Theoretical and Applied Essays on the Instrumental Variable Method

Davood Souri

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Djavad Salehi-Isfahani, Chair

Richard Ashley

Dennis Yang

G. Geoffrey Vining

Bradford Mills

August 2004
Blacksburg, Virginia

Keywords:Instrumental variable method, Weak instruments, Instrument relevancy, Specification, Return to schooling, Natural experiment

Theoretical and Applied Essays on the Instrumental Variable Method

Davood Souri

(Abstract)

This dissertation is intended to provide a statistical foundation for the IV models and shed lights on a number of issues related to the IV method. The first chapter shows that the theoretical Instrumental Variable model can be derived by reparameterization of a well-specified statistical model defined on the joint distribution of the involved random variables as the actual (local) data generation process. This reveals the covariance structure of the error terms of the usual theory-driven instrumental variable model. The revealed covariance structure of the IV model have important implications, particularly, for designing simulation studies.

Monte Carlo simulations are used to reexamine the Nelson and Startz (1990a) findings regarding the performance of IV estimators when the instruments are weak. The results from the simulation exercises indicate that the sampling distribution of $\hat{\beta}_{IV}$ is concentrated around $\hat{\beta}_{OLS}$.

The second chapter considers the underlying joint distribution function of the instrumental variable (IV) model and presents an alternative definition for the exogenous and relevant instruments. The paper extracts a system of independent and orthogonal equations that covers up a non-orthogonal structural model and argues that the estimated IV regression is well-specified if the underlying system of equations is well-specified. It proposes a new instrument relevancy measure that does not suffer from the first-stage R^2 deficiencies.

Third chapter argues the application of the IV method in estimation of models with omitted variable. The paper considers the implicit parametrization of statistical models and presents five conditions for an appropriate instruments. Two of them are empirically measurable and can be tested. This improves the literature by adding one more objective criterion for the selection of instruments. This chapter applies the IV method to estimate the rate of return to education in Iran. It argues that the education of two cohorts of Iranians was delayed or cut short by the Cultural Revolution. Therefore, the Cultural Revolution, as an exogenous shock to the supply of education, establishes the year of birth as the exogenous and relevant instrument for education. Using the standard Mincerian earnings function with control for experience, ethnicity, location of residence and sector of employment, the instrumental variable estimate of the return to schooling is equal to 5.6%. The estimation results indicate that the Iranian labor market values degrees more than years of schooling.

Acknowledgments

I would like to express my sincere gratitude and appreciation to all those who have offered support and assistance in the completion of this dissertation . Foremost, I thank my parents for their prayers and support during my years at VT.

I also would like to acknowledge and thank the members of my committee: Djavad Salehi-Isfahani (chair), Richard Ashley, Dennis Yang, Geoffery Vining, and Bradford Mills. I am also indebted to many teachers in past, particularly S.A. Hashemi (Shahid-Beheshti University in Iran) and S.H. Zonnoor (Shiraz University in Iran).

I also extend my appreciation to the staff of the department of economics, Barbara Barker, Mike Cutlip, Jamie Lucero, and Sherry Williams for helping with all paperwork and details of my graduate program.

Contents

List of Figures

List of Tables

Chapter 1

The Statistical Parametrization of the Instrumental Variable Model:

More on the Small Sample Properties of the Instrumental Variable Estimators

1.1 Introduction

Instrumental Variable Method (IV) was initially proposed by Philips G Wright in 1928 to deal with the problems arising from using endogenous regressors (See Stock and Trebbi (2003)). The frequent use of endogenous regressors in economic models, brought the IV method in to the heart of econometric techniques. However, despite its increasing importance, the existing literature of the IV method is still deeply rooted in traditional econometrics where econometrics is seen as the empirical estimation of the models used in economic theory. The IV method can be summarized by using the following linear structural model

$$\begin{aligned} y_i &= \beta' X_i + \epsilon_i \qquad (1.1) \\ X_i &= \boldsymbol{B}' Z_i + \boldsymbol{u}_i \end{aligned}$$

$$\begin{bmatrix} \epsilon_i \\ \boldsymbol{u}_i \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon\epsilon} & \boldsymbol{\sigma}_{\epsilon u} \\ \boldsymbol{\sigma}_{u\epsilon} & \Sigma_{uu} \end{bmatrix} \right] \qquad i \in \mathtt{N}$$

$$\begin{aligned} &i) \frac{X'Z}{N} \xrightarrow{p} \Sigma_{23} \neq 0 && ii) \frac{Z'y}{N} \xrightarrow{p} \boldsymbol{\sigma}_{31} \neq 0 \\ &iii) \frac{Z'Z}{N} \xrightarrow{p} \Sigma_{33} > 0 && iv) E(Z'\boldsymbol{\epsilon}) = 0 \\ &v) E(Z'\boldsymbol{u}) = 0 \end{aligned}$$

where y_i refers to i^{th} observation of outcome variable of interest and X_i denote the $k \times 1$ vector of i^{th} observation of explanatory variables. Let Z_i denote the $p \times 1$ vector of i^{th} observation of instrumental variables ($p \geq k$) that, for theoretical reason, are not allowed to be used as explanatory variables for the outcome variable of interest, y_i. Also let $\boldsymbol{y} = (y_1, y_2, ..., y_N)'$, $X = (X_1, X_2, ..., X_N)'$, $Z = (Z_1, Z_2, ..., Z_N)'$, $\frac{\boldsymbol{y}'\boldsymbol{y}}{N} \xrightarrow{p} \sigma_{11}$, $\frac{X'\boldsymbol{y}}{N} \xrightarrow{p} \boldsymbol{\sigma}_{21}$, and $\frac{X'X}{N} \xrightarrow{p} \Sigma_{22}$. Clearly, if $\boldsymbol{\sigma}_{\epsilon u} = 0$ ($\Rightarrow E(X'\epsilon) = 0$), the Least Square method (LS) will provide the most efficient and unbiased estimators for the parameters of interest $(\beta, \sigma^2_\epsilon)$ using the sample counterparts : $\beta_{LS} = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$ and $\sigma^2_{LS} = \sigma_{11} - \boldsymbol{\sigma}_{12}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$. The instrumental variable (IV) method is the appropriate estimation method if $\boldsymbol{\sigma}_{\boldsymbol{\epsilon} u} \neq 0$ ($\Rightarrow E(X'\epsilon) \neq 0$), in which case it is well known that the least square estimators are biased and inconsistent. The key idea of the IV method is that, even though X and ϵ are correlated, Z–as characterized by 1.1[i] to 1.1[v]– can be used as an instrumental variable to estimate β and σ^2 consistently[1]. The consistent IV estimators of β and σ^2_{IV} are defined by

$$\hat{\beta}_{IV} = (X'P_ZX)^{-1}X'P_Z\boldsymbol{y}$$

where $P_Z = Z(Z'Z)^{-1}Z'$, and

$$\hat{\sigma}^2_{IV} = \frac{1}{N-k}(\boldsymbol{y} - X\hat{\beta}_{IV})'(\boldsymbol{y} - X\hat{\beta}_{IV})$$

Moreover, if $N^{-1/2}Z'\epsilon \xrightarrow{p} N(0, \sigma^2\Sigma_{33})$ then,

$$N^{1/2}(\hat{\beta}_{IV} - \beta) \sim N(0, \sigma^2(\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1})$$

(see Bowden and Turkington (1984) for more details)

Model (1.1) is by far the most widely used model in analysis of the instrumental variable method. It has been used for different purposes including: investigation of the finite sample properties of the IV estimators (see Blomquist and Dahlberg

[1]IV estimators can be used to estimate parameters of interest consistently even if $E(X'\epsilon) = 0$, the resulting estimators are, however, less efficient than the least square estimators.

(1999)), analysis of the power of instrument relevancy measures (see Flores-Lagunes (2000)), and to study the behavior of the IV estimators with weak instruments (see Staiger and Stock (1997) and Stock, Wright, and Yogo (2002) among others).

Model (1.1) assumes the process of data generation (DGP) consists of two independent parts: deterministic and stochastic. It treats the endogenous variables (y, X) as being "caused" by the un-modelled variables Z: Zs is taken to cause X, which, in turn, causes y. The error terms $(\epsilon, \boldsymbol{u})$ are the only sources of the stochastic behavior of the model that, additively, with deterministic part constitute the variable of interest $\boldsymbol{y}$. This notion is consistent with the experimental data where output y_i is "caused" by the inputs X_i.

$$y_i = f(X_i) + u_i$$

In this view, $f(X_i)$ generates the left-hand side such that repeating the experiment with the given X_i produces the same outputs except for some small experimental errors u_i. In a similar manner, in observational data, any observed variable can be decomposed into two components: explained $h(X_i)$ and error ϵ_i

$$y_i = h(X_i) + \epsilon_i$$

However, such a decomposition is possible even if y_i does not depend on $h(X_i)$ (see Hendry (2000)). Therefore, in contrast with the experimental data, the error term is, indeed, a by-product of the modeler's choice of X_i not an 'autonomous input' that, additively, with $h(X_i)$ constitute the observed output variable y_i. In this regard, the structural model (1.1) bears little resemblance to the actual DGP that gives rise to the economic observations (y_i, X_i, Z_i). In reality, however, model (1.1) is specified only to represent a theoretical view on how the observed variables (y_i, X_i, Z_i) are interrelated; it is highly unlikely to coincide with the actual DGP that gave rise to the observed variables.

Spanos (1986) advocates a modeling strategy where the actual DGP is more general than the theoretical model. This strategy considers the observed economic data $W_i = (y_i, X_i, Z_i)$ as a realization of a vector of stochastic processes $\{W_i, i \in \mathtt{N}\}$ and specifies the joint distribution of $\{W_i, i \in \mathtt{N}\}$ as the actual DGP of the observed data. The joint distribution of $\{W_i, i \in \mathtt{N}\}$ will be simplified by imposing probabilistic assumption from three broad categories: Distribution, Dependance and Heterogeneity to form statistical models. In this view, the theoretical model is a reparametrization/restriction of a well-defined statistical model intending to provide a consistent framework to understand some observable phenomenon of interest within the scope of the actual DGP.

In an attempt to illustrate the above procedure let $W_i = (y_i, X_i)$ and assume $\{W_i, i \in \mathtt{N}\}$ is an *iid* normal stochastic process with a distribution indicated by,

$$W_i \equiv \begin{pmatrix} y_i \\ X_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \Sigma_{22} \end{pmatrix}\right), i \in \mathtt{N}$$

The probabilistic assumptions of Normality, Independence and Identical Distribution for W_i help us to concentrate exclusively on the conditional process $\{(y_i|X_i),\ i \in \mathtt{N}\}$, which can be modeled by

$$y_i = \alpha + X_i'\beta + u_i,\ E(X_i'u_i) = 0$$

The model parameters $\varphi = (\alpha, \beta, \sigma^2)$ are defined in terms of the distribution parameters of $W_i = (y_i, X_i)$ by

$$\alpha = \mu_1 - \boldsymbol{\mu}_2'\beta \quad \beta = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21} \quad \sigma_u^2 = \sigma_{11} - \boldsymbol{\sigma}_{12}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$$

In practice, the model parameters φ could have direct economic interpretation, or some functions of them, $f(\varphi)$, coincide with the economic parameters (see Spanos (1986) for more details). A distinctive aspect of this approach is the definite relation between the estimable model and the distribution of the observable random variables. Moreover, the model parameters are interrelated through the variances and covariances of the model's random variables.

The goal of this paper is to provide a consistent statistical foundation for the structural model (1.1) in terms of the observable random variables (y, X, Z). In section (2), the joint distribution of (y_i, X_i, Z_i) will be considered as the actual DGP that gave rise to the observed variables. Section (2) will show that how the structural model (1.1) constitutes a reparameterized form of a statistical model defined on the joint distribution of (y_i, X_i, Z_i). This section uncovers the relation between the parameters of the distribution of $\begin{bmatrix}\epsilon_i & \boldsymbol{u}_i\end{bmatrix}'$ and the regression parameters β, $\boldsymbol{B}$ in (1.1). Section (3) applies the findings of the paper to reexamine the Shea (1997) and Nelson and Startz (1990a) results regarding weak instruments. Concluding remarks are provided in the last section of the paper.

1.2 The implicit statistical parameterization of the IV model

Consider $W_i = (y_i, X_i', Z_i')'$ as an *iid* vector of observable stochastic processes defined on the probability space $(S, \Im, P(.))$ where y_i is a scalar, X_i and Z_i are $k \times 1$ and $p \times 1$

$(p \geq k)$ vectors of observed values at $i = (1, ..., \mathtt{N})$. Assume W_i has a distribution of the form: $W_i \sim N(0, \Sigma), \det(\Sigma) > 0$, where

$$\begin{bmatrix} y_i \\ X_i \\ Z_i \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_{13} \\ \boldsymbol{\sigma}_{21} & \Sigma_{22} & \Sigma_{23} \\ \boldsymbol{\sigma}_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \right], i \in \mathtt{N} \tag{1.2}$$

Spanos (1986) showed that this distribution can be considered as the underlying distribution of the following multivariate regression model where y and X are treated as endogenous variables.

$$\begin{aligned} y_i &= \boldsymbol{\beta}_1' Z_i + u_{1i} \\ X_i &= \boldsymbol{B}' Z_i + \boldsymbol{u}_i \end{aligned} \tag{1.3}$$

$$\begin{bmatrix} u_{1i} \\ \boldsymbol{u}_i \end{bmatrix} \sim N(\mathbf{0}, \Omega), \Omega = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \Omega_{22} \end{bmatrix} \qquad i \in \mathtt{N}$$

where, by definition, $E(Z_i' u_{1i}) = 0$, $E(Z_i' \boldsymbol{u}_i) = 0$. The parameters of the statistical model (1.3) are related to the distribution primary parameters via

$$\begin{aligned} \beta_1 &= \Sigma_{33}^{-1} \boldsymbol{\sigma}_{31} & \boldsymbol{B} &= \Sigma_{33}^{-1} \Sigma_{32} \\ \omega_{11} &= \sigma_{11} - \boldsymbol{\sigma}_{13} \Sigma_{33}^{-1} \boldsymbol{\sigma}_{31}, & \Omega_{22} &= \Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32} \\ \boldsymbol{\omega}_{21} &= \boldsymbol{\sigma}_{21} - \Sigma_{23} \Sigma_{33}^{-1} \boldsymbol{\sigma}_{31} \end{aligned}$$

Assume $\beta_1 \neq 0$, $\boldsymbol{B} \neq 0$ and multiply (1.3) by a non-singular matrix $M_1 = \begin{bmatrix} 1 & -\boldsymbol{\omega}_{12}\Omega_{22}^{-1} \\ 0 & I_k \end{bmatrix}$ to reparametrize[2] the first equation of (1.3) as

$$\begin{aligned} y_i &= \beta_0' X_i + \gamma_0' Z_i + u_{0i} \\ X_i &= \boldsymbol{B}' Z_i + \boldsymbol{u}_i \end{aligned} \tag{1.4}$$

$$\begin{bmatrix} u_{0i} \\ \boldsymbol{u}_i \end{bmatrix} \sim N(\mathbf{0}, \Psi), \Psi = \begin{bmatrix} v_{11} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} \end{bmatrix} \qquad i \in \mathtt{N}$$

where

$$\beta_0 = \Omega_{22}^{-1}\boldsymbol{\omega}_{21} \qquad \gamma_0 = \beta_1 - \boldsymbol{B}\Omega_{22}^{-1}\boldsymbol{\omega}_{21}$$
$$v_{11} = \omega_{11} - \boldsymbol{\omega}_{12}\Omega_{22}^{-1}\boldsymbol{\omega}_{21}$$

and $E(X_i'u_{0i}) = E(Z_i'u_{0i}) = 0$. The difference between the statistical model (1.4) and the structural model (1.1) is in their first equations where $\beta_0 \neq \beta$, $\gamma_0 \neq 0$ and, therefore, $u_{0i} \neq \epsilon$. In order to accommodate the structural model (1.1), the statistical model (1.4) must be *reparameterized* such that the resulting statistical model preserves all properties of the structural model (1.1), including $E(X'\epsilon) \neq 0$, $E(Z'\epsilon) = 0$, and the absence of Z from the first equation.

Theorem 1. *In the context of the structural model (1.1), the conditional covariance of Z and $\boldsymbol{y}$ given X is not zero.*

Proof: Let $\boldsymbol{\sigma}_{31.2} = \boldsymbol{\sigma}_{31} - \Sigma_{32}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$ be the conditional covariance of Z and $\boldsymbol{y}$ given X. Considering (1.1[i]-1.1[ii]), $\boldsymbol{\sigma}_{31.2}$ can be zero only if $\boldsymbol{\sigma}_{31} = \Sigma_{32}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$ that is

[2] See Spanos (1986) ch 24-25.

in contradiction with the assumptions of $E(X'\epsilon) \neq 0$, $E(Z'\epsilon) = 0$. The contradiction comes from the facts that

$$E(X'\epsilon) = E(X'[y - X\beta]) = E(X'y) - E(X'X\beta) = 0 \Leftrightarrow \beta = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$$

$$E(Z'\epsilon) = E(Z'[y - X\beta]) = E(Z'y) - E(Z'X\beta) = 0 \Leftrightarrow \beta = (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31}$$

and the fact that if $\boldsymbol{\sigma}_{31} = \Sigma_{32}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$ then

$$\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21} = (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31}$$

Q.E.D

Theorem (1) says that one can not drop Z_i from the first equation of the statistical model (1.4) simply by imposing the restriction $\gamma_0 = 0$ [3], because

$$\gamma_0 = \beta_1 - \boldsymbol{B}\Omega_{22}^{-1}\boldsymbol{\omega}_{21} = (\Sigma_{33} - \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23})^{-1}(\boldsymbol{\sigma}_{31} - \Sigma_{32}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}) \neq 0$$

Theorem (2) will provide the necessary reparameterization in the statistical model (1.4) to make it similar to the structural model (1.1).

Theorem 2. *The structural parameters* $(\beta_{IV}, \sigma^2_{IV})$ *are related to the parameters of the statistical model (1.4) via:*
i)$\beta_{IV} = (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31} = \beta_0 + \boldsymbol{B}^{-}\gamma_0$
ii) If $E(u_{0i}^2) = v_{11}I_N$ *and* $E(\boldsymbol{u}_i'\boldsymbol{u}_i) = \Omega_{22} \otimes I_N$ *then* $\sigma^2_{IV} = v_{11} + (\beta_{IV} - \beta_0)'\Omega_{22}(\beta_{IV} - \beta_0)$

Proof: let

$$\begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} = \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\sigma}_{21} \\ \boldsymbol{\sigma}_{31} \end{bmatrix}$$

[3]See Spanos (1986) pp: 637-644 and Spanos (2000).

and

$$\begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}^{-1} = \begin{bmatrix} D & -D\Sigma_{23}\Sigma_{33}^{-1} \\ -\Sigma_{33}^{-1}\Sigma_{32}D & \Sigma_{33}^{-1} - \Sigma_{33}^{-1}\Sigma_{32}D\Sigma_{23}\Sigma_{33}^{-1} \end{bmatrix}$$

where $D = (\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}$ then

$$\beta_0 + (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\gamma_0 = D\boldsymbol{\sigma}_{21} - D\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31} - (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}$$
$$D\boldsymbol{\sigma}_{21} + (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31} + (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}D\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31} =$$
$$(\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31} = \beta_{IV}$$

Using $\beta_0 = \beta_{IV} - \boldsymbol{B}^{-}\gamma_0$ from (i), the error term of the first equation of (1.4) will change to $\epsilon_{IV} = u_{0i} - \boldsymbol{u}_i\boldsymbol{B}^{-}\gamma_0$ with variance equal to $\sigma_{IV}^2 = v_{11} + (\beta_{IV} - \beta_0)'\Omega_{22}(\beta_{IV} - \beta_0)$.

Q.E.D

By Theorem (2), one can show that the structural model (1.1) constitutes a reparameterization of the statistical model (1.4)

$$\begin{aligned} y_i &= \beta_{IV}'X_i + \epsilon_i \\ X_i &= \boldsymbol{B}'Z_i + \boldsymbol{u}_i \end{aligned} \tag{1.5}$$

where $\epsilon_i = u_{0i} - \gamma_0'\boldsymbol{B}'^{-}\boldsymbol{u}_i$ and

$$\begin{bmatrix} \epsilon_i \\ \boldsymbol{u}_i \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} v_{11} + \gamma_0'\boldsymbol{B}'^{-}\Omega_{22}\boldsymbol{B}^{-}\gamma_0 & -\gamma_0'\boldsymbol{B}'^{-}\Omega_{22} \\ -\Omega_{22}\boldsymbol{B}^{-}\gamma_0 & \Omega_{22} \end{bmatrix}\right] \quad i \in \mathbb{N}$$

$\beta_{IV} = \boldsymbol{B}^{-}\beta_1$, $\boldsymbol{B}^{-} = (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}$, $\sigma_{\epsilon\epsilon} = v_{11} + \gamma_0'\boldsymbol{B}'^{-}\Omega_{22}\boldsymbol{B}^{-}\gamma_0$, $\sigma_{\epsilon\boldsymbol{u}} = \gamma_0'\boldsymbol{B}'^{-}\Omega_{22}$, and $\Sigma_{\boldsymbol{uu}} = \Omega_{22}$. To reach the structural/statistical model (1.5) one must multiply

(1.4) by a non-singular matrix $M_2 = \begin{bmatrix} 1 & -\gamma_0'\boldsymbol{B}'^{-} \\ 0 & I_k \end{bmatrix}$ or multiply (1.3) by a non-singular matrix $M_{IV} = M_1 * M_2 = \begin{bmatrix} 1 & -\beta_1'\boldsymbol{B}'^{-} \\ 0 & I_k \end{bmatrix}$. The structural/statistical model (1.5) preserves all properties of the structural model (1.1)

$$\begin{aligned} E(X_i'\epsilon_i) &= E(X_i'[u_{i0} - \gamma_0'\boldsymbol{B}'^{-}\boldsymbol{u}_i]) \neq 0 \\ E(Z_i'\epsilon_i) &= E(Z_i'[u_{i0} - \gamma_0'\boldsymbol{B}'^{-}\boldsymbol{u}_i]) = 0 \end{aligned}$$

$$\begin{aligned} \beta_1 \neq 0 &\Rightarrow \sigma_{31} \neq 0 \\ \boldsymbol{B} \neq 0 &\Rightarrow \Sigma_{32} \neq 0 \\ det(\Sigma) > 0 &\Rightarrow \Sigma_{33} > 0 \end{aligned}$$

and provides a sound statistical foundation with direct connection with the distribution of the observed variables. The parametrization of model (1.5) should come as no surprise because under conditions 1[i]-1[iii], the IV estimators converge in probability to β_{IV} and σ_{IV}, i.e.

$$\hat{\beta}_{IV} = (X'P_Z X)^{-1} X'P_Z \xrightarrow{P} (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}\Sigma_{23}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31}$$

$$\hat{\sigma}_{IV}^2 = \frac{1}{N-k}(\boldsymbol{y} - X\hat{\beta}_{IV})'(\boldsymbol{y} - X\hat{\beta}_{IV}) \xrightarrow{P} v_{11} + \gamma_0'\boldsymbol{B}'^{-}\Omega_{22}\boldsymbol{B}^{-}\gamma_0$$

Numerical Example: Consider $w_i = (y_i, x_i, z_i)'$, which is a normally distributed vector of random variables with mean zero and covariance matrix Σ, where

$$\Sigma = \begin{bmatrix} 4.61 & 2.71 & 1 \\ 2.71 & 1.81 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Assume y and x are endogenous variables. The goal is to estimate the causal effect of x on y. According to (1.3), the joint distribution of w_i implies the following statistical model, in which, y and x are treated as endogenous variables.

$$\begin{aligned} y_i &= 0.5z_i + u_{1i} \\ X_i &= 0.5z_i + u_i \end{aligned}$$

$$\begin{bmatrix} u_{1i} \\ u_i \end{bmatrix} \sim N(0, \Omega), \Omega = \begin{bmatrix} 4.11 & 2.21 \\ 2.21 & 1.31 \end{bmatrix} \qquad i \in \mathtt{N}$$

In order to get the endogenous variable x into the first equation, purporting to explain y, the statistical model can be reparameterized to

$$\begin{aligned} y_i &= 1.687x_i - 0.343z_i + u_{0i} \\ x_i &= 0.5z_i + u_i \end{aligned}$$

$$\begin{bmatrix} u_{0i} \\ u_i \end{bmatrix} \sim N(0, \Psi), \Psi = \begin{bmatrix} 0.381 & 0 \\ 0 & 1.31 \end{bmatrix} \qquad i \in \mathtt{N}$$

This model can be reparameterized more to get the structural model, where x is the only determinant of y

$$\begin{aligned} y_i &= x_i + \epsilon_i \\ x_i &= 0.5z_i + u_i \end{aligned}$$

$$\begin{bmatrix} \epsilon_i \\ u_i \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1.31 \end{bmatrix} \right] \qquad i \in \mathtt{N}$$

To investigate the consistency of the IV estimators, a thousand samples of 100 observations of $w_i = (y_i, x_i, z_i)'$ were generated by the joint distribution of w_i, and, for each sample $y_i = \beta x_i + \epsilon_i$ was estimated by the IV method using z_i as instrument. The empirical distribution of the IV estimators are presented[4] in table 1. Table 1.1 shows that the empirical distributions of $\hat{\beta}_{IV}$ and $\hat{\sigma}^2_{IV}$ are concentrated on population values of β and σ^2. It is also clear that the concentration point of $\hat{\beta}_{OLS}$ is far away from the true value of β.

$\square$

Model (1.5) reveals some hidden aspects of the structural model (1.1). It specifies that, in general, variances of the error terms, $(\epsilon_i, \boldsymbol{u}_i)$, are not equal. The conventional wisdom in econometrics studies is to normalize σ^2_{IV} and Ω_{22} to the same values (see, for example, Hahn and Hausman (2003)). However, the covariance matrix of model (1.5) shows, explicitly, that the elements of the covariance matrix of ϵ_i and $\boldsymbol{u}_i$ are interrelated. Therefore, normalization without considering these interrelations could impose unwanted restrictions on the data generating process. In fact, in some cases, this restriction could be highly unrealistic. For example, appendix 3 finds out an implicit restriction in the Maddala and Jeong (1992)'s DGP, which is inconsistent

[4]The program was written in STATA . It is available in appendix 1.

with their goal to argufy the Nelson and Startz (1990a) findings.

The other common practice in the IV literature is to investigate the small sample properties of the IV estimators or instrument relevancy measures by Monte-Carlo simulations. For the Monte-Carlo simulations of the structural model (1.1), one has to assign some values to the unknown parameters $(\beta, \boldsymbol{B}, \sigma_{\epsilon\epsilon}, \sigma_{\epsilon\boldsymbol{u}}, \Sigma_{\boldsymbol{uu}})$ as population parameters. Model (1.5) displays the implicit relationship between the population parameters, which are, usually, ignored in simulation studies. For instance, keeping $\boldsymbol{B}$ constant while increasing $\sigma_{\epsilon\boldsymbol{u}}$ or leaving $\sigma_{\epsilon\boldsymbol{u}}$ unchanged while reducing $\boldsymbol{B}$ to investigate weak instrument models are unjustified practices. As an example, the population r-square between x and z in the above numerical example is

$$R^2_{xz} = 1 - \frac{\sigma_{xx} - \sigma_{xz}\sigma_{zz}^{-1}\sigma_{zx}}{\sigma_{xx}} = 0.276$$

To have a DGP with lower population R^2_{xz}, say 0.10, the appropriate covariance matrix for $w = (y, x, z)'$ would be

$$\Sigma^* = \begin{bmatrix} 5 & 3 & 0.5 \\ 3 & 2 & 0.5 \\ 0.5 & 0.5 & 1.25 \end{bmatrix}$$

that is consistent with

$$\begin{aligned} y_i &= x_i + \epsilon_i \\ x_i &= 0.4z_i + u_i \end{aligned}$$

$$\begin{bmatrix} \epsilon_i \\ u_i \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1.8 \end{bmatrix} \right] \qquad i \in \mathtt{N}$$

as the structural model. Note that to the simultaneous changes in $\boldsymbol{B}$ and the covariance matrix of the error terms.

1.3 Weak Instruments and statistical parametrization of the IV model

This section presents two example from the IV literature to demonstrate the importance of the implicit relationships between parameters of the IV model in properly designing simulation studies.

1- Shea (1997) relevancy measure: Shea (1997) has proposed a measure for the relevance of instruments that considers the possible correlation among the instruments. Godfrey (1999) in a clarification note on Shea's paper has demonstrated that Shea's statistic may be expressed as partial-$R_i^2 = \frac{(X'X)_{ii}^{-1}}{(X'P_ZX)_{ii}^{-1}}$, where $P_Z = Z(Z'Z)^{-1}Z'$ and $(X'X)_{ii}^{-1}$ and $(X'P_ZX)_{ii}^{-1}$ refer to the ith element of $(X'X)^{-1}$ and $(X'P_ZX)^{-1}$, respectively. Shea (1997) has used the following data generating process to investigate the finite-sample behavior of IV in a multivariate model

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \lambda u_1 + (1-\lambda) u_2 \qquad (1.6) \\ x_1 &= \gamma u_1 + (1-\gamma) e_1 \\ x_2 &= \gamma u_2 + (1-\gamma) e_2 \\ z_1 &= \delta e_1 + (1-\delta) e_2 + \phi v_1 \\ z_2 &= (1-\delta) e_1 + \delta e_2 + \phi v_2 \end{aligned}$$

where u_1, u_2, e_1, e_2, v_1 and v_2 are unobserved disturbances, standard normal and jointly orthogonal. The parameter δ is designed to govern correlation among the instruments. According to Shea (1997), partial-R_i^2 approaches zero and IV estimators behave badly as δ moves toward 0.5. Considering the joint distribution of observable variables $(y,\ x_1,\ x_2,\ ,z_1,\ z_2)$, the population value of partial-R_i^2 would be

$$\frac{\gamma^2 - 2\gamma + 1}{2\gamma^2 - 2\gamma + \phi^2 - 2\gamma\phi^2 + 2\gamma^2\phi^2 + 1} \quad i = 1, 2$$

Since this is not a function of δ, Shea's partial-R^2 is, in fact, unrelated to the correlation between the instruments.

Table 1.2 presents a series of simulation experiments conducted to study the finite-sample behavior of Shea's statistic in a multivariate model. In each experiment a thousand samples of 100 observations were generated, where β_1 and β_2 were set equal to zero, λ equal to 0.9, ϕ equal to 0.1, γ equal to 0.3, and δ varies across experiments. The first column reproduces the first block of table(1) in Shea (1997)[5]. The sample of this column were generated by model (1.6). As Shea indicates partial-R^2 declines much more faster than standard-R^2 as δ approaches 0.5. The second column uses the joint distribution of the model's random variables to generate samples. Appendix 3 reports the joint distribution of (y, x_1, x_2, z_1, z_2), implied by (1.6), and the computer programs that generate table 2. As expected, both standard-R^2 and partial-R^2 are insensitive to variations in δ. Although this argument has nothing to do with Shea's relevancy measure, it argues that the chosen DGP is not qualified to study the finite-sample behavior of partial-R^2.

2- Weak Instruments: Nelson and Startz(1990a,b), examined the distribution of the IV estimators $(\hat{\beta}_{IV},\ \hat{\sigma}^2_{IV})$ in the weak instrument situation. They used the

[5]The difference is only in number of replications; that is 10000 in Shea's study.

following model as data generating mechanism

$$\begin{aligned} y &= \beta x + u \qquad (1.7)\\ x &= \epsilon + \lambda u \\ z &= \theta\epsilon + v \end{aligned}$$

where y is a $N \times 1$ vector of observations on the dependent variable, x is a $N \times 1$ vector of observed values of explanatory variable, and z is a $N \times 1$ vector of observations on an instrumental variable. ϵ, v and u all are distributed as independent drawings from a multivariate normal distribution with mean $\mathbf{0}$ and identity covariance matrix. Their analysis show substantial differences between the exact and asymptotic distributions of the IV estimators as the square correlation between x and z decreases, i.e. as $R^2_{xz} \to 0$:

1 The IV estimator $\hat{\beta}_{IV}$ concentrates around $\beta + \lambda^{-1}$, which is greater than the plim of the least square estimator, $\beta + \frac{\lambda}{1+\lambda^2}$. The ratio of the IV to LS estimator biases falls as degree of endogeneity λ rises.

2 The IV estimator $\hat{\sigma}_{IV}$ concentrates around λ^{-2}.

3 The "t" statistic based on asymptotic standard errors concentrates around a value greater than 1. The concentration point of "t" statistic increases as endogeneity degree rises.

4 The exact distribution of the IV estimator $\hat{\beta}_{IV}$ is bimodal[6].

[6] Maddala and Jeong (1992) argue that the bimodality of $\hat{\beta}_{IV}$ is a result of setting $\lambda = 1$ in model (2.4). Appendix 3 reviews their work and shows that their model is incomparable with model (2.4).

Table 1.3 reviews the Nelson and Startz's analysis by simulating model (2.4). Two separate simulation for $\lambda = 1$ and $\lambda = 10$ with samples of 100 observations, $\beta = 0$, and $\theta = 0.001$ has been performed. The simulation results are consistent with Nelson and Startz's findings. Figure 1-1 depicts the empirical distribution of $\hat{\beta}_{IV}$, which is a bimodal distribution. Do these results hold if data would be generated by the joint distribution of the observed random variables?

Below, a series of simulation experiments are described, which reexamine Nelson and Startz's findings. To be consistent with Nelson and Startz, the data generating process is specified as the conditional distribution of y and x given z

$$\begin{bmatrix} y_i & \\ & \mid \quad z_i \\ x_i & \end{bmatrix} \sim N(\mu, \Omega) \tag{1.8}$$

$$\mu = \begin{bmatrix} \beta_1 z_i \\ B_2 z_i \end{bmatrix} \quad \Omega = \begin{bmatrix} \beta^2 + (1+\beta\lambda)^2 - \frac{\beta^2\theta^2}{\theta^2+1} & \beta + (\lambda + \beta\lambda^2) - \frac{\beta\theta^2}{\theta^2+1} \\ \beta + (\lambda + \beta\lambda^2) - \frac{\beta\theta^2}{\theta^2+1} & 1 + \lambda^2 - \frac{\theta^2}{\theta^2+1} \end{bmatrix}$$

where $\beta_1 = \frac{Cov(y,z)}{Var(z)} = \frac{\beta\theta}{\theta^2+1}$, $B_2 = \frac{Cov(x,z)}{Var(z)} = \frac{\theta}{\theta^2+1}$. Like Nelson and Startz z will be drawn once and remains fixed as we sample from the distribution of y and x given z [7]. Each experiment consists of 10,000 trials. In all cases `N` is equal to 100, and β equal to zero; θ and λ vary across experiments.

Table 1.4 reports the empirical distributions of $\hat{\beta}_{IV}$, $\hat{\sigma}^2$ and t-statistic for the test of $H_0 : \beta = 0$. The following results can be derived from the table 4:

1 Reading down the blocks of table 1.4 one can find that decreases in θ moves the center of the empirical distribution of $\hat{\beta}_{IV}$ toward the plim of the least square estimator $\frac{\lambda}{1+\lambda^2}$ (see figure 1-2). For example, in $\lambda = 1$, the mean and median

[7] Staiger and Stock (1997) showed that many results that hold in the fixed-instrument model can be reinterpreted asymptotically in a random-instrument model.

of the distribution of $\hat{\beta}_{IV}$ approach to the plim of the least square estimator 0.5 as θ falls. The bias of $\hat{\beta}_{IV}$, on average, is always less than the asymptotic bias of $\hat{\beta}_{LS}$ even in an extremely low value for θ. This result is consistent with Woglom (2001) and Bound, Jaeger, and Baker (1995), that is, with very poor instrument IV bias would be close to the bias of least square and it never concentrates on a value more biased than the least square estimator.

2 The distribution of the IV estimator $\hat{\beta}_{IV}$ is less dispersed around the plim of the least square estimator in models with high degree of endogeneity λ (see figure 1-3).

3 The central tendency of the empirical distribution of $\hat{\sigma}^2_{IV}$ approaches $\frac{2}{1+\lambda^2}$ as instrument relevance θ falls (see figure 1-4). An increase in the degree of endogeneity λ moves the central tendency of the distribution of $\hat{\sigma}^2_{IV}$ toward zero.

4 The third row of each block shows the empirical distribution of the t-ratio $\frac{\hat{\beta}_{IV}-\beta}{\hat{\sigma}_{\beta_{IV}}}$. As column (1)-(6) indicate the median(mean) of the empirical distribution of $t_{\hat{\beta}_{IV}}$ increases as θ falls, but for moderate degrees of endogeneity it is still less than one. The central tendency of the empirical distribution of $t_{\hat{\beta}_{IV}}$ increases as degree of endogeneity λ rises, however, in models with large λ the empirical distribution of $t_{\hat{\beta}_{IV}}$ is far away from the asymptotic student's t_{N-1} distribution (see figure 1-5).

5 Figure 1-6 draws the empirical distribution of $\hat{\beta}_{IV}$ for a case where $\lambda = 1$ and $\theta = 0.001$. Clearly, it does not exhibit bimodality. As $Cov(x, z)$ decreases, the distribution of $\hat{\beta}_{IV}$ will be much more similar to the t-distribution than its asymptotic distribution (see Han and Schmidt (2001) for unidentified case $\theta = 0$). However, for large values of λ, the distribution of $\hat{\beta}_{IV}$ is bimodal.

As table 1.4 shows the consequences of high degree of endogeneity λ for IV inference is more severe than low relevance θ of instruments. Low relevance misleads statistical inferences only when it is combined with high endogeneity. In this regard, the square correlation coefficient of x and z

$$R_{xz}^2 = \frac{\theta^2}{(1+\lambda^2)(1+\theta^2)} \tag{1.9}$$

as a combinational measure of degree of endogeneity λ and instrument relevance θ has a limited applicability in applied studies. The limitation of R_{xz}^2 comes from the fact that the upper bound of $R_{xz}^2 \in [0, \frac{1}{1+\lambda^2})$ is determined by the degree of endogeneity λ, which is unknown in practice. For example, assume R_{xz}^2 of x and z, generated by model (1.8), is equal to 0.019. This R_{xz}^2 is consistent with two set(of many) of values for parameters of (1.8): a) $\lambda = 0.1,\ \theta = 0.141$, and b)$\lambda = 5,\ \theta = 1$. In case (b) – first block of fifth column of table 4– $R_{xz}^2 = 0.019$ indicates that z is a strong instrument, however, in case (a) it signals that z is a weak instrument. In summary, if a practitioner believes that the degree of endogeneity of x is high she should not expect to find a valid instrument $E(z'\epsilon) = 0$ with high correlation with x. Nelson and Startz (1990a) argue that if R_{xz}^2 is less than $\frac{1}{N}$ the practitioners must be worried about spurious statistical inference. Test of hypothesis $H_0 : \boldsymbol{B} = 0$ vs. $H_0 : \boldsymbol{B} \neq 0$ is another recommended way to identify weak instruments. In the context of the Nelson and Startz's model this test is equivalent to testing $H_0 : \theta = 0$ vs. $H_0 : \theta \neq 0$, which considers the effect of high degree of endogeneity only through the variance of $\boldsymbol{B}$. Considering theorem (1), next chapter shows that the problem of weak instrument can be assessed by testing the hypothesis $H_0 : \gamma_0 = 0$ vs. $H_0 : \gamma_0 \neq 0$. This hypothesis can be tested by likelihood ratio test or some variation thereof. If the null hypothesis is not rejected then z is a weak instrument and therefore IV does not have any advantage over least square. In the context of the Nelson and Startz's model, γ_0 is equivalent to $\frac{-\theta\lambda}{1+\lambda^2+\lambda^2\theta^2} \neq 0$.

1.4 Conclusion

The existing instrumental variable models are formed in terms of the theoretical consideration. The aim of this chapter was to provide a statistical background for the IV models where the actual data generation process (DGP) is more general than the theoretical model. The chapter showed that the IV model is embedded in a well-specified statistical model defined on the joint distribution of the involved random variables. It also showed that the theoretical parameters of interest are functions of the parameters of the statistical model, which, in turn, are functions of the parameters of the actual DGP.

Monte Carlo simulation exercises were used to reexamine the Nelson and Startz (1990a) findings regarding the small sample performance of the IV estimators when the instruments are weak. Simulation results indicate that IV estimators are bias if instruments are weak, however, the biases are not more than OLS estimators. In addition, the empirical distribution of $\hat{\beta}_{IV}$ does not show bimodality. In fact, it is more similar to Student's t-distribution.

Table 1.1: Estimated Percentiles for OLS and IV estimators with n=100
(Based on 1000 Replications)

Percentiles	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widehat{\sigma}^2_{IV}$	$t_{\widehat{\beta}_{IV}}$
1	1.36	0.5	0.52	1.47
10	1.43	0.8	0.68	3.77
50	1.50	1.00	0.99	7.16
90	1.57	1.16	1.50	11.13
99	1.62	1.30	2.47	15.48
Mean	1.5	0.99	1.06	7.38

Table 1.2: Estimated Standard and Shea's Partial R-square
(Based on 1000 Replications)

		Data is generated by the model		Data is generated by the joint distribution	
δ	γ	R^2_*	R^2_p	R^2_*	R^2_p
1	0.3	0.84	0.84	0.84	0.83
0.53	0.3	0.54	0.36	0.84	0.83
0.52	0.3	0.48	0.21	0.84	0.83
0.51	0.3	0.44	0.08	0.84	0.83
0.50	0.3	0.42	0.02	*	*

R^2_*: Standard R-square

R^2_p: Shea's partial R-square

*: The covariance matrix is not a positive definite matrix.

Table 1.3: Estimated Percentiles for IV estimators with γ=0.001, n=100
(Duplication of Nelson & Startz results)

Percentiles	λ=1			λ=10		
	$\widehat{\beta}$	$\widehat{\sigma}^2$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$\widehat{\sigma}^2$	$t_{\widehat{\beta}}$
1	-10.3	0.39	-0.04	0.09	0.008	0.11
10	0.17	0.50	0.02	0.10	0.009	1.13
50	0.89	0.97	0.71	0.10	0.009	7.40
90	1.58	4.97	1.87	0.10	0.010	16.80
99	7.78	1404	3.18	0.11	0.030	25.87
Mean	-1.67	10883	0.85	0.10	0.011	8.31

Table 1.4: Estimated percentiles for $\widehat{\beta}_{IV}$ based on 10,000 replications with n=100 and varying λ and θ .

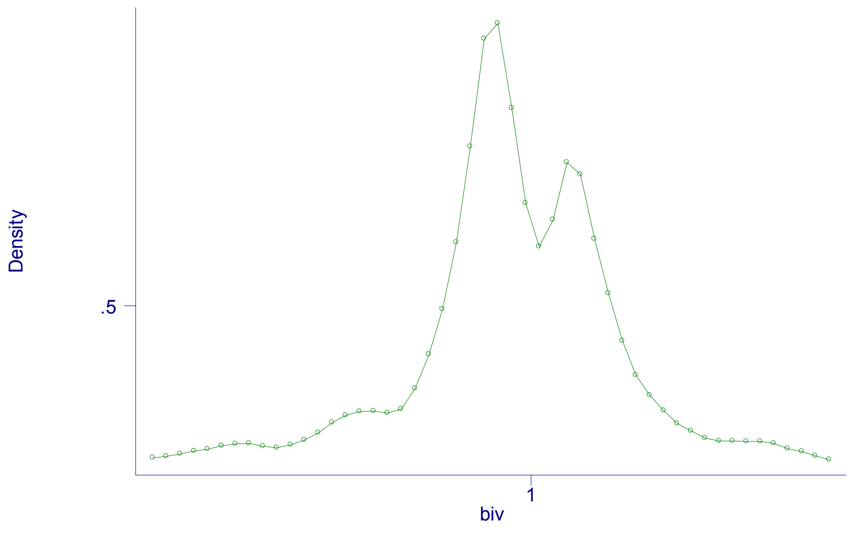
λ		0.1						0.5						0.75					
θ	Percentiles	1	10	50	90	99	Mean	1	10	50	90	99	Mean	1	10	50	90	99	Mean
1.00	$\widehat{\beta}$	-0.28	-0.16	0.00	0.16	0.27	0.00	-0.32	-0.17	0.00	0.15	0.24	0.00	-0.34	-0.18	0.00	0.15	0.23	-0.01
	$\widehat{\sigma}^2$	0.73	0.83	1.00	1.20	1.34	1.00	0.67	0.79	0.99	1.28	1.53	1.02	0.62	0.75	0.99	1.38	1.75	1.03
	$t_{\widehat{\beta}}$	-1.99	-1.21	0.00	1.26	2.07	0.00	-1.76	-1.13	0.00	1.34	2.30	0.05	-1.62	-1.07	0.00	1.40	2.43	0.08
0.05	$\widehat{\beta}$	-14.1	-2.44	0.06	2.43	15.3	0.05	-11.7	-1.77	0.32	2.45	11.98	0.32	-9.80	-1.35	0.41	2.16	9.77	0.37
	$\widehat{\sigma}^2$	0.82	0.98	1.78	26.1	831.9	32.49	0.66	0.80	1.47	24.87	770.8	30.84	0.53	0.64	1.24	20.59	558.9	23.67
	$t_{\widehat{\beta}}$	-0.99	-0.58	0.03	0.75	1.23	0.06	-0.66	-0.36	0.19	1.13	1.81	0.30	-0.52	-0.26	0.29	1.39	2.24	0.44
0.01	$\widehat{\beta}$	-14.9	-2.74	0.09	2.93	16.32	0.15	-12.7	-1.75	0.40	2.76	13.35	0.45	-9.09	-1.29	0.48	2.30	10.20	0.48
	$\widehat{\sigma}^2$	0.82	0.99	1.98	36.04	854.3	36.69	0.67	0.80	1.60	29.37	837.5	33.21	0.54	0.65	1.29	20.56	596.2	24.92
	$t_{\widehat{\beta}}$	-0.94	-0.54	0.03	0.71	1.17	0.06	-0.60	-0.33	0.20	1.08	1.76	0.30	-0.48	-0.24	0.32	1.37	2.22	0.46
0.001	$\widehat{\beta}$	-15.5	-2.72	0.09	2.95	17.11	0.16	-12.4	-1.77	0.40	2.68	12.9	0.44	-9.75	-1.31	0.48	2.31	10.34	0.47
	$\widehat{\sigma}^2$	0.83	0.99	1.97	33.15	1073	40.46	0.67	0.80	1.61	29.33	777.8	33.09	0.53	0.65	1.28	20.50	507.9	22.47
	$t_{\widehat{\beta}}$	-0.94	-0.54	0.03	0.70	1.17	0.06	-0.60	-0.32	0.21	1.08	1.76	0.30	-0.47	-0.23	0.32	1.36	2.22	0.46

Table 1.4: Continued

θ	λ / Percentiles	1: 1	1: 10	1: 50	1: 90	1: 99	1: Mean	5: 1	5: 10	5: 50	5: 90	5: 99	5: Mean	10: 1	10: 10	10: 50	10: 90	10: 99	10: Mean
1.00	$\widehat{\beta}$	-0.38	-0.19	0.00	0.14	0.22	-0.01	-1.78	-0.33	0.02	0.11	1.97	-0.02	-0.97	-0.13	0.04	0.29	1.30	0.06
	$\widehat{\sigma}^2$	0.57	0.70	1.00	1.49	2.08	1.05	0.17	0.30	1.00	13.12	406.2	15.56	0.07	0.14	0.73	17.28	484.1	19.7
	$t_{\widehat{\beta}}$	-1.48	-1.02	0.00	1.44	2.57	0.11	-0.41	-0.35	0.14	2.30	4.77	0.60	-0.20	-0.16	0.50	3.70	7.86	1.22
0.05	$\widehat{\beta}$	-7.82	-1.03	0.43	1.86	8.86	0.45	-0.46	0.06	0.19	0.31	0.91	0.19	-0.10	0.07	0.098	0.13	0.28	0.097
	$\widehat{\sigma}^2$	0.42	0.50	1.00	17.61	537.0	21.54	0.03	0.04	0.09	1.66	51.65	1.98	0.008	0.10	0.02	0.44	15.73	0.62
	$t_{\widehat{\beta}}$	-0.41	-0.19	041	1.69	2.71	0.59	-0.06	0.05	2.18	7.22	11.40	2.97	-0.01	0.19	4.40	14.28	22.51	5.94
0.01	$\widehat{\beta}$	-7.49	-0.92	0.49	1.81	8.03	0.47	-0.43	0.07	0.19	0.30	0.76	0.19	-0.07	0.07	0.099	0.13	0.26	0.098
	$\widehat{\sigma}^2$	0.42	0.50	1.01	16.03	414.8	17.95	0.03	0.04	0.08	1.38	36.66	1.47	0.008	0.009	0.02	0.36	11.58	0.48
	$t_{\widehat{\beta}}$	-0.38	-0.17	0.44	1.68	2.68	0.61	-0.05	0.07	2.30	7.35	11.43	3.06	-0.01	0.23	4.59	14.59	22.68	6.13
0.001	$\widehat{\beta}$	-7.70	-0.90	0.49	1.85	7.99	0.46	-0.44	0.07	0.192	0.30	0.79	0.19	-0.07	0.07	0.099	0.13	0.26	0.098
	$\widehat{\sigma}^2$	0.42	0.50	1.00	16.74	490.9	19.49	0.03	0.04	0.08	1.40	39.07	1.59	0.008	0.009	0.02	0.36	10.66	0.42
	$t_{\widehat{\beta}}$	-0.37	-0.17	0.43	1.67	2.69	0.61	-0.05	0.07	2.30	7.34	11.42	3.06	-0.01	0.23	4.61	14.68	22.68	6.13

Figure 1-1: Empirical Distribution of $\widehat{\beta}_{IV}$ with γ=0.001, n=100
(Duplication of Nelson & Startz results)

a) λ=1



b) λ=10

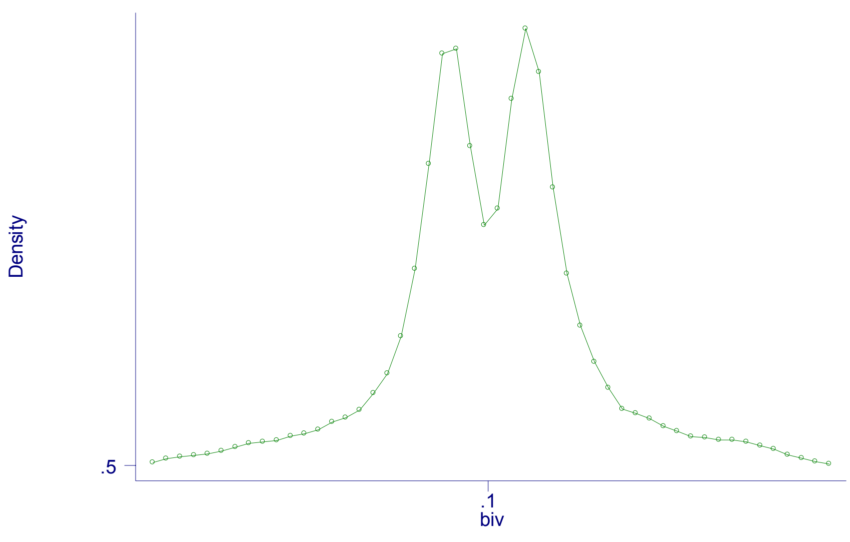
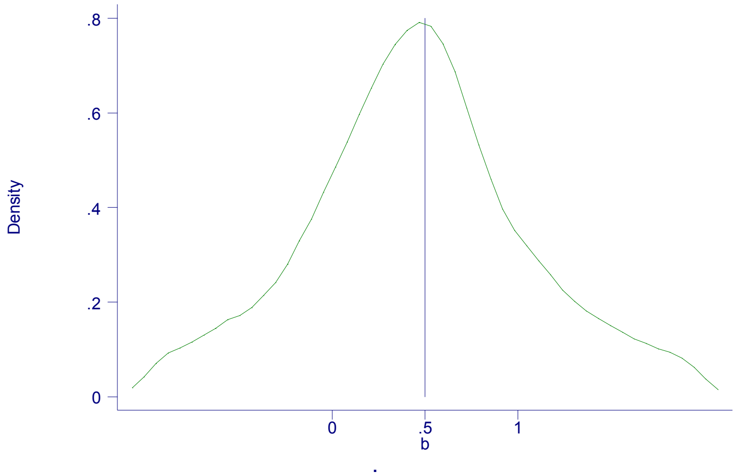
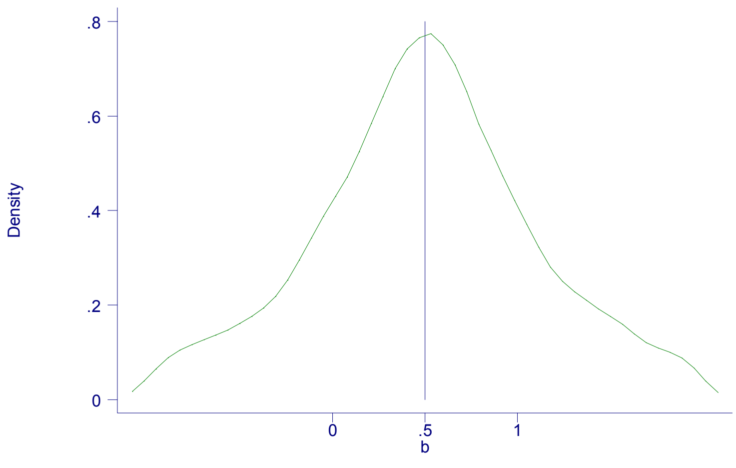


Figure 1-2 – Empirical distribution of $\widehat{\beta}_{IV}$ (N=100, $\lambda = 1$)

a) $\theta = 0.05$



b) $\theta = 0.01$



c) $\theta = 0.001$

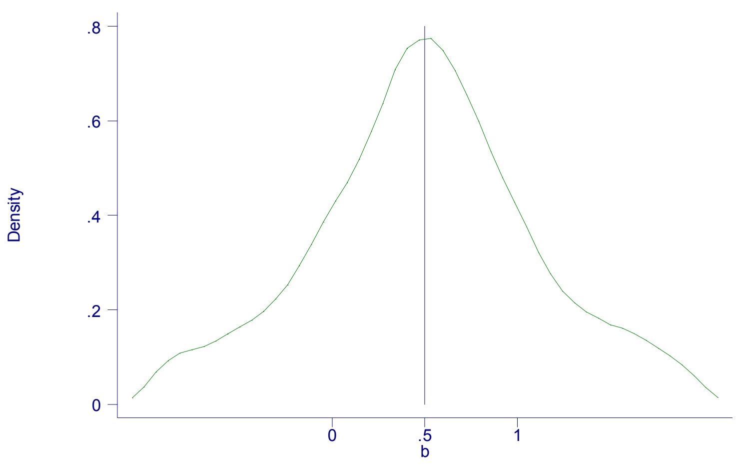
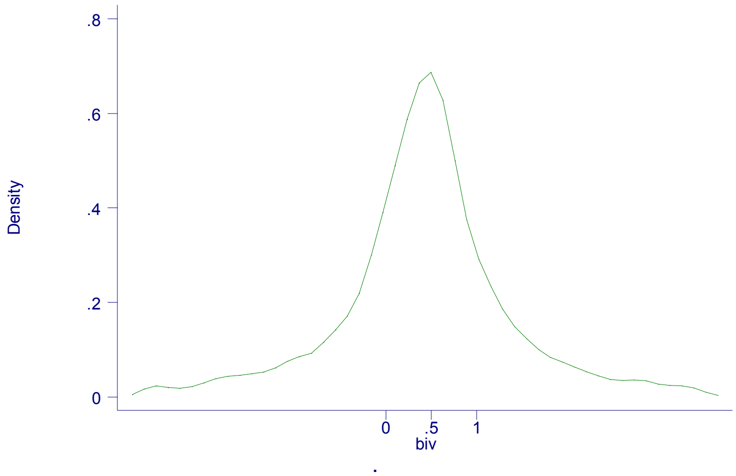
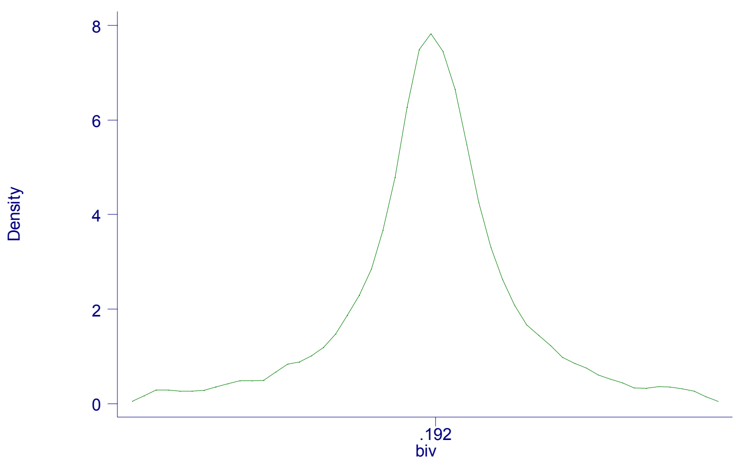


Figure 1-3 – Empirical distribution of $\widehat{\beta}_{IV}$ (N=100, $\theta = 0.05$)

a) $\lambda = 1$



b) $\lambda = 5$



c) $\lambda = 10$

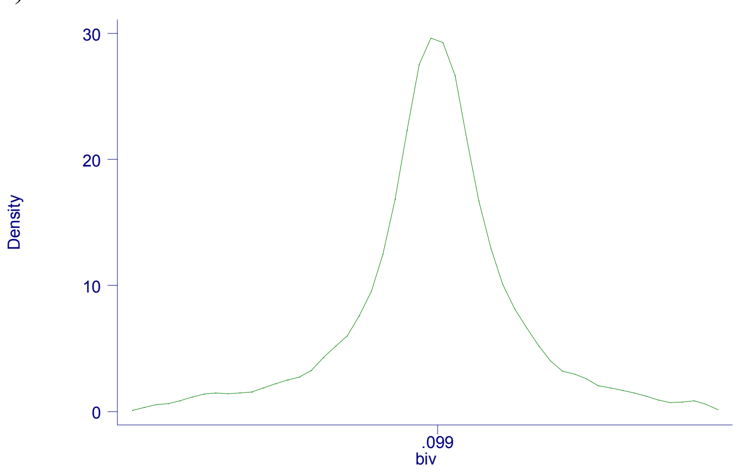
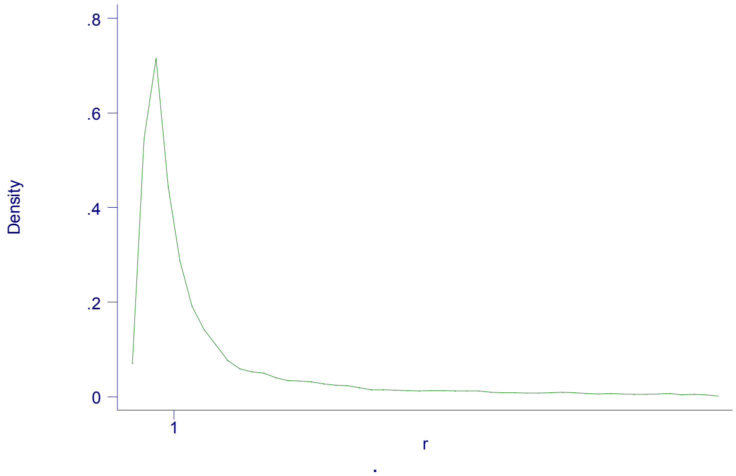


Figure 1-4 – Empirical distribution of $\widehat{\sigma^2}_{IV}$ (N=100, $\lambda = 5$)

a) $\theta = 1$



b) $\theta = 0.05$

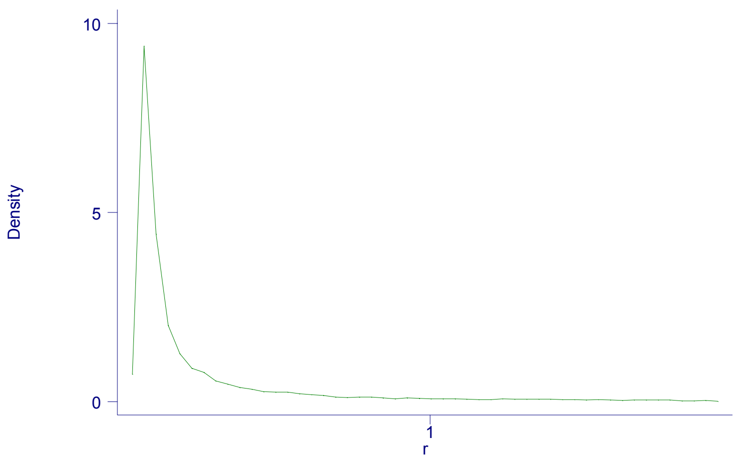


Figure 1-5 – Empirical distribution of $t_{\widehat{\beta}_{IV}}$ (N=100, $\lambda = 10$)

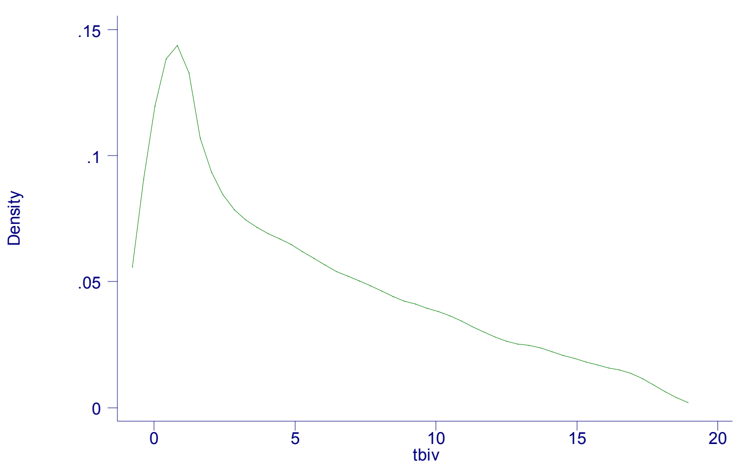
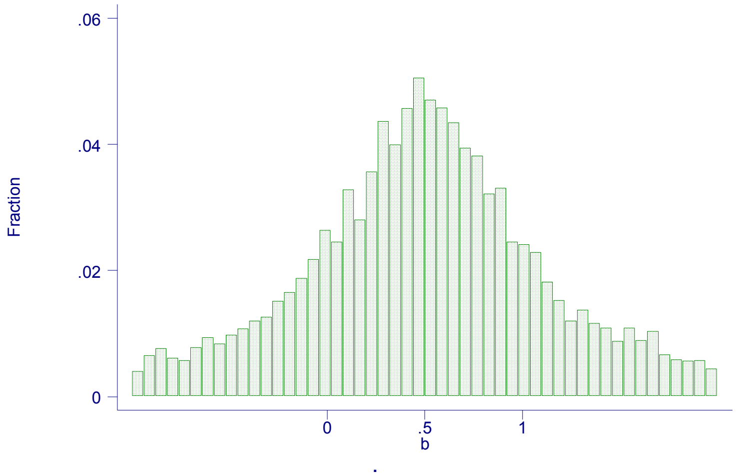
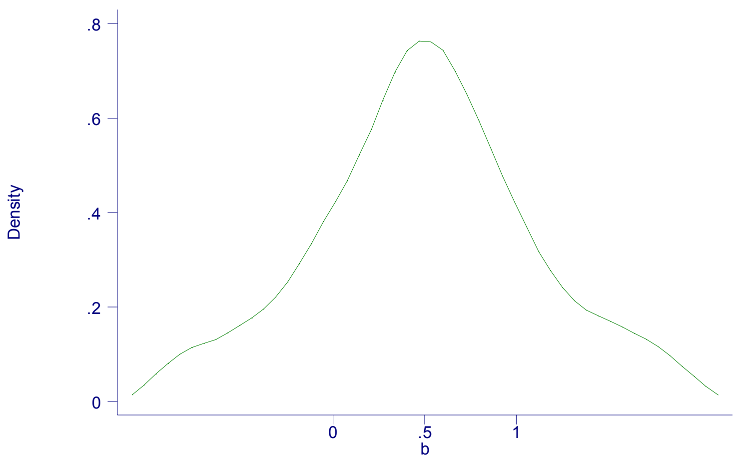


Figure 1-6- Empirical distribution of $\hat{\beta}_{IV}$ (N=100, $\lambda = 1$, $\theta = 0.001$)

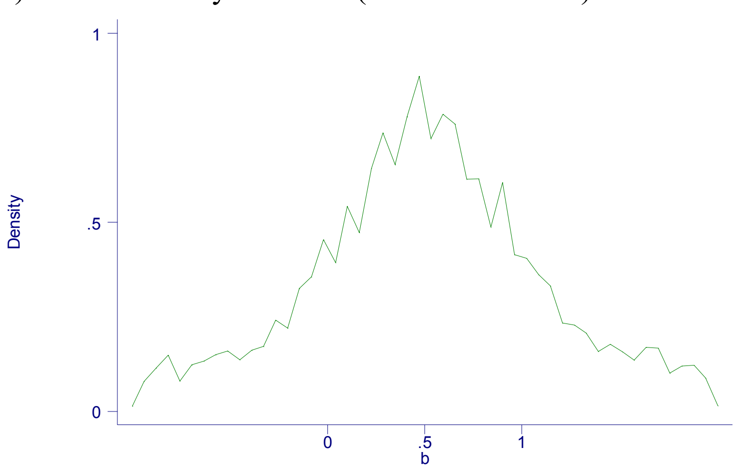
a) Histogram



b) Kernel Density Estimate (Bandwidth=0.10)



c) Kernel Density Estimate (Bandwidth=0.01)



1.5 Appendix 1

```
**********************************************************
Stata code used to simulate the numerical example in the text
**********************************************************

set matsize 800
set more off
program drop _all
set obs 100
mat sigma=[4.61, 2.71, 1/ 2.71, 1.81, 1/ 1, 1, 2]
program define iv
version 7.0
if ”‘1’” == ”?”{
global S_1 ”bols biv s2iv tiv”
exit
}
quietly {
set obs 100
drawnorm y x z, cov(sigma)
reg y x
scalar bols=_b[x]
ivreg y (x=z)
scalar biv=_b[x]
scalar s2iv=e(rmse)^2
scalar tiv=_b[x]/_se[x]
post ‘1’ (bols) (biv) (s2iv) (tiv)
```

```
drop _all
}
end
simul iv,reps(1000)
```

1.6 Appendix 2

1-The covariance matrix of the observable random variables in the Shea's model:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

where

$$\begin{aligned}
\sigma_{11} &= (\beta_1\gamma + \lambda)^2 + (1 - \lambda + \beta\gamma)^2 + \beta_1^2(1-\gamma)^2 + \beta_2^2(1-\gamma)^2 \\
\sigma_{21} &= \gamma(\beta_1\gamma + \lambda) + (1-\gamma)^2\beta_1 \\
\sigma_{31} &= \gamma(1 - \lambda + \beta_2\gamma) + \beta_2(1-\gamma)^2 \\
\sigma_{41} &= \delta\beta_1(1-\gamma) + (1-\delta)\beta_2(1-\gamma) \\
\sigma_{51} &= (1-\delta)\beta_1(1-\gamma) + \delta\beta_2(1-\gamma) \\
\sigma_{22} &= \gamma^2 + (1-\gamma)^2\sigma_{32} = 0 \\
\sigma_{42} &= \delta(1-\gamma)\sigma_{52} = (1-\gamma)(1-\delta) \\
\sigma_{33} &= \gamma^2 + (1-\gamma)^2 \\
\sigma_{43} &= (1-\gamma)(1-\delta)\sigma_{53} = \delta(1-\gamma) \\
\sigma_{44} &= \delta^2 + (1-\delta)^2 + \phi^2 \\
\sigma_{54} &= 2\delta(1-\delta) + \phi^2 \\
\sigma_{55} &= \delta^2 + (1-\delta)^2 + \phi^2
\end{aligned}$$

2-The stata code used to simulate Shea's model:

****

a) Generating random variables by the model

****

```
program drop _all
set matsize 500
global i=100
scalar gam=0.3
scalar del=0.50
program define iv
version 7.0
if ”‘1’” == ”?”{
global §_1 ”biv1 shea tz1 tz2 f r2”
exit
}
quietly {
set obs $i
mat sigma=I(6)
drawnorm u1 u2 e1 e2 v1 v2,cov(sigma) n($i)
g y=0.9*u1+0.1*u2
g x1=gam*u1+(1-gam)*e1
g x2=gam*u2+(1-gam)*e2
g z1=del*e1+(1-del)*e2+0.1*v1
g z2=(1-del)*e1+del*e2+0.1*v2
reg y x1 x2 z1 z2
mat var=e(V)
```

```
scalar tz1=_b[z1]/sqrt(var[3,3])
scalar tz2=_b[z2]/sqrt(var[4,4])
test z1 z2
scalar f=r(F)
ivreg2 y (x1 x2=z1 z2) , first
scalar shea=e(first)
scalar biv1=_b[x1]
regress x1 z1 z2
}
post `1' (biv1) (shea) (tz1) (tz2) (f) (e(r2))
end
simul iv,reps(1000)
```

****

a) Generating random variables by the joint distribution of the observable random variables

****

```
program drop _all
set matsize 500
global i=100
global b1=0
global b2=0
global lam=0.90
global gam=0.3
global del=0.50
global phi=0.1
mat sig11=($lam+$b1*$gam)^2+(1-$lam+$b2*$gam)^2+$b1^2*(1-$gam)^2+$b2^2*(1-
$gam)^2
mat sig12=$gam*($lam+$b1*$gam)+$b1*(1-$gam)^2
mat sig21=sig12
mat sig13=$gam*(1-$lam+$b2*$gam)+$b2*(1-$gam)^2
mat sig31=sig13
mat sig14=$del*$b1*(1-$gam)+$b2*(1-$del)*(1-$gam)
mat sig41=sig14
mat sig15=$del*$b2*(1-$gam)+$b1*(1-$del)*(1-$gam)
mat sig51=sig15
mat sig22=$gam^2+(1-$gam)^2
mat sig23=0
```

```
mat sig32=sig23
mat sig24=(1-$gam)*$del
mat sig42=sig24
mat sig25=(1-$gam)*(1-$del)
mat sig52=sig25
mat sig33=$gam^2+(1-$gam)^2
mat sig34=(1-$gam)*(1-$del)
mat sig43=sig34
mat sig35=$del*(1-$gam)
mat sig53=sig35
mat sig44=$del^2+$phi^2+(1-$del)^2
mat sig45=2*$del*(1-$del)+$phi^2
mat sig54=sig45
mat sig55=(1-$del)^2+$del^2+$phi^2
mat sigma=[sig11,sig12,sig13,sig14,sig15 / sig21,sig22,sig23,sig24,sig25 /
sig31,sig32,sig33,sig34,sig35 / sig41,sig42,sig43,sig44,sig45 / sig51,sig52,sig53,sig54,sig55]
program define iv
version 7.0
if ”‘1’” == ”?”{
global §_1 ”biv1 biv2 r2 t1 t2 f shea”
exit
}
set obs $i
drawnorm y x1 x2 z1 z2,cov(sigma) n($i)
reg y x1 x2 z1 z2
mat var=e(V)
```

```
scalar tz1=_b[z1]/sqrt(var[3,3])
scalar tz2=_b[z2]/sqrt(var[4,4])
test z1 z2
scalar f=r(F)
ivreg2 y (x1 x2=z1 z2) ,first
scalar shea=e(first)
scalar biv1=_b[x1]
scalar biv2=_b[x2]
regress x1 z1 z2
post `1' (biv1) (biv2) (e(r2)) (tz1) (tz2) (f) (shea)
end
simul iv,reps(1000)
```

1.7 Appendix 3

Maddala and Jeong (1992) showed that bimodality is a result of the perfect conditional correlation between structural and first stage error terms ($\lambda = 1$) considered by Nelson and Startz. They used the following model as data generating process and showed that it is, algebraically, equal to the Nelson and startz's model

$$\begin{aligned} y &= \beta x + \epsilon \\ x &= \delta z + u \end{aligned} \tag{1.10}$$

where $z \sim N(0, 1)$ and

$$\begin{bmatrix} \epsilon \\ u \end{bmatrix} \sim N(\mathbf{0}, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}$$

To compare models (2.4) and (1.10), consider the joint distribution of the observable variables $w = (y, x, z)'$ implied by Nelson and Startz's model

$$W \sim (0, \Sigma), \quad \Sigma = \begin{bmatrix} \beta^2 + (1+\beta\lambda)^2 & \lambda(1+\beta\lambda)+\beta & \beta\theta \\ \lambda(1+\beta\lambda)+\beta & 1+\lambda^2 & \theta \\ \beta\theta & \theta & \theta^2+1 \end{bmatrix} \tag{1.11}$$

The joint distribution of observable variables $w = (y, x, z)'$ contains all properties of model (2.4) including $\beta_{LS} = \beta + \frac{\lambda}{1+\lambda^2}$, $\beta_{IV} = \beta$, $\sigma^2_{IV} = 1$ and $Cov(y, z|x) = \frac{-\theta\lambda}{1+\lambda^2} \neq 0$. The reparameterized structural/statistical model defined on the joint distribution of

$w = (y, x, z)'$ is

$$y = \beta x + \epsilon \tag{1.12}$$
$$x = \delta z + u$$

where

$$\delta = \frac{Cov(x,z)}{Var(z)}, \quad \beta = \frac{Cov(y,z)}{Cov(x,z)}$$

$$\begin{bmatrix} \epsilon \\ u \end{bmatrix} \sim N(\mathbf{0}, \Omega), \quad \Omega = \begin{bmatrix} 1 & \lambda \\ \lambda & \frac{1+\lambda^2(\theta^2+1)}{\theta^2+1} \end{bmatrix}, \ det(\Omega) = \frac{1}{1+\theta^2} \in (0,1]$$

Obviously, Maddala and Jeong's model (1.10) is equal to the Nelson and Startz's only if $\lambda^2 = \frac{\theta^2}{1+\theta^2}$ that is not considered by Nelson and Startz.

Chapter 2

IV Regressions: Specification and Instrument Selection

2.1 Introduction

The method of Instrumental Variables(IV) is generally designed to deal with the inconsistency problem of the least square method in the presence of correlation between regressors and the error term. In other words, in the context of the regression model,

$$y_i = \mathbf{x}_i\beta + \epsilon_i,\ E(x'\epsilon) \neq 0,\ \epsilon_i \sim N(0, \sigma^2\mathbf{I}_{\mathtt{N}})\quad i \in \mathtt{N} \tag{2.1}$$

where $\mathbf{x}_i$ is a $1 \times k$ vector of stochastic explanatory variables and β is a vector of unknown parameters, the instrumental variable method provides consistent estimators for β and σ^2 if there exists a $1 \times p, (p \geq k)$ vector of variables $\mathbf{z}_i$ that is correlated with $\mathbf{x}_i$ but uncorrelated with the disturbance term. Specifically, an appropriate set of instruments should satisfy the following two criteria:

$$\begin{aligned} &a)\ \ \mathtt{N}^{-1}(\mathbf{Z}'\epsilon) \xrightarrow{p} 0 \\ &b)\ \mathtt{N}^{-1}(\mathbf{Z}'\mathbf{X}) \xrightarrow{p} \Sigma_{ZX} \neq 0 \end{aligned} \tag{2.2}$$

where Σ_{ZX} is a $p \times k$ matrix of covariances between $\mathbf{z}_i$ and $\mathbf{x}_i$. These conditions ensure the consistency of the IV estimators of $\beta,\ \sigma^2$. Given $\mathbf{z}_i$ as an admissible vector of instrumental variables, the IV estimators can be defined as,

$$\hat{\beta}_{iv} = (\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\mathbf{y}, \hat{\sigma}^2_{iv} = (\mathtt{N} - k)^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_{iv})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{iv}) \tag{2.3}$$

where, $P_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. The IV estimator $\hat{\beta}_{iv}$ is consistent if $\mathtt{N}^{-1}\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \Sigma_{ZZ}$, where Σ_{ZZ} is a positive definite matrix. Moreover, if $\mathtt{N}^{-1/2}\mathbf{Z}'\epsilon \xrightarrow{p} N(0, \sigma^2\Sigma_{ZZ})$ then, (see Bowden and Turkington (1984) for more details)

$$\mathtt{N}^{1/2}(\hat{\beta}_{iv} - \beta) \sim N(0, \sigma^2(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX})^{-1})$$

The major difficulty of the IV estimation is the choice of appropriate instruments. The first condition is purely subjective and unverifiable. The second condition is useful, only if the first one has already been satisfied. However, a low correlation between the endogenous variable $\mathbf{x}$ and the instrument $\mathbf{z}$, known as a *weak instrument*, will exacerbate any problems associated with a correlation between the instrument and the error term ϵ (see Bound, Jaeger, and Baker (1995)). Unfortunately, except for some rules of thumb, there is no practical guidance to identify weak instruments. Specification of the estimated IV regression is another subject that has its own difficulties. Conventionally, specification tests are used to investigate the quality of the regression model and residuals of the estimated regression have a vital role in performing specification tests. However, Pagan and Hall (1983) and Pesaran and Taylor (1999) have argued that the IV regression residuals are appropriate only for serial correlation and normality tests. In particular, Pesaran and Smith (1994) have shown that the residuals of the IV regression cannot be used in constructing measures of goodness-of-fit and model selection criteria.

The plan of this chapter is as follows. Section 2 considers the underlying joint distribution function of an instrumental variable model and shows that it has two properties that are in practice more useful than condition (2.2). This section establishes a consistent statistical foundation for the instrumental variable method and shows that a non-orthogonal structural model is hidden in a system of independent equations. Section 3, deals with the specification of the estimated IV regression and provides a simple way to ensure the specification assumptions. Section 4 poses a question on the usefulness of the first stage R-square (F-statistic) as a relevancy measure and introduces a new instrument relevancy measure. This paper considers a "one independent–one instrument" variables model; however, the results are readily extendable to the general case. Monte-carlo simulation are quoted to illustrate the results in section 3 and 4.

2.2 Estimation

2.2.1 Static Model

Let the data generating mechanism be described by (see Nelson and Startz (1990a))

$$\begin{aligned} \mathbf{y} &= \beta \mathbf{x} + u \\ \mathbf{x} &= \zeta + \lambda u \\ \mathbf{z} &= \theta_1 \zeta + \theta_2 v \end{aligned} \tag{2.4}$$

where $\mathbf{y}$ is a $N \times 1$ vector of observations on the dependent variable and, $\mathbf{x}$ is a $N \times 1$ vector of observed values of the independent variable. ζ is a $N \times 1$ vector of the unobservable component of $\mathbf{x}$, which can be considered as the exogenous variable that is suggested by economic theory to describe the stochastic behavior of y. Let $\mathbf{z}$ denote a $N \times 1$ vector of observations on an instrumental variable. The goal is to estimate β and test $H_0 : \beta = \beta_*$, which needs an unbiased and efficient estimator for σ_u^2. Assume u, v and ζ all are distributed as independent drawings from a multivariate normal distribution with mean $\mathbf{0}$ and identity covariance matrix. The assumptions of zero means and identity covariance matrix are assumptions of convenience.

It follows that $(\mathbf{y}, \mathbf{x}, \mathbf{z})$ is normally distributed with zero mean and covariance matrix (see Carter and Fuller (1980))

$$\Sigma = \begin{pmatrix} \beta^2\sigma_\zeta^2 + (1+\beta\lambda)^2\sigma_u^2 & \lambda(1+\beta\lambda)\sigma_u^2 + \beta\sigma_\zeta^2 & \beta\theta_1\sigma_\zeta^2 \\ \lambda(1+\beta\lambda)\sigma_u^2 + \beta\sigma_\zeta^2 & \sigma_\zeta^2 + \lambda^2\sigma_u^2 & \theta_1\sigma_\zeta^2 \\ \beta\theta_1\sigma_\zeta^2 & \theta_1\sigma_\zeta^2 & \theta_1^2\sigma_\zeta^2 + \theta_2^2\sigma_v^2 \end{pmatrix} \tag{2.5}$$

The structure of the above covariance matrix conveys valuable information regarding the appropriate instrumental variable. Note that in the absence of non-orthogonality($\lambda = 0$), where $\mathbf{x} = \zeta$

$$Cov(\mathbf{y}, \mathbf{z}|\zeta) = Cov(\mathbf{y}, \mathbf{z}) - Cov(\mathbf{y}, \zeta)Var(\zeta)^{-1}Cov(\zeta, \mathbf{z}) = 0 \tag{2.6}$$

and if non-orthogonality exists ($\lambda \neq 0$)

$$Cov(\mathbf{y}, \mathbf{z}|\mathbf{x}) = Cov(\mathbf{y}, \mathbf{z}) - Cov(\mathbf{y}, \mathbf{x})Var(\mathbf{x})^{-1}Cov(\mathbf{x}, \mathbf{z}) \neq 0 \tag{2.7}$$

Condition (2.6) states that instruments must be *redundant* in the context of the theoretical model (see Breusch, Qian, Schmidt, and Wyhowski (1999)), and condition (2.7) specifies that instruments must be *relevant* in the context of the observed variables. *Redundancy* is exactly equivalent to the definition of the excluded exogenous regressors in the context of the simultaneous regression models. *Relevancy* evaluates the practical ability of the instrument in describing the stochastic behavior of y given x. In view of the usual motivation for using IV estimators (e.g. $\lambda \neq 0$ or errors in observations for ζ), (2.6) is a non-testable condition. Condition (2.7), which is a testable condition, is more general than (2.2b)because it considers the information content of $\mathbf{z}$ regarding the stochastic behavior of y, not only its correlation with $\mathbf{x}$.

Lemma (See Muirhead (1982), theorem 1.2.11): Let W be $N_m(\mu, \Sigma)$ and partition W, μ and Σ as

$$W = \begin{pmatrix} y \\ X \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Let, without loss of generality, Σ_{22} be a non-singular matrix and let $\sigma_{11.2} = \sigma_{11} - \sigma_{12}\Sigma_{22}^{-1}\sigma_{21}$, $\beta = \Sigma_{22}^{-1}\sigma_{21}$. Then, the conditional distribution of y given X is

$$N_k(\mu_1 + (X - \mu_2)\beta, \sigma_{11.2})$$

which implies

$$E(y|X) = (\mu_1 - \mu_2\beta) + X\beta$$

as the regression function.

$\square$

In view of the above lemma, the appropriate regression function for (2.5) is

$$\mathbf{y} = \beta_0\mathbf{x} + \gamma_0\mathbf{z} + \epsilon_0,\ \ \epsilon_0 \sim N(0, \sigma_0^2) \tag{2.8}$$

$$\beta_0 = \frac{Cov(\mathbf{y}, \mathbf{x}|\mathbf{z})}{Var(\mathbf{x}|\mathbf{z})},\ \ \gamma_0 = \frac{Cov(\mathbf{y}, \mathbf{z}|\mathbf{x})}{Var(\mathbf{z}|\mathbf{x})}$$

$$\sigma_0^2 = Var(y|\mathbf{x}, \mathbf{z}) = Var(y) - Cov(y, \mathbf{x})\beta_0 - Cov(y, \mathbf{z})\gamma_0$$

where

$$\beta_0|_{\lambda=0} = \beta \qquad \gamma_0|_{\lambda=0} = 0 \qquad \sigma_0^2|_{\lambda=0} = 1$$

$$- - - - - - - - - - -$$

$$\beta_0|_{\lambda\neq 0} = \beta + \frac{\lambda\theta_1^2 + \lambda\theta_2^2}{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2} \qquad \gamma_0|_{\lambda\neq 0} = \frac{-\lambda\theta_1}{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2}$$

$$\sigma_0^2|_{\lambda\neq 0} = \frac{\theta_2^2}{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2}$$

and, by definition, $E(\mathbf{x}\epsilon_0) = 0, E(\mathbf{z}\epsilon_0) = 0$. Clearly, in the non-orthogonal case ($\lambda \neq 0$) the above lemma can not provide an unbiased estimator for the causal effect of $\mathbf{x}$ on y[1]; however, the primary parameters of (2.5) can still be used to estimate β and σ_u^2. The Instrumental Variable estimator, defined as $\frac{Cov(\mathbf{y},\mathbf{z})}{Cov(\mathbf{x},\mathbf{z})}$, is the known consistent estimator of β that does not change with the non-orthogonality factor λ. The Instrumental Variable estimator is, indeed, a weighted sum of the parameters of (2.8). The weighted sum comes in the form of

$$\beta_{IV} = \frac{Cov(\mathbf{y},\mathbf{z})}{Cov(\mathbf{x},\mathbf{z})} = \beta_0 + \frac{\gamma_0}{\delta}, \quad \sigma_u^2 = \sigma_0^2 + \sigma_1^2(\beta - \beta_0)^2 \tag{2.9}$$

[1]In other words, if $\mathbf{x}$ and u are non-orthogonal, the structural model is different with the mean expectation of y.

where δ and σ_1^2 are the parameters of the linear projection of $\mathbf{x}$ on $\mathbf{z}$, defined as

$$\begin{aligned}
\mathbf{x} &= \delta\mathbf{z} + \epsilon_1 \quad \epsilon_1 \sim N(0, \sigma_1^2) \\
\delta &= \frac{Cov(\mathbf{x}, \mathbf{z})}{Var(\mathbf{z})} = \frac{\theta_1}{\theta_1^2 + \theta_2^2} \\
\sigma_1^2 &= Var(x|z) = \frac{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2}{\theta_1^2 + \theta_2^2}
\end{aligned} \tag{2.10}$$

proof:

$$\beta_0 + \frac{\gamma_0}{\delta} = \frac{Cov(y, x|z)}{Var(x|z)} + \frac{Var(z)}{Cov(x,z)} \frac{Cov(y, z|x)}{Var(z|x)} =$$

$$\frac{\frac{Var(z)}{Cov(x,z)} Var(x)Cov(y,z) - Cov(y,z)Cov(z,x)}{Var(z)Var(x) - Cov(x,z)^2} = \frac{Cov(y,z)}{Cov(x,z)} = \beta_{IV}$$

Using (2.9) one can rewrite (2.8) as

$$y = \beta x + u, \quad u = \epsilon_0 - \frac{\gamma_0}{\delta}\epsilon_1 \ u \sim N(0, \sigma_0^2 + \sigma_1^2(\beta - \beta_0)^2)$$

where, $E(\mathbf{x}u) \neq 0$ and $E(\mathbf{z}u) = 0$.

In other words, a typical non-orthogonal structural model is embedded into a system of equations consisting of (2.8) and (2.10).

In fact, model (2.4) can be shown by two other representations: *a)*

$$\begin{aligned}
y &= \beta\mathbf{x} + u \\
\mathbf{x} &= \delta\mathbf{z} + \epsilon_1
\end{aligned} \tag{2.11}$$

where

$$\delta = \frac{Cov(\mathbf{x}, \mathbf{z})}{Var(\mathbf{z})}$$

$$\begin{pmatrix} u \\ \epsilon_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & \frac{\theta_2^2+\lambda^2(\theta_1^2+\theta_2^2)}{\theta_1^2+\theta_2^2} \end{pmatrix}\right)^2$$

and *b)*

$$y = \beta_0 \mathbf{x} + \gamma_0 \mathbf{z} + \epsilon_0 \tag{2.12}$$

$$\mathbf{x} = \delta \mathbf{z} + \epsilon_1$$

where

$$\beta_0 = \beta + \frac{\lambda\theta_1^2 + \lambda\theta_2^2}{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2}$$

$$\gamma_0 = \frac{-\lambda\theta_1}{\theta_2^2 + \lambda^2\theta_1^2 + \lambda^2\theta_2^2}$$

$$\begin{pmatrix} \epsilon_0 \\ \epsilon_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\theta_2^2}{\theta_2^2+\lambda^2(\theta_1^2+\theta_2^2)} & 0 \\ 0 & \frac{\theta_2^2+\lambda^2(\theta_1^2+\theta_2^2)}{\theta_1^2+\theta_2^2} \end{pmatrix}\right)$$

Note that γ_0 is, indeed, δ times $(\beta - \beta_0)$, which in turn, converges to the bias of the least square estimator as θ_1 approaches zero.[3]

The above argument formalized the statistical foundation of the method of instrumental variables and is general enough to address all kinds of endogenous regressor models. To show this, consider $f(y, \mathbf{x}, \mathbf{z}; \Phi)$ as the underlying joint distribution function of (3.8). In view of (2.9) and the fact that the parameters of interest are (β, σ^2),

[2]this representation was derived by Maddala and Jeong (1992); however, they failed to consider the necessary restriction on the covarainace matrix of u and ϵ_1.

[3]Spanos (2000) looks at the optimal instrumental variable as a variable that satisfies $\gamma_0 = 0$ in (2.8), where $Var(\mathbf{z}) > 0$, and $Cov(y, \mathbf{z}) \neq 0$(See Spanos (2000), p.9). However, (2.12) shows that $\gamma_0 = 0$ if and only if $\delta = 0$, or $\beta = \beta_0$, which are not consistent with the assumptions of the instrumental variable method.

the appropriate reduction[4] of $f(y, \mathbf{x}, \mathbf{z})$ would be

$$f(y, \mathbf{x}, \mathbf{z}; \Phi) = f(y, \mathbf{x}|\mathbf{z}; \psi_1) f(\mathbf{z}; \psi_2)$$

where, assuming normality, the probabilistic structure of $\mathbf{z}$ can be ignored (See Engle, Hendry, and Richard (1983)). $f(y, \mathbf{x}|\mathbf{z}; \psi_1)$, which indicates that $\mathbf{x}$ is a vector of endogenous variables, is the underlying distribution of the following multivariate regression model

$$\begin{aligned} y &= \pi_1 \mathbf{z} + u_1 \\ \mathbf{x} &= \pi_2 \mathbf{z} + u_2 \end{aligned} \tag{2.13}$$

where $\pi_1 = Var(\mathbf{z})^{-1} Cov(y, \mathbf{z}), \pi_2 = Var(\mathbf{z})^{-1} Cov(\mathbf{x}, \mathbf{z})$ and

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Var(y|\mathbf{z}) & Cov(y, \mathbf{x}|\mathbf{z}) \\ Cov(\mathbf{x}, y|\mathbf{z}) & Var(\mathbf{x}|\mathbf{z}) \end{pmatrix} \right)$$

As Spanos (1986) shows, by reparametrisation of (2.13) one can get a system of equations consisting of (2.8) and (2.10), where conditioning set of (2.8) and (2.10) are $(\sigma(\mathbf{x}), \mathbf{z})$ and $(\mathbf{z})$, respectively.

2.2.2 Dynamic Model

The above argument can easily be extended to the autoregressive distributed lag (ARDL) models. Without loss of generality, consider the following ARDL(1,1) model

$$y_t = \alpha^* y_{t-1} + \beta_0^* \zeta_t + \beta_1^* \zeta_{t-1} + \epsilon_{*t} \quad t \in \mathtt{T} \tag{2.14}$$

[4]Condition (2.7) restricts further reduction.

where

$$\begin{aligned}
\beta_0^* &= D_5^* \\
\alpha_1^* &= D_4^* - D_3^* \beta_0^* \\
\beta_1^* &= D_2^* - D_1^* \beta_0^* - D_0^* \alpha_1^*
\end{aligned}$$

and

$$\begin{aligned}
D_0^* &= \frac{Cov(y_{t-1}, \zeta_{t-1})}{Var(\zeta_{t-1})} & D_1^* &= \frac{Cov(\zeta_t, \zeta_{t-1})}{Var(\zeta_{t-1})} \\
D_2^* &= \frac{Cov(y_t, \zeta_{t-1})}{Var(\zeta_{t-1})} & D_3^* &= \frac{Cov(y_{t-1}, \zeta_t|\zeta_{t-1})}{Var(y_{t-1}|\zeta_{t-1})} \\
D_4^* &= \frac{Cov(y_t, y_{t-1}|\zeta_{t-1})}{Var(y_{t-1}|\zeta_{t-1})} & D_5^* &= \frac{Cov(y_t, \zeta_t|\zeta_{t-1} y_{t-1})}{Var(\zeta_t|\zeta_{t-1} y_{t-1})}
\end{aligned}$$

Let $\mathbf{z}_t$ be a redundant variable vector for y_t given ζ_t, such that:

$$\begin{aligned}
a) Cov(y_t, z_t|\zeta_t) &= 0, & d) Cov(y_{t-1}, z_{t-1}|\zeta_{t-1}) &= 0 \\
b) Cov(\zeta_{t-1}, z_t|\zeta_t) &= 0, & e) Cov(y_t, z_{t-1}|\zeta_{t-1}) &= 0 \\
c) Cov(y_{t-1}, z_t|\zeta_t) &= 0, & f) Cov(\zeta_t, z_{t-1}|\zeta_{t-1}) &= 0
\end{aligned} \tag{2.15}$$

where

$$\begin{array}{ccc}
Cov(y_t, Z_t) \neq 0 & Cov(\zeta_t, Z_t) \neq 0 & Cov(y_{t-1}, Z_t) \neq 0 \\
Cov(\zeta_{t-1}, Z_t) \neq 0 & Cov(y_t, Z_{t-1}) \neq 0 & Cov(\zeta_t, Z_{t-1}) \neq 0
\end{array}$$

If for any reason we have to use $\mathbf{x}_t$ instead of ζ that is correlated with ϵ_{*t}, and if $\mathbf{z}_t$ is not redundant for y_t given $\mathbf{x}_t$[5], the appropriate linear projection of y changes from (2.14) to

$$y_t = \alpha y_{t-1} + \beta_0 \mathbf{x}_t + \beta_1 \mathbf{x}_{t-1} + \gamma_0 \mathbf{z}_t + \gamma_1 \mathbf{z}_{t-1} + \epsilon_{0t} \quad t \in \mathtt{T} \tag{2.16}$$

but still the parameters of (2.14) can consistently estimated by the sample counterpart of

$$\beta_0^* = \beta_0 + \frac{\gamma_0}{\delta_0} \tag{2.17}$$

and D^*s by

$$D_0 = \frac{Cov(x_{t-1}, y_{t-1}|z_{t-1})}{Var(x_{t-1}|z_{t-1})} \qquad D_1 = \frac{Cov(x_{t-1}, x_t|z_{t-1})}{Var(x_{t-1}|z_{t-1})}$$

$$D_2 = \frac{Cov(x_{t-1}, y_t|z_{t-1})}{Var(x_{t-1}|z_{t-1})} \qquad D_3 = \frac{Cov(y_{t-1}, x_t|z_{t-1}x_{t-1})}{Var(y_{t-1}|z_{t-1}x_{t-1})}$$

$$D_4 = \frac{Cov(y_{t-1}, y_t|z_{t-1}x_{t-1})}{Var(y_{t-1}|z_{t-1}x_{t-1})}$$

As (2.17) shows β_0^* is a weighted sum of β_0 and γ_0 where weights are 1 and δ_0 from the following dynamic linear regression model

$$\mathbf{x}_t = \delta_0 \mathbf{z}_t + \theta_0 y_{t-1} + \theta_1 \mathbf{x}_{t-1} + \theta_2 \mathbf{z}_{t-1} + \epsilon_{1t} \quad t \in \mathtt{T} \tag{2.18}$$

2.3 Specification

Section (1) showed that the residual of the IV regression is, in fact, a weighted sum of the residuals of (2.8) and (2.10). This property may help us avoid using IV regression residuals to investigate the specification of the estimated model. As an alternative,

[5]If there exists at least one inequality in (2.15).

it proposed that the modeler concentrate exclusively on the specification of (2.8) and (2.10)[6]. To illustrate the importance of the specification of (2.8) and (2.10) in the process of IV estimation, consider the following simulation exercises.

1-Dependence. Let $\beta = \lambda = \theta_1 = \theta_2 = 1$ in (2.5) and assume $E(y_t) = 2 + 0.8t$ which implies $y_t = 2 + 0.8t + x_t + u_t$ as the relevant regression function. Suppose that the modeler fails to notice the time-heterogeneity of y_t and begins to estimate $y_t = \alpha + \beta x_t + u_t$ using z_t as an appropriate instrumental variable. The average estimates (IV method) from 1000 samples of 100 observations of (y_t, x_t, z_t) is

$$y_t = constant + \underset{(0.147)}{2.35} x_t + \hat{u}_t \qquad \sigma^2 = \underset{(4.70)}{533.3}$$

where (average) standard errors are in parentheses. The average p-value of serial correlation[7] and heteroscedasticity[8] tests are 0.001 and 0.432, respectively. Although the estimated coefficients for β and σ^2 are highly significant, the existence of serial correlation indicates that the above model is misspecified, and so the estimated parameters are unreliable. Re-specification of the statistical model is the most general method to deal with serial correlation; however, how to respecify the model is still a data-dependent question. At this stage, the specification of (2.8) and (2.10) help the modeler not only to detect the misspecified model, but also to choose the right

[6]Comparable with (2.8) and (2.10), Spanos (2000) argues that (2.13) should be a well-specified multivariate system of equations. However, his first purpose of estimating (2.13) is to ensure that $Cov(\mathbf{x},\mathbf{z})$ and $Cov(\mathbf{z}, y)$ are not zero.

[7]F-test for λ in $y_t = \beta_0 x_t + \lambda \hat{u}_{t-1} + \epsilon_0$ where $\hat{u}_{t-1}$ is the lagged residuals from the first stage IV estimation. x_t has been instrumented by the same set of instrumental variables in both stages.

[8]Pagan and Hall (1983) heteroscedasticity test with fitted value of the dependent variable and its square as indicator variables have been used to test heteroscedasticity.

specification. Results reported in table (1) show the percent of rejection of each specification assumption for (2.8) and (2.10) in 1000 replication. The first two columns show the misspecification analysis for the underlying (2.8) and (2.10) models of the above estimated model. These results indicate very clearly that the independence assumption of (2.10) is completely irrelevant; it has been rejected in almost all of the replications. By re-specifying the model to include a trend, one can obtain the following (average) model[9]

$$\underset{}{y_t = constant} + \underset{(0.003)}{0.80t} + \underset{(0.144)}{0.987x_t} + \hat{u}_t \qquad \underset{(0.38)}{\sigma^2 = 1.03}$$

where the average p-value of serial correlation and heteroscedasticity tests are 0.88 and 0.53, respectively. Columns 3 and 4 of table (1) show that the corresponding (2.8) and (2.10) models are statistically adequate models.

2-Parameter shift. Let $\lambda = \theta_1 = \theta_2 = 1$ in (2.5) and assume $\beta = 1$ for the first 100 observations and $\beta = 2$ for the second 100 observations where the relevant population regression function is $E(y_i|x_i) = constant + x_i + Dumx_i$ and $Dumx_i = \begin{cases} 0 & i = 1, \cdots, 100 \\ x_i & i = 101, \cdots, 200 \end{cases}$. One is likely to begin by estimating $y_i = \alpha + \beta x_i + u_i$ using z_i as instrument. The average of 1000 IV-regressions of 200 observations for (y_i, x_i, z_i)

[9] Re-specifying the model by adding lag values of y_t and x_t can not resolve the dependence problem of (2.10)

Test	Misspecified		Well-Specified	
Model	(2.8)	(2.10)	(2.8)	(2.10)
R-square	0.28		0.28	
Normality:				
Jarque -Bera	0.05	1	0.05	0.05
Doornik-Hansen	0.05	0.98	0.05	0.05
Dependence:				
Breusch-Godfrey(1)	0.05	1	0.05	0.05
Breusch-Godfrey(2)	0.05	1	0.05	0.04
White's Homosced.	0.05	0.02	0.05	0.05
RESET(2)	0.05	0.03	0.05	0.04
Joint Mean:				
Trend in mean	0.05	0.15	0.05	0.04
RESET(2)	0.05	0.05	0.05	0.07
Correlation	0.05	0.05	0.05	0.05
Joint Variance:				
Trend in variance	0.07	0.7	0.06	0.05
RESET(2)	0.05	0.09	0.05	0.05
ARCH(1)	0.04	0.18	0.03	0.03

Table 2.1: Percent of rejection of specification tests ($\alpha = 0.05$)

is:

$$y_i = constant + \underset{(0.132)}{1.49} x_i + \hat{u}_i \qquad \sigma^2 = \underset{(0.304)}{1.52}$$

where (average) standard errors are in parentheses. The average p-value of serial correlation and heteroscedasticity tests are 0.44 and 0.42, respectively[10]. The estimated model looks very reasonable; significant parameters, no heteroscedasticity and no serial correlation all are the dream of econometricians. Perhaps, if one did not know the actual data generating mechanism the estimated model would be convincing. Following our argument, let's take a look at the results of misspecification tests for (2.8) and (2.10). The first two columns of table (2) report the percent of rejection of each specification assumption of (2.8) and (2.10) in 1000 estimated regressions. The estimated regression for (2.8) is, on average, statistically adequate, but White heteroscedasticity test and RESET linearity test indicate that (2.10) suffers from variance heterogeneity and inconstancy of coefficients. By re-specifying the model to include $Dumx_i$ as another explanatory variable, which should be instrumented by $Dumz_i = \begin{cases} 0 & i = 1, \cdots, 100 \\ z_i & i = 101, \cdots, 200 \end{cases}$ one can obtain the following (average) regression model

$$y_i = constant + \underset{(0.163)}{0.977} x_i + \underset{(0.236)}{1.002} Dumx_i + \hat{u}_i \qquad \sigma^2 = \underset{(0.293)}{1.06}$$

[10]Pagan-Hall heteroscedasticity test has rejected homoscedasticity assumption in only 12% of regressions.

Test	Misspecified		Well-Specified	
Model	(2.8)	(2.10)	(2.8)	(2.10)
R-square	0.25		0.25	
Normality:				
Jarque -Bera	0.05	0.05	0.05	0.04
Doornik-Hansen	0.05	0.05	0.05	0.04
Dependence:				
Breusch-Godfrey(1)	0.04	0.05	0.04	0.05
Breusch-Godfrey(2)	0.05	0.05	0.05	0.05
White's Homosced.	0.04	1	0.05	0.06
RESET(2)	0.05	0.3	0.05	0.05
Joint Mean:				
Trend in mean	0.05	0.05	0.05	0.05
RESET(2)	0.05	0.28	0.05	0.05
Correlation	0.04	0.05	0.04	0.05
Joint Variance:				
Trend in variance	0.05	0.14	0.05	0.06
RESET(2)	0.04	1	0.04	0.05
ARCH(1)	0.04	0.04	0.04	0.04

Table 2.2: Percent of rejection of specification tests ($\alpha = 0.05$)

where the average p-value of serial correlation and heteroscedasticity tests are 0.50 and 0.56, respectively. Columns 3 and 4 of table (2) show that the corresponding (2.8) and (2.10) are statistically adequate models.

3-ARDL(1,1). Monte Carlo experiments have been performed to examine the established framework for the dynamic models and to analyze the sensitivity of the estimated model to the specification of (2.18) and (2.16). Consider the following dynamic model[11]

$$\begin{aligned} y_t &= \beta_0 x_t + \beta_1 x_{t-1} + \alpha y_{t-1} + \epsilon_{1t} \\ x_t &= \gamma_1 x_{t-1} + \gamma_2 z_t + \lambda \epsilon_{1t} + \epsilon_{2t} \\ z_t &= \gamma_3 z_{t-1} + \epsilon_{3t} \end{aligned}$$

where $\epsilon_{it} \sim (0,1)$ $i = 1, 2, 3$. To take into account the relation between the model and distribution parameters, the data must be generated by the following vector of first order Markov dependence and stationary normal process

$$W_t = AW_{t-1} + \mathbf{U}_t \qquad\qquad \mathbf{U}_t \sim N(\mathbf{0}, \Omega) \tag{2.19}$$

where, $W_t = (y_t, x_t, z_t)', \mathbf{U}_t = (u_1, u_2, u_3)'$ and

$$A = \begin{pmatrix} \alpha & \beta_0\gamma_1 + \beta_1 & \beta_0\gamma_2\gamma_3 \\ 0 & \gamma_1 & \gamma_2\gamma_3 \\ 0 & 0 & \gamma_3 \end{pmatrix},$$

[11]Note that if $\lambda \neq 0$, which indicates x_t and ϵ_{1t} are non-orthogonal, $\hat{\beta}_0^{ols} = \beta_0 + \frac{\lambda}{1+\lambda^2}$, $\hat{\alpha}^{ols} = \alpha$, $\hat{\beta}_1^{ols} = \beta_1 - \frac{\gamma_1\lambda}{1+\lambda^2}$.

$$\Omega = \begin{pmatrix} (1+\beta_0\lambda)^2+\beta_0^2+\beta_0^2\gamma_2^2 & (1+\beta_0\lambda)\lambda+\beta_0+\beta_0\gamma_2^2 & \beta_0\gamma_2 \\ (1+\beta_0\lambda)\lambda+\beta_0+\beta_0\gamma_2^2 & 1+\gamma_2^2+\lambda^2 & \gamma_2 \\ \beta_0\gamma_2 & \gamma_2 & 1 \end{pmatrix}$$

Using (2.19) as a data generating mechanism, 1000 samples of various sizes have generated where $\beta_0 = 1, \alpha = 0.25, \beta_1 = 0.5, \lambda = 1(-1), \gamma_1 = 0.8, \gamma_2 = 0.9, \gamma_3 = 1$, and its parameters estimated by Instrumental Variable method; the results are shown in the table (3). A glance at the estimated parameters confirms the quality, even in the small sample sizes, of the above established framework for using instrumental variable method in time series models. Table (4) considers the same parameter's values for (2.19) except for sample size that is T=200 and $\lambda = 1$, and assumes three likely specifications for regression of y_t on x_t, ***1)*** Static model, $y_t = \beta_0 x_t + \epsilon_t$ ***2)*** Dynamic but misspecified model, $y_t = \alpha_1 y_{t-1} + \beta_0 x_t + \epsilon_t$ and ***3)*** Dynamic and well-specified model, $y_t = \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$. The estimated coefficients of each model and specification tests of the underlying (2.18) and (2.16) equations of each model are given in table (4). The results of the static and mis-specified dynamic models in table(4) emphasize the guiding role played by (2.18) and (2.16) in the process of IV estimation. Correlation tests of both (2.18) and (2.16) for static model reject the null hypothesis of non-independence in almost 100 percent of trials. Misspecified dynamic model is also inappropriate, because non-independence of (2.18) has been rejected in almost 90 percent of replications. One may argue that the R^2 of (2.18) could lead the modeler to the right specification. As McGuirk and Driscoll (1995) argued R^2 is a misleading association measure when data are trended, in fact, the population R^2 for the above parametrization is 0.288 and all estimated R^2s are highly biased.

Sample Size	$\lambda = -1$				$\lambda = 1$			
	α_1	β_0	β_1	σ^2	α_1	β_0	β_1	σ^2
$T = 50$	0.249 (0.07)	1.05 (0.25)	0.48 (0.23)	1.06 (0.32)	0.23 (0.11)	0.96 (0.23)	0.55 (0.27)	1.05 (0.28)
$T = 100$	0.25 (0.04)	1.01 (0.12)	0.49 (0.13)	1.02 (0.15)	0.24 (0.06)	0.98 (0.12)	0.52 (0.17)	1.02 (0.15)
$T = 250$	0.25 (0.02)	1.00 (0.07)	0.50 (0.07)	1.00 (0.09)	0.24 (0.04)	0.99 (0.07)	0.51 (0.10)	1.00 (0.08)
$T = 500$	0.25 (0.02)	1.00 (0.05)	0.50 (0.05)	1.00 (0.06)	0.25 (0.03)	1.00 (0.05)	0.50 (0.07)	1.00 (0.06)

Table 2.3: Mean (standard deviation) of α_1, β_0, β_1, and σ^2

4-Weak instrument: Although, Staiger and Stock (1997) argue that the first stage F-statistic, the number of instruments, and the amount of bias in the least square estimates are the only determinants of the IV estimators properties, ***specification*** of (2.8) and (2.10) must be the first concern of practitioners. To illustrate that, consider the following regression

$$y = constant + \underset{(0.115)}{0.594}\mathbf{x} + \hat{u} \qquad F - statistic = 1.15$$

which is the average of 1000 regressions of y on $\mathbf{x}$, where $\mathbf{x}$ has been instrumented by $\mathbf{z}$. For each regression 100 observations of $(y, \mathbf{x}, \mathbf{z})$ were generated by (2.5), where $\beta = \lambda = \theta_1 = \theta_2 = 1, E(y) = E(\mathbf{z}) = 0, E(\mathbf{x}) = 1 + 0.4t$. According to Staiger and Stock (1997), $\mathbf{z}$ is a weak instrument because the F-statistic is lower than 10. A closer look at the specification assumption of the underlying (2.8) and (2.10) models reveals that the estimated model is, indeed, misspecified. Independence assumption in (2.8) and (2.10) and normality in (2.10) have been rejected in 100% of trials. By

Model	Static Misspecified		Dynamic Misspecified		Dynamic Well-Specified	
β_0	0.879(0.39)		0.99(0.09)		0.99(0.08)	
α	—		0.49(0.04)		0.25(0.05)	
β_1	—		—		0.51(0.12)	
σ^2	4.03(1.05)		1.09(0.123)		1.01(0.10)	
Test	(2.18)	(2.16)	(2.18)	(2.16)	(2.18)	(2.16)
R-square	0.06		0.81		0.85	
Normality:						
Jarque -Bera	0.20	0.05	0.05	0.06	0.05	0.05
Doornik-Hansen	0.22	0.05	0.04	0.06	0.05	0.06
Dependence:						
Breusch-Godfrey(1)	1.0	0.99	0.89	0.10	0.05	0.04
Breusch-Godfrey(2)	1.0	0.97	0.83	0.09	0.05	0.04
White's Homosced.	0.05	0.05	0.04	0.06	0.05	0.06
RESET(2)	0.06	0.07	0.04	0.05	0.04	0.05
Joint Mean:						
Trend in mean	0.06	0.08	0.02	0.05	0.05	0.05
RESET(2)	0.05	0.04	0.03	0.06	0.04	0.06
Correlation	0.17	0.98	0.88	0.11	0.05	0.04
Joint Variance:						
Trend in variance	0.05	0.05	0.06	0.05	0.06	0.05
RESET(2)	0.04	0.04	0.05	0.04	0.05	0.04
ARCH(1)	0.98	0.09	0.08	0.04	0.04	0.04

Table 2.4: Estimated model and the % of rejection of their specification tests ($\alpha = 0.05$)

re-specifying the model to include a trend, one can obtain the following (average) model:

$$y = constant + \underset{(0.144)}{0.987\mathbf{x}} - \underset{(0.05)}{0.395t} + \hat{u} \qquad F - statistic = 4587$$

which does not suffer from the so-called weak instrumental variable bias.

2.4 Relevancy

As argued earlier an appropriate instrumental variable must satisfy two properties of *Redundancy and Relevancy.* Traditionally, relevancy may be readily tested by examining the fit of a regression of endogenous regressor on the instrumental variable(s). F-static and R^2 are among the most recommended goodness of fit measures that usually utilize to assess the relevancy of the instruments. The higher R^2 or F-statistic indicates the more relevant instrumental variable(s). Partial R^2 suggested by Bound, Jaeger, and Baker (1995) is a measure that should be used instead, where along x_t there exist other exogenous variables in (3.8). However, all of these measures are applicable only in the single endogenous regressor models. For models with multiple endogenous variables, Shea (1997) has proposed a version of partial R^2 that considers the possible correlation among the instruments. Godfrey (1999) in a clarification note on Shea's paper has demonstrated that Shea's statistic may be expressed as $PR_i^2 = \frac{(X'X)_{ii}^{-1}}{(X'P_ZX)_{ii}^{-1}}$, where $(X'X)_{ii}^{-1}$ and $(X'P_ZX)_{ii}^{-1}$ refer to the ith element of $(X'X)^{-1}$ and $(X'P_ZX)^{-1}$, respectively. Although, standard R^2 is a number between 0 and 1, the upper bound of R^2 in (2.10) is an unknown and variable number, yet less than one. Specifically, consider the following relation

$$\beta_{LS} = \frac{Cov(y,x)}{Var(x)} = \beta_0 + \frac{Cov(\mathbf{x},\mathbf{z})}{Var(\mathbf{x})}\gamma_0 \tag{2.20}$$

where β_0 and γ_0 are defined in (2.8).

Proof:

$$\beta_0 + \frac{Cov(\mathbf{x},\mathbf{z})}{Var(\mathbf{x})}\gamma_0 = \frac{Cov(y,x|z)}{Var(x|z)} + \frac{Cov(\mathbf{x},\mathbf{z})}{Var(\mathbf{x})}\frac{Cov(y,z|x)}{Var(z|x)} =$$

$$\frac{Var(z)Cov(y,x) - \frac{Cov(x,z)^2}{Var(x)}Cov(y,x)}{Var(z)Var(x) - Cov(x,z)^2} = \frac{Cov(y,x)}{Va(x)} = \beta_{LS}$$

Considering R^2_{xz} as $\frac{Cov(\mathbf{x},\mathbf{z})^2}{Var(\mathbf{z})Var(\mathbf{x})}$, the instrumental variable estimator β_{IV} can be rewritten as

$$\beta_{IV} = \frac{Cov(y,\mathbf{z})}{Cov(\mathbf{x},\mathbf{z})} = \beta_0 + \frac{Cov(\mathbf{x},\mathbf{z})}{Var(\mathbf{x})}\frac{\gamma_0}{R^2_{xz}} \tag{2.21}$$

which implies as R^2_{xz} approaches to one, the instrumental variable estimator moves toward least square estimator. This outcome virtually never happens; the condition of positive definite covariance matrix for $(y,\mathbf{x},\mathbf{z})$ does not allow R^2_{xz} to be greater than q, where $0 < q < 1$. For example, in the presented model by (2.5) the upper limit of R^2_{xz} is $q = \frac{1}{1+\lambda^2}$, which is unknown in practice, and is also independent of the instruments. The lack of a given upper bound makes it difficult for practitioners to judge instruments solely in terms of their correlations with endogenous regressors. Moreover, sample R^2 is a biased estimator of its population counterpart, in general. Given redundancy, condition (2.7) shows a simple way to test relevancy. Relevancy can be simply tested by testing $H_0 : \gamma_0 = 0$ in $y = \beta_0\mathbf{x} + \gamma_0\mathbf{z} + \epsilon$. The other advantages of this measure over already mentioned measures are its simplicity, and known sample distribution. In the following a series of experiments were undertaken

Pop. R^2	$\hat{\beta}_{iv}$	Emp. size of $t_{\hat{\gamma}_0}$	R^2	F-statistic
Model 1:				
0.001	0.440(-3.490)	0.08	0.005(0.01)	0.54(1.2)
0.05	0.013(0.265)	0.65	0.05(0.06)	5.32(6.6)
0.15	0.003(-0.038)	0.99	0.15(0.16)	17.9(19.1)
0.25	-0.006(-0.021)	1.00	0.25(0.25)	33.1(34.8)
Model 2:				
0.001	0.360(0.880)	0.11	0.005(0.01)	0.53(1.24)
0.05	0.007(-0.002)	1.00	0.05(0.06)	5.23(6.61)
0.15	0.006(-0.032)	1.00	0.15(0.15)	17.9(19.3)

Table 2.5: Median (mean) of the estimated $\hat{\beta}_{iv}$, sample R^2_{xz}, F-statistic and the empirical size of the introduced relevancy test ($\alpha = 0.05$)

to investigate the empirical distribution of $\hat{\beta}_{iv}$, t-statistic of $\hat{\gamma}_0$, and goodness of fit measures for regression of x_i on z_i.

Experiment 1: The data were generated by (2.5). The parameters for model 1 are: $\lambda = 1, \beta = 0, \theta_2 = 1$, and for model 2: $\lambda = 2, \beta = 0, \theta_2 = 1$. θ_1 varies across experiments to generate instrument z, with different population correlation coefficient with the endogenous regressor x. Each experiment consists of 1000 trials, and in each trial 100 observation were generated. Note that, the maximum possible R^2 for model 1 and model 2 are 0.5 and 0.2, respectively. Table (5) shows that while an instrument with mild correlation with $\mathbf{x}$ ($R^2 \approx 0.05$, F-statistic around 6) is perfect for model 2, the appropriate instrument in the model 1 must be more correlated with x ($R^2 \approx 0.15$, F-statistic around 18). Due to unknown λ in practice, R^2 and F-statistic have limited applicability , but test of $H_0 : \gamma_0 = 0$ can– as shown by its empirical size– help practitioners to choose instruments carefully.

Experiment 2: Let $\lambda = \beta = \theta_1 = \theta_2 = 1$, but consider another instrument for $\mathbf{x}$ defined as $\mathbf{z}^+ = \theta_1\mathbf{x} + \theta_2\epsilon_1$. Clearly, $\mathbf{z}^+$ is not uncorrelated with the error term,

$Pop.R^2$	$\hat{\beta}_{iv}$	Emp. size of $t_{\hat{\gamma}_0}$	R^2	F-statistic
0.001	1.50(0.329)	0.06	0.005(0.01)	0.50(1.21)
0.20	1.50(1.50)	0.06	0.20(0.20)	25.11(26.56)

Table 2.6: Median(mean) of the estimated $\hat{\beta}_{iv}$,sample R^2_{xz}, F-statistic and empirical size of the introduced relevancy test($\alpha = 0.05$).

$E(\mathbf{z}^{+}u) \neq 0$. In the context of the existent literature there is not an objective way to distinguish between $\mathbf{z}$ and $\mathbf{z}^{+}$. Table (6) shows that in 94 percent of trials[12] and regardless of $Cov(\mathbf{x}, \mathbf{z}^{+})$ the introduced relevancy measure can readily identify the invalid instrument, $\mathbf{z}^{+}$.

In this regard the other example would be the rank of $\mathbf{x}$ as an instrumental variable that suggested by Sargan(1954) (see Judge, Griffiths, Hill, Lutkepohl, and Lee (1985)). Because rank of x doesn't convey any information more than $\mathbf{x}$, it cannot be used as instrument even though it has good correlation with $\mathbf{x}$. The same may apply to the study of the rate of return to education by Angrist and Krueger (1991), where quarter of birth used as an instrumental variable. It seems, given years of schooling, quarter of birth is not enough informative to describe the stochastic behavior of the earnings. This has confirmed by insignificant differences between the IV and LS estimates of the rate of return to education in all different estimated models.

2.5 Conclusion

The main objective of this chapter has been to study the statistical foundation of the instrumental variables method. The chapter extracts two properties in the underlying

[12] The data was generated using the joint distribution of the first two equations of the (2.4) and $\mathbf{z}^{+}$ as defined in the text. Each trial has performed by 100 observations, and trials repeated 1000 times.

joint distribution of the instrumental variables model. It shows that behind every non-orthogonal structural equation, there is a system of independent and orthogonal equations. Specification of the IV regression depends entirely on the specification of the underlying system of equations. The chapter uses one of the properties of the joint distribution function of the IV models to test instruments relevancy.

Chapter 3

Using the IV method to estimate omitted variable models:

An application of the IV method to estimate the rate of return to education in Iran

3.1 Introduction

Many economic models are formulated in terms of variables that cannot be observed or measured directly. The statistical estimation of these models is always a difficult task for researchers because a statistical model requires the specification of the form of the relationship between observable variables. In some cases, there exist good proxy variables that could be replaced with unobservable variables. However, in many cases researcher has to consider a specification that excludes the unobservable variable. The primary effect of having omitted variables in a statistical model is bias and inconsistent statistical inference. The IV method is the widely used method to estimate models with omitted variables. In the IV literature, usually, instruments are defined in terms of the model disturbances. Since disturbances are unobservable, the only practical guidance for selection of a good instrument is a variable that is highly correlated with the observable variables of the model. Section 2 describes the IV method. It considers the parametrization of a well-specified three-variable statistical model and introduces the appropriate instruments that could be used if a variable of the model is omitted. Two extra conditions for a good instrument are introduced in section 2. One of them is empirically testable, which could help to reduce the uncertainty involved in the selection of instruments. Section 3 reviews the literature of the rate of return to education. The models of returns to education are a classical example of models with omitted variable. This section surveys the usual instruments that are used to estimate returns to schooling. Section 4 is an application of the IV method to estimate the rate of returns to education in Iran. In search for an appropriate instrument for schooling, this section concludes with the year of birth of individuals. It argues that the year of birth is correlated with schooling because of the Cultural Revolution in 1980. The Cultural Revolution, as a natural experiment, has affected the education of four identical groups of Iranians differently. These groups are distinguishable only by their year of birth. Section 4 shows that the year of birth barely satisfies all conditions to be a good instrument. Section 5 contains the main conclusions.

3.2 Omitted variable models and the IV method

Consider the following regression model

$$W = \beta S + \gamma A + e \tag{3.1}$$

Where S and A are $n \times 1$ vector of observations; β and γ are the parameters to be estimated. The disturbance term, e, is assumed to be uncorrelated with S and A. If one has access to the error-free observations for W, S, and A then the estimators for β and γ can be defined by[1]

$$\hat{\beta} = \frac{Cov(W,S|A)}{Var(S|A)} \tag{3.2}$$

$$\hat{\gamma} = \frac{Cov(W,A|S)}{Var(A|S)} \tag{3.3}$$

Such that $E(\hat{\beta}) = \beta$ and $E(\hat{\gamma}) = \gamma$. Let assume, however, that the researcher lacks data on the variable A and propose to estimate

$$W = \beta S + v \tag{3.4}$$

[1]Covariance of A and B given C is defined as

$$Cov(A,B|C) = Cov(A,B) - Cov(A,C)Var(C)^{-1}Cov(B,C)$$

.

In that case, the estimator of β will be

$$\hat{\beta}_{OLS} = \frac{Cov(W,S)}{Var(S)} = \beta + \gamma\frac{Cov(S,A)}{Var(S)} + \frac{Cov(S,e)}{Var(S)} \tag{3.5}$$

Since, in general, $\gamma \neq 0$, $Cov(S, A) \neq 0$, and $Cov(S, e) = 0$, one has $E(\hat{\beta}_{OLS}) \neq \beta$ which is to say that $\hat{\beta}_{OLS}$ is a biased estimator.

Instrumental variable(IV) method is, by far, the most widely used statistical method to deal with the inconsistency problem of least square estimator in the omitted variable models. The IV estimator of β is defined by

$$\hat{\beta}_{IV} = \frac{Cov(W,Z)}{Cov(S,Z)} \tag{3.6}$$

where Z as an instrumental variable must have certain properties. As might be expected, the main object of the instrumental variable method is to estimate β, consistently. In other words, the selection of instrument is restricted by the following equality

$$\frac{Cov(W,Z)}{Cov(S,Z)} = \frac{Cov(W,S|A)}{Var(S|A)}$$

One can show that the necessary conditions for Z to have a consistent IV estimator are[2]:

$$Cov(W, Z|S, A) = 0 \tag{3.7}$$

$$Cov(A, Z) = 0 \tag{3.8}$$

[2]See appendix 1 for proof.

Moreover, IV estimator is defined only if

$$Cov(S, Z) \neq 0 \tag{3.9}$$

In addition to the separate effect of each condition (3.7–3.9) on determining a good instrument, their combination has also some consequences for instrument selection. The following implications stem immediately from (3.7–3.9)

$$Cov(A, Z|S) \neq 0 \tag{3.10}$$

$$Cov(W, Z|S) \neq 0 \tag{3.11}$$

To show them consider (3.10), which is

$$Cov(A, Z|S) = Cov(A, Z) - Cov(A, S)Var(S)^{-1}Cov(S, Z) \tag{3.12}$$

Given (3.8) and (3.9), it is clear that $Cov(A, Z|S)$ can not be zero. Now consider (3.7)

$$\begin{aligned} Cov(W, Z|S, A) &= Cov(W, Z|S) - Cov(W, A|S)Var(A|S)^{-1}Cov(A, Z|S) \\ &= Cov(W, Z|S) - \gamma Cov(A, Z|S) = 0 \end{aligned}$$

which, given (3.10), implies (3.11) as a necessary condition. The above argument brought out (3.10) and (3.11) as two implicit conditions of the appropriate instrumental variable that are not usually taken into account when researchers use instrumental variable method to resolve omitted variable bias. Condition (3.8) states that the appropriate instrumental variable should not be correlated with the omitted variable A. However, (3.10) requires non-zero correlation between A and Z given specified values for S. Condition (3.11) specifies that when A is omitted from the model the instrumental variable must convey some information regarding the stochastic behavior of W, yet be uninformative in the presence of A by (3.7). If (3.11) does not hold IV estimator is a biased estimator, and the bias is in the same direction as the least square estimator:

$$Cov(W, Z|S) = 0 \implies Cov(W, Z) = Cov(W, S)Var(S)^{-1}Cov(S, Z)$$

$$\hat{\beta}_{IV} = \frac{Cov(W, Z)}{Cov(S, Z)} = \frac{Cov(W, S)Var(S)^{-1}Cov(S, Z)}{Cov(S, Z)} = \frac{Cov(W, S)}{Var(S)} = \hat{\beta}_{OLS}$$

Fortunately, this condition can be readily tested by significance of the coefficient of Z in a regression of W on S and Z.

An explanatory point seems necessary at this point. Condition (3.7) specifies that the instrument should not have a direct effect on W in the presence of S and A. In other words, in a theoretical regression of W on S, A, and Z the coefficient of Z must be zero. The insignificance of Z in the model is conditional to the presence of S and A and is not general. In the presence of A, the conditional correlation of W and Z given S will be offset by conditional correlation of A and Z given S adjusted by γ, such that Z, in the end, remains un-informative for W. Due to indeterminacy of (3.9), condition (3.11) must be complemented with (3.9) to identify irrelevant instruments.

There exists another possibility for IV and least square estimators to be equal. Suppose we have a candidate instrument Z with the properties that $Cov(W, Z|S, A) = 0$, $Cov(S, Z) \neq 0$ but $Cov(A, Z) \neq 0$, so Z is an invalid instrument. Then

$$\hat{\beta}_{IV} = \beta + \gamma\frac{Cov(Z, A)}{Cov(Z, S)} \tag{3.13}$$

Now, if $Cov(A, Z|S) = 0$ then

$$Cov(A, Z) = Cov(A, S)Var(S)^{-1}Cov(Z, S)$$

$$\frac{Cov(A, Z)}{Cov(Z, S)} = \frac{Cov(A, S)}{Var(S)}$$

which means (3.5) and (3.13) are equal. However, in general, an invalid IV estimator is not equal to least square estimator.

The above argument has introduced conditions (3.7–3.11) as the necessary conditions for Z to be considered as a valid instrument. Among them, (3.10) and (3.11) are secondary conditions; that is to say if (3.7-3.9) hold then (3.10) and (3.11) are true . Since A is not observable, one can not be confident that a given Z satisfies all the primary conditions (3.7-3.9). The non-verifiability of (3.7) and (3.8) renders the implementation of the IV method as a central concern of modern applied microeconomics, because the choice of instruments depends, entirely, on the theoretical considerations and the modeler's subjective beliefs. Using secondary conditions could help a researcher to reduce the risk of employing a bad instrument. Having one more testable condition will improve the chance of selecting an appropriate instrument.

Numerical Example: Consider $\omega_i = (W_i, S_i, A_i, Z_i)'$, which is a normally distributed vector of random variables with mean zero and covariance matrix Σ, where

$$\Sigma = \begin{bmatrix} 7.3125 & 2.65 & 2.795 & 0.8 \\ 2.65 & 2 & 0.7 & 1 \\ 2.795 & 0.7 & 1.49 & 0 \\ 0.8 & 1 & 0 & 2 \end{bmatrix}$$

Let W_i be the random variable whose behavior is of interest. The population regression function for W_i would be

$$W_i = 0.8S_i + 1.5A_i + e_i \tag{3.14}$$

Assume, for some reason, a researcher has to estimate

$$W_i = \beta S_i + v_i \tag{3.15}$$

In this context, Z could serve as a valid instrument to estimate β because Z holds all necessary conditions, (3.7–3.11).

To investigate the consistency of the IV estimators, a thousand samples of 100 observations of $\omega_i = (W_i, S_i, A_i, Z_i)'$ were generated by the joint distribution of ω_i, and, for each sample $W_i = \beta S_i + v_i$ was estimated by the IV method using Z_i as instrument. The empirical distribution of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ are presented[3] in table 3.1. In addition, table 1 reports the empirical distribution of t-statistic of $\hat{\delta}$ from the estimation of $W_i = \theta S_i + \delta Z_i + u_i$. This t-statistic can be used to test condition(3.11).

Table 3.1 shows that the empirical distributions of $\hat{\beta}_{IV}$ is concentrated around the population value of β. However, the concentration point of $\hat{\beta}_{OLS}$ is far away from the true value of β. The empirical distribution of $t_{\hat{\delta}}$ indicates that $Cov(W, Z|S)$ is not equal to zero.

Now consider $\omega_i^* = (W_i, S_i, A_i, Z_i)'$, as a normally distributed vector of random

[3]The STATA program is available in appendix 2.

variables with mean zero and covariance matrix Σ^*,

$$\Sigma^* = \begin{bmatrix} 7.3125 & 2.65 & 2.795 & 2.65 \\ 2.65 & 2 & 0.7 & 2 \\ 2.795 & 0.7 & 1.49 & 0.7 \\ 2.65 & 2 & 0.7 & 3 \end{bmatrix}$$

The corresponding population regression function of ω_i^* is

$$W_i = 0.8S_i + 1.5A_i + e_i \tag{3.16}$$

Obviously Z is not a valid instrument to estimate β. The variable Z satisfies condition (3.7) and has non-zero correlation with S, but does not hold conditions (3.8), (3.10), and (3.11). Since, in practice, (3.7) and (3.8) are unverifiable one may be tempted by

$$R^2_{SZ} = 1 - \frac{Var(S) - Cov((S,Z)Var(Z)^{-1}Cov(Z,S)}{Var(S)} = 0.66 \tag{3.17}$$

to employ Z as a valid instrument. Table 3.2 shows the empirical distribution of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ from a thousand samples of 100 observations of ω_i^*. For each sample $W_i = \beta S_i + v_i$ was estimated by the IV method using Z_i as instrument. It is clear that the concentration point of the empirical distribution of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ are far away from the population value of β. In fact the empirical distribution of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ are similar. In the context of the primary conditions (3.7–3.9), it is impossible to identify Z an an invalid instrument. However, the empirical distribution of $t_{\hat{\delta}}$ indicate that in 95 percent of samples Z is not a good instrument because the coefficient of Z in a regression of W on S and Z is insignificant.

3.3 Returns to education: Literature review

This section reviews two general framework to understand distortions in returns to schooling and reexamines the mostly used instruments to estimate the rate of return to education. The first model, developed by Card (1999), provides an analytically tractable framework to analyze the classical model of human capital of Becker (1967). Card's model considers an individual who chooses an optimal amount of schooling S, to maximize her utility function

$$U(S, W) = logW - h(S) \tag{3.18}$$

where h(S) is an increasing convex function and W refer to the level of earning associated to schooling by W=W(S). The first order condition of the individual's optimization problem implies equality between marginal benefit and marginal cost of schooling

$$\frac{W'(S)}{W(S)} = h'(S) \tag{3.19}$$

A simple way to appreciate individual heterogeneity is to let

$$\frac{W'(S)}{W(S)} = \beta \tag{3.20}$$

$$h'(S) = r_i + kS \tag{3.21}$$

where r_i is a random variable with mean $\overline{r}$, and both β and k are constant parameters. Equation (3.21) is justified by imperfect financial markets or by different tastes for schooling such that higher r_i is associated with individuals with higher difficulties to finance schooling or with lower preferences for studying.

Considering equations (3.20) and (3.21), the optimal schooling would be

$$S_i = \frac{\beta - r_i}{k} \tag{3.22}$$

and wage equation can be obtained by integrating equation (3.20)

$$log(W_i) = \alpha_i + \beta S_i \tag{3.23}$$

where α_i is a person-specific constant of integration and refers to the capacity of individuals to earn different wages at the given level of schooling. Equations (3.22) and (3.23) constitute a system of equations for determining schooling and earnings in terms of the underlying random variables α_i and r_i. Equation (3.23) indicates that individuals with the same schooling level might have different earnings. This difference can be modeled by

$$\alpha_i = \delta_0 + \gamma A_i \tag{3.24}$$

where A_i, *"Ability"*, is a general index for talent, motivation and any other individual's attribute that might affect wages. Using (3.24), the wage equation can be rewritten as

$$log(W_i) = \delta_0 + \beta S_i + \gamma A_i \tag{3.25}$$

The stochastic form of this equation, known as Mincerian earning function, has been used extensively by many researchers in the context of human capital theory. The Mincerian earning function has been derived by Blundell, Dearden, and Sianesi (2003) in a different fashion. They consider the problem of measuring the impact of education on earnings in the context of the causal inference literature. In this framework any given individual faces a finite and exhaustive set of highest attainable schooling levels(treatments). Let $S_{ij} = 1$ to show that the j is the highest schooling level attained by individual i where $j \in (0, 1, 2, ..., J)$, and W_i^j as the potential (log) earning of individual i if $S_{ij} = 1$.The problem of estimating the rate of return to education can be summarized as evaluating $W_i^j - W_i^{j-1}$, averaged over some population of interest. However, the fundamental problem of causal inference (see Holland (1986)) is the impossibility of observing W_i^j and W_i^{j-1} simultaneously. The observed earning of individual i can be written as

$$W_i = W_i^0 + \sum_{j=1}^{J} (W_i^j - W_i^0) S_{ij} \tag{3.26}$$

Let W_i^j be, in an additive manner, a function of observable individual's attributes X_i, and unobservable factors U_i^j, which are specific to the individual and schooling level

$$W_i^j = m_j(X_i) + U_i^j \tag{3.27}$$

where $E(W_i^j | X_i) = m_j(X_i)$. Assume that an individual's potential earning and schooling level are independent from the schooling choices of other individuals in the population and let

$$U_i^j = \alpha_i + \epsilon_i + b_{ij} \tag{3.28}$$

where α_i represents the unobservable individual attributes, ϵ_i is the standard error term, and b_{ij} shows the individual-specific unobserved component of earning associated to the schooling level j. Using equations (3.27) and (3.28) and by normalizing b_{i0} to 0, observed earning equation (3.26) becomes

$$W_i = m_0(X_i) + \sum_{j=1}^{J} \beta_{ij} S_{ij} + \alpha_i + \epsilon_i \tag{3.29}$$

with $\beta_{ij} = (m_j(X_i) - m_0(X_i)) + b_{ij}$. In the homogeneous returns framework where $\beta_{ij} = \beta_j$ equation (3.29) becomes

$$W_i = m_0(X_i) + \sum_{j=1}^{J} \beta_j S_{ij} + \alpha_i + \epsilon_i$$

By assuming that each additional year of schooling has the same marginal return $\beta_{ij} = j\beta$, one can get

$$W_i = m_0(X_i) + \beta S_i + \alpha_i + \epsilon_i$$

which is equivalent to equation (3.23).

The empirical estimation of the Mincerian earning function is linked to econometrics of the omitted variable model because, in practice, ability is a non-observable variable and must be omitted from the regression function. As argued in section 1, the IV method is a way to cope with the inconsistency problem of the omitted variable model. To employ the IV method, one has to find an instrument, Z, correlated with S, uncorrelated with A and given A and S uncorrelated with W. The first condition is the only one that can be analytically argued and practically measured. Based on Card's model, correlation between Z and S is possible only through r_i. In other words, an appropriate instrument must be correlated with individual's "attitude toward schooling" or individual's "access to funds" (family wealth). Some studies use family background information–such as mother's and father's education–as instrumental variable for education. Table 3.3 uses NLSY79-2000 database[4] to investigate the correlation between children's discount rate (tastes for education) and their parents education. In a hypothetical question, respondents were asked if they were ready to accept an offer for a full time job at 2.5, 3.5, and 5 dollar per hour.

[4]This data set and its documentations are publicly accessible from the web-site of Bureau of Labor Statistics: http://www.bls.gov/nls/home.htm.

The job offered was dish washing which, of course, is not an interesting job for youth. Table 1 considers those who were enrolled in regular schools at 1979 and shows their (average) education as well as their mother's and father's (average) education in 2000. As table 3 indicates, children of higher-educated parents have lower discount rates and lower marginal dis-utilities of schooling. Therefore, parent's education are correlated with schooling of children.

However, there exist some theoretical criticism against using family background variables as instruments for schooling. According to these studies, given that ability is partially inheritable and directly affected by family background, children of higher-educated parents– who presumably have also higher ability and higher income– are more able than children who grew up in lower-educated families. Plug and Vijverberg (2003), using an intergenerational sample of families and by comparing biological and adopted children, conclude that ability is partly responsible for the education attainment of children, and that the largest part of ability for education is inherited. Carneiro and Heckman (2002), using AFQT[5] as a measure of ability, find that the correlation between parents education and child ability is significantly greater than zero (0.31). Surprisingly, parents' education has the highest correlation with ability among the instruments considered by Carneiro and Heckman (2002).

Beside the correlation of family background and ability, it is argued that family background has direct effect on wages. For instance, Becker and Tomes (1986) developed a human capital model that specifies a wage function containing family background variables as regressors. Montgomery (1991), in what he called *networking effect*, argued that high level families have friends and connections which enable them to find a good job for their children. Direct relation between wage and family background variables is in contradiction with (3.7) and does not allow to use family

[5]Armed Forces Qualification Test

background variables as instruments for education.

College proximity (see Card (1995)) is another variable that recently used as instrument for schooling. The idea for using college proximity as instrument comes from the fact that those who are growing up near a college can be educated much cheaper than people who are living far away from a school. The lower cost might comes in the form of saving in transportation cost and living with family. Obviously, this variable has a good correlation with Becker's interpretation of r_i as a measure of access to funds necessary to finance schooling. However, there are some convincing reasons, as mentioned by Card (1995), for the existence of a direct effect of college proximity on the level of earnings. Direct effect of instrument on wages is not allowed by (3.7), and disqualifies college proximity as a good instrument.

The recent studies, inspired by the causal inference literature, propose *Natural Experiment* as a new resource for the optimal instruments. The basic idea of the approach is to take advantage of a situation in which two identical groups (or time periods) are affected differently by a "natural" event that is exogenous to the relationship between W_i and S_i and causes a considerable change in both. In schooling studies, perhaps, the famous example of natural experiment is Angrist and Krueger (1991)), where quarter of birth was used as an instrument for education to estimate the rate or return to education for men. They find significant association between quarter of birth, schooling and earnings and argue that the observed associations are generated by compulsory school attendance law. According to compulsory school attendance law all children must to start the first grade in the fall of the year in which they turn 6 and to stay in school until their 16th birthday. Therefore, two boys who were randomly born at the different quarter of a same year may achieve different levels of schooling and, therefore, different levels of income. They used the observed association between quarter of birth, schooling and wages to justify quarter of birth as an appropriate instrument for schooling. They estimated the rate of return to schooling, using quarter of birth as an instrument, in a sample of 329,500 men born between 1930-1939 from the 1980 census. As an example, in one version of their model, there are 30 instruments created by interacting quarter and year of birth. Their IV estimation of schooling coefficient is 0.089, which is not significantly different from the ordinary least square estimation, 0.071 (see Angrist, Imbens, and Krueger (1999)). Despite the observed correlation, their study is criticized by Bound, Jaeger, and Baker (1995) as an application of weak instruments. In fact, the equality of IV and OLS estimates could be a result of bad instrument or non-existence of omitted variable bias. In this case, condition (3.11) provides another testable condition to examine quality of quarter of birth as an appropriate instrument. In a regression of the log of weekly wage on education, year effects and instruments none of the instruments are significant, which indicates that (3.11) is not satisfied.[6] It seems, given years of schooling, quarter of birth is not enough informative to describe the stochastic behavior of the earnings.

3.4 Returns to Education in Iran: Using the Cultural Revolution as a Natural Experiment

Despite the importance of the estimates of return to schooling in modern labor economies, there have been few attempts to estimate the rate of return to education in the Iranian economy. The available studies are also limited to either geographical areas or economic sectors. In addition, neither of them has addressed the endogeneity problem of education in the earning function. Henderson (1983) using data from

[6] Angrist and Krueger's data set is available on the data archive of the Journal of Applied Econometrics.

a 1975 socioeconomic survey has estimated the return to education only among the self-employed workers in Tehran. Sadegi (1999) has estimated the return to the education of heads of households in Isfahan, and Naderi and Mace (2003) have estimated the return to education in manufacturing sector. Salehi-Isfahani (2002) has used a nation-wide data set to estimate the returns of schooling; however, he has not considered the endogeneity of schooling. This section is intended to estimate the rate of return to education in Iran by the IV method. It utilizes year of birth as an appropriate instrument for years of schooling. It is argued that the cultural Revolution in 1980 generates a reasonable association between individual's year of birth and schooling.

3.4.1 Scope and Extent of the Cultural Revolution

In the recent history of Iran, the most important event regarding education is the Cultural Revolution in late April 1980. Following the overthrow of the Shah in 1979, Iran's education system underwent a process of Islamization to institutionalize the values of the leaders of the revolution. In the spring of 1980, Moslem activists demand's for a fundamental reorganization of the educational system led to closure of all higher education institutions for the subsequent 17 months. Ayatollah Khomeini then appointed the Cultural Revolution Panel (Setad-e Enqelab-e Farhangi), a board of Islamic educational experts, who were to restructure the educational system. The panel changed the curricula and student admission procedures and reopened the universities in October 1981.

As a result of the Cultural Revolution the schooling of two cohorts of Iranian citizens was delayed or cut short. The first cohort contains those who were in college when colleges shot down. The second cohort includes those who turned 17-18 years old during the Cultural Revolution.

After the Cultural Revolution ended, universities began restricted recruitment. New students drawn only from those who had Islamic or at least neutral political attitudes with known family backgrounds. Those who entered the university before the Cultural Revolution and had not completed their degrees were, partly, allowed to return and finish their education. However, some of them due to political attitudes and some due to family responsibilities never got the chance to go back to school. Figure 3-1 depicts the ratio of individuals with college degree to the total individuals by age for men and women using 1996 census data. It is readily discernible that those who were have 30-39 years old in 1996 are, relatively, less college educated than other age groups. This pattern is observable for both men and women, however the gender gap is going to turn in favor of women after the Cultural Revolution.

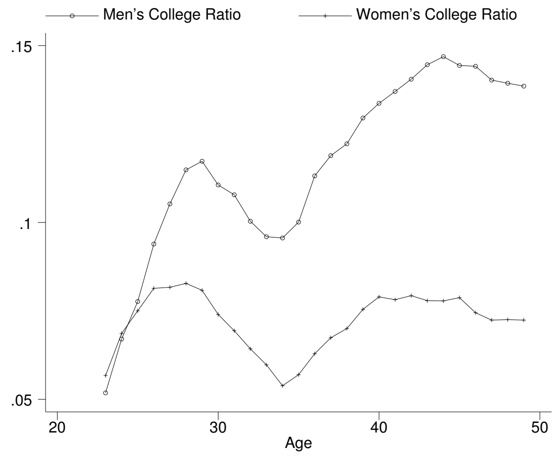


Figure 3-1: Percent with college education (census 1996)

3.4.2 Data and Empirical Results

This section utilizes data from the Iranian Socio-Economic Survey conducted by the Statistical Center of Iran in 2001. This is the most reliable collection of personal micro-data representative of the Iranian population. The Socio-Economic Survey is a longitudinal survey of the Iranian households randomly selected at 2001 and followed, subsequently, for four years. A total of 6960 households were interviewed, generating information on 30715 individuals. Among them 3064 individuals are wage and salary workers with the valid information regarding wage, hours of work, and education. Some descriptive statistics for the original survey and the sub-sample used in this study are presented in Table 4.

As is apparent from Table 3.4, the measure of education is not the actual years spent at school but the highest degree attained by the individual. These degrees include: Primary school, middle school, High School and College. The statutory number of years required to obtain a primary and middle school certificate is 5 and 8 years, respectively. For high school diploma, students must spend 4 more years in school. The statuary number of years required to complete college education varies with the field of specialization form 4 years in engineering and science to 6 years for medical studies. Figure 3-2 depicts the distribution of hourly wages (log) by education categories where the solid line connects the conditional means. Clearly, after middle school, wage increases linearly as education rises. The lower wage (mean) for those who did not complete a degree is an indicator for the existence of sheepskin effect in Iran's labor market.

To obtain years of education from levels, this study assigns 5 years to primary school, 8 years to middle school, and 11 and 12 years to high school diploma depending on the high school type. Finally, college degrees are considered as equivalent

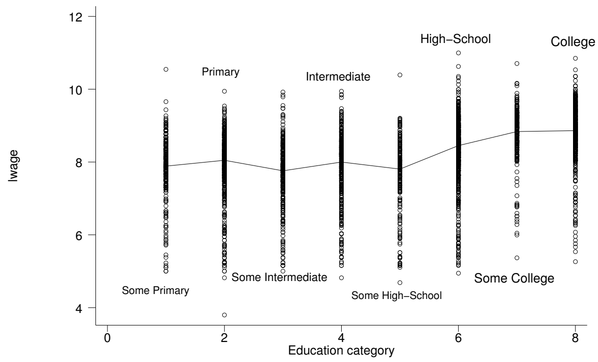


Figure 3-2: Wage (log) distribution and education categories

to 16 years of education. For those who have dropped out without the degree, the average statuary years of their lowest and uncompleted degree is considered as the years of education. It must be emphasized that the statutory number of years can be different from the actual number of years spent to obtain a degree for some individuals. Moreover, the lack of detailed data for college graduates does not allow to distinguish them by their field of specialization and even post graduate degrees.

Table 3.5 reports the OLS estimates of wage equation by four different specifications. Years of education (S), experience (Exp), Experience square ($Expsq$), area of residence ($Tehran$, $Urban$), sector of employment ($Private$) and ethnicity[7] are the explanatory variables used to explain the stochastic behavior of wages. Experience is measured by age minus age at the first job. In the first specification, only schooling, experience, experience square and area of residence are included. Second specification controls for ethnicity of individuals and third specification adds sector

[7]Iran is a nation of multi-language cultures: Persian, Azari, Kurdish, Lori, Baluchi, Armenian, and others.

of employment to the explanatory variables. To test for the existence of “Sheepskin effect”, fourth specification considers a binary variable (*Diploma*), which is equal to one for those who dropped out without a degree: some primary, some intermediate, some high school and some college. All of the estimates are significant and have the expected signs. The percentage increase in hourly earning (log) attributed to an additional year of schooling ranges from 9.3% to 8.2%. The significant and negative coefficient of *Diploma* in the fourth specification is an indication of the existence of “sheepskin” effect in Iran’s labor market. Given schooling and other control variables, a private sector worker’s wage is less than his/her public sector counterparts. In fact, this effect is much higher than -8.8%, because public sector workers are enjoying from variety of non-monetary benefits that do not reflect in their wages.

As discussed earlier, OLS estimate of the returns to education β is not consistent either because of measurement errors in schooling variable or because of omitted variable bias. To improve efficiency of the estimates, schooling must be instrumented by a variable that satisfies conditions 3.7–3.9. The instruments used in this study are a series of binary variables designed to exploit cohort effects due to the Cultural Revolution. In regard to the effect of the Cultural Revolution on education four age cohorts are distinguishable. The first cohort are those who were born before 1959. This cohort had a chance to graduate from college before the Cultural revolution in 1980. The second group are those who were born in 1959-1962. This group were in school during the Cultural Revolution. The third group are those who were turned to college age during the Cultural Revolution. These individuals were born from 1963 to 1965. Finally, the fourth cohort are those who were born after 1965. Like the first cohort, these individuals had a chance to have a normal education, albeit with different curricula and quality. Figure 3-3 shows how the education of each cohort was affected by the Cultural Revolution.[8] According to the figure 3, 32% of first cohort, 28% of second cohort, 23% of third cohort, and 27% of fourth cohort had a chance to go to college. Obviously, for no reason than the unlucky time of birth the education of second and third cohorts is less than other cohorts. The correlation between year of birth and years of schooling is, indeed, generated by the Cultural revolution. Since year of birth is a random variable, it, intuitively, could not be correlated with ability and in a theoretical view it is not also a determinant of wages. In this view, the year of birth satisfies conditions (3.7–3.9) and, therefore is appropriate instrument for schooling in a wage function.

Three dummy variables, defined as

$$\begin{aligned} Z_1 = 1 \;&:\; 1959 < \text{year of birth} \leq 1962,\ Z_1 = 0 : \text{ otherwise} \\ Z_2 = 1 \;&:\; 1963 \leq \text{year of birth} \leq 1965,\ Z_2 = 0 : \text{otherwise} \\ Z_3 = 1 \;&:\; \text{year of birth} > 1966,\ Z_3 = 0 : \text{otherwise} \end{aligned}$$

are used as instruments for years of schooling. Table 3.6 presents the IV estimates of the wage equation. First stage R^2, Partial R^2 and F-test of excluded instruments suggested by Bound, Jaeger, and Baker (1995) are reported at the bottom of the table for each specification. The relatively high first stage R-square and significant F-statistic are indication of good correlation between year of birth and years of schooling.

The IV estimates of the return to schooling ranges from 8.2% to 5.6% depending on the specification. The IV estimate of the "sheepskin effect" is higher and more significant than OLS estimate. As predicted by (3.5) the IV estimates of the return

[8]Revolution and war, as other natural events, are additional factors that may have impact on the education of Iranians. In fact, these factors are another reasons to use year of birth as instrument for schooling.

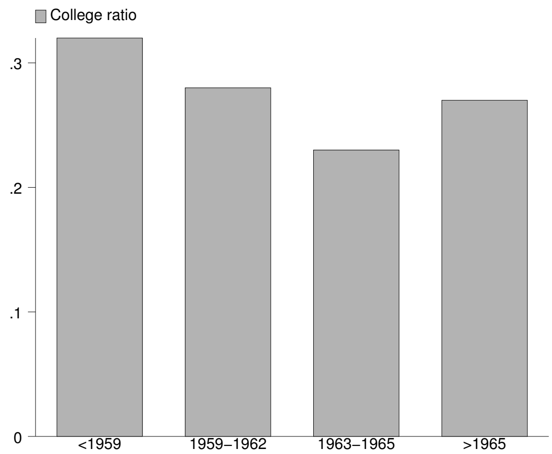


Figure 3-3: Percent with college education (Sample 2001)

to schooling is smaller than the OLS estimates. However, the difference is significant at $\alpha = 0.15$ as the level of significance. Similarly, in a test for condition (3.11), the coefficients of Z_1, Z_2, and Z_3 in a regression of hourly wages on schooling, control variables and instruments are 0.073(0.065), 0.056(0.08), and 0.10(0.06) respectively. The numbers inside the parenthesis are the estimated standard errors for the estimated coefficients. The high standard errors are consistent with the weak differences between the IV and OLS estimates of the rate of return to schooling.

3.5 Conclusion

This chapter argued the estimation of models with omitted variable by the IV method. The consistent IV estimation is conditioned to using appropriate instruments. Five condition has been derived for an appropriate instrument. This adds two extra conditions to the usual IV conditions. One of these extra conditions is empirically testable, which is useful to reduce the uncertainty in selection of instruments.

The chapter reviews the literature of returns to schooling and estimates the rate of return to education in Iran. In line with the literature, to estimate the unknown parameters of the earning function, the endogeneity of schooling is taken into account by using instrumental variable Method. It has been argued that the Cultural Revolution, as an exogenous supply shock to the schooling, establishes year of birth as a determining factor of education. The paper find that the IV estimates of returns to schooling, 5.6%, is less than OLS estimate 8.2%. However, the difference is barley significant. In addition, the paper reports evidence of sheepskin effect in Iran's labor market.

Table 3.1: Estimated percentiles for OLS and IV estimators (good instrument) with n=100.

Percentiles	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$t_{\widehat{\delta}}$
1	0.96	-0.15	-4.54
10	1.14	0.37	-3.52
50	1.32	0.78	-2.25
90	1.50	1.13	-1.03
99	1.65	1.41	-0.11
Mean	1.32	0.77	-2.22

Table 3.2: Estimated percentiles for OLS and IV estimators (bad instrument) with n=100.

Percentiles	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$t_{\widehat{\delta}}$
1	0.96	0.95	-2.27
10	1.13	1.11	-1.22
50	1.32	1.32	-0.03
90	1.51	1.54	1.20
99	1.56	1.72	2.17
Mean	1.33	1.32	-0.02

Table 3.3: Attitudes toward work and parent’s education in NLSY data.

Minimum acceptable wage	Education	Mother Education	Father Education
2.5	13.05	9.66	8.51
3.5	13.64	10.60	9.70
5.0	14.10	11.00	10.60
>50	14.55	11.28	11.00

Table 3.4- Sample Characteristics for Overall and this paper subset of Socio-Economic Survey of Iran 2001.

	Overall	Subset	
		Valid education	Valid education and wage
Education (%)			
Some Primary	--	22.11	7.77
Primary	--	14.85	14.56
Some Intermediate	--	15.51	10.38
Intermediate	--	10.92	14.23
Some High School	--	11.94	6.89
High School	--	16.85	23.37
Some College	--	2.24	7.34
College	--	5.58	15.47
Gender (%)			
Male	50.14	53.91	83.58
Female	49.86	46.09	16.42
Age Cohort (%)			
Older than 42	21.43	12.72	20.56
39-42	3.44	3.32	7.70
36-38	3.74	3.85	7.28
Younger than 36	71.40	80.11	64.46
Sector of Employment (%)			
Public	--	--	44.91
Private	--	--	55.09
Live in Tehran Province(%)	17.43	20.39	26.93
Live in urban area (%)	57.61	62.59	75.55
Sample Size	30,715	21,867	3,064

Table 3.5 – Estimated Regression Models (OLS) for Log Hourly Earnings

	1	2	3	4
1- Schooling	0.093	0.094	0.084	0.082
	(24.10)	(24.48)	(19.64)	(18.66)
2- Experience	0.09	0.09	0.086	0.086
	(22.45)	(22.71)	(21.45)	(21.38)
3- Experience Square/100	-0.154	-0.154	-0.148	-0.148
	(-14.09)	(-14.20)	(-13.74)	(-13.71)
4- Live in Tehran	0.106	0.136	0.14	0.137
	(3.17)	(4.05)	(4.18)	(4.07)
5- Live in Urban Area	0.095	0.08	0.076	0.077
	(2.59)	(2.21)	(2.08)	(2.13)
6- Work in Public Sector			-0.088	-0.088
			(-5.20)	(-5.24)
7- Diploma				-0.062
				(-1.93)
8- Ethnicity	No	Yes	Yes	Yes
R-Square	0.33	0.35	0.36	0.37

Table 3.6 – Estimated Regression Models (IV) for Log Hourly Earnings

	1	2	3	4
1- Schooling	0.083	0.083	0.060	0.056
	(6.68)	(6.67)	(3.46)	(3.10)
2- Experience	0.090	0.091	0.085	0.084
	(22.47)	(22.75)	(20.16)	(19.87)
3- Experience Square/100	-0.156	-0.156	-0.15	-0.15
	(-13.93)	(-14.06)	(-13.78)	(-13.75)
4- Live in Tehran	0.114	0.146	0.157	0.151
	(3.28)	(4.18)	(4.41)	(4.30)
5- Live in Urban Area	0.126	0.12	0.130	0.134
	(2.45)	(2.31)	(2.48)	(2.53)
6- Work in Public Sector			-0.131	-0.133
			(-3.83)	(-3.86)
7- Diploma				-0.105
				(-2.41)
8- Ethnicity	No	Yes	Yes	Yes
First stage R-Square	0.22	0.24	0.37	0.40
Partial R-Squared of Excluded instruments	0.097	0.097	0.061	0.06
Test of Excluded Instruments	109.37	109.34	66.71	63.77

3.6 Appendix 1

To show if $Cov(W,Z|A,S)=0$ then $\frac{Cov(W,S|A)}{Var(S|A)} = \frac{Cov(W,S|A,Z)}{Var(S|A,Z)}$ or

$$Cov(W,S|A)Var(S|A,Z) = Cov(W,S|A,Z)Var(S|A)$$

Let

$$Cov(W,S|A,Z) = Cov(W,S|A) - Cov(W,Z|A)V(Z|A)^{-1}Cov(Z,S|A)$$

$$Var(S|A,Z) = Var(S|A) - Cov(S,Z|A)V(Z|A)^{-1}Cov(Z,S|A)$$

Therefore,

$$\begin{aligned} Cov(W,S|A)Var(S|A) - Cov(W,S|A)Cov(S,Z|A)V(Z|A)^{-1}Cov(Z,S|A) &= \\ Var(S|A)Cov(W,S|A) - Var(S|A)Cov(W,Z|A)V(Z|A)^{-1}Cov(Z,S|A) \\ Cov(W,S|A)Cov(S,Z|A) = Var(S|A)Cov(W,Z|A) \end{aligned}$$

Now if $Cov(W,Z|A,S)=0$ which implies

$$Cov(W,Z|A) = Cov(W,S|A)V(S|A)^{-1}Cov(S,Z|A)$$

The above identity can be written as,

$$Cov(W,S|A)Cov(S,Z|A) = Cov(W,S|A)Cov(S,Z|A)$$

So

$$\frac{Cov(W,S|A)}{Var(S|A)} = \frac{Cov(W,S|A,Z)}{Var(S|A,Z)}$$

.

Now if $Cov(A, Z) = 0$ then

$$\beta = \frac{Cov(W, S|A, Z)}{Var(S|A, Z)} = \frac{Cov(W, S|A) + Cov(W, S|Z) - Cov(W, S)}{Var(S|A) + Var(S|Z) - Var(S)}$$

$$\beta \left(Var(S|A) + Var(S|Z) - Var(S)\right) = \left(Cov(W, S|A) + Cov(W, S|Z) - Cov(W, S)\right)$$

Considering

$$\beta = \frac{Cov(W, S|A)}{Var(S|A)} \rightarrow \beta Var(S|A) = Cov(W, S|A)$$

$$\beta Var(S|A) + \beta Var(S|Z) - \beta Var(S) = Cov(W, S|A) + Cov(W, S|Z) - Cov(W, S)$$

$$Cov(W, S|A) + \beta Var(S|Z) - \beta Var(S) = Cov(W, S|A) + Cov(W, S|Z) - Cov(W, S)$$

$$\beta Var(S|Z) - \beta Var(S) = Cov(W, S|Z) - Cov(W, S)$$

$$\beta = \frac{Cov(W, S|Z) - Cov(W, S)}{Var(S|Z) - Var(S)} = \frac{Cov(W, S) - Cov(W, S|Z)}{Var(S) - Var(S|Z)} = \frac{Cov(W, Z)}{Cov(S, Z)} = \beta_{IV}$$

3.7 Appendix 2

The STATA code used to generate Table 3.1 and 3.2.

```
set matsize 800
set more off
program drop _all
set obs 100
*mat sigma=[7.3125,2.65,2.795,0.8/2.65, 2,0.7,1/2.795,0.7,1.49,0/0.8,1,0,2]
mat sigma=[7.3125,2.65,2.795,2.65/2.65, 2,0.7,2/2.795,0.7,1.49,0.7/2.65,2,0.7,3]

program define iv
version 7.0
if "`1'" == "?" {
global S_1 "bols biv s2iv tz covsz"
exit
}
quietly {
set obs 100
drawnorm w s a z, cov(sigma)

reg w s
scalar bols=_b[s]
ivreg w (s=z)
scalar biv=_b[s]
scalar s2iv=e(rmse)^2
reg w s z
```

```
scalar tz=_b[z]/_se[z]
factor w s a z, pc cov
mat def asd=get(Co)
scalar covsz=asd[2,4]

    post `1' (bols) (biv) (s2iv) (tz) (covsz)
drop _all
}

    end
simul iv,reps(1000)
```

Bibliography

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

Angrist, J. D., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, CVI(4), 979–1014.

Becker, G. (1967): *Human Capital and the Personal Distribution of Income*. University of Michigan Press, Ann Arbor.

Becker, G., and N. Tomes (1986): "Human capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), S1–S39.

Blomquist, S., and M. Dahlberg (1999): "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics*, 14, 69–88.

Blundell, R., L. Dearden, and B. Sianesi (2003): "Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from NCDS," Discussion paper, IFS.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

Bowden, R. J., and D. A. Turkington (1984): *Instrumental Variables.* Cambridge University Press.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999): "Redundancy of Moment Conditions," *Journal of Econometrics*, 91, 89–111.

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour:Essay in Honour of John Vanderkampt*, ed. by E. Christofides, E. K. Grant, and R. Swidinsky. University of Toronto press, Toronto.

Card, d. (1999): "The Causal Effect of Education on Earning," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, chap. 30, pp. 1801–1863. North-Holland, Amesterdam.

Carneiro, P., and J. Heckman (2002): "The Evidence on Credit Constraints in Post-Secondary Schooling," *The Economic Journal*, 112, 705–734.

Carter, R. L., and W. A. Fuller (1980): "Instrumental Variable Estimation of the Simple Errors-in-Variables Model," *Journal of American Statistical Association*, 75(371), 687–692.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983): "Exogeniety," *Econometrica*, 51, 227–304.

Flores-Lagunes, A. (2000): "Tests for Instrument Relevance: A Monte Carlo Investigation and a New Bootstrap Approach," Unpublished paper.

Godfrey, L. G. (1999): "Instrument Relevance in Multivariate Linear Models," *The Review of Economics and Statistics*, 81(3), 550–552.

Hahn, J., and J. Hausman (2003): "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review*, 93(2), 118–125.

Han, C., and P. Schmidt (2001): "The Asymptotic Distribution of the Instrumental Variable Estimators when the Instruments are not Correlated with the Regressors," *Economics Letters*, 74(1), 61–66.

Henderson, J. W. (1983): "Earnings Functions For the Self-Employed:Comment.," *Journal of Development Economics*, 12(3), 97–102.

Hendry, D. F. (2000): "Econometric Modlling," Lecture Notes for Economic Modelling and Economic Forecasting, Economic Department, University of Oslo.

Holland, P. W. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–970.

Judge, G. G., R. Griffiths, C. Hill, H. Lutkepohl, and T. Lee (1985): *The Theory and Practice of Econometrics*. John Wiley & Sons, New York, second edn.

Maddala, G. S., and J. Jeong (1992): "On the Exact Small Distribution of the Instrumental Variable Estimator," *Econometrica*, 60(1), 181–183.

McGuirk, A. M., and P. Driscoll (1995): "The Hot Air in R^2 and Consistent Measures of Explained Variation," *American Journal of Agricultural Economics*, 77, 319–328.

Montgomery, J. (1991): "Social Networks and Labor-Market Outcomes:Toward an Economic Analysis," *American Economic Review*, 81(5), 1408–1428.

Muirhead, R. J. (1982): *Aspects of Multivariate Statistical Theory*. Jhon Wiley & Sons, INC.

Naderi, A., and J. Mace (2003): "Education and Earnings: A Multilevel Analysis a Case Study of the Manufacturing Sector in Iran," *Economics of Education Review*, 22, 143–156.

Nelson, C. R., and R. Startz (1990a): "The Distribution of the Instrumental Variable Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63(1), S125–S140.

Nelson, C. R., and R. Startz (1990b): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58(4), 967–976.

Pagan, A. R., and A. D. Hall (1983): "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, 2(2), 159–218.

Pesaran, M. H., and R. Smith (1994): "A Generalized R^2 Criterion for Regression Models Estimated by the Instrumental Variables Method," *Econometrica*, 62(3), 705–710.

Pesaran, M. H., and L. W. Taylor (1999): "Diagnostics for IV Regressions," *Oxford Bulletin of Economics and Statistics*, 61, 255–281.

Plug, E., and W. Vijverberg (2003): “Schooling, Family Background, and Adoption: Is It Nature or is it Nurture?,” *Journal of Political Economy*, 111(3), 611–641.

Sadegi, J. M. (1999): “Private Rate of Return to Education of Male & Female Household Heads in Isfahan, Iran,” ERF Working Paper, 9915.

Salehi-Isfahani, D. (2002): “Microeconomics of Growth in Iran: The role of Households,” ERF Working Paper.

Shea, J. (1997): “Instrument Relevance in Multivariate Linear Models:A Simple Measure,” *Review of Economics and Statistics*, 79(2), 348–352.

Spanos, A. (1986): *Statistical Foundation of Econometric Modelling*. Cambdridge University Press.

——— (2000): “The Instrumental Variables Method Revisited: On the Nature and Choice of Optimal Instruments,” Virginia Tech, mimeo.

Staiger, D., and J. H. Stock (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business & Economic Statistics*, 20(4), 518–529.

Stock, M. J., and F. Trebbi (2003): “Who Invented IV Regression?,” *Journal of Economic Perspectives*.

Woglom, G. (2001): “More Results on the Exact Small Sample Properties of the Instrumental Variable Estimator,” *Econometrica*, 69, 1381–1389.

Vita

Davood Souri

Date of Birth: Feb 6, 1967

Education and Training:

1) Education

- B.A in Economics, Shahid Beheshti University, Tehran, Iran 1988.
- Ms. in Economic Systems Planning, Shiraz University, Shiraz, Iran, 1992.
- MA. in Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA, U.S.A, 2002.
- Ph.D. in Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA, U.S.A, 2004

2) Training

- ERF Summer Institute 1: Training in the use of Household Budget Surveys in Poverty and Income Distribution Analysis, Koc University, Istanbul, Turkey, 1998.
- Poverty Alleviation and Human Resource Development: Methodological Techniques on the Use of Household Surveys, Al-Akhawayn University, Ifrane, Morocco, 2000

Employment

- 1993-1994 Economic Analyst, Plan and Budget Organization, Tehran, Iran

- 1994-2000 Instructor, Dept. of Economics, Institute for Research in Planning and Development (IRPD), Tehran, Iran.

Professional experience

1) Teaching Experience

- Elementary Statistics (1994-96)
- Econometrics (1994-2000)

2) Research

- Research associate on "Evaluation of Foreign loan affect on Iranian Economy", Central Bank of Iran 1995-96.
- Research associate on "Poverty and Macroeconomics Policy", IRPD 1997-1998.
- Research associate on "Poverty, its Measurement and Dynamics in Iran", Economic Research Forum for Arab Countries, Iran and Turkey 1998-1999.
- Research associate on "Female Labor Supply in Egypt, Iran and Turkey", Economic Research Forum for Arab Countries, Iran and Turkey 1998-1999.
- Research associate on "Business Cycle in Selected MENA Region Countries", Economic Research Forum for Arab Countries, Iran and Turkey 1998-1999.
- Research associate on "Income Distribution and Macroeconomics Policy", IRPD 1999-2000.