W1

8626

UNIVERSITY
OF
WISCONSIN-
MADISON

SSRI
Workshop Series

SEMIPARAMETRIC IDENTIFICATION
AND ESTIMATION OF POLYNOMIAL
ERRORS-IN-VARIABLES MODELS

Jerry Hausman, Hidehiko Ichimura,
Whitney Newey, James Powell

8626

SOCIAL SYSTEMS RESEARCH INSTITUTE

dP

Social Systems Research Institute

University of Wisconsin

SEMIPARAMETRIC IDENTIFICATION AND ESTIMATION OF POLYNOMIAL ERRORS-IN-VARIABLES MODELS

Jerry Hausman, Hidehiko Ichimura, Whitney Newey, James Powell

8626

August 1986

Social Systems Research Institute

University of Wisconsin

SEMIPARAMETRIC IDENTIFICATION AND ESTIMATION OF POLYNOMIAL ERRORS-IN-VARIABLES MODELS

Jerry Hausman, Hidehiko Ichimura,
Whitney Newey, James Powell

8626

August 1986

SEMIPARAMETRIC IDENTIFICATION AND ESTIMATION

OF POLYNOMIAL ERRORS-IN-VARIABLES MODELS*

by

Jerry A. Hausman, Hidehiko Ichimura,

Whitney K. Newey, and James L. Powell

1. Introduction

A particularly challenging statistical problem is the construction of consistent estimators of the parameters of nonlinear regression models when the regressors as well as the dependent variable are subject to measurement errors. As Griliches and Ringstad (1970) illustrate, the bias of classical least squares estimates can be exacerbated when the regression function is nonlinear; in view of the increasing number of applications of nonlinear models, and of the implausibility of assumptions which confine errors of measurement to the "dependent" variable, consistent estimation in this context is of more than theoretical interest. In the linear errors-in-variables model, this problem can be easily solved if additional observations on instrumental variables are available; since the linear model with measurement error is isomorphic to a linear simultaneous equations model, nonlinear two-stage least squares (T. Amemiya (1974)) yields a consistent estimator for linear regression functions. However, as recently noted by Y. Amemiya (1985), this correspondence breaks down when the regression function is nonlinear, and the standard application of instrumental variables estimation does not yield

----

* This research was supported by grants from the National Science Foundation. An earlier version of this paper was presented at the Fifth World Congress of the Econometric Society at M.I.T. in August, 1985.

consistent estimators for nonlinear models.

Thus, the few results available for nonlinear models require quite strong restrictions on the distribution of the measurement errors for identification (here used synonymously with consistent estimability) of the unknown regression coefficients. Knowledge of the parametric form of the distribution function of the measurement errors is not sufficient in general, unless the "true" values of the regressors (i.e., values without measurement errors) are also assumed to be random drawings from a distribution with known (or suitably restricted) parametric form. If, as assumed below, the "true" regressors are treated as fixed but unknown constants, then maximum likelihood estimation for the nonlinear model suffers from the notorious "incidental parameters" problem (Neyman and Scott (1948)) which renders maximum likelihood estimators inconsistent.

Most results on consistent estimation of nonlinear errors-in-variables models assume that the covariance matrix of the measurement errors for the regressors is shrinking toward zero as the sample size increases. Such a condition might be appropriate when a large number of measurements (relative to the sample size) on each "true" regressor are available, so that the average of these measurements more closely approximates the "true" regressors (in probability). Alternatively, the "shrinking covariance matrix" approximation may be appropriate if the measurement errors are thought to be small relative to the sample size. Examples of consistency results under this assumption can be found in Villegas (1969), Dolby and Lipton (1972), Wolter and Fuller (1982b), Powell and Stoker (1984), and Y. Amemiya (1985). An exception is Wolter and Fuller (1982a), which proposes a consistent estimator for a quadratic regression model with normal errors and requires neither additional measurements nor instrumental variables.

In this paper, we impose more structure on the form of the (nonlinear) regression function of interest, in order to eliminate the assumption of a shrinking error variance. Specifically, we assume that the regression function is a polynomial in the "true" regressors; for this special case, we can construct a consistent and asymptotically normal estimator for the unknown polynomial coefficients if either instrumental variables or an additional measurement of each "true" regressor is available (and satisfies appropriate regularity conditions). While a polynomial regression function is far from the most general forms of nonlinearity which arise in practice, it is an important starting point, and may lead to a more general theory of estimation based upon polynomial approximation of smooth nonlinear regression functions.

In the next section, identification and estimation of the coefficients of a polynomial regression function is considered when the explanatory variable is measured with error and when a single additional replicated measurement of the regressor (also with error) is available. Section 3 considers identification and estimation when additional information is available in the form of a structural model for the unobserved regressor in terms of observable instrumental variables, rather than a replicated measurement. The paper concludes with some remarks on possible extensions and limitations of the present approach.

2. The Functional Model with a Single Replicated Measurement

2.1 Identification

In this and the following section, we consider estimation of the parameters of a behavioral equation

$$(2.1) \qquad y_i = \sum_{j=0}^{K} \beta_j (z_i)^j + \varepsilon_i , \qquad i = 1, \ldots, n$$

which is a K^{th}-order polynomial in the variable z_i, which is assumed to be unobserved. In this section we treat z_i as a random variable with distribution function which is completely unknown (but satisfies certain regularity conditions). Alternatively, the $\{z_i\}$ can be viewed as a sequence of fixed constants, with appropriate modifications in the regularity conditions (as described at the end of this subsection). The variable z_i is assumed to be measured with error, with the observed variable x_i being related to z_i by the measurement equation

$$(2.2) \qquad x_i = z_i + \eta_i , \qquad i = 1, \ldots, n,$$

defining the measurement error η_i. In the absence of further knowledge of the functional form of the distribution of the unobservables (e.g., joint normality of e_i and η_i and nonnormality of z_i, or known moments of η_i), more information than is contained in equations (2.1) and (2.2) will generally be required in order to consistently estimate the regression coefficients $\beta \equiv (\beta_0, \ldots, \beta_K)'$. In this section we will consider identification and estimation when this extra information takes the form of a single repeated measurement w_i of z_i, with an additional measurement error v_i defined as

$$(2.3) \qquad w_i = z_i + v_i , \qquad i = 1, \ldots, n.$$

To guarantee that this equation contains sufficient additional information that can be used to identify and estimate β, it will be necessary to impose some conditions on the joint distribution of ε_i, η_i, v_i, and z_i. Defining the matrix norm $\|A\| \equiv \max_{i,j}|a_{ij}|$ (where $A = [a_{ij}]$), we impose

<u>Assumption 1</u>: The random variables ε_i, η_i, v_i, and z_i are jointly i.i.d. with

(i) $E(\varepsilon_i \mid z_i, v_i) = E(\eta_i \mid z_i, v_i) = 0$;

(ii) v_i is independent of z^i;

(iii) $E\|(\varepsilon_i, \eta_i, v_i^{2K}, z_i^{2K})\|^2 < \infty$.

(iv) All necessary moment matrices are nonsingular.

Of these conditions, (ii) is the most crucial for the estimation scheme discussed below. The i.i.d. assumption is made simply for convenience; since the identification argument is based only on the properties of the marginal distribution of the random variables for a given index i, dependence and/or heterogeneity can be introduced at the cost of some additional complexity. Only first moments need to exist in (iii) for the identification of $\bar{\beta}$, but second moments are assumed to obtain the asymptotic distribution of the corresponding estimator. Also, the conditional moment restriction in (i) is imposed instead of independence in order to derive results under the weakest possible assumptions which permit identification; while this introduces an asymmetry which is somewhat unnatural in the measurement equations (2.2) and (2.3), the results will obviously hold if the stronger assumption of independence of η_i and z_i is imposed.

Note also that the usual restriction $E(v_i) = 0$ is not imposed here. Allowing v_i to have a nonzero mean is equivalent to allowing the presence of a constant term in the measurement equation (2.3). It would also be useful to allow for a coefficient of z_i in (2.3) which is not unity, but it can be shown that this coefficient and the regression coefficients of interest would not be identified if v_i is allowed to have nonzero mean. Thus the restriction that the coefficient of z_i in (2.3) is equal to one is a normalization that is essential for identification of the remaining parameters. It will be apparent from the analysis to follow that the measurement equations (2.2) and (2.3) together with the restrictions of Assumption 1 place no additional restrictions on the moments of the observable variables.

Turning now to the identification question, if the moments $\xi_\ell \equiv E[y_i(z_i)^\ell]$ and $\zeta_m \equiv E[(z_i)^m]$ are identified for $\ell = 0, \ldots, K$ and $m = 0, \ldots, 2K$, then the coefficient vector $\beta \equiv (\beta_0, \ldots, \beta_K)'$ would be identified as the solution to the linear projection equations

$$(2.4) \qquad \xi_\ell = \sum_{j=0}^{K} \beta_j \cdot \zeta_{j+\ell}, \qquad \ell = 0, \ldots, K.$$

(These are just the population analogues of the "normal equations.") Although the moments in this projection involve the unobservable variable z_i, they are related to the moments $E[x_i(w_i)^\ell]$, $E[(w_i)^\ell]$, and $E[y_i(w_i)^\ell]$ of the observable variables. First, note that $1 \equiv E[(w_i)^0] \equiv E[(z_i)^0] \equiv E[(v_i)^0]$. Also, it follows from Assumption 1 that the observable moments satisfy the following relations:

$$(2.5) \qquad E\left[x_i(w_i)^{j-1}\right] = E\left[\sum_{\ell=0}^{j-1} \binom{j-1}{\ell} \cdot (z_i + \eta_i) \cdot (z_i)^\ell \cdot (v_i)^{j-1-\ell}\right]$$

$$= \sum_{\ell=0}^{j-1} \binom{j-1}{\ell} \zeta_{\ell+1} \cdot \nu_{j-\ell-1}, \text{ for } j = 1, \ldots, 2K;$$

(2.6) $$E\left[(w_i)^j\right] = E\left[\sum_{\ell=0}^{j} \binom{j}{\ell} (z_i)^{\ell} \cdot (v_i)^{j-\ell} \right]$$

$$= \sum_{\ell=0}^{j} \binom{j}{\ell} \zeta_{\ell} \cdot \nu_{j-\ell} \text{ , for } j = 1, \ldots, 2K;$$

and

(2.7) $$E\left[y_i (w_i)^j\right] = E\left[\sum_{\ell=0}^{j} \binom{j}{\ell} y_i \cdot (z_i)^{\ell} \cdot (v_i)^{j-\ell} \right]$$

$$= \sum_{\ell=0}^{j} \binom{j}{\ell} \xi_{\ell} \cdot \nu_{j-\ell} \text{ , for } j = 0, \ldots, K.$$

These (5K + 1) equations yield a one-to-one relationship between the moments of the observable variables and the (5K + 1) elements of the "unobservable" moment vectors $\zeta \equiv (\zeta_1, \ldots, \zeta_{2K})'$, $\nu \equiv (\nu_1, \ldots, \nu_{2K})'$, and $\xi \equiv (\xi_0, \ldots, \xi_K)'$. Moreover, the relationships can be used to solve recursively for the parameter vector $\theta \equiv (\zeta', \nu', \xi')'$. The recursion starts with $\zeta_0 = 1$, $\nu_0 = 1$, and $\xi_0 = E[y_i]$ (which follows from (2.7) above). Then, from equations (2.5) and (2.6), the 2K values of ζ and the nuisance parameters ν can be obtained from

(2.8) $$\zeta_j = E\left[x_i (w_i)^{j-1}\right] - \sum_{\ell=1}^{j-1} \binom{j-1}{\ell-1} \zeta_{\ell} \cdot \nu_{j-\ell}$$

and

(2.9) $$\nu_j = E\left[(w_i)^j\right] - \sum_{\ell=1}^{j} \binom{j}{\ell} \zeta_{\ell} \cdot \nu_{j-\ell}$$

for j = 1, ..., 2K. Finally, the remaining ξ coefficients can be obtained from

$$(2.10) \qquad \xi_j = E\left[y_i(w_i)^j\right] - \sum_{\ell=1}^{j} \binom{j}{\ell} \xi_\ell \cdot \nu_{j-\ell} ,$$

using the previously-obtained ν coefficients. Thus θ can be computed from the observable $E[x_i(w_i)^\ell]$, $E[(w_i)^\ell]$, and $E[y_i(w_i)^\ell]$, and β is then identifiable as a solution to the normal equations (2.4).

Note that if the $\{z_i\}$ are assumed to be fixed rather than random, then expectations over z_i can be interpreted as sample averages (e.g., $\zeta_\ell \equiv n^{-1} \sum z_i^\ell \equiv \zeta_{\ell,n}$) in the derivations above. Provided the matrix $D \equiv [d_{ij}] \equiv [\zeta_{i+j-2}, i,j = 1, \ldots, K+1]$ which characterizes the linear projection equations has mimimum characteristic root which is bounded away from zero for all n suitably large, the "observable" moments used in the calculation of ζ_ℓ, ν_ℓ, and ξ_ℓ can clearly be replaced by the corresponding sample averages in the arguments above.

2.2 Estimation

To estimate the "structural parameters" β in equation (2.1), the moments in the projection equation (2.4) can be estimated and β can then be obtained as the solution to these equations. As shown above, the moments in the projection equation are related to moments of the observable variables, which can be estimated by sample moments. Let

$$m_i \equiv \left(x_i, \ldots, x_i(w_i)^{2K-1}, w_i, \ldots, (w_i)^{2K}, y_i, \ldots, y_i(w_i)^K\right)$$

denote the (5K + 1)-dimensional data vector and let the corresponding vector of population moments be $\mu \equiv E[m_i]$. The moment vector μ can be consistently estimated by the sample moment vector $\hat{m} \equiv \frac{1}{n} \sum_i m_i$, and, by the Lindeberg-Lévy central limit theorem, the asymptotic distribution of $\hat{m}_i$ is given by

$$(2.11) \qquad \sqrt{n}\,(\hat{m} - \mu) \xrightarrow{d} \mathbf{N}(0,\ \Omega),$$

where $\Omega \equiv E[m_i m_i'] - \mu\mu'$. The covariance matrix Ω can clearly be consistently estimated by $\hat{\Omega} \equiv \left[\frac{1}{n} \sum_i m_i m_i' \right] - \hat{m}\hat{m}'$.

The recursion formulae (2.8)-(2.10) yield a one-to-one relationship between the moments μ of the observable data vector m_i and the moment vector $\theta \equiv (\zeta', \nu', \xi')'$ needed for the projection formula (2.4). Since the mapping $\theta = h(\mu)$ given by (2.8)-(2.10) is clearly continuous (and continuously differentiable), θ can be consistently estimated by $\hat{\theta} = h(\hat{m})$, i.e., as a solution to the recursion equations using the estimated moments of the data vector. The "delta method" gives the asymptotic distribution of $\hat{\theta}$ to be

$$(2.12) \qquad \sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{d} \mathbf{N}(0,\ H\Omega H'),$$

where $H = \partial h(\mu)/\partial \mu'$ is the Jacobian matrix for the transformation $\theta = h(\mu)$. The elements of H can also be calculated recursively: starting with $\partial \nu_0/\partial \mu_k = \partial \zeta_0/\partial \mu_k = 0$ and $\partial \xi_0/\partial \mu_k = 1[k = 4K+1]$ (with "1[A]" denoting the indicator function of the statement "A"), direct differentiation of (2.8)-(2.10) yields

$$(2.13) \qquad \frac{\partial \zeta_j}{\partial \mu_k} = 1[k = j] - \sum_{\ell=1}^{j-1} \binom{j-1}{\ell-1} \left[\frac{\partial \zeta_\ell}{\partial \mu_k} \cdot \nu_{j-\ell} + \frac{\partial \nu_{j-\ell}}{\partial \mu_k} \cdot \zeta_\ell \right],$$

$$(2.14)\qquad \frac{\partial \nu_j}{\partial \mu_k} = 1[k = j+2K] - \sum_{\ell=1}^{j} \binom{j}{\ell} \left[\frac{\partial \zeta_\ell}{\partial \mu_k} \cdot \nu_{j-\ell} + \frac{\partial \nu_{j-\ell}}{\partial \mu_k} \cdot \zeta_\ell \right] \text{, and}$$

$$(2.15)\qquad \frac{\partial \xi_j}{\partial \mu_k} = 1[k = j+4K+1] - \sum_{\ell=1}^{j} \binom{j}{\ell} \left[\frac{\partial \xi_\ell}{\partial \mu_k} \cdot \nu_{j-\ell} + \frac{\partial \nu_{j-\ell}}{\partial \mu_k} \cdot \xi_\ell \right]$$

for $j \geq 1$. A consistent estimator of H is given by $\hat{H} \equiv \partial h(\hat{m})/\partial \mu'$.

Finally, the structural coefficients β can be consistently estimated by solving the normal equations (3.2) with elements of the estimated vector of moments $\hat{\theta}$ used in place of the corresponding elements of θ. To give an algebraic representation of this solution, let D denote the second moment matrix of $(1, z_i, \ldots, (z_i)^K)'$ and $\hat{D} = D(\hat{\theta})$ its estimator based on the preceeding calculations, and write $\hat{\theta} \equiv (\hat{\zeta}', \hat{\nu}', \hat{\xi}')'$. Then the solution to the normal equations (2.4) is given by

$$(2.16)\qquad \hat{\beta} = \hat{D}^{-1}\hat{\xi} .$$

We can obtain the asymptotic distribution of $\hat{\beta}$ by using the fact that

$$(2.17)\qquad \sqrt{n}\ (\hat{\beta} - \beta) = \hat{D}^{-1}\sqrt{n}\ [\hat{\xi} - \hat{D}\beta]$$

$$\equiv \hat{D}^{-1}\sqrt{n}\ [S_\xi\hat{\theta} - (\beta' \otimes I_{K+1})S_\zeta\hat{\theta}],$$

where $\otimes$ denotes the Kronecker product, I_{K+1} is a K+1-dimensional identity matrix, and the matrices S_ξ and S_ζ are the selection matrices which yield $S_\xi\hat{\theta} = \hat{\xi}$ and $S_\zeta\hat{\theta} = \mathrm{vec}(\hat{D})$, the usual column vectorization of $\hat{D}$. It follows that

$$\text{(2.18)} \qquad \sqrt{n}\,(\hat{\beta} - \beta) \overset{d}{\to} N(0, V),$$

where

$$\text{(2.19)} \qquad V \equiv D^{-1}[S_\xi - (\beta' \otimes I_{K+1})S_\zeta]\cdot H\Omega H'\cdot[S_\xi - (\beta' \otimes I_{K+1})S_\zeta]'D^{-1} .$$

This asymptotic covariance matrix will be consistently estimated by

$$\text{(2.20)} \qquad \hat{V} \equiv \hat{D}^{-1}[S_\xi - (\hat{\beta}' \otimes I_{K+1})S_\zeta]\cdot \hat{H}\hat{\Omega}\hat{H}'\cdot[S_\xi - (\hat{\beta}' \otimes I_{K+1})S_\zeta]'\hat{D}^{-1} ,$$

where each of the component estimators is defined above.

3. <u>The Structural Model with Instrumental Variables</u>

3.1 Identification

In this section, we consider identification and estimation when the identifying information takes the form of "instrumental variables" which can be used to predict the unobserved regressor z_i. The polynomial behavioral equation and measurement equations are the same as in (2.1) and (2.2) above: that is, (3.1) $$y_i = \sum_{j=0}^{K} \beta_j (z_i)^j + \varepsilon_i \text{ , and}$$

(3.2) $$x_i = z_i + \eta_i \text{ , for } i = 1, \ldots, n.$$

In this section, though, we assume that z_i is related to a p-dimensional vector of instrumental variables q_i by the "causal equation"

(3.3) $$z_i = q_i'\alpha + v_i \text{ , } i = 1, \ldots, n.$$

Unlike the previous section, where the latent variable z_i was independent of the error term v_i, here we will assume v_i and the instruments q_i are independent. Thus (3.3) can be viewed as an auxiliary behavioral equation for the unobserved z_i.

Again, it is necessary to impose sufficient conditions on the disturbance terms ε_i, η_i, and v_i to ensure that these equations contain information that can be used to identify the regression coefficients β. Here we impose

<u>Assumption 2</u>: The random variables ε_i, η_i, v_i, and q_i are jointly i.i.d. with

(i) $E(\varepsilon_i \mid q_i, v_i) = E(\eta_i \mid q_i, v_i) = 0$, $E(\varepsilon_i \cdot \eta_i \mid q_i, v_i) = \sigma_{\varepsilon\eta}$;

(ii) v_i is independent of q_i with $E[v_i] = 0$;

(iii) $E[\|(\varepsilon_i, \eta_i)\|^2] < \infty$, $E[\|(v_i, q_i')\|^{2(K+1)}] < \infty$; and

(iv) All necessary moment matrices are nonsingular.

As before, the assumption of independently and identically-distributed data can be relaxed, and Assumption 2 (ii) is crucial for the scheme described below. Unlike the previous section, we need here to rule out any dependence of the conditional covariance of ε_i and η_i on the conditioning variates q_i and v_i; also, we impose a mean zero restriction of v_i, though relaxation of this restriction is discussed below.

If equation (3.3) is substituted into equation (3.2) we obtain

$$(3.4) \qquad x_i = q_i'\alpha + \eta_i + v_i , \qquad\qquad i = 1, \ldots, n,$$

so that α is identified as the coefficient vector of the least squares projection of x_i on q_i. We will consider the estimation of α below, but since α is identified we can assume without loss of generality that α is known when considering the identification of the behavioral parameters β below.

Let $w_i \equiv q_i'\alpha$ (again, assumed observable) and $\nu_j = E[(v_i)^j]$. As in the previous section, identification of β will also involve identification of the nuisance parameters $\{\nu_j,\ j = 2, \ldots, K\}$ (note that $\nu_0 \equiv 1$, and by Assumption 2 (ii), $\nu_1 = 0$). First, substitution of (3.3) into (3.1) yields

$$\begin{aligned}(3.5) \qquad y_i &= \sum_{j=0}^{K} \beta_j (w_i + v_i)^j + \varepsilon_i \\ &= \sum_{j=0}^{K} (w_i)^j \left[\sum_{\ell=j}^{K} \binom{\ell}{j} \beta_\ell \cdot (v_i)^{\ell-j} \right] + \varepsilon_i \\ &\equiv \sum_{j=0}^{K} \gamma_j (w_i)^j + e_i \quad ,\end{aligned}$$

where the second inequality follows from a binomial expansion with suitable reindexing, and where

$$(3.6) \qquad \gamma_j \equiv \sum_{\ell=j}^{K} \binom{\ell}{j} \beta_\ell \cdot \nu_{\ell-j}, \qquad j = 0, \ldots, K,$$

and

$$(3.7) \qquad e_i \equiv \varepsilon_i + \sum_{j=0}^{K} \sum_{\ell=j}^{K} \binom{\ell}{j} \beta_\ell \cdot \left[(v_i)^{\ell-j} - \nu_{\ell-j} \right] \cdot (w_i)^j ,$$

which implies $E(e_i \mid w_i) \equiv 0$. Because the disturbance term e_i is uncorrelated with any function of w_i, a least-squares projection of y_i on the vector $(1, w_i, \ldots, (w_i)^K)'$ of powers of w_i gives the coefficients $\gamma \equiv (\gamma_1, \ldots, \gamma_K)'$.

It is interesting to note that it follows from $\nu_0 = 1$ and $\nu_1 = 0$ that $\gamma_K = \beta_K$ and $\gamma_{K-1} = \beta_{K-1}$, so that the coefficients on the two highest order terms are identified from (3.5) alone. For example, if equation (3.1) is quadratic ($K = 2$), then the coefficients of the nonconstant variables will be identical to the corresponding coefficients in the least-squares regression of y_i on $(1, w_i, (w_i)^2)'$. In general, though, the coefficients $\gamma_0, \ldots, \gamma_{K-2}$ will be functions of both the structural coefficients β and the K-dimensional vector $\nu \equiv (\nu_1 \ldots, \nu_K)'$ of nuisance parameters, so not all the structural coefficients will be identified from (3.5) alone (unless $K = 1$, i.e., the model is linear).

To identify the remaining structural coefficients we use the restriction $E[\varepsilon_i \cdot \eta_i \mid w_i, v_i] = \sigma_{\varepsilon\eta}$ of Assumption 2(i). We make use of this restriction by considering the regression equation for $x_i \cdot y_i$, which can also be written in terms of β and ν. Multiplying equation (3.1) by x_i and substituting $w_i + v_i$ for z_i as in (3.5) above, we have

$$(3.8)\qquad x_i \cdot y_i = \sum_{j=0}^{K} \beta_j (w_i + v_i)^{j+1} + \eta_i \cdot y_i + (w_i + v_i)\cdot \varepsilon_i$$

$$= \sum_{j=0}^{K+1} (w_i)^j \left[\sum_{\ell=j}^{K+1} \binom{\ell}{j} \beta_{\ell-1} \cdot (v_i)^{\ell-j} \right] + \eta_i y_i + z_i \varepsilon_i$$

$$\equiv \sum_{j=0}^{K+1} \delta_j (w_i)^j + u_i ,$$

where $\beta_{-1} \equiv 0$,

$$(3.9)\qquad \delta_0 \equiv \sum_{\ell=0}^{K} \beta_\ell \cdot \nu_{\ell+1} + \sigma_{\xi\eta} ,$$

$$\delta_j \equiv \sum_{\ell=j}^{K+1} \binom{\ell}{j} \beta_{\ell-1} \cdot \nu_{\ell-j} , \qquad \text{for } j = 1, \ldots, K+1,$$

and

$$(3.10)\qquad u_i \equiv \sum_{j=0}^{K+1} \sum_{\ell=j}^{K+1} \binom{\ell}{j} \beta_{\ell-1} \cdot \left[(v_i)^{\ell-j} - \nu_{\ell-j} \right] \cdot (w_i)^j$$

$$+ [\eta_i y_i - \sigma_{\xi\eta}] + z_i \varepsilon_i , \qquad i = 1, \ldots, n.$$

Because of the moment restrictions imposed in Assumption 2 and the independence of w_i and v_i, the disturbance term u_i has $E[u_i | w_i, v_i] = 0$ by construction, so that the projection of $x_i \cdot y_i$ on the vector $(1, w_i, \ldots, (w_i)^{K+1})'$ yields the coefficients $\delta \equiv (\delta_0, \ldots, \delta_{K+1})'$. Here also the coefficients on the two highest-order terms in this projection equation are equal to the corresponding highest-order coefficients in the polynomial; that is, $\delta_{K+1} = \beta_K$ and $\delta_K = \beta_{K-1}$.

The (2K + 3) "reduced form" coefficients $\gamma_0, \ldots, \gamma_K, \delta_0, \ldots, \delta_{K+1}$ are thus identified as coefficients of the linear projections of y_i and $x_i y_i$ on the appropriate powers of w_i. The K + 1 "structural" coefficients $\beta_0, \ldots, \beta_K$, and the K nuisance parameters $\nu_1, \ldots, \nu_K$ can then be identified from the reduced form coefficients γ and δ. First, note that γ_K, γ_{K-1}, δ_{K+1}, and δ_K identify only β_K and β_{K-1}, and provide no information concerning the other structural coefficients. Thus, the remaining 2K − 1 structural and nuisance parameters must be obtained as functions of the remaining 2K − 1 reduced form coefficients. Equations (3.6) and (3.9) can be solved to obtain recursion relationships for the remaining parameters. For j > 1, these formulae are given as follows:

CALCULATION OF ν_j: Assume $\{\nu_\ell, \ \ell = 0, \ldots, j-1\}$ and $\{\beta_{K-\ell}, \ \ell = 0, \ldots, j-1\}$ are known from previous calculations. Then by (3.6) and (3.9),

$$\text{(3.11)} \qquad \delta_{K-j+1} - \gamma_{K-j} = \sum_{\ell=K-j}^{K} \left[\binom{\ell+1}{K-j+1} - \binom{\ell}{K-j} \right] \beta_\ell \cdot \nu_{\ell-K+j}$$

$$= \left[\binom{K}{K-j+1} \beta_K \cdot \nu_j \right] + \sum_{\ell=K-j+1}^{K-1} \binom{\ell}{K-j+1} \beta_\ell \cdot \nu_{\ell-K+j} ,$$

so that

$$\text{(3.12)} \qquad \nu_j = \left[\binom{K}{K-j+1} \beta_K \cdot \nu_j \right]^{-1}$$

$$\cdot \left[\delta_{K-j+1} - \gamma_{K-j} - \sum_{\ell=K-j+1}^{K-1} \binom{\ell}{K-j+1} \beta_\ell \cdot \nu_{\ell-K+j} \right] ,$$

where the right-hand side of this equation depends only on parameters that have been assumed to be previously identified.

CALCULATION OF β_j: Now assume $\{\nu_\ell,\ \ell=0,\dots,j\}$ and $\{\beta_{K-\ell},\ \ell=0,\dots,j-1\}$ are known; then

$$(3.13)\qquad \gamma_{K-j} = \beta_{K-j} + \sum_{\ell=K-j+1}^{j-1} \binom{\ell}{K-j} \beta_\ell \cdot \nu_{\ell-K+j}\ ,$$

and solving for β_{K-j} gives

$$(3.14)\qquad \beta_{K-j} = \gamma_{K-j} - \sum_{\ell=K-j+1}^{K} \binom{\ell}{K-j} \beta_\ell \cdot \nu_{\ell-K+j}\ ,$$

which also depends only on previously-known coefficients. Equation (2.9) could also be used to solve for β_{K-j} here.

The recursion relationships (3.12) and (3.14) can thus be used to identify all of the parameters of the original polynomial equation. Note that the intercept term δ_0 of equation (3.9) is not used in the identification of β; if the value of $\sigma_{\varepsilon\eta}$ were known, δ_0 could be used to identify ν_{K+1} (or vice versa).

3.2 Estimation

It is useful to consider estimation in two stages, where the first stage consists of estimation of the "reduced form" parameters γ and δ and the second stage consists of solving for the "structural" and nuisance parameters β and ν from γ and δ. For the estimation of the reduced form parameters, we follow the identification results given above and consider a two-step least squares procedure for their estimation. This type of estimator has the virtue of ease of computation, with the drawback that there may be other, more efficient estimators of γ and δ.

As noted above, the parameter vector α can be consistently estimated by a least squares regression of x_i on q_i; letting $\hat{\alpha}$ denote this estimator, a sequence of estimated values $\hat{w}_i$ of w_i can be formed as $\hat{w}_i \equiv q_i'\hat{\alpha}$. The coefficients $\gamma \equiv (\gamma_0, \ldots, \gamma_K)'$ can then be estimated with a least squares regression of y_i on $\hat{s}_i \equiv (1, \hat{w}_i, \ldots, (\hat{w}_i)^K)'$, and the coefficients $\delta \equiv (\delta_0, \ldots, \delta_{K+1})'$ can similarly be estimated from a regression of $x_i \cdot y_i$ on $t_i \equiv (1, \hat{w}_i, \ldots, (\hat{w}_i)^{K+1})'$.

To conduct inference and to form estimates of the structural parameters from the reduced form parameters, it is useful to have an estimator of the asymptotic covariance matrix of $(\hat{\gamma}', \hat{\delta}')'$. Let

$$r_i \equiv x_i - w_i = \eta_i + v_i$$

and define the matrices

$$Q \equiv E[q_i q_i'] \text{ and } R \equiv E[(r_i)^2 q_i q_i'].$$

Then it immediately follows from our assumptions that

$$\text{(3.15)} \qquad \sqrt{n}\,(\hat{\alpha} - \alpha) = \left[\frac{1}{n} \sum_{i=1}^{n} q_i q_i' \right]^{-1} \cdot \left[\frac{1}{n} \sum_{i=1}^{n} q_i r_i \right]$$

$$\xrightarrow{d} N\,(0,\, Q^{-1} R Q^{-1}).$$

Since the reduced form estimators depend on the estimator $\hat{\alpha}$ through the estimated regressors $\hat{w}_i$, the asymptotic covariance matrix of $(\hat{\gamma}', \hat{\delta}')'$ will include terms resulting from estimation of α. To give an explicit algebraic form for this asymptotic covariance matrix, define

$$s_i \equiv (1,\ w_i,\ \ldots,\ (w_i)K)',\quad t_i \equiv (1,\ w_i,\ \ldots,\ (w_i)^{K+1}),$$

$$(\Delta s_i)' \equiv \frac{\partial s_i'}{\partial(q_i'\alpha)} \equiv (0,\ \ldots,\ K(q_i'\alpha)^{K-1}), \qquad \text{and}$$

$$(\Delta t_i)' \equiv \frac{\partial t_i'}{\partial(q_i'\alpha)} \equiv (0,\ \ldots,\ (K+1)\cdot(q_i'\alpha)^{K}) \ .$$

Also, let

$$S \equiv E[s_i s_i'],\ T \equiv E[t_i t_i'],$$

$$F \equiv E[(\Delta s_i)'\gamma\cdot s_i q_i'],\ \text{and}\ G \equiv E[(\Delta t_i)'\delta\cdot t_i q_i].$$

Then straightforward calculations (see Newey (1984) for details) yield

$$(3.16) \qquad \sqrt{n}\begin{bmatrix} \hat{\gamma} - \gamma \\ \hat{\delta} - \delta \end{bmatrix} \overset{d}{\to} \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} M_0 & M_2' \\ M_2 & M_1 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix}^{-1} \right),$$

for

$$M_0 \equiv E[(s_i e_i - FQ^{-1}q_i r_i)(s_i e_i - FQ^{-1}q_i r_i)'],$$

$$M_1 \equiv E[(t_i u_i - GQ^{-1}q_i r_i)(t_i u_i - GQ^{-1}q_i r_i)'],\ \text{and}$$

$$M_2 \equiv E[(t_i u_i - GQ^{-1}q_i r_i)(s_i e_i - FQ^{-1}q_i r_i)'].$$

The matrices S, T, F, G, M_0, M_1, and M_2 can be consistently estimated by their finite-sample analogues, replacing expectations, error terms, and parameters by the corresponding sample averages, residuals, and estimates. If α were known (equivalently, if w_i were directly observable), the matrices F and G would be replaced by conformable zero matrices in the formulae above.

With these preliminary estimators of the reduced form parameters, we turn now to the estimation of the structural parameters β of interest. Given the particular estimators $\hat{\gamma}$ and $\hat{\delta}$ of the reduced form parameters, an efficient method of obtaining estimates of β (and ν) is optimal minimum distance estimation, i.e., minimum chi-square estimation. To define these estimators, first note that the coefficient δ_0 is not useful in the identification of the β parameters, and thus the estimate $\hat{\delta}_0$ is not used in the estimation of β. For later convenience, we write the remaining estimated reduced form parameters as the $(2K + 2)$-dimensional vector

$$\hat{\pi} \equiv (\hat{\gamma}_K, \gamma_{K-1}, \hat{\delta}_{K+1}, \hat{\delta}_K, \hat{\gamma}_{K-2}, \ldots, \hat{\gamma}_0, \hat{\delta}_{K-1}, \ldots, \hat{\delta}_1)'$$
$$\equiv (\hat{\pi}_1', \hat{\pi}_2')', \text{ with } \hat{\pi}_1 \equiv (\hat{\gamma}_K, \hat{\gamma}_{K-1}, \hat{\delta}_{K+1}, \hat{\delta}_K)',$$

with the corresponding vector of estimands denoted by $\pi \equiv (\pi_1', \pi_2')'$. Similarly, we write the 2K-dimensional vector of structural and nuisance parameters as

$$\theta \equiv (\beta_K, \beta_{K-1}, \beta_{K-2}, \ldots, \beta_0, \nu_K, \ldots, \nu_2)'$$
$$\equiv (\theta_1', \theta_2')', \text{ where } \theta_1 \equiv (\beta_K, \beta_{K-1})'.$$

The parameters π can be written in terms of θ,

$$\pi \equiv h(\theta),$$

which denotes the relationships given in (3.6) and (3.9) above. Finally, we let V denote the asymptotic covariance matrix of $\hat{\pi}$ (obtained by suitable rearrangement of the matrix given in (3.16)), and $\hat{V}$ denote its consistent estimator based on sample averages and estimated reduced form parameters.

With these definitions, the minimum chi-square (MCS) estimator $\hat{\theta}$ of θ is given as

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, [\hat{\pi} - h(\theta)]'\hat{V}^{-1}[\hat{\pi} - h(\theta)] \, . \tag{3.17}$$

In the context of the particular $h(\theta)$ function of the present problem, this

MCS estimator takes a particularly simple form. Because θ_2 is "just identified" given θ_1 and π_2, i.e.,

$$\theta_2 = g(\theta_1, \pi_2)$$

for a particular invertible function $g(\cdot)$, it can be shown that the MCS estimator $\hat{\theta}_1$ of $\theta_1 \equiv (\beta_K, \beta_{K-1})'$ depends on $\hat{\pi}$ only through $\hat{\pi}_1 \equiv (\hat{\gamma}_K, \hat{\gamma}_{K-1}, \hat{\delta}_{K+1}, \hat{\delta}_K)'$, i.e.,

$$(3.18) \qquad \hat{\theta}_1 = \underset{\theta}{\mathrm{argmin}}\, [\hat{\pi}_1 - h_1(\theta_1)]'\hat{V}_{11}^{-1}[\hat{\pi}_1 - h_1(\theta_1)] \ ,$$

where $\hat{V}_{11}$ is the submatrix of $\hat{V}$ corresponding to $\hat{\pi}_1$ and

$$h_1(\theta_1) = (\theta_1', \theta_1')' = (\beta_K, \beta_{K-1}, \beta_K, \beta_{K-1})'.$$

Solving this minimization problem, we find that $\hat{\theta}_1$ can be written as a matrix-weighted average,

$$(3.19) \qquad \hat{\theta}_1 \equiv \begin{bmatrix} \hat{\beta}_K \\ \hat{\beta}_{K-1} \end{bmatrix} = \hat{W} \cdot \begin{bmatrix} \hat{\gamma}_K \\ \hat{\gamma}_{K-1} \end{bmatrix} + [I - \hat{W}] \cdot \begin{bmatrix} \hat{\delta}_{K+1} \\ \hat{\delta}_K \end{bmatrix} ,$$

where

$$\hat{W} \equiv [\hat{W}_{\gamma\gamma} + \hat{W}_{\gamma\delta} + \hat{W}_{\delta\gamma} + \hat{W}_{\delta\delta}]^{-1}\cdot[\hat{W}_{\gamma\gamma} + \hat{W}_{\gamma\delta}],$$

for $\hat{V}_{11}^{-1}$ written in the partitioned form

$$\hat{V}_{11}^{-1} \equiv \begin{bmatrix} \hat{W}_{\gamma\gamma} & \hat{W}_{\gamma\delta} \\ \hat{W}_{\delta\gamma} & \hat{W}_{\delta\delta} \end{bmatrix} .$$

The MCS estimator of the remaining parameters θ_2 can be obtained by substitution of $\hat{\theta}_1$ into the minimand of (3.17) and minimization with respect to θ_2. The solution to this problem is given by

(3.20) $$\hat{\theta}_2 = g(\hat{\theta}_1, \tilde{\pi}_2),$$

where

$$\tilde{\pi}_2 \equiv \hat{\pi}_2 - \hat{V}_{21}\hat{V}_{11}^{-1}[\hat{\pi}_1 - h_1(\hat{\theta}_1)],$$

for $\hat{V}_{21}$ the appropriate submatrix of $\hat{V}$. In terms of the recursion formulae (3.12) and (3.14) given above, tis equation is equivalent to solving for the estimates of the remaining structural parameters using the MCS estimators $\hat{\beta}_K$ and $\hat{\beta}_{K-1}$ and the modified reduced form estimators

$$\tilde{\pi}_2 \equiv (\tilde{\gamma}_{K-2}, \ldots, \tilde{\gamma}_0, \tilde{\delta}_{K-1}, \ldots, \tilde{\delta}_1)'.$$

The asymptotic covariance matrix of the MCS estimator $\hat{\theta}$ will have the usual form $(H'V^{-1}H)^{-1}$, where $h = \partial\pi/\partial\theta' = \partial h(\theta)/\partial\theta'$, so that

(3.21) $$\sqrt{n}\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\nu} - \nu \end{bmatrix} \xrightarrow{d} \mathbf{N}(0, (H'V^{-1}H)^{-1}) .$$

To estimate the asymptotic covariance matrix of $\hat{\theta}$, an estimate of the Jacobian matrix H is required; recursion formulae for computation of the components of this matrix (similar to (2.13)-(2.15) of the previous section) can be obtained by differentiation of (3.6) and (3.9), and evaluating these formulae at $\hat{\theta}$ yields a consistent estimate of H.

In addition to providing efficient estimators of the structural parameters given the particular estimator of γ and δ used, the minimum chi-square estimator also allows one to test the overidentification of β_K and β_{K-1} in a convenient way. The model (3.1)-(3.3) can be viewed as a special case of a more general model, in which the measurement equation (3.2) is replaced by

$$(3.2') \qquad x_i = \tau + \rho z_i + \eta_i , \qquad i = 1, \ldots, n .$$

It can be shown that the structural parameters β and the nuisance parameters ν, ρ, and τ are just identified given the reduced form parameters γ and δ above, so the overidentification of β_K and β_{K-1} can be viewed as the imposition of the null hypothesis H_0: $\tau = 0$, $\rho = 1$. Because of the structure of the minimand in (3.17), its minimized value is equal to the minimized value of the criterion in (3.18), so by the general theory of minimum chi-square estimation,

$$(3.22) \qquad n[\hat{\pi}_1 - h_1(\hat{\theta}_1)]'\hat{V}_{11}^{-1}[\hat{\pi}_1 - h_1(\hat{\theta}_1)] \xrightarrow{d} \chi^2(2)$$

under H_0. Large values of the statistic in (3.22) provide evidence that the overidentifying restrictions implied by our assumptions are not satisfied, or indicates some other departure from the assumptions of the model.

4. Extensions and Limitations

The approaches taken above to estimation of regression coefficients for a polynomial equation of a single latent regressor can be extended to multivariate versions of polynomial regression functions, with each of the regressors being measured with error. While requiring a considerable increase in notation, identification and estimation results analogous to those in the previous sections can readily be obtained. For example, for estimation of a multivariate quadratic model with "structural equations" for the latent regressors as auxiliary identifying information, the coefficients of the nonconstant variables can be consistently estimated by a least squares regression which replaces the unobserved regressors by their fitted values from the estimated causal equations. Similarly, concommitant variables which are measured without error can be introduced into the regression equation; these can be viewed as special cases of the multivariate polynomial regression model, in which the appropriate regressors have measurement errors which are identically zero.

It should be noted that the estimation strategies given above will not in general yield asymptotically efficient estimators, even under the relatively weak moment restrictions imposed on the measurement errors. We have focused on the usual linear projection equations in the development of our proposed estimators, but these need not be the most efficient subset of the(infinite) class of unconditional moment restrictions which follow from the conditional moment restrictions on the error terms; thus, no claim for efficiency of the proposed procedures is made. The question of best attainable efficiency under the conditions imposed above, and the construction of feasible efficient estimators in these cases, are interesting questions for futher research.

A related caveat concerns the robustness of the proposed methods. The approaches outlined above require the existence of higher-order moments of latent variables and error terms, and the precision of the estimators is dependent on the precision with which these high-order moments can be estimated. Though this dependence is a direct consequence of the nature of the polynomial regression model considered, it does suggest caution in the application of this approach to polynomial models of high degree when the measured variables are though to be particularly "noisy."

REFERENCES

Amemiya, T. [1974], "The Nonlinear Two Stage Least Squares Estimator," Journal of Econometrics, 2: 105-110.

Amemiya, Y. [1985], "Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model," Journal of Econometrics, 28: 273-289.

Dolby, G.R. and S. Lipton [1972], "Maximum Likelihood Estimation of the Generalized Nonlinear Functional Relationship with Replicated Observations and Correlated Errors," Biometrika, 59: 121-129.

Griliches, Z. and V. Ringstad [1970], "Error-in-the-Variables Bias in Nonlinear Contexts," Econometrica, 38: 368-370.

Newey, W.K. [1984], "A Method of Moments Interpretation of Sequential Estimators," Economics Letters, 14: 201-206.

Neyman, J. and E.L. Scott [1948], "Consistent Estimates Based on Partially Consistent Observations," Econometrica, 38: 368-370.

Powell, J.L. and T.M. Stoker [1986], "The Estimation of Complete Aggregation Structures," Journal of Econometrics, 30: 317-341.

Villegas, C. [1969], "On the Least Squares Estimation of Non-Linear Relations," Annals of Mathematical Statistics, 11: 284-300.

Wolter, K.M. and W.A. Fuller [1982a], "Estimation of the Quadratic Error-in-Variables Model," Biometrika, 69: 175-182.

Wolter, K.M. and W.A. Fuller [1982b], "Estimation of Nonlinear Errors-in-Variables Models," Annals of Statistics, 10: 539-548.