a}_j, \widehat{b}_j, \widehat{\mathbf{c}}'_j)$ ($j = 1, \dots, J$).

Step 2: Estimation of $\boldsymbol{\theta}_2$: Form, from the equations in Blocks [1] and [2],

$$\begin{aligned} [\mathbf{E}(Q_j) - a_j - \mathbf{D}_0 \mathbf{c}_j] / b_j &= C_1, \\ [\mathbf{E}(Q_j X) - C_1 a_j - \mathbf{D}_1 \mathbf{c}_j] / b_j &= C_2, \\ [\mathbf{E}(Q_j X^2) - \mathbf{E}(X^2) a_j - \mathbf{E}(X^2 \mathbf{Z}) \mathbf{c}_j] / b_j &= C_3 + C_1 \lambda_2, \\ [\mathbf{E}(Q_j X^3) - \mathbf{E}(X^3) a_j - \mathbf{E}(X^3 \mathbf{Z}) \mathbf{c}_j] / b_j &= C_4 + 3C_2 \lambda_2 + C_1 \lambda_3, \\ [\mathbf{E}(Q_j X^4) - \mathbf{E}(X^4) a_j - \mathbf{E}(X^4 \mathbf{Z}) \mathbf{c}_j] / b_j &= C_5 + 6C_3 \lambda_2 + 4C_2 \lambda_3 + C_1 \lambda_4, \\ [\mathbf{E}(Q_j X^5) - \mathbf{E}(X^5) a_j - \mathbf{E}(X^5 \mathbf{Z}) \mathbf{c}_j] / b_j &= C_6 + 10C_4 \lambda_2 + 10C_3 \lambda_3 + 5C_2 \lambda_4 + C_1 \lambda_5, \end{aligned}$$

and let $Q_j = \psi_j + w_j$, where $\psi_j \equiv a_j + \xi b_j + \mathbf{Z} \mathbf{c}_j \iff \xi \equiv (\psi_j - a_j - \mathbf{Z} \mathbf{c}_j) / b_j$. The full Block [2] and the first equation of Block [1] can be rearranged into

$$\begin{aligned} \begin{bmatrix} \mathbf{E}(X^{i-1} \xi) \\ \mathbf{E}(X^i) \\ \mathbf{E}(X^i \mathbf{Z}') \end{bmatrix} &= \begin{bmatrix} C_i \\ C_i + \lambda_i \\ \mathbf{D}'_i + \mathbf{D}'_0 \lambda_i \end{bmatrix}, & i = 1, 2, \\ \begin{bmatrix} \mathbf{E}(X^{i-1} \xi) \\ \mathbf{E}(X^i) \\ \mathbf{E}(X^i \mathbf{Z}') \end{bmatrix} &= \begin{bmatrix} C_i \\ C_i + \lambda_i \\ \mathbf{D}'_i + \mathbf{D}'_0 \lambda_i \end{bmatrix} + \begin{bmatrix} \sum_{r=1}^{i-1} \binom{i-1}{i-r-1} C_{i-r} \lambda_r \\ \sum_{r=1}^{i-1} \binom{i}{i-r} C_{i-r} \lambda_r \\ \sum_{r=1}^{i-1} \binom{i}{i-r} \mathbf{D}'_{i-r} \lambda_r \end{bmatrix}, & i = 3, 4, 5, \\ \mathbf{E}(X^{i-1} \xi) &= C_i + \sum_{r=1}^i \binom{i-1}{i-1-r} C_{i-r} \lambda_r, & i = 6. \end{aligned}$$

Since the transformation

$$\begin{bmatrix} C_i \\ \lambda_i \\ \mathbf{D}'_i \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ \mathbf{D}'_0 & -\mathbf{D}'_0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} C_i \\ \lambda_i + C_i \\ \mathbf{D}'_i + \mathbf{D}'_0 \lambda_i \end{bmatrix},$$

where \mathbf{I} is the identity matrix and $\mathbf{0}$ is the zero row-vector, has lower-triangular ‘coefficient matrix’, these equations are equivalent to:

$$\begin{aligned}
(41) \quad C_1 &= \mathbf{E}(\xi) = \mathbf{E}(X), \\
C_2 &= \mathbf{E}(X\xi), \\
\lambda_2 &= \mathbf{E}(X^2) - C_2, \\
C_3 &= \mathbf{E}(X^2\xi) - C_1\lambda_2, \\
\lambda_3 &= \mathbf{E}(X^3) - C_3 - 3C_1\lambda_2, \\
C_4 &= \mathbf{E}(X^3\xi) - 3C_2\lambda_2 - C_1\lambda_3, \\
\lambda_4 &= \mathbf{E}(X^4) - C_4 - 6C_2\lambda_2 - 4C_1\lambda_3, \\
C_5 &= \mathbf{E}(X^4\xi) - 6C_3\lambda_2 - 4C_2\lambda_3 - C_1\lambda_4, \\
\lambda_5 &= \mathbf{E}(X^5) - C_5 - 10C_3\lambda_2 - 10C_2\lambda_3 - 5C_1\lambda_4, \\
C_6 &= \mathbf{E}(X^5\xi) - 10C_4\lambda_2 - 10C_3\lambda_3 - 5C_2\lambda_4 - C_1\lambda_5,
\end{aligned}$$

$$\begin{aligned}
(42) \quad D_0 &= \mathbf{E}(\mathbf{Z}), \\
D_1 &= \mathbf{E}(X\mathbf{Z}), \\
D_2 &= \mathbf{E}(X^2\mathbf{Z}) - D_0\lambda_2, \\
D_3 &= \mathbf{E}(X^3\mathbf{Z}) - 3D_1\lambda_2 - D_0\lambda_3, \\
D_4 &= \mathbf{E}(X^4\mathbf{Z}) - 6D_2\lambda_2 - 4D_1\lambda_3 - D_0\lambda_4, \\
D_5 &= \mathbf{E}(X^5\mathbf{Z}) - 10D_3\lambda_2 - 10D_2\lambda_3 - 5D_1\lambda_4 - D_0\lambda_5.
\end{aligned}$$

In the empirical illustration to be given in Section 5.4, three procedures for implementing this step will be considered:

Procedure 1: Let $\hat{v}_m \equiv X_m - \hat{X}_m$ and estimate C_i and λ_i by the analogous empirical moments, $\sum \hat{X}_m^i/M$ and $\sum \hat{v}_m^i/M$ ($i = 1, 2, \dots$). For estimation of the D_i s utilize (42) with $\mathbf{E}(X^i\mathbf{Z})$ estimated by $\sum X_m^i\mathbf{Z}_m/M$. Equations (41) are not used.

Procedure 2: Represent ξ by $\hat{\xi} = \hat{X}$ and estimate in (41) and (42), $\mathbf{E}(X^i\xi)$ and $\mathbf{E}(X^i\mathbf{Z})$ by, respectively, $\sum X_m^i\hat{X}_m/M$ and $\sum X_m^i\mathbf{Z}_m/M$ to obtain $\hat{C}_2, \dots, \hat{C}_6$ and $\hat{\lambda}_2, \dots, \hat{\lambda}_5$. For estimation of D_i use (42) as in Procedure 1.

Procedure 3: Instead of representing ξ by (the j -invariant) \hat{X} , use $\tilde{X}_j = (Q_j - \hat{a}_j - \mathbf{Z}\hat{\mathbf{c}}_j)/\hat{b}_j$. Otherwise, follow Procedure 2. Letting $\hat{Q}_j \equiv \hat{a}_j + \hat{X}\hat{b}_j + \mathbf{Z}\hat{\mathbf{c}}_j \iff \hat{X} \equiv (\hat{Q}_j - \hat{a}_j - \mathbf{Z}\hat{\mathbf{c}}_j)/\hat{b}_j$ and writing Q_j as $Q_j = \hat{Q}_j + \hat{w}_j$, the representations of ξ in Procedures 2 and 3 are related by $\tilde{X}_j = \hat{X} + \hat{w}_j/\hat{b}_j$. Under the assumptions made, \hat{X} and \tilde{X}_j are, like ξ , uncorrelated with all u_1, \dots, u_N and v asymptotically.

When *symmetry of the error distribution* ($\lambda_3 = \lambda_5 = 0$), is imposed, (41)–(42) imply

$$\begin{aligned}
(43.a) \quad C_1 &= \mathbf{E}(\xi) = \mathbf{E}(X), \\
(43.b) \quad C_2 &= \mathbf{E}(X\xi), \\
(43.c) \quad \lambda_2 &= \mathbf{E}(X^2) - C_2, \\
(43.d) \quad C_3 &= \mathbf{E}(X^2\xi) - C_1\lambda_2 = \mathbf{E}(X^3) - 3C_1\lambda_2, \\
(43.e) \quad C_4 &= \mathbf{E}(X^3\xi) - 3C_2\lambda_2 = \mathbf{E}(X^4) - 6C_2\lambda_2, \\
(43.f) \quad \lambda_4 &= \mathbf{E}(X^4) - C_4 - 6C_2\lambda_2, \\
(43.g) \quad C_5 &= \mathbf{E}(X^4\xi) - 6C_3\lambda_2 - C_1\lambda_4 = \mathbf{E}(X^5) - 10C_3\lambda_2 - 5C_1\lambda_4, \\
(43.h) \quad C_6 &= \mathbf{E}(X^5\xi) - 10C_4\lambda_2 - 5C_2\lambda_4,
\end{aligned}$$

(43.c)–(43.g) implying the restrictions

$$\begin{aligned}
(43.i) \quad \lambda_2 &= \mathbf{E}(X^2) - C_2 = [\mathbf{E}(X^3) - \mathbf{E}(X^2\xi)]/[2C_1] = [\mathbf{E}(X^4) - \mathbf{E}(X^3\xi)]/[3C_2], \\
(43.j) \quad \lambda_4 &= \mathbf{E}(X^4) - C_4 - 6C_2\lambda_2 = [\mathbf{E}(X^5) - \mathbf{E}(X^4\xi) - 4C_3\lambda_2]/[4C_1],
\end{aligned}$$

and

$$\begin{aligned}
(44.a) \quad & \mathbf{D}_0 = \mathbf{E}(\mathbf{Z}), \\
(44.b) \quad & \mathbf{D}_1 = \mathbf{E}(X\mathbf{Z}), \\
(44.c) \quad & \mathbf{D}_2 = \mathbf{E}(X^2\mathbf{Z}) - \mathbf{D}_0\lambda_2, \\
(44.d) \quad & \mathbf{D}_3 = \mathbf{E}(X^3\mathbf{Z}) - 3\mathbf{D}_1\lambda_2, \\
(44.e) \quad & \mathbf{D}_4 = \mathbf{E}(X^4\mathbf{Z}) - 6\mathbf{D}_2\lambda_2 - \mathbf{D}_0\lambda_4, \\
(44.f) \quad & \mathbf{D}_5 = \mathbf{E}(X^5\mathbf{Z}) - 10\mathbf{D}_3\lambda_2 - 5\mathbf{D}_1\lambda_4.
\end{aligned}$$

Step 3: Estimation of θ_3 : Define the estimators

$$\begin{aligned}
[\widehat{\mathbf{E}}(\widehat{Y}_n), \widehat{\mathbf{E}}(\widehat{Y}_n X), \widehat{\mathbf{E}}(\widehat{Y}_n X^2), \widehat{\mathbf{E}}(\widehat{Y}_n X^3)] &= \frac{1}{M} \sum_{m=1}^M [Y_{nm}, Y_{nm}X_m, Y_{nm}X_m^2, Y_{nm}X_m^3], \\
[\widehat{\mathbf{E}}(\widehat{\mathbf{Z}}), \widehat{\mathbf{E}}(\widehat{X}\widehat{\mathbf{Z}}), \widehat{\mathbf{E}}(\widehat{Y}_n\widehat{\mathbf{Z}}), \widehat{\mathbf{E}}(\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})] &= \frac{1}{M} \sum_{m=1}^M [\mathbf{Z}_m, X_m\mathbf{Z}_m, Y_{nm}\mathbf{Z}_m, \mathbf{Z}_m'\mathbf{Z}_m],
\end{aligned}$$

and replace C_1, \dots, C_6 , $\mathbf{D}_0, \dots, \mathbf{D}_3$, λ_2 and λ_3 by their estimates of from Step 2. The system defining for estimators of the coefficients of the N polynomials then becomes:¹⁶

$$(45) \quad \begin{bmatrix} 1 & \widehat{C}_1 & \widehat{C}_2 & \widehat{C}_3 & \widehat{D}_0 \\ \widehat{C}_1 & \widehat{C}_2 & \widehat{C}_3 & \widehat{C}_4 & \widehat{D}_1 \\ \widehat{C}_2 & \widehat{C}_3 & \widehat{C}_4 & \widehat{C}_5 & \widehat{D}_2 \\ \widehat{C}_3 & \widehat{C}_4 & \widehat{C}_5 & \widehat{C}_6 & \widehat{D}_3 \\ \widehat{D}_0 & \widehat{D}_1 & \widehat{D}_2 & \widehat{D}_3 & \widehat{\Sigma} \end{bmatrix} \begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_{n1} \\ \widehat{\beta}_{n2} \\ \widehat{\beta}_{n3} \\ \widehat{\gamma}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 1 & 0 & 0 & \mathbf{0} \\ 0 & -\widehat{\lambda}_2 & 1 & 0 & \mathbf{0} \\ 0 & -\widehat{\lambda}_3 & -3\widehat{\lambda}_2 & 1 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{E}}(\widehat{Y}_n) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X^2) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X^3) \\ \widehat{\mathbf{E}}(\widehat{Y}_n \widehat{\mathbf{Z}}) \end{bmatrix},$$

with solution

$$(46) \quad \begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_{n1} \\ \widehat{\beta}_{n2} \\ \widehat{\beta}_{n3} \\ \widehat{\gamma}_n \end{bmatrix} = \begin{bmatrix} 1 & \widehat{C}_1 & \widehat{C}_2 & \widehat{C}_3 & \widehat{D}_0 \\ \widehat{C}_1 & \widehat{C}_2 & \widehat{C}_3 & \widehat{C}_4 & \widehat{D}_1 \\ \widehat{C}_2 & \widehat{C}_3 & \widehat{C}_4 & \widehat{C}_5 & \widehat{D}_2 \\ \widehat{C}_3 & \widehat{C}_4 & \widehat{C}_5 & \widehat{C}_6 & \widehat{D}_3 \\ \widehat{D}_0 & \widehat{D}_1 & \widehat{D}_2 & \widehat{D}_3 & \widehat{\Sigma} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 1 & 0 & 0 & \mathbf{0} \\ 0 & -\widehat{\lambda}_2 & 1 & 0 & \mathbf{0} \\ 0 & -\widehat{\lambda}_3 & -3\widehat{\lambda}_2 & 1 & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{E}}(\widehat{Y}_n) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X^2) \\ \widehat{\mathbf{E}}(\widehat{Y}_n X^3) \\ \widehat{\mathbf{E}}(\widehat{Y}_n \widehat{\mathbf{Z}}) \end{bmatrix}, \quad n = 1, \dots, N.$$

Steps 1 and 2 may be interpreted as providing the estimates of the moments of the latent variables, $(\widehat{C}_2, \widehat{C}_3, \widehat{C}_4, \widehat{C}_5, \widehat{C}_6)$, $(\widehat{D}_2, \widehat{D}_3)$ and $(\widehat{\lambda}_2, \widehat{\lambda}_3)$ needed in Step 3 to estimate the coefficients of the N polynomials.

5.4 Illustration: system of Engel functions

The data set for the following empirical illustration, for $J = 1$, $N = 5$, is an extract of consumption data from the Norwegian Surveys of Consumer Expenditures for the years 1975–1977 and income data from tax returns. This is an admittedly old data set, but nevertheless suitable for a case study.¹⁷ It contains observations as two-year *averages* from $M = 408$ households, observed twice, one half in 1975 and 1976 and one half in 1976 and 1977. Owing to this averaging, unobserved household-specific heterogeneity

¹⁶We here disregard disturbances/measurement errors in modelling discrepancies between $\mathbf{E}(Y_n)$, $\mathbf{E}(Y_n X^i)$, $\mathbf{E}(Y_n \mathbf{Z}')$, $\mathbf{E}(\mathbf{Z}'\mathbf{Z})$ and $\widehat{\mathbf{E}}(\widehat{Y}_n)$, $\widehat{\mathbf{E}}(\widehat{Y}_n X^i)$, $\widehat{\mathbf{E}}(\widehat{Y}_n \widehat{\mathbf{Z}})$, $\widehat{\mathbf{E}}(\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})$ and between $(\widehat{C}_i, \widehat{\lambda}_i)$ and (C_i, λ_i) .

¹⁷See Aasness, Biørn, and Skjerpen (1993, Section 3, Appendix A) for detailed information.

becomes part of the equations' disturbances, assuming latent heterogeneity uncorrelated with ξ .¹⁸ Specifically, Y_n is the observed expenditure on commodity n , X is total expenditure and \mathbf{Z} is a (1×2) -vector containing the number of adults and the number of children in the household and Q_1 , the only indicator for total consumption used, is an income measure recorded in the tax code from the tax returns of the households' tax payers.¹⁹ The five commodities, exhausting total consumption: $X = \sum_{n=1}^5 Y_n$, are:

1. Food, beverages and tobacco
2. Clothing and footwear
3. Housing, fuel and furniture
4. Travel and recreation
5. Other goods and services

The Step 1-result gave the following 2SLS-estimate of the measurement equation (19):

$$(\hat{a}_1, \hat{b}_1, \hat{c}_{11}, \hat{c}_{12}) = (-0.6041, 0.6795, -0.0023, 1.1909)$$

To obtain \hat{X} in this step, only Y_1, \dots, Y_4 are used as regressors, as inclusion also of Y_5 would, because of the adding-up condition, have given an unsuitable IV for Step 2.

Table 4 gives, as a benchmark, OLS estimates for linear, quadratic and cubic equations with measurement errors neglected. There are weak signs of non-linearity: some quadratic and cubic terms have coefficient estimates differing significantly from zero at common levels. The $\hat{\beta}_{n1}$ s add to one, the estimate of the intercepts and the other coefficients add to zero (see the \sum -rows), which confirms a well-known property of OLS applied on systems with identical regressors.²⁰ Not surprisingly, the standard errors of $\hat{\beta}_{n1}$ from the quadratic model exceed those from the linear model, and the standard errors of $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$ from the cubic model exceed those from the quadratic model.

Consider first the C_i s, \mathbf{D}_i s and λ_i s as *free parameters*, implying that non-symmetric error distribution is allowed. Results are given in Tables 5 through 8. A primary intention is to illustrate the sensitivity of the results to (a) the choice between Procedures 1, 2 or 3 to implement Step 2 and (b) the imposition of symmetry of the distribution of the measurement errors. Throughout, the (conditional) standard errors reported are based on residuals $\hat{u}_{nm} = Y_{nm} - \hat{\alpha}_n - \sum_{i=1}^3 \hat{X}_m^i \hat{\beta}_{ni} - \mathbf{Z}_m \hat{\mathbf{c}}_n$, as obtained from (46). Table 5 gives the results for the linear equations, as estimated by Procedure 1 or 2 (which give the same result when $I=1$). All $\hat{\beta}_{n1}$ estimates in this table exceed those in Table 4, which concurs with the well-known 'attenuation' of OLS in one-regressor EIV models. The 'child effect' of commodity 3 Housing etc. changes sign when EIV is allowed for.

¹⁸The expenditure data are recorded by a combination of two-week bookkeeping and two interviews. Housing expenses are measured by rent (including maintenance and repairs), other durable goods are represented by the value of last year's purchases. As deflators for the expenditures are used price indexes constructed from the basic data used in calculating the official Norwegian consumer price index.

¹⁹It combines information on taxable income for the central government tax assessment (net of taxes) and the reported income base underlying the calculation of social security premiums and pension rights in the public social security system.

²⁰It is to be understood that similar restrictions are not imposed on the coefficients; see below.

Table 4: POLYNOMIAL EQUATIONS: NO MEASUREMENT ERROR

n	α_n	β_{n1}	β_{n2}	β_{n3}	γ_{n1}	γ_{n2}
<i>Coefficient estimates, linear</i>						
1	0.2446	0.1950			0.0485	-0.0140
2	-0.1038	0.2378			0.0059	0.1092
3	-0.2130	0.2753			-0.0180	0.0017
4	-0.0653	0.1146			-0.0081	-0.0135
5	0.1375	0.1772			-0.0283	-0.0834
Σ	0.0000	1.0000			0.0000	0.0000
<i>Standard errors</i>						
1	0.0481	0.0099			0.0149	0.0228
2	0.0568	0.0116			0.0176	0.0269
3	0.0638	0.0131			0.0198	0.0302
4	0.0298	0.0061			0.0092	0.0141
5	0.0410	0.0084			0.0127	0.0194
<i>Coefficient estimates, quadratic</i>						
1	0.2112	0.2153	-0.0020		0.0468	-0.0160
2	-0.1910	0.2907	-0.0053		0.0014	0.1040
3	-0.1181	0.2177	0.0057		-0.0131	0.0074
4	0.0199	0.0630	0.0052		-0.0036	-0.0085
5	0.0780	0.2132	-0.0036		-0.0314	-0.0869
Σ	0.0000	1.0000	0.0000		0.0000	0.0000
<i>Standard errors</i>						
1	0.0664	0.0294	0.0028		0.0151	0.0230
2	0.0782	0.0347	0.0033		0.0178	0.0271
3	0.0878	0.0389	0.0037		0.0200	0.0304
4	0.0407	0.0180	0.0017		0.0093	0.0141
5	0.0565	0.0250	0.0024		0.0128	0.0195
<i>Coefficient estimates, cubic</i>						
1	0.0823	0.3370	-0.0294	0.0017	0.0464	-0.0199
2	-0.0460	0.1538	0.0255	-0.0019	0.0018	0.1084
3	-0.1359	0.2345	0.0020	0.0002	-0.0131	0.0068
4	-0.0185	0.0992	-0.0030	0.0005	-0.0038	-0.0097
5	0.1181	0.1754	0.0049	-0.0005	-0.0313	-0.0857
Σ	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
<i>Standard errors</i>						
1	0.0961	0.0720	0.0150	0.0009	0.0151	0.0230
2	0.1131	0.0848	0.0177	0.0011	0.0177	0.0271
3	0.1275	0.0955	0.0199	0.0012	0.0200	0.0305
4	0.0591	0.0443	0.0092	0.0006	0.0093	0.0141
5	0.0820	0.0614	0.0128	0.0008	0.0129	0.0196

Table 5: LINEAR EQUATIONS: ESTIMATES FROM PROCEDURE 1

n	α_n	β_{n1}	γ_{n1}	γ_{n2}
<i>Coefficient estimates</i>				
1	0.2373	0.2025	0.0446	-0.0231
2	-0.1127	0.2469	0.0012	0.0980
3	-0.2233	0.2858	-0.0235	-0.0111
4	-0.0696	0.1190	-0.0104	-0.0189
5	0.1309	0.1840	-0.0319	-0.0917
Σ	-0.0375	1.0382	-0.0199	-0.0468
<i>Standard errors</i>				
1	0.0473	0.0099	0.0147	0.0225
2	0.0557	0.0116	0.0173	0.0265
3	0.0625	0.0130	0.0194	0.0297
4	0.0293	0.0061	0.0091	0.0140
5	0.0402	0.0084	0.0125	0.0191

Table 6 gives results obtained by applying *Procedure 1* on the quadratic and cubic equations. The $\hat{\beta}_{n2}$ s have the same signs as when measurement errors are disregarded, and for $\hat{\beta}_{n1}$ we find a similar ‘attenuation’ as in the linear model. The estimates and their cross-commodity sums differ substantially from those in the error-free case, in particular for the cubic equations. For example, $\sum_n \hat{\beta}_{n1}$ obtained from linear, quadratic and cubic equations is 1.04, 1.18, and 0.31, respectively. The $\hat{\beta}_{n3}$ estimates for commodities 2, 3 and 5 appear as significantly negative at common levels (according to ‘*t*-tests’). Again,

unsurprisingly, the standard errors of $\widehat{\beta}_{n1}$ obtained from the quadratic model exceed those from the linear model, while the standard errors of $\widehat{\beta}_{n1}$ and $\widehat{\beta}_{n2}$ in the cubic model exceed those for the quadratic model. The estimated moments of the latent regressor X and the error v are all positive:

$$\begin{aligned} (\widehat{C}_2, \widehat{C}_3, \widehat{C}_4, \widehat{C}_5, \widehat{C}_6) &= (21.82, 136.93, 976.71, 7694.61, 65460.64), \\ (\widehat{\lambda}_2, \widehat{\lambda}_3, \widehat{\lambda}_4, \widehat{\lambda}_5) &= (0.12, 0.04, 0.08, 0.07), \end{aligned}$$

but the estimated fourth and fifth-order moments of the error are quite small. All eigenvalues of the design matrix, *i.e.*, the leftmost matrix in (45), are also positive. In the cubic case, the range is (0.02, 66383.76; for the quadratic and linear models it is much smaller, (0.06, 1000.63) and (0.12, 28.46), respectively.

When comparing and interpreting the (sometimes substantial) differences between the coefficient estimates (notably the coefficient sums), across polynomial order and moment estimation procedure, it is essential to recall that adding-up of *observed* expenditures does not imply adding-up of the equation system and its coefficients when expressed in the latent variables. To see this, notice that (19), (34) and (35) imply

$$(47) \quad \sum_n \mathbf{E}(Y_n) - \mathbf{E}(X) = \sum_n \alpha_n + C_1(\sum_n \beta_{n1} - 1) + \sum_{i=2}^I C_i \sum_n \beta_{ni} + \mathbf{D}_0 \sum_n \gamma_n,$$

and hence

$$\sum_n \mathbf{E}(Y_n) \equiv \mathbf{E}(X) \iff \sum_n \beta_{n1} = 1, \sum_n \alpha_n = \sum_n \beta_{n2} = \dots = \sum_n \beta_{nI} = 0, \sum_n \gamma_n = \mathbf{0}.$$

Therefore, if the latter sum restrictions are not met, there will be a ‘trade-off’ between the estimated coefficient sums for different models and procedures, which restricts the comparability of their results.

Table 6: QUADRATIC AND CUBIC EQUATIONS: ESTIMATES FROM PROCEDURE 1

n	α_n	β_{n1}	β_{n2}	β_{n3}	γ_{n1}	γ_{n2}
<i>Coefficient estimates, quadratic</i>						
1	0.1580	0.2506	-0.0047		0.0401	-0.0288
2	-0.2562	0.3340	-0.0085		-0.0070	0.0878
3	-0.1922	0.2669	0.0019		-0.0217	-0.0089
4	-0.0106	0.0832	0.0035		-0.0070	-0.0147
5	0.0295	0.2454	-0.0060		-0.0376	-0.0989
Σ	-0.2715	1.1802	-0.0139		-0.0332	-0.0634
<i>Standard errors</i>						
1	0.0654	0.0292	0.0027		0.0149	0.0227
2	0.0767	0.0342	0.0032		0.0174	0.0266
3	0.0867	0.0387	0.0036		0.0197	0.0301
4	0.0405	0.0181	0.0017		0.0092	0.0141
5	0.0553	0.0247	0.0023		0.0126	0.0192
<i>Coefficient estimates, cubic</i>						
1	0.2242	0.1920	0.0085	-0.0008	0.0398	-0.0288
2	0.1261	-0.0043	0.0676	-0.0047	-0.0090	0.0877
3	0.0476	0.0548	0.0496	-0.0030	-0.0230	-0.0090
4	0.0599	0.0209	0.0175	-0.0009	-0.0074	-0.0147
5	0.2573	0.0439	0.0393	-0.0028	-0.0388	-0.0990
Σ	0.7150	0.3073	0.1825	-0.0122	-0.0384	-0.0639
<i>Standard errors</i>						
1	0.0998	0.0728	0.0152	0.0009	0.0149	0.0227
2	0.1143	0.0834	0.0175	0.0011	0.0170	0.0260
3	0.1314	0.0959	0.0201	0.0012	0.0196	0.0299
4	0.0617	0.0450	0.0094	0.0006	0.0092	0.0140
5	0.0831	0.0606	0.0127	0.0008	0.0124	0.0189

Powers of $\widehat{X}_i, \widehat{v}_i$ used for moment estimation

Table 7: QUADRATIC AND CUBIC EQUATIONS: ESTIMATES FROM PROCEDURE 2

n	α_n	β_{n1}	β_{n2}	β_{n3}	γ_{n1}	γ_{n2}
<i>Coefficient estimates, quadratic</i>						
1	0.1872	0.2327	-0.0030		0.0420	-0.0262
2	-0.2227	0.3133	-0.0066		-0.0047	0.0912
3	-0.1444	0.2381	0.0047		-0.0193	-0.0062
4	0.0114	0.0701	0.0049		-0.0060	-0.0138
5	0.0547	0.2299	-0.0046		-0.0359	-0.0964
Σ	-0.1138	1.0842	-0.0046		-0.0240	-0.0515
<i>Standard errors</i>						
1	0.0662	0.0296	0.0028		0.0149	0.0227
2	0.0777	0.0348	0.0033		0.0175	0.0266
3	0.0873	0.0391	0.0037		0.0196	0.0300
4	0.0407	0.0182	0.0017		0.0092	0.0140
5	0.0560	0.0251	0.0023		0.0126	0.0192
<i>Coefficient estimates, cubic</i>						
1	0.0622	0.3482	-0.0288	0.0016	0.0416	-0.0298
2	-0.0401	0.1446	0.0311	-0.0024	-0.0042	0.0963
3	-0.1351	0.2296	0.0066	-0.0001	-0.0193	-0.0059
4	-0.0183	0.0975	-0.0013	0.0004	-0.0061	-0.0147
5	0.1153	0.1740	0.0079	-0.0008	-0.0358	-0.0947
Σ	-0.0160	0.9939	0.0156	-0.0013	-0.0238	-0.0488
<i>Standard errors</i>						
1	0.0985	0.0738	0.0153	0.0009	0.0148	0.0227
2	0.1155	0.0864	0.0180	0.0011	0.0174	0.0266
3	0.1305	0.0977	0.0203	0.0012	0.0197	0.0301
4	0.0608	0.0455	0.0095	0.0006	0.0092	0.0140
5	0.0836	0.0626	0.0130	0.0008	0.0126	0.0193

(C_i, λ_i) -recursions based on \hat{X} used for moment estimation

Corresponding estimates obtained by *Procedure 2*, *i.e.*, using (C_i, λ_i) -recursions with ξ represented by \hat{X} , are given in Table 7, and imply $\sum_{n=1}^5 \hat{\beta}_{n1} = 1.08$ for the former and $= 0.99$ for the latter model. Both are much closer to unity than obtained by Procedure 1. The $\hat{\beta}_{n1}$ -estimates from the quadratic model are smaller than obtained by Procedure 1. The corresponding sum for the quadratic term, $\sum_{n=1}^5 \hat{\beta}_{n2}$ is -0.005 in the quadratic model and 0.016 in the cubic model, while for the latter $\sum_{n=1}^5 \hat{\beta}_{n3} = -0.001$. The estimated moments change to

$$\begin{aligned} (\hat{C}_2, \hat{C}_3, \hat{C}_4, \hat{C}_5, \hat{C}_6) &= (21.82, 136.56, 967.04, 7531.59, 63171.10), \\ (\hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5) &= (0.12, 0.04, -0.82, -3.19), \end{aligned}$$

giving negative fourth- and fifth-order error moments. These values, however, do not enter the estimation of the polynomials' coefficients in Step 3, see (46). Again, all eigenvalues of the design matrix are positive, with a considerable range, $(0.02, 64088.02)$, much larger than the range for the quadratic model, $(0.06, 991.09)$.

Turning to *Procedure 3*, *i.e.*, using (C_i, λ_i) -recursions with ξ represented by \tilde{X}_1 , we obtain the results in Table 8. Assuming linearity the change in the estimates are modest, for the cubic model, the change is substantial. All five functions come out as convex in (latent) total expenditure ($\hat{\beta}_{n2} > 0$), two commodities have negative first-order coefficients ($\hat{\beta}_{n1} < 0$), while all have large positive intercept estimates. Overall, the departure from linearity of the Engel functions appears as stronger when following Procedure 3 than when using Procedures 1 and 2. The estimated moments of the latent regressor X and the error v now are, respectively:

$$\begin{aligned} (\hat{C}_2, \hat{C}_3, \hat{C}_4, \hat{C}_5, \hat{C}_6) &= (22.02, 135.39, 916.31, 6642.66, 50572.28), \\ (\hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5) &= (-0.08, 3.69, 16.07, 6.34). \end{aligned}$$

Still all eigenvalues of the design matrix are positive, with ranges (0.02, 51466.07), (0.06, 941.41), and (0.12, 28.61) when assuming a cubic, quadratic and linear model, respectively. This procedure, however, gives a negative estimate of the error variance, λ_2 , while it gives a much larger estimate of λ_3 than Procedures 1 and 2 and estimates of λ_4 and λ_5 , substantially exceeding those from Procedure 1.

The moments of the error distribution have so far been assumed unrestricted. To implement the restrictions on the C_i s, D_i s and λ_i s to concur with symmetry, we (a) for Procedure 1, set $\lambda_3 = \lambda_5 = 0$ and estimate the D_i s according to (44.a)–(44.f); (b) for Procedures 2 and 3, set $\lambda_3 = \lambda_5 = 0$, estimate λ_2 and λ_4 as the average of the three, respectively two, values obtained from (43.i) and (43.j), estimate C_3, C_4, C_5 as the average of the resulting two values obtained from (43.d), (43.e), and (43.g), respectively, and finally estimate the D_i s according to (44.a)–(44.f).

Table 8: POLYNOMIAL EQUATIONS: ESTIMATES FROM PROCEDURE 3

n	α_n	β_{n1}	β_{n2}	β_{n3}	γ_{n1}	γ_{n2}
<i>Coefficient estimates, linear</i>						
1	0.2497	0.1899			0.0512	-0.0077
2	-0.0977	0.2316			0.0092	0.1168
3	-0.2059	0.2681			-0.0143	0.0106
4	-0.0624	0.1116			-0.0065	-0.0099
5	0.1421	0.1725			-0.0259	-0.0777
Σ	0.0258	0.9738			0.0137	0.0321
<i>Standard errors</i>						
1	0.0487	0.0099			0.0151	0.0230
2	0.0575	0.0117			0.0178	0.0272
3	0.0646	0.0131			0.0200	0.0305
4	0.0301	0.0061			0.0093	0.0142
5	0.0416	0.0084			0.0129	0.0196
<i>Coefficient estimates, quadratic</i>						
1	0.5350	0.0249	0.0187		0.0558	-0.0165
2	0.2005	0.0591	0.0195		0.0140	0.1076
3	0.3490	-0.0529	0.0363		-0.0052	-0.0065
4	0.2177	-0.0504	0.0183		-0.0020	-0.0185
5	0.3701	0.0406	0.0149		-0.0222	-0.0847
Σ	1.6722	0.0214	0.1078		0.0404	-0.0187
<i>Standard errors</i>						
1	0.0639	0.0271	0.0029		0.0144	0.0220
2	0.0763	0.0323	0.0034		0.0172	0.0262
3	0.0795	0.0337	0.0036		0.0179	0.0273
4	0.0362	0.0154	0.0016		0.0082	0.0125
5	0.0548	0.0232	0.0025		0.0124	0.0189
<i>Coefficient estimates, cubic</i>						
1	-0.1208	0.9799	-0.2381	0.0188	0.0559	-0.0969
2	-0.1640	0.5899	-0.1232	0.0105	0.0141	0.0629
3	-0.5862	1.3089	-0.3299	0.0269	-0.0051	-0.1211
4	-0.2766	0.6695	-0.1752	0.0142	-0.0019	-0.0791
5	0.0079	0.5680	-0.1269	0.0104	-0.0222	-0.1291
Σ	-1.1397	4.1161	-0.9934	0.0808	0.0408	-0.3632
<i>Standard errors</i>						
1	0.0621	0.0597	0.0152	0.0011	0.0110	0.0174
2	0.0928	0.0892	0.0227	0.0017	0.0164	0.0260
3	0.0679	0.0652	0.0166	0.0012	0.0120	0.0190
4	0.0235	0.0226	0.0058	0.0004	0.0042	0.0066
5	0.0636	0.0611	0.0156	0.0011	0.0112	0.0178

(C_i, λ_i) -recursions based on \tilde{X}_1 used for moment estimation

The resulting coefficient estimates are given in Table 9, from the cubic model, that for which the differences are most notable. Imposing error symmetry, the estimates from Procedure 1 are moderately changed (bottom half of Table 6 versus top third of Table 9). For example $\sum_n \hat{\beta}_{n1}$ increases from 0.31 to 0.34. Imposition of error symmetry on Procedure 2 induces notable changes (bottom half of Table 7 versus middle third of Table 9). For example for $n = 1$, commodity Food etc., $(\hat{\beta}_{n1}, \hat{\beta}_{n2}, \hat{\beta}_{n3})$ change from $(0.35, -0.03, 0.001)$ to $(0.40, -0.04, 0.002)$, while $(\sum_n \hat{\beta}_{n1}, \sum_n \hat{\beta}_{n2}, \sum_n \hat{\beta}_{n3})$ change from $(0.99, 0.02, -0.001)$ to $(1.12, 0.01, 0.000)$ and the estimated intercept sum $\sum_n \hat{\alpha}_n$, changes from -0.02 to -0.16 . More sensitive to imposition of error symmetry are the results from Procedure 3 (bottom third of Table 8 versus bottom third of Table 9). The polynomial coefficients for Food change from $(0.98, -0.24, 0.019)$ to $(0.12, -0.02, 0.035)$, the coefficient sums changes from $(4.12, -0.99, 0.08)$ to $(0.49, -0.06, 0.01)$, and the intercept sum changes from -1.14 to 1.68 . Note also the substantial reduction in the standard errors of the $(\hat{\beta}_{n1}, \hat{\beta}_{n2}, \hat{\beta}_{n3})$ s when error symmetry is imposed on Procedure 3.

Table 9: CUBIC EQUATIONS: ERROR SYMMETRY IMPOSED

n	α_n	β_{n1}	β_{n2}	β_{n3}	γ_{n1}	γ_{n2}
<i>Coefficient estimates: Procedure 1 with symmetry</i>						
1	0.2133	0.2015	0.0063	-0.0007	0.0399	-0.0287
2	0.1161	0.0035	0.0658	-0.0046	-0.0087	0.0884
3	0.0377	0.0629	0.0477	-0.0028	-0.0228	-0.0085
4	0.0557	0.0244	0.0167	-0.0008	-0.0073	-0.0146
5	0.2478	0.0517	0.0375	-0.0027	-0.0387	-0.0985
Σ	0.6706	0.3440	0.1741	-0.0116	-0.0377	-0.0619
<i>Standard errors</i>						
1	0.0997	0.0729	0.0152	0.0009	0.0149	0.0227
2	0.1143	0.0836	0.0175	0.0011	0.0171	0.0260
3	0.1313	0.0960	0.0201	0.0012	0.0196	0.0299
4	0.0616	0.0450	0.0094	0.0006	0.0092	0.0140
5	0.0831	0.0607	0.0127	0.0008	0.0124	0.0189
<i>Coefficient estimates: Procedure 2 with symmetry</i>						
1	0.0039	0.3999	-0.0397	0.0023	0.0409	-0.0322
2	-0.0526	0.1564	0.0283	-0.0022	-0.0040	0.0964
3	-0.1666	0.2580	0.0005	0.0003	-0.0195	-0.0069
4	-0.0330	0.1107	-0.0041	0.0006	-0.0063	-0.0152
5	0.0918	0.1952	0.0033	-0.0005	-0.0359	-0.0954
Σ	-0.1565	1.1202	-0.0118	0.0004	-0.0247	-0.0532
<i>Standard errors</i>						
1	0.1042	0.0795	0.0165	0.0010	0.0148	0.0227
2	0.1226	0.0936	0.0194	0.0012	0.0174	0.0267
3	0.1384	0.1057	0.0219	0.0013	0.0197	0.0302
4	0.0645	0.0492	0.0102	0.0006	0.0092	0.0141
5	0.0887	0.0677	0.0141	0.0009	0.0126	0.0194
<i>Coefficient estimates: Procedure 3 with symmetry</i>						
1	0.5994	0.1187	-0.0215	0.0035	0.0647	-0.0363
2	0.0924	0.1385	0.0037	0.0008	0.0154	-0.1150
3	0.3788	0.0920	-0.0198	0.0048	0.0071	-0.0227
4	0.2698	0.0249	-0.0150	0.0030	0.0059	-0.0321
5	0.3492	0.1131	-0.0078	0.0017	-0.0183	-0.0902
Σ	1.6897	0.4872	-0.0604	0.0138	0.0748	-0.0663
<i>Standard errors</i>						
1	0.0626	0.0195	0.0046	0.0005	0.0141	0.0219
2	0.0778	0.0243	0.0057	0.0006	0.0175	0.0272
3	0.0785	0.0245	0.0058	0.0006	0.0177	0.0275
4	0.0341	0.0107	0.0025	0.0003	0.0077	0.0119
5	0.0556	0.0173	0.0041	0.0004	0.0125	0.0195

Procedure 1: powers of \hat{X}_i, \hat{v}_i used for moment estimation.

Procedure 2: (C_i, λ_i) -recur., using \hat{X}_i , for moment estimation.

Procedure 3: (C_i, λ_i) -recur., using \hat{X}_1 , for moment estimation.

On the treatment of symmetry, see text

6 CONCLUSION

Our primary concern in this paper has been identification and estimation of polynomial single- and multi-equation measurement error models from moments of observable variables when (a) rather weak distributional assumptions are made for the latent regressor and its error part while (b) linear ‘measurement equations’ represent the relationships between the latent regressor and each of its indicators. A simple order-condition involving the number of equations, the polynomial order, and the number of instruments is presented. An essential element of the approach is that endogenous variables are included as instruments for the error-ridden regressor.

The structure of the identification problem has been shown to motivate method of moment procedures for uncovering the coefficients of the polynomial and of the measurement equations jointly with the moments of the latent regressor and the measurement errors. In the stepwise estimation procedure given most attention in the empirical application, three ways of obtaining empirical counterparts to the moments of the latent variable and the error are considered. Violation of the assumed stochastic independence of the latent regressor and the measurement error, and the orthogonality between the disturbances of the polynomial and the measurement equations, which are prerequisites for this method, may impair the ‘quality’ of the coefficient estimates, which might to some extent have been explored by simulation studies and bootstrapping of standard errors. The empirical illustration for a five equation-system of Engel functions in the final part shows that the procedure for estimating moments and whether or not error symmetry is imposed may have strong effect on the coefficient estimates, and the equations’ curvature, for the quadratic and cubic Engel functions. For the cubic models, the procedures for obtaining moments may strongly ‘interfere with’ the curvature of the estimated functions obtained. Assuming linear equations, the impact is less strong.

REFERENCES

- Aasness, J., Bjørn, and Skjerpen, T. (1993): Engel Functions, Panel Data, and Latent Variables. *Econometrica* **61**, 1395–1422.
- Amemiya, T. (1974): The Nonlinear Two-Stage Least-Squares Estimator. *Journal of Econometrics* **2**, 105–110.
- Amemiya, T. (1983): Non-linear Regression Models. Chapter 6 in: Z. Griliches and M.D. Intriligator (eds.) *Handbook of Econometrics, Volume I*. Amsterdam: North-Holland Publishing Company.
- Balakrishnan, N. and Nevzorov, V.B. (2003): *A Primer on Statistical Distributions*. Hoboken (NJ): Wiley.
- Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006): *Measurement Error in Non-linear Models. A Modern Perspective. Second Edition*. London: Chapman & Hall.
- Cheng, C.-L. and Schneeweiss, H. (1998): Polynomial Regression with Errors in the Variables. *Journal of the Royal Statistical Society: Series B* **60**, 189–199.
- Dagenais, M.G. and Dagenais, D.L. (1997): Higher Moment Estimators for Linear Regression Models with Errors in the Variables. *Journal of Econometrics* **76**, 193–221.
- Davidson, J. (1994): *Stochastic Limit Theory*. Oxford: Oxford University Press.

- Davidson, R. and MacKinnon, J.G. (1993): *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997): *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.
- Erickson, T. and Whited, T.M. (2002): Two-Step GMM Estimation of the Errors-In-Variables Model Using High-Order Moments. *Econometric Theory* **18**, 775–799.
- Evans, M., Hastings, N., and Peacock, B. (1993): *Statistical Distributions*, Second Edition. New York: Wiley.
- Hall, A.R. (2005): *Generalized Method of Moments*. Oxford: Oxford University Press.
- Harman, H.H. (1976): *Factor Analysis*. Chicago: University of Chicago Press.
- Harris, D. and Mátyás, L. (1999): Introduction to the Generalized Method of Moments Estimation. Chapter 1 in *Generalized Method of Moments Estimation*, ed. by L. Mátyás. Cambridge: Cambridge University Press.
- 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.
- Hausman, J.A., Newey, W.K., and Powell, J.L. (1995): Nonlinear Errors in Variables: Estimation of Some Engel Curves. *Journal of Econometrics* **65**, 205–233.
- Hu, Y. and Schennach, S.M. (2008): Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica* **76**, 195–216.
- Kuha, J. and Temple, J. (2003): Covariate Measurement Error in Quadratic Regression. *International Statistical Review* **71**, 131–150.
- Lewbel, A. (1997): Constructing Instruments for Regressions with Measurement Error When no Additional Data are Available, with an Application to Patents and R & D. *Econometrica* **65**, 1201–1213.
- Montfort, K. van, Mooijaart, A., and Leeuw, J. de (1987): Regression with Errors in Variables: Estimates Based on Third Order Moments. *Statistica Neerlandica* **41**, 223–239.
- Moon, M.-S. and Gunst, R.F. (1995): Polynomial Measurement Error Modeling. *Computational Statistics & Data Analysis* **19**, 1–21.
- Newey, W.K. (2001): Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models. *Review of Economics and Statistics* **83**, 616–627.
- Reiersøl, O. (1950): Identifiability of a Linear Relation Between Variables Which are Subject to Error. *Econometrica* **18**, 375–389.
- Satorra, A. (2002): Asymptotic Robustness in Multiple Group Linear-Latent Variable Models. *Econometric Theory* **18**, 297–312.
- Schennach, S.M. (2004): Estimation of Nonlinear Models with Measurement Error. *Econometrica* **72**, 33–75.
- Schennach, S.M. (2007): Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models. *Econometrica* **75**, 201–239.
- Wansbeek, T., and Meijer, E. (2000): *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North-Holland.
- Wolter, K.M. and Fuller, W.A. (1982): Estimation of the Quadratic Errors-in-Variables Model. *Biometrika* **69**, 175–182.