NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

MODELS CONTAINING UNOBSERVABLE VARIABLES AND TRADITIONAL REGRESSION ANALYSIS

V. Fedorov

March 1984
WP-84-16

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

The objective of the Forest Sector Project at IIASA is to study long-term development alternatives for the forest sector on a global basis. One of the key problems of analysing long-term development is to predict the changing patterns of demand, supply capacity and international trade. In the modeling of these structural change processes econometric estimation plays a central role.

In this paper Professor Fedorov draws attention to the fact that some of the results from econometric analysis of estimation in situations with unobservable variables can be readdressed within the framework of traditional regression analysis. These observations are of importance in the choice of methods to be used in estimating demand, supply or import equations.

Markku Kallio
Project Leader
Forest Sector Project

ACKNOWLEDGMENT

I am very grateful to Lucy Tomsits for editing and typing the paper.

CONTENTS

MODELS CONTAINING UNOBSERVABLE VARIABLES AND TRADITIONAL REGRESSION ANALYSIS

V. Fedorov

1. INTRODUCTION

At present, models with unobservable variables are widely spread in econometric investigations. In this paper we deal with two of the most popular models of this kind. The first model (compare with Robinson and Ferrara, 1977) which we shall consider is described in the following.

$$y_i = \vartheta_0^T z_i + \varepsilon_i$$

$$z_i = B_0^T x_i + \nu_i, \qquad i = \overline{1,n}. \tag{1}$$

In (1) we suggest that the response y_i is observable, the vector $x_i \in R^l$ describes the conditions of observation and is supposed to be known; the vector $z_i \in R^m$ corresponds to unobservable variables; ε_i are i.i.d. random values with zero means and the variance σ^2; ν_i are i.i.d. random vectors with zero means and the variance matrix d; vector ϑ_0 and matrix B_0 contain unknown components (some of their components can be known *a priori*, e.g., they can be zeros). "T" stands for transposition, "0" points out the true value of the parameters.

The second model under consideration is described by the following set of equations (compare with Zellner, 1977):

$$
\begin{aligned}
y_i &= \vartheta_0^T z_i + \varepsilon_i \\
z_i &= B_0^T x_i, \\
u_i &= z_i + \nu_i, \qquad i = \overline{1,n}.
\end{aligned}
\tag{2}
$$

In (2) all notations and suggestions which take place for (1) are fulfilled. However, the relation between z_i and x_i does not contain random values, and the vector z_i can be observed with some additive random errors.

Both models can be generalized for multi-response cases (y_i can be a vector). The generalization of estimators is straight-forward and this is the reason why we consider the scalar case in this paper. Our main purpose is to show that the models (1) and (2) can be transformed to some well-known regression models and corresponding estimators for which the properties and the numerical procedures are well studied, and can be efficiently used after the appropriate adjustment.

2. REDUCING TO A REGRESSION MODEL WITH A VARIANCE CONTAINING UNKNOWN PARAMETERS

Let us start with the first model. It is obvious that variables z_i can be eliminated by constructing

$$y_i = \vartheta_0^T B_0^T x_i + \xi_i \tag{3}$$

where

$$E[\xi_i] = 0$$

and

$$E[\xi_i \xi_j] = \delta_{ij}(\sigma^2 + \vartheta_0^T d \vartheta_0)$$

In other words (2) is equivalent to regression problem (nonlinear, if both elements B and ϑ are unknown) with a variance depending on parameters ϑ_0.

CASE A. In the most trivial case when ϑ_0 and d_0 are given and some elements of the matrix B are unknown, model (3) can be transformed:

$$y_1 = \gamma_0^T x_i + \xi_i \tag{4}$$

where $\gamma = B\,\vartheta_0$. This model is a traditional linear regression model with unknown variance $s^2 = \sigma^2 + \vartheta_0^T d_0\, \vartheta_0$. Therefore no more than l linear combinations of the elements of B can be estimated and the least squares method provides the best linear unbiased estimators of these linear combinations.

It should be noted that the least square estimator of γ will be consistent if (see for instance, Wu 1981)

$$\limsup_{n \to \infty} \frac{\left(\text{max eigenvalue of } \sum_{i=1}^{n} x_i x_i^T\right)^{(1+c)/2}}{\text{min eigenvalue of } \sum_{i=1}^{n} x_i x_i^T} \le C < \infty \tag{5}$$

for some $c > 0$. In other words, the possibility of the consistent estimation is defined not only by the structure of the model but also by the structure of a design (conditions under which observations are made).

In the linear case (see Schmidt, 1976), the consistent estimator of variance $E[\xi_i^2] = \sigma^2 + \vartheta_0^T d_0 \vartheta_0$ can be constructed without any condition on the design.

CASE B. When all elements of the matrices B_0 and d_0 are known, one would also have to deal with the rather trivial situation which nevertheless presents some interest in applications. For the sake of simplicity, let us suppose that rank $B_0 = m$. Then model (3) can be transformed to the linear regression model with an unknown variance depending on parameters ϑ.

$$y_i = \vartheta_0^T w_i + \xi_i \tag{6}$$

where $w_i = B_0 x_i$, ξ_i is defined in the comments to (3).

When $B_0 = I$, where I is identity matrix, model (6) becomes a particular case of the well-known regression model with controllable variables subject to error (see for example, Fedorov 1974).

CASE C. The estimation problem for model (3) becomes more difficult when both ϑ and B depend upon the same unknown parameters. Consider the most simple case with $l = m = 1$, and $\vartheta = \nu\gamma$, $B = b\gamma$, where ν and b are known. Then model (1) can be reduced to the following model.

$$y_i = a\gamma^2 x_i + \xi_i \tag{7}$$

where

$$a = \nu b\,,\ E[\xi_i] = 0,\ E[\xi_i \xi_j] = \delta_{ij}[\sigma^2 + \vartheta^2\gamma^2 d]$$

It is obvious that the parameter γ for (1) is consistently estimated if at least a sign of γ_0 is known *a prior*.

In the general case, when $\vartheta = \vartheta(\gamma)$, $B = B(\gamma)$, $l, m \geq 1$ and $\gamma \in R^k$, one has

$$y_i = \vartheta^T(\gamma_0) B^T(\gamma_0) x_i + \xi_i = \Psi^T(\gamma) x_i + \xi_i \tag{8}$$

where

$$E[\xi_i] = 0, E[\xi_i \xi_j] = \delta_{ij}[\sigma^2 + \vartheta^T(\gamma_0) d \vartheta(\gamma_0)]$$

It is known (Wu, 1981: Theorem 1) that if there exist a consistent estimator for all $\gamma_0 \in \Gamma \subset R^k$ where Γ is compact, then (under very mild conditions on the distribution of ξ_i)

$$[\Psi(\gamma) - \Psi(\gamma_0)]^T M_n [\Psi(\gamma) - \Psi(\gamma_0)] \to \infty \tag{9}$$

as $n \to \infty$ for all $\gamma \neq \gamma_0$ in Γ. Here $M_n = \sum_{i=1}^{n} x_i x_i^T$. Moreover, the condition (9) provides the consistency of the least squares estimator of γ. In other words, an experimenter should appropriately choose a design $\{x_i\}_1^n$ as well as Γ. Note that the condition (9) is not fulfilled if the set Γ includes $\gamma = -\gamma_0$ in (7). It is clear from (9) that no more than l parameters can be estimated if either σ^2 or d is unknown.

CASE D. As it was noted in *Case B*, model (3) is closely connected with the regression models when controllable variables are subject to error. Moreover, the method suggested in Fedorov (1974) can be used for approximate estimation of parameters of the generalized version of (1):

$$\begin{aligned} y_i &= \eta(\gamma_0, z_i) + \varepsilon_i \\ z_i &= \rho(\gamma_0, x_i) + \delta\nu_i = \rho_i + \delta\nu_i \qquad i = \overline{1,n} \end{aligned} \tag{1*}$$

where γ stands for both parameters ϑ and B, $\eta \in R^1$, $\rho \in R^l, \gamma \in R^k$, ε_i are i.i.d. random values with variance σ^2, ν_i are i.i.d. random vectors with unit variance. Matrix d and δ are some constants.

In the following, it will also be assumed that the function $\eta(\gamma, z)$ has derivatives in correspondence to z up to the third ones for all $z_i = \rho(\gamma, x_i)$ and $\gamma \in \Gamma \subset R^k$ where Γ is compact, and

$$E[|\nu_{ip}\nu_{iq}\nu_{ir}|] \leq c < \infty, \qquad i = \overline{1,n}, \quad p, q, r = \overline{1,m}.$$

Similarly to Fedorov (1974), one can calculate that

$$E[y_i] = E[\eta(\gamma_0, \rho_i + \delta\nu_i) + \varepsilon_i] = \varphi(\gamma_0, x_i) + 0(\delta^3)$$
$$E[(y_i - E(y_i))^2] = \lambda^{-1}(\gamma_0, x_i) + 0(\delta^3) \qquad (10)$$

where

$$\varphi(\gamma,x) = \eta(\gamma,\rho) + \frac{\delta^2}{2}\, tr\ d\ \frac{\partial^2\eta(\gamma,\rho)}{\partial\rho\partial\rho^T}\Big|_{\rho=\rho(\gamma,x)},$$
$$\lambda^{-1}(\gamma,x) = \sigma^2 + \delta^2\ \frac{\partial\eta(\gamma,\rho)}{\partial\rho^T}\, d\, \frac{\partial\eta(\gamma,\rho)}{\partial\rho}\Big|_{\rho=\rho(\gamma,x)}.$$

Therefore the model (1*) can be approximated by the following regression model

$$y_i = \varphi(\gamma_0, x_i) + \mu_i \qquad (11)$$

where $\lambda^{\frac{1}{2}}(\gamma_0, x_i)\mu_i$ are i.i.d. random values. This regression problem is well studied (see for example: Fedorov 1974; Jobson & Fuller, 1980, Carrol, 1982) and the simple estimator closely related to the least squares method can be used here. This estimator is defined as a limit point of the following iterative procedure:

$$\hat{\gamma} = \lim_{s\to\infty} \gamma_s\ , \qquad (12)$$
$$\gamma_s = \mathrm{Arg}\min_{\gamma\in\Gamma} \sum_{i=1}^{n} \lambda(\gamma_{s-1}, x_i)[y_i - \varphi(\gamma, x_i)]^2$$

The estimator (12) will be consistent within the frame of approximation (10) under very mild assumptions (see for example, Fedorov 1974; Wu 1981), the main one of which is

$$\sum_{i=1}^{n} \lambda(\gamma_0, x_i)[\varphi(\gamma, x_i) - \varphi(\gamma_0, x_i)]^2 \to \infty$$

as $n\to\infty$ for all $\gamma \neq \gamma_0$ in Γ. For "sufficiently" smooth functions $\varphi(\gamma, x_i)$, the estimator (12) is normally asymptotically distributed:

$$\tau_n(\hat{\gamma}_n - \gamma) \to N(0, M^{-1}) \tag{13}$$

where $\tau_n \to \infty$ as $n \to \infty$, and

$$M = \lim_{n \to \infty} \tau_n^{-1} \sum_{i=1}^{n} \lambda(\gamma_0, x_i) \frac{\partial \varphi(\gamma, x_i)}{\partial \gamma} \frac{\partial \varphi(\gamma, x_i)}{\partial \gamma^T} \Big|_{\gamma=\gamma_0} \tag{14}$$

Naturally, the existence of the limit in (14) is suggested. The last result is more general than a similar one from Fedorov (1974), and it is the obvious corollary of theorem 5 from Wu (1981). It is useful to bear in mind that the estimator

$$\tilde{\gamma} = \operatorname{Arg} \min_{\gamma \in \Gamma} \sum_{i=1}^{n} \lambda(\gamma, x_i)[y_i - \varphi(\gamma, x_i)]^2$$

is not consistent for regression problem (11) (Fedorov, 1974).

In conclusion of this section, we point out that the models from the *Cases A - C* can be treated as the particular cases of model (1*) and for all of them the residuals terms in (10) equal to zero.

3. REDUCING TO A MULTIRESPONSE REGRESSION MODEL

The model (2) can be treated as the specific case of the multiresponse regression model

$$w_i = \eta(\gamma, x_i) + \mu_i \tag{15}$$

where $\eta(\gamma, x)$ is a given vector function; $\gamma \in R^k$, random vectors μ_i are i.i.d. with zero means; and the variance matrix

$$E[\mu_i \mu_i^T] = \Sigma = \begin{Bmatrix} b^2 & 0 \\ 0 & d_s \end{Bmatrix}$$

Consider the case when all elements of B and ϑ are unknown. Model (2) can be transformed to model (15) if one assumes that

$$\gamma^T = (\vartheta_1, \cdots, \vartheta_m, B_{11}, \cdots, B_{l1}, \ldots, B_{1m}, \cdots, B_{lm})$$

$$\eta_1(\gamma, x) = \sum_{\alpha=1}^{m} \vartheta_\alpha \sum_{\beta=1}^{l} B_{\beta\alpha} x_\beta$$

$$\eta_2(\gamma, x) = \sum_{\beta=1}^{l} B_{\beta 1} x_\beta \tag{16}$$

$$\eta_{m+1}(\gamma, x) = \sum_{\beta=1}^{l} B_{\beta m} x_\beta$$

Model (15) was studied for instance by Philips (1976) and Fedorov (1977). Several slightly different estimators were suggested for the estimation of γ and Σ (the case when Σ is given is too well-known to be discussed here). Similar to the estimator defined by (12), one of them can also be found with the help of the following iterative procedure:

$$\hat{\gamma} = \lim_{s \to \infty} \gamma_s, \quad \hat{\Sigma} = \lim_{s \to \infty} \Sigma_s,$$

$$\gamma_s = \operatorname{Arg} \min_{\gamma \in \Gamma} \sum_{i=1}^{n} [w_i - \eta(\gamma, x_1)]^T \Sigma_s^{-1} [w_i - \eta(\gamma, x_i)] \tag{17}$$

$$\Sigma_s \doteq \sum_{i=1}^{n} [w_i - \eta(\gamma_{s-1}, x_i)][w_i - \eta(\gamma_{s-1}, x_1)]^T$$

Unlike model (1), in the case under consideration, all elements of B and ϑ can be consistently estimated if the sequences $x_1, \cdots, x_n$ is appropriately chosen. The estimator (17) asymptotically coincides with the maximum likelihood estimator when μ_i are normally distributed.

The estimator (17) can be improved if the structure of the variance matrix Σ will be taken into account, and in the iterative procedure instead of the matrices Σ_s, one will use

$$\Sigma_s^1 = \begin{Bmatrix} b^2 & 0 \\ 0 & d_s \end{Bmatrix} = \sum_{i=1}^{n} \begin{Bmatrix} (y_i - \vartheta_s^T z_{si})^2 & 0 \\ 0 & (u_i - z_{si})(u_i - z_{si})^T \end{Bmatrix} \tag{18}$$

where $z_{si} = B_s^T x_i$; $\tilde{\gamma}$ (or $\tilde{\vartheta}$ and $\tilde{B}$) and $\tilde{\Sigma}$ (or $\tilde{\sigma}^2$ and $\tilde{d}$) will denote below the improved estimators of γ (or ϑ and B) and Σ (or σ^2 and d).

Let us introduce matrices

$$M_{11}=\sum_{i=1}^{n} \tilde{z}_i \tilde{z}_i^T, \quad M_{12}=\sum_{i=1}^{n} x_i \tilde{z}_i^T, \quad M_{22}=\sum_{i=1}^{n} x_i x_i^T,$$

where $\tilde{z}_i = \tilde{B}^T x_i$. Using the standard techniques of regression analysis, it is possible to check that the consistent estimator $\tilde{D}$ of the variance matrix $D = E[\tilde{\gamma}\tilde{\gamma}^T]$ can be calculated in the following way:

$$\tilde{D} = (J_1 + J_2)^{-1},$$

where

$$J_1 = \tilde{\sigma}^{-2} \begin{bmatrix} M_{11} & \tilde{\vartheta}^T \otimes M_{12}^T \\ \tilde{\vartheta} \otimes M_{12} & \tilde{\vartheta}\tilde{\vartheta}^T \otimes M_{22} \end{bmatrix} \quad \text{and} \quad J_2 = \begin{bmatrix} 0 & 0 \\ 0 & \tilde{d}^{-1} \otimes M_{22} \end{bmatrix}$$

The matrix J_1 can be interpreted as the information matrix corresponding to observations of y_i and the matrix J_2 can be interpreted as the information matrix corresponding to observations of z_i.

When rank B_0 = rank $d_0 = m$, then similarly to (5), the parameters ϑ and B will be consistently estimated by (17) if

$$\limsup_{n \to \infty} \frac{(\text{max eigenvalue of} M_{22})^{(1+c)/2}}{\text{min eigenvalue of} M_{22}} \le c < \infty,$$

for some $c > 0$. This fact is derived from the results of Wu (1981) and from the structure of the matrix $\tilde{D}$ which is mainly defined by the matrix M_{22}:

$$M_{11}=\tilde{B}^T M_{22}\tilde{B} \quad \text{and} \quad M_{12} = M_{22}\tilde{B}\ .$$

In conclusion, it should be emphasized that (12), (17), and (18) not only describe the estimators with some admissable statistical properties;

it does, moreover, deliver the effective numerical procedures which are based on the well-studied standard least square techniques.

REFERENCES

Carrol, R.J. (1982) Adapting for Heteroscedasticity in Linear Models. *Ann. Statis*. 10:1224-1233

Fedorov, V.V. (1974) Regression Problems with Controllable Variables Subject to Errors. *Biometrika*, 61:49-56.

Jobson, J.D., and W.A. Fuller (1980) Least Squares Estimation when the Covariance Matrix and Parameter Vector are Functionally Related. *JAAS* 75:176-181.

Phillips, P.C.B. (1976) The Iterated Minimum Distance Estimator and Quasi-Maximum Likelihood Estimator. *Econometrica* 44:449-460.

Robinson, P.M., and M.C. Ferarra (1977) The Estimation of Model for an Unobservable Variable with Endogenous Causes, in *Latent Variables in Socio-Economic Models*: 131-142. Amsterdam: North Holland.

Schmidt, W.H. (1976) Strong Consistency of Variance Estimation and Asymptotic Theory of the Linear Hypotheses in Multivariate Linear models. *Math. Operationsforsch. Statist.* 7:701-705.

Wu, C.F. (1981) Asymptotic Theory of Nonlinear Least Squares Estimation. *Ann. Stat.* 9:501-513.

Zellner, A. (1977) Estimation of Regression Relationships Containing Unobservable Independent Variables, in *Latent Variables in Socio-Economic Models*:67-83. Amsterdam: North-Holland.