Bureau d'économie théorique et appliquée (BETA)
UMR 7522

Documents de travail

« Increasing returns to scale in U.S. manufacturing industries: evidence from direct and reverse regression »

Auteurs

Xi Chen

Document de Travail n° 2011 - **11**

Juin 2011

Faculté des sciences économiques et de gestion
Pôle européen de gestion et d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secétariat du BETA
Géraldine Manderscheidt
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.manderscheidt@unistra.fr
http://cournot2.u–strasbg.fr/beta

cnrs
Nancy-Université Université Nancy 2
UNIVERSITÉ DE STRASBOURG

Increasing Returns to Scale in U.S. manufacturing industries: evidence from direct and reverse regression

Xi Chen*

June 23, 2011

Abstract

In this paper, I compare the OLS and IV estimators for the direct and reverse regression models in the context of estimating returns to scale and technical progress. It shows that the direct and reverse OLS estimators are inconsistent, that the direct OLS is always more precise than the reverse OLS under the normality assumption, and that the direct IV estimator and its reverse counterpart are consistent and asymptotically equivalent. Working with data from U.S. manufacturing industries over the last half-century, the estimation results show that in most industries increasing returns to scale are important and technical progress is small when it comes to explaining productivity growth.

JEL classification: C13, D24.

1 Introduction

Increasing returns to scale are of great importance for various macroeconomic models (see e.g. Farmer and Gou 1994, and by Jones, 2004). However, there is a lack of consensus on whether the assumption of increasing returns to scale is empirically plausible. Different methods of estimating returns to scale have been used in the literature, and produced divergent results. Therefore, it is important to understand exactly what each method does and when it might be preferable to use one over others.

This paper contributes to the existing literature in several ways. First, I conducted a comparative study between the *Ordinary Least Squares* (OLS) and *Instrumental Variable* (IV) estimators for the direct and reverse regression models within the linear *Error-in-Variable* (EIV) framework, where both the dependent and explanatory variables are assumed to be subject to errors. Second, the statistical properties of direct and reverse OLS are derived and analyzed. I prove that the direct OLS is always more precise than the reverse OLS under the normality assumption. Third, this paper shows that, in contrast to OLS estimators, both the direct IV estimator and its reverse counterpart are consistent and asymptotically equivalent in the context of linear EIV models. Furthermore, a complete Monte Carlo simulation is proposed to sketch the finite-sample properties of the direct/reverse OLS and IV estimators in a controlled environment. The Monte Carlo experiments show that the finite-sample bias of IV estimator is essentially dependent on the strength of instruments but not the true value of the parameter.

An important part of the literature on the estimation of returns to scale relates the output growth index linearly to the input growth index. The intercept and the slope of the linear equation

*Xi Chen (xichen@unistra.fr), BETA, Faculty of Economics and Management, University of Strasbourg, 61 Av de la Forêt Noire, 67085 Strasbourg, France. I would like to thank Bertrand Koebel and François Laisney for helpful comments.

appear as a measurement of technical progress and a measurement of returns to scale, respectively. Whereas the theory provides a deterministic relationship between both variables of interest (output and input growth rates), from an empirical perspective it is necessary to decide which variable, the input or the output, is stochastic and therefore measured with errors from the true population regression line. Suppose that the cloud of observations is represented in a coordinate system, where the x-axis (abscissa) is the input variable and the y-axis (ordinate) is the output variables. The *direct regression* model assumes that the output variable is stochastic and fits a line that minimizes the squared vertical distance between the data and the regression line in the direction of y-axis. By contrast, the *reverse regression* model thinks of the "reversed" situation, where the input variable becomes the dependent variable. The model should now be fitted by minimizing the squared horizontal distance in the direction of x-axis. Both regression models could be supported for some reasons and have been applied in empirical works. Depending on which regression model is chosen, the estimating results are often very different. Apart from the lack of consensus on regression models, researchers are also debating the choice of estimator. Two estimators are considered in the literature, i.e. the OLS and IV estimators. For instance, Hall (1988, 1990) presents the estimation results of returns to scale and price-cost markup coefficients using annual two-digit sectoral data for 1953-1984 by using the IV estimator. His estimated returns to scale are often unreasonably large for the reverse regression model and even negative for the direct regression model. Bartelsman (1995) was one of first authors to question the IV estimator used by Hall (1990). Bartelsman provided a series of Monte Carlo experiments to illustrate that the bias is likely to be large when estimating coefficients from the reverse approach, and that the IV estimator suffers from a finite-sample bias. An influential article by Basu and Fernald (1997) has compared OLS and IV estimation strategies for the direct regression model. Their OLS-results for thirty-four U.S. private business industries (1959-1989) show that estimated coefficients are often much smaller than one (decreasing returns). Returns to scale are larger in the instrumented regression, but their average value is still close to one, and cannot confirm the increasing returns to scale hypothesis. Recently, by applying the OLS estimation to a larger database (for 1949-2000) of two-digit U.S. industries, Diewert and Fox (2008) also obtained contradictory results between the direct and reverse approach.

Working with data for twenty one sectors of U.S. manufacturing industries (with two-digit SIC classification) over the last half-century, I found strong evidence of increasing returns to scale. On the other hand, the technical progress has made little contribution to U.S. economic growth. The weak instrument test is employed to assess the quality of instruments and the reliability of the estimated coefficients. Compared with prior empirical results on the estimation of returns to scale, such as Hall (1990), our results are more plausible and support a growing body of theoretical models emphasizing the importance of increasing returns to scale in explaining the productivity growth.

The remainder of this paper is organized as follows: I first present the econometric model and the identification issues in Section 2. More attention is given to the discussion of direct/reverse OLS and IV estimation in Section 3. The empirical application to the U.S. manufacturing industries data is reported in Section 4. In Section 5, Monte Carlo simulations are performed to confirm the empirical findings in a controlled environment. Section 6 concludes.

2 Econometric Model

Based on the prior works of Diewert (1976) and Diewert and Fox (2008), this paper follows the Diewert-Fox method of measuring technical progress and returns to scale, where a (multiple inputs and multiple outputs) firm's technology is represented by a non-constant returns to scale translog cost function. The Diewert and Fox framework not only relaxes a series of simplifying restrictions of prior approaches, i.e. single-output, constant returns to scale and perfect competition, but also establishes a very practical relationship between aggregate inputs index and aggregate outputs index. A measurement of technical progress and a measurement of returns to scale appear respectively in this equation as the intercept and the slope, which seem easy to identify. However, I will demonstrate that the identification issue is not straightforward, after a broadly acceptable stochastic specification is chosen. Under a series of restriction on the translog cost function and the neutral technical change assumption, the deterministic relationship between the log-Törnqvist input growth index denoted by x_t^* and the log-Törnqvist output growth index denoted by y_t^*, is

$$y_t^* = \alpha + \beta x_t^* \,, \tag{1}$$

where α is the constant rate of cost reduction; β is the degree of returns to scale. When input increases, if output increases by the same proportional change, i.e. $\beta = 1$, then the technology exhibits constant returns to scale. If output increases by less than that proportional change, i.e. $\beta < 1$, the technology exhibits decreasing returns to scale. If output increases by more than that proportional change, i.e. $\beta > 1$, the technology exhibits increasing returns to scale.

The intercept and the slope of equation (1) are the two parameters of interest. Since these factors can never be measured or observed perfectly in the real world, a common practice is to introduce additive error terms. Suppose that there are T observations in the sample, where the observable values are denoted by (x_t, y_t). They are measured with additive random errors, u and v. Let

$$x = x^* + u \quad \text{and} \quad y = y^* + v \,. \tag{2}$$

The model (1)-(2) is a linear EIV model, which can be also rewritten in a more compact form with only the observable variables, $y = \alpha + \beta x + \varepsilon$, where $\varepsilon \equiv v - \beta u$. We made some statistical assumptions to restrict our stochastic framework.

Assumptions A: u and v are two i.i.d. zero-mean mutually independent variables. Error variances do not vary over time (homogeneity of variance). Formally, let

$$V[u] \equiv \sigma_u^2 \quad \text{and} \quad V[v] \equiv \sigma_v^2$$

and $V[\varepsilon] = E[(v - \beta u)^2] \equiv \sigma_\varepsilon^2$. The latent variables (x^*, y^*) are uncorrelated with error terms; suppose that the first and second moments exist. Let

$$E[x^*] \equiv \mu \quad \text{and} \quad V[x^*] \equiv \sigma^2 \,.$$

The set of parameters that we want to estimate in this model is $\theta \equiv (\alpha, \beta, \mu, \sigma, \sigma_u, \sigma_v)$.

The symmetric treatment of x and y seems to be a simple extension of classical stochastic specification, where only one variable is assumed to be subject to error. The introduction of the second error term increases dramatically the difficulty of estimation. A surprising consequence is that the unique intercept and slope of the fitting line cannot be identified from the bivariate data set (x, y) alone. This is the well-known identification problem of EIV model which was firstly highlighted by Adcock (1878) who tried to handle it by using the *Orthogonal regression* (which is a consistent estimation method only if both variables are subject to errors that have the same variance, i.e. $\sigma_u^2 = \sigma_v^2$). Adcock's intuition is the origin of *Total least squares* (TLS) estimation, which was generalized one hundred years later by Golub and Van Loan (1980). Another idea on the estimation of measurement error models was introduced by Wald (1940) with the objective of proposing a method in which strong assumptions regarding the error structure are not required (e.g. $\sigma_u^2 = \lambda \sigma_v^2$, $\forall \lambda \in \mathbb{R}$ with λ is known). Unfortunately, this class of estimators is not feasible. Since the publication of Wald's method, the problem of estimating EIV models, has received increasing attention from researchers. There have been several surveys, including Madansky (1959), Stefanski (2000) and Gillard (2006). Two books, Fuller (1987) and Carroll et al (2006) cover the linear and nonlinear measurement error models, respectively. In the following subsection the identification problem of EIV model is formally described; some special cases show how additional information can allow us to identify the model.

Identification In the context of the EIV model, identification means that there is a unique set of parameters θ consistent with the observed data. Formally, identification involves being able to obtain unique estimates of structural parameters given the sample moments of data. Under Assumptions A, Kendall and Stuart (1973) derived the five first and second order moment equations by using the Law of Large Numbers:

$$\operatorname{plim} \bar{x} - \mu = 0; \tag{3}$$
$$\operatorname{plim} \bar{y} - \alpha - \beta\mu = 0; \tag{4}$$
$$\operatorname{plim} s_{xx} - \sigma^2 - \sigma_u^2 = 0; \tag{5}$$
$$\operatorname{plim} s_{yy} - \beta^2\sigma^2 - \sigma_v^2 = 0; \tag{6}$$
$$\operatorname{plim} s_{xy} - \beta\sigma^2 = 0. \tag{7}$$

The sample moments of x and y are computed as

$$\bar{x} = T^{-1}\sum_{t=1}^{T} x_t; \quad \bar{y} = T^{-1}\sum_{t=1}^{T} y_t \,.$$

$$s_{xx} = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})^2; \quad s_{yy} = T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})^2$$

and

$$s_{xy} = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y}) \,.$$

The identification problem of EIV models is apparent from an examination of the system (3)-(7), where there are only five moment equations but six unknown parameters. Thus, we have not enough moment conditions to "fix" the fitting line in order to identify θ. Under Assumptions A, the bivariate data set which contains only the input and output growth index does not provide enough information for consistent estimation of returns to scale and technical progress. Prior knowledge can help to overcome the identification problem. For instance, given a fixed α the consistent estimator of returns to scale can be simply computed as $\hat{\beta} = -\alpha/\bar{x} + \bar{y}/\bar{x}$. In the same spirit, I give here two examples of consistent estimator given prior knowledge concerning true variances of the error terms, further examples are given in Judge et al. (1980, p.509-531).

Example 1: the error variance σ_u^2 or σ_v^2 is known. Let $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ be consistent estimators of α β, μ, σ^2, σ_v^2 and σ_u^2, respectively. Simply by using the second order moment conditions, (5), (6) and (7), we can either estimate the slope β by $\hat{\beta} = s_{xy}/(s_{xx} - \sigma_u^2)$ (if σ_u is known) or $\hat{\beta} = (s_{yy} - \sigma_v^2)/s_{xy}$ (if σ_v is known). Once we have the consistent estimate of β, the other the parameters can be also identified as follows.

- $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$;
- $\hat{\sigma}^2 = s_{xy}/\hat{\beta}$
- $\hat{\sigma}_u^2 = s_{xx} - \hat{\sigma}^2$ (if σ_v is known)
- $\hat{\sigma}_v^2 = s_{yy} - \hat{\beta}^2\hat{\sigma}^2$ (if σ_u is known)

Example 2: the variance ratio $\lambda \equiv \sigma_v^2/\sigma_u^2$ is known. Multiplying (5) by λ,

$$\lambda s_{xx} = \lambda \hat{\sigma}^2 + \hat{\sigma}_v^2 \, . \tag{8}$$

by using (6), (7) and (8) we obtain a quadratic equation of $\hat{\beta}$.

$$\hat{\beta}^2 s_{xy} - \hat{\beta}(s_{yy} - \lambda s_{xx}) - \lambda s_{xy} = 0 \, . \tag{9}$$

Given the restriction that $\hat{\beta}$ has the same sign as s_{xy} (because $s_{xy} = \hat{\beta}\hat{\sigma}^2$), the unique relevant solution of this quadratic equation is:

$$\hat{\beta} = \frac{(s_{yy} - \lambda s_{xx}) + \sqrt{(s_{yy} - \lambda s_{xx})^2 - 4\lambda s_{xy}^2}}{2 s_{xy}} \, . \tag{10}$$

Once again all parameters are uniquely determined by solving a system of five equations with five unknowns. Adcock's orthogonal regression (or Total least squares) estimation method is a special case of Example 2, which assumes that $\lambda = 1$.

The class of estimators based on prior knowledge of the parameters is called by Wansbeek and Meijer (2000) the *Consistent Adjusted Least Squares* estimation method. Example 1 and 2 show that the strategy for identifying key model parameters is to obtain the variance-covariance structure of the error terms from a preliminary study (e.g. replication measurement), which in general is impossible in economics. There are other propositions in the literature to deal with the identification in EIV models. Wald (1940) suggested splitting the observations into two groups according to a valid grouping rule.[1] However, a crucial drawback of the grouping method, noticed by Wald himself and many others, such as Pakes (1982), is that in general the grouping criteria cannot be learned from the observations alone, we need extraneous information on grouping criteria; the estimator based on an invalid grouping criterion has the same bias as the classical OLS estimator. The use of third moment based estimators has been also considered in the literature, but it is rarely the case that the observed values of x and y are sufficiently asymmetric to allow third order moment equations to be used with any degree of confidence.[2] In addition, sample sizes needed to identify third order moments are somewhat larger than is the case for first and second order moments. Therefore, estimates are often unstable in finite samples, see e.g. Drion (1951). So far, we can conclude that the identification of the linear EIV model is possible only if additional *a priori* assumptions are made. Paul Samuelson (1942, p.80), in his essay, "*A note on alternative regressions*", already emphasized this point:

"*These assumptions must be in the nature of postulates; by no possible method can they be determined inductively from an examination of the data, even in an infinitely large sample.*"

3 Estimation

Two types of estimators are commonly used in empirical studies estimating returns to scale. Some authors, such as Hall (1988, 1990) have suggested using IV estimators. Others like Basu and Fernald (1997), Diewert and Fox (2008) emphasize the OLS approach. This section presents a theoretical comparison between the two methods for both direct and reverse regressions.

[1] A valid grouping rule satisfies two conditions:
i) The limit inferior of $| \bar{x}_{G1} - \bar{x}_{G2} | > 0$, the subscripts G1 and G2 denote Group 1 and Group 2, respectively.
ii) The grouping rule should be independent of measurement errors.

[2] the third moment of latent variable exists and is not zero.

3.1 Direct and reverse OLS estimation

This section shows that OLS is biased and that the direct and reverse OLS regressions produce upper and lower bounds on the true value of returns to scale. The direct OLS estimator is defined as $\hat{\beta}_{ols} = s_{xy}/s_{xx}$ with:

$$\text{plim}\,(\hat{\beta}_{ols} - \beta\frac{\sigma^2}{\sigma^2+\sigma_u^2}) = 0.$$

Since the ratio $\hat{\sigma}^2/(\hat{\sigma}^2+\hat{\sigma}_u^2)$ (which is called the *reliability ratio* in the literature) is always less than one, the OLS estimate is downward biased w.r.t. the consistent estimator of the true slope. The regression coefficient is *attenuated* by the measurement error. The reverse OLS estimator is defined as $\hat{\beta}_{rols} = s_{yy}/s_{xy}$ with:

$$\text{plim}\,(\hat{\beta}_{rols} - \frac{\beta^2\sigma^2+\sigma_v^2}{\beta\sigma^2}) = 0$$

and it is upward biased.

The direct and reverse OLS regressions are widely used in empirical studies. For instance, labor economists typically try to answer a pair of dual questions: "Are men paid more than equally qualified women?" and "Are men less qualified than equally paid women?" by running respectively the regression of an earnings variable on a qualification variable and the reverse counterpart, see Goldberger (1984). However, little is known about the distribution of direct and reverse OLS estimators in the context of EIV models, especially the second-order moment of the reverse estimator, which is important for the construction of confidence regions and test statistics.[3] I follow Richardson and Wu (1970), and investigate the statistical properties of the direct OLS estimator and extend the results to the reverse OLS estimator under the assumption of joint normality, i.e.,

Assumption B: Suppose that

$$(x^*, u, v) \sim N[(\mu, 0, 0)', diag(\sigma^2, \sigma_u^2, \sigma_v^2)]\,.$$

It follows from the structural model (1)-(2) and Assumption B that the vector (y, x) is distributed as a bivariate normal vector:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\begin{bmatrix} \mu & , & \begin{pmatrix} \sigma^2+\sigma_u^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2+\sigma_v^2 \end{pmatrix} \\ \alpha+\beta\mu & & \end{bmatrix}.$$

Under the normality assumption, Anderson (1985) showed that the statistics s_{xx}, s_{xy} and s_{yy} are jointly distributed as a non-central Wishart distribution. The density function of the direct OLS estimator $\hat{\beta}_{ols} = s_{xy}/s_{xx}$ is then obtained by replacing the variable s_{xy} of the joint density function by $\hat{\beta}_{ols}s_{xx}$ and integrating w.r.t. (s_{xx} ,s_{yy}). This way we obtain the density function of the direct and reverse OLS estimators, as well as their mean and variance. Before stating the result formally, let us define the following quantities:

- $h_{20} = \sigma^2 + \sigma_u^2$
- $h_{02} = \beta^2\sigma^2 + \sigma_v^2$
- $h_{11} = \beta\sigma^2$
- $H = h_{20}h_{02} - h_{11}^2$
- Γ denotes the Gamma function.

[3]Richardson and Wu (1970) derived exact results for the first two moments of the direct OLS estimator in the context of EIV models.

Proposition 1: *Under Assumption B, the exact density function of the direct OLS regression estimator is*

$$f_{\hat{\beta}_{ols}}(\beta) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{h_{20}^{\frac{T-2}{2}}(h_{20}\beta^2 - 2h_{11}\beta + h_{02})^{T/2}} \,. \tag{11}$$

The mean and variance of $\hat{\beta}_{ols}$ are respectively

$$E(\hat{\beta}_{ols}) = \beta\left(1 - \frac{\sigma_u^2}{\sigma^2 + \sigma_u^2}\right) \;; \tag{12}$$

$$V(\hat{\beta}_{ols}) = \frac{1}{T-3}\left[\frac{\sigma_v^2}{\sigma^2 + \sigma_u^2} + \frac{\beta^2\sigma^2\sigma_u^2}{(\sigma^2 + \sigma_u^2)^2}\right] \,. \tag{13}$$

Proposition 2: *Under Assumption B, the exact density function of the reverse OLS regression estimator is*

$$f_{\hat{\beta}_{rols}}(\beta) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{h_{02}^{\frac{T-2}{2}}\beta^2(h_{02}\beta^{-2} - 2h_{11}\beta^{-1} + h_{20})^{T/2}} \,. \tag{14}$$

The mean is

$$E(\hat{\beta}_{rols}) = \beta\left(1 + \frac{\sigma_v^2}{\beta^2\sigma^2}\right) \,. \tag{15}$$

The approximating variance is

$$V(\hat{\beta}_{rols}) \simeq \frac{1}{T-3}[\frac{(\beta^2\sigma^2 + \sigma_v^2)^2(\beta^2\sigma^2\sigma_u^2 + \sigma_v^2\sigma_u^2 + \sigma^2\sigma_v^2)}{\beta^4\sigma^8}] \,. \tag{16}$$

Some interesting observations emerge from the above propositions. First, the direct OLS estimator's bias is only due to $\sigma_u^2 \neq 0$ (variance of the error term which affects x) and the reverse OLS estimator's bias is due to $\sigma_v^2 \neq 0$ (variance of the error term which affects y). Second, one of the important results of this paper, is that the variance of the reverse OLS estimator is always larger than the variance of direct OLS estimator under the normality assumption.

Corollary 1: *Under Assumption B, the direct OLS estimator is always more precise than the reverse OLS estimator.*

The intuition behind the Corollary 1 is simply that the variance of the reverse estimator is mechanically amplified by the reciprocal mapping. The empirical recommendation stemming from Corollary 1 is that the comparison of direct and reverse OLS estimates may be not always appropriate, because the two estimators have different proprieties. Formal proofs of Propositions 1, 2 and Corollary 1 are given in Appendix A.

3.2 Direct and reverse IV estimation

In the previous subsection, we have seen how the direct and reverse OLS estimators behave in the EIV context under the normality assumption. In particular, both estimators are biased and not consistent. Now, I will show that both direct and reverse IV estimators are feasible and consistent. This result should support the use of IV estimator in empirical research, despite the early reluctance raised by Bartelsman (1995) or Basu and Fernald (1997).

Let us discuss the issue of identification in the case where there is one valid instrument z, which is correlated with the latent variable x^*:

$$x_t^* = \pi_0 + \pi_1 z_t + e_t \; , \tag{17}$$

where e is the error term of the prediction equation, which has zero mean and is uncorrelated with z. By construction of the model, the instrument z is also linearly correlated with the observed variables x and y:

$$x_t = \pi_0 + \pi_1 z_t + e_t + u_t \; , \tag{18}$$

$$y_t = \alpha + \beta(\pi_0 + \pi_1 z_t + e_t) + v_t \; . \tag{19}$$

The mean and variance of z are respectively denoted by μ_z and σ_z^2. The instrumental variable z is independent from the error terms e, u and v. By including the additional variable z into our model (1) and (2), we win the following four additional moment equations:

$$\text{plim}\, \bar{z} - \mu_z = 0; \tag{20}$$

$$\text{plim}\, s_{zz} - \sigma_z^2 = 0; \tag{21}$$

$$\text{plim}\, s_{yz} - \beta\pi_1\sigma_z^2 = 0; \tag{22}$$

$$\text{plim}\, s_{xz} - \pi_1\sigma_z^2 = 0, \tag{23}$$

where

$$\bar{z} = T^{-1}\sum_{t=1}^{T} z_t \, ; \quad s_{zz} = T^{-1}\sum_{t=1}^{T}(z_t - \bar{z})^2$$

and

$$s_{xz} = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})(z_t - \bar{z}) \, ; \quad s_{yz} = T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})(z_t - \bar{z}) \; .$$

Together with the five moment equations (3)-(7), we now have nine non-redundant moment conditions to estimate a set of nine parameters, $(\alpha, \beta, \pi_1, \mu, \mu_z, \sigma, \sigma_z, \sigma_u, \sigma_v)$ which are just-identified. Under the assumption that the coefficient $\pi_1 \neq 0$, the consistent estimator of β is simply the ratio s_{yz}/s_{xz}, which is a classical linear IV estimator. The intercept π_0 of the prediction equation cannot be identified, which is not worrying as our interest is focused on the structural relationship and the measurement error structure. When more instruments are added into the data set, we earn more moment conditions and the parameters of interest can be estimated more efficiently by *Two-Stage Least Squares* (2SLS).

Two-Stage Least Squares Now, assume that we have more than one instrumental variable for the input index. Let $Z \equiv [z_1, ..., z_L]$ be a vector of L instruments, each of these instruments correlated with latent variables x^* and observable variable x. A compact form of the direct 2SLS estimator of β can be written as

$$\hat{\beta}_{2sls} = (\tilde{x}' P_Z \tilde{x})^{-1} \tilde{x}' P_Z \tilde{y} \; . \tag{24}$$

Similarly, the reverse 2SLS estimator is

$$\hat{\beta}_{r2sls} = (\tilde{y}' P_Z \tilde{x})^{-1} \tilde{y}' P_Z \tilde{y} \ , \tag{25}$$

where $\tilde{x} \equiv x - \bar{x}$, $\tilde{y} \equiv y - \bar{y}$ and $P_Z \equiv Z(Z'Z)^{-1}Z'$ is the orthogonal projection on the column space of Z. A key assumption of consistency of 2SLS is the *rank condition*, namely $rankE[Z'\tilde{x}] = 1$, which can be tested using a standard F or LM statistic for the Null hypothesis H_0: $\pi_1 = 0,...,\pi_L = 0$. Rejection of this hypothesis confirms the choice of instruments provided the instruments are not endogenous.

Before applying 2SLS to the data, I summarize here the asymptotic properties of the direct and reverse 2SLS estimator for our EIV model. The formal proofs are given in Appendix B. Here, we need three assumptions which were already introduced informally.

Assumption C.1: orthogonality conditions
$E[Z'u] = E[Z'v] = E[Z'e] = 0$.

Assumption C.2: rank conditions
$rankE[Z'Z] = L$ with $L \geq 1$, and $rankE[Z'\tilde{x}] = rankE[Z'\tilde{y}] = 1$.

Assumption C.3: homoskedasticity assumption
$E[\varepsilon^2 Z'Z] = \sigma_\varepsilon^2 E[Z'Z]$, where $\varepsilon = v - \beta u$.

Proposition 3: *Under Assumptions C.1-C.3, the direct 2SLS estimator defined in (24) is consistent for β and asymptotically normally distributed as*

$$\sqrt{T}(\hat{\beta}_{2sls} - \beta) \sim N(0, \sigma_\varepsilon^2 A^{-1}) \ , \tag{26}$$

where $A \equiv E(\tilde{x}'Z)E(Z'Z)^{-1}E(Z'\tilde{x})$.

Proposition 4: *Under Assumptions C.1-C.3, the reverse 2SLS estimator defined in (25) is consistent for β and asymptotically normally distributed as*

$$\sqrt{T}(\hat{\beta}_{r2sls} - \beta) \sim N(0, \sigma_\varepsilon^2 A^{-1}) \ . \tag{27}$$

A direct corollary to Propositions 3 and 4 is that $\sqrt{T}(\hat{\beta}_{2sls} - \beta) \xrightarrow{d} \sqrt{T}(\hat{\beta}_{r2sls} - \beta)$. This result complements Bartelsman's (1995, p.61) finding on the relationship between the direct and reverse 2SLS, which are related to the squared correlation between the projections of x and y on the instruments Z (equation (28) below). The next corollary states that this correlation converges asymptotically to 1, and implies that the direct and reverse 2SLS are asymptotically equivalent.

Corollary 2: *The relationship between $\hat{\beta}_{2sls}$ and$\hat{\beta}_{r2sls}$ is*

$$\hat{\beta}_{2sls} = \hat{\beta}_{r2sls}[(\tilde{y}' P_Z \tilde{x})^{-1} \tilde{y}' P_Z \tilde{x} (\tilde{x}' P_Z \tilde{x})^{-1} \tilde{x}' P_Z \tilde{y}]. \tag{28}$$

Asymptotically

$$(\tilde{y}' P_Z \tilde{x})^{-1} \tilde{y}' P_Z \tilde{x} (\tilde{x}' P_Z \tilde{x})^{-1} \tilde{x}' P_Z \tilde{y} \xrightarrow{p} 1. \tag{29}$$

Bartelsman (1995) concluded that the direct and reverse 2SLS estimates are equal, even in finite-samples, if Z contains only one instrument. Corollary 2 extends his result, asymptotically, to the case of $L \geq 1$ instruments. The problem when using only one instrument, is that estimation is less precise than estimation results from over-identified 2SLS, given small sample size. Bartelsman also claimed that the bias of reverse 2SLS increases as the true parameter β becomes smaller. I will show by performing Monte Carlo simulation that the actual cause of this bias is the weak instrument, and in fact the bias increases as the instrument becomes weaker.

4 Evidence of Increasing returns to Scale

4.1 Presentation of data

The data set used in this work comes from the U.S. Bureau of Labor Statistics (BLS), especially the historical KLEMS database for 1949-2001.[4] The BLS's Multifactor Productivity (MFP) program publishes annual measures of output and combined input quantity indexes, which are calculated using the Törnqvist index formula. The combined input is an aggregate of capital, labor, energy, material and purchased business services inputs. The associated output and input prices are also provided. Twenty one manufacturing sectors including three aggregate sectors are considered in this study. The input and output growth indexes x and y correspond to the first-difference of inputs and output quantity indexes, respectively.

Choice of instruments Choosing instruments is often a difficult task in practice. A valid instrument must satisfy two conditions, the exogeneity and the rank condition, which are formally stated in Assumptions C.1 and C.2. In empirical production analysis, prices are typically assumed to be valid instruments. Studies by Staiger and Stock (1997), Shea (1997), Stock and Yogo (2002a, 2002b) and Hahn and Hausman (2002) show that, in addition to the two classical requirements, a valid instrument must also be highly correlated with the endogenous variable. Instruments which do not have a high degree of explanatory power, can magnify the bias of 2SLS. Fortunately, the degree of the correlation between instruments and endogenous variables can be measured, and the quality of instruments is statistically testable. However, there is no formal test of instrument exogeneity without imposing subjective decisions. For example, usually we assume that one set of instruments which guarantees the just-identification are exogenous and test only the over-identifying restrictions. I consider input prices to be exogenous.[5] The changes (first-differences) in the price of capital, labor, energy, material and services make up my set of instruments, which is denoted by Z. The prices expressed in level as instrument set have also been considered, but the first-stage F statistics show that the prices expressed in first-difference are more powerful, when its come to explaining the regressor.

4.2 Estimation results

Equation (1) is estimated using OLS and 2SLS with the instrument set in which the weakest instrument is dropped. The main outcomes of these estimations are summarized in Table 1 and 2, which report the estimates of returns to scale and technical progress parameters, respectively. The values in the parenthesis are the t-ratio for testing the null hypothesis that $\beta = 1$ and $\alpha = 0$.

[4] See BLS website http://www.bls.gov/mfp/. The data for 1950, 1951 and 1952 are missing. An older version of this data set was used by Diewert and Fox (2008).

[5] I do not include the output prices in the instrument set, for two reasons. First, the output prices cannot be directly observed, the prices provided by database are computed as the ratio of annual profit and output quantity. Thus some errors can affect the output price index. Second, the exogeneity of output price is particular questionable when the market is imperfect.

Table 1: Estimates of returns to scale

Sector	D-OLS	R-OLS	D-2SLS	F	WI	R-2SLS	F	WI
Manufacturing	1.315 (5.01)	1.456 (9.54)	1.467 (5.06)	11.95	z4	1.510 (7.94)	16.69	z4
Nondur. Goods	1.193 (1.77)	1.659 (7.23)	1.736 (2.56)	3.08	z5	1.856 (4.93)	6.08	z5
Food & Kindred Prod.	0.468 (-4.19)	2.083 (3.99)	0.834 (-0.66)	4.68	z3	1.719 (2.02)	7.31	z1
Textile Mill Prod.	0.951 (-1.05)	1.060 (1.21)	1.056 (0.74)	8.30	z4	1.070 (0.97)	10.23	z4
Apparel & Related Prod.	0.973 (-0.53)	1.096 (1.86)	0.880 (-0.53)	0.62	z3	2.076 (1.94)	1.07	z2
Paper & Allied Prod.	1.253 (2.40)	1.669 (7.96)	1.318 (2.25)	13.83	z4	1.414 (3.91)	10.86	z4
Printing & Publishing	1.224 (2.21)	1.616 (7.46)	1.331 (1.93)	6.05	z4	1.763 (5.44)	8.00	z4
Chem. & Allied Prod.	1.166 (1.37)	1.758 (7.30)	1.498 (2.10)	4.78	z3	2.249 (6.34)	10.91	z3
Petroleum Refining	1.210 (6.17)	1.255 (9.06)	1.245 (4.46)	7.09	z3	1.311 (6.71)	8.01	z3
Rubber & Plastic Prod.	1.109 (2.28)	1.207 (4.78)	1.296 (2.73)	3.83	z2	1.367 (4.04)	5.56	z5
Durable Goods	1.250 (5.72)	1.322 (9.21)	1.326 (5.65)	17.48	z5	1.351 (7.97)	21.82	z5
Lumber & Wood Prod.	0.827 (-2.41)	1.120 (1.38)	0.965 (-0.35)	13.65	z3	0.983 (-0.17)	14.39	z5
Furniture & Fixtures	1.150 (4.10)	1.204 (6.43)	1.305 (3.79)	4.39	z2	1.369 (5.43)	6.41	z2
Stone, Clay & Glass	1.277 (4.01)	1.452 (8.37)	1.668 (4.40)	5.86	z3	1.778 (7.88)	13.92	z5
Primary Metal Ind.	1.215 (5.02)	1.285 (8.11)	1.277 (4.57)	11.91	z4	1.296 (6.19)	13.49	z5
Fabricated Metal Prod.	1.123 (4.01)	1.163 (5.95)	1.213 (4.38)	9.93	z4	1.222 (5.48)	12.83	z4
Ind. Machinery, Comp.Eq.	1.136 (2.43)	1.265 (5.39)	1.253 (2.53)	5.67	z5	1.312 (3.79)	7.01	z5
Electric & Electr. Eq.	1.195 (4.12)	1.283 (7.14)	1.226 (2.61)	4.80	z5	1.297 (4.16)	5.05	z5
Transportation Equip.	1.151 (4.67)	1.193 (6.89)	1.185 (3.59)	7.33	z4	1.238 (5.26)	8.21	z4
Instruments	1.012 (0.25)	1.113 (2.45)	1.083 (1.34)	16.42	z4	1.140 (2.37)	20.46	z2
Misc. Manufacturing	1.050 (0.47)	1.555 (5.49)	0.436 (-1.39)	1.46	z3	5.362 (3.24)	1.75	z5
Average	1.117	1.377	1.278	7.77	-	1.434	10.00	-

Note: D-OLS denotes the direct OLS estimator, $\hat{\beta}_{ols}$; R-OLS denotes the reverse OLS estimator, $\hat{\beta}_{rols}$; D-2SLS denotes the direct 2SLS estimator, $\hat{\beta}_{2sls}$; R-2SLS denotes the reverse 2SLS estimator $\hat{\beta}_{r2sls}$. WI: weakest instrument; z1: price of capital; z2: price of labor; z3: price of energy; z4: price of material; z5: price of service. The average of estimates is computed by excluding two sectors, "Apparel&Related Prod" and "Misc.Manufacturing".

Table 2: Estimates of technical progress

Sector	D-OLS	R-OLS	D-2SLS	R-2SLS
Manufacturing	0.006 (2.11)	0.002 (-1.04)	0.003 (0.92)	0.002 (0.62)
Nondur. Goods	0.002 (0.72)	-0.002 (1.79)	-0.008 (-1.29)	-0.011 (-1.90)
Food & Kindred Prod.	0.014 (3.54)	-0.004 (1.93)	0.007 (0.01)	-0.010 (-1.05)
Textile Mill Prod.	0.024 (8.94)	0.021 (-7.80)	0.024 (8.50)	0.024 (7.17)
Apparel & Related Prod.	0.011 (4.83)	0.008 (-3.93)	0.011 (4.07)	0.003 (0.23)
Paper & Allied Prod.	0.000 (-0.01)	-0.004 (1.90)	-0.002 (-0.31)	-0.004 (-0.75)
Printing & Publishing	-0.010 (-2.46)	-0.008 (5.60)	-0.013 (-0.01)	-0.025 (-5.10)
Chem. & Allied Prod.	0.005 (0.70)	-0.004 (1.69)	-0.006 (-0.59)	-0.029 (-2.36)
Petroleum Refining	0.000 (-0.27)	-0.001 (0.86)	-0.001 (-0.64)	-0.002 (-1.33)
Rubber & Plastic Prod.	0.003 (0.88)	0.000 (0.15)	-0.004 (-0.73)	-0.007 (-1.17)
Durable Goods	0.011 (3.57)	0.006 (-2.86)	0.010 (2.87)	0.009 (2.51)
Lumber & Wood Prod.	0.013 (2.93)	0.008 (-1.78)	0.011 (2.44)	0.011 (2.18)
Furniture & Fixtures	0.003 (1.26)	0.001 (-0.69)	-0.001 (-0.22)	-0.002 (-0.63)
Stone, Clay & Glass	0.002 (0.54)	0.000 (0.26)	-0.005 (-0.90)	-0.006 (-1.18)
Primary Metal Ind.	-0.001 (-0.25)	-0.001 (0.39)	-0.002 (-0.37)	-0.002 (-0.40)
Fabricated Metal Prod.	0.000 (0.10)	-0.001 (0.28)	-0.002 (-0.68)	-0.002 (-0.76)
Ind. Machinery, Comp.Eq.	0.018 (3.52)	0.008 (-2.44)	0.014 (2.41)	0.012 (1.81)
Electric & Electr. Eq.	0.025 (5.49)	0.013 (-4.16)	0.024 (4.41)	0.021 (3.10)
Transportation Equip.	0.005 (1.33)	0.003 (-1.01)	0.004 (1.04)	0.003 (0.65)
Instruments	0.015 (3.56)	0.009 (-2.34)	0.012 (2.48)	0.009 (1.77)
Misc. Manufacturing	0.010 (1.90)	0.001 (-0.36)	0.018 (2.15)	-0.053 (1.20)
Average	0.007	0.003	0.004	-0.003

Compared with Diewert and Fox's (2008) estimation results, our OLS estimator produces similar results which show that most sectors exhibit increasing returns to scale. Our results are also similar to those found by Koebel and Laisney (2010) who use the same data. The estimates rate of returns to scale from the direct OLS regression averages to about 1.117 and the reverse OLS regression averages to about 1.377, the average difference between direct and reverse estimates is 0.26. As already noted in Section 3, this difference is mainly due to the opposite bias of the direct and reverse OLS. Diewert and Fox (2008) claimed the legitimacy of OLS estimation, but both type of OLS estimates are obviously inconsistent in the EIV framework. Despite the fact that the null hypothesis of constant returns to scale can generally be rejected at 5% level of significance. For few sectors the results are contradictory. For instance, the direct OLS regression suggests a statistically significant decreasing returns to scale of 0.468 for the "Food & Kindred Prod" sector. On the other hand, the reverse OLS reports an increasing returns to scale of 2.083 for the same sector. Similarly to Diewert and Fox (2008), Table 2 shows that the estimates of technical progress are low and insignificantly different from zero at the 5% level of significance for a majority of estimates.

A simple response to the weak instruments problem is to limit the number of instruments used, for reference see Donald and Newey (2001) and Cameron and Trivedi (2008). Dropping the weakest instrument (selected by examining the first-stage regression) does not change dramatically the estimation outcomes, but the average first-stage F statistic is significantly increased. The direct and reverse 2SLS estimates are closer than those of the OLS estimates, except for two sectors, "Appeared & Related prod" and "Misc. Manufacturing". Theoretically, according to Propositions 3 and 4, the direct and reverse 2SLS are asymptotically equivalent. Nevertheless, a relatively small difference is due to the finite sample bias, which is formally given by Hahn and Hausman (2002).

The 2SLS estimates suggest larger degree of returns to scale than the OLS, all 2SLS estimates of returns to scale are significantly non-decreasing, and these outcomes are much more plausible than Hall's (1990) results. For instance, Hall reports direct and reverse estimated returns to scale of -0.21 and 139.47, respectively for "Chem & Allied prod" sector and I obtain 1.50 and 2.25. We can also notice that, for example, the "Lumber & Wood Prod" sector exhibits constant returns to scale and positive technical progress effects (according to the 2SLS estimate). Therefore, we can say that productivity growth of this sector is mainly driven by technical progress. The same results are also found for the "Textile Mill Prod" and "Instruments" sectors.

Four different estimation outcomes are reported in Tables 1 and 2, they are the direct OLS, the reverse OLS, the direct 2SLS, and the reverse 2SLS. Thus, we have to ask here the question of which estimator should be preferred in practice. In order to answer this question, we firstly consider the OLS estimators versus the 2SLS estimators, then the direct regressions versus the reverse regressions.

4.2.1 OLS estimators versus 2SLS estimators

Proposition 1-4 already pointed out the consistency of 2SLS estimations and the inconsistency of OLS estimations in the EIV framework. However, weak instruments lead to a 2SLS estimator with non-normal asymptotic distribution, so that conventional IV inference is misleading. Two consequences of the weak instruments problem are the introduction of bias and the loss in precision. Cameron and Trivedi (2008, p106, 107) illustrate these problems with two simple relationships. First, the 2SLS relative bias, which is the ratio of the 2SLS bias to the OLS bias, can be expressed as

$$\frac{\text{plim}\hat{\beta}_{2sls} - \beta}{\text{plim}\hat{\beta}_{ols} - \beta} = \frac{Cor[\hat{x}, \varepsilon]}{Cor[x, \varepsilon]} \cdot \frac{1}{R^2}\,, \tag{30}$$

where $\hat{x}$ and R^2 denote respectively the fitted value and the coefficient of determination from the regression of $\tilde{x}$ on Z. In practice, the term $Cor[\hat{x}, \varepsilon]/Cor[x, \varepsilon]$ is close to zero (but probably not null) which represents a relatively small finite-sample bias. When an instrument set is weak, the R_p^2 is low, say 0.1, for example. As a consequence, this finite-sample bias is multiplied by 10. Second, the variance of the 2SLS estimator is

$$V[\hat{\beta}_{2sls}] = \frac{\sigma_\varepsilon^2/\tilde{x}'\tilde{x}}{[(Z'\tilde{x})^2/(Z'Z)(\tilde{x}'\tilde{x})]}\,. \tag{31}$$

We notice that the numerator of (29) is the variance of a classical unbiased OLS estimator; the denominator is the squared sample correlation coefficient. In the case where the squared sample correlation coefficient between Z and $\tilde{x}$ equals 0.1, the standard error of the 2SLS estimator is 10 times larger than the standard error of OLS.

There is not a unique definition of the weak instrument problem. Cameron and Trivedi (2008) define a weak instrument set as one for which the F statistic for the regression of $\tilde{x}$ on Z (which is a function of R^2) is small. A more precise definition is given by Stock and Yogo (2002b), who consider that a set of instruments is strong when the 2SLS relative bias is at most 10% (or 20%). Otherwise, the instruments are weak. For example, when four instruments are used, Stock and Yogo report that the null hypothesis of a relative bias larger or equal to 20% can be rejected in favor of the alternative at the 5% level of significance, if $F \geq 6.72$. The threshold value of F is 10.29 for a more rigorous test with the null hypothesis that the relative bias is larger or equal to 10%. On the other hand, the 2SLS estimation with a F statistic less than 5, produces problematic estimates (more bias together with less precision w.r.t. OLS estimates). The weak instrument issue explains the contradictory results of 2SLS estimation by Hall (1990) and Basu and Fernald (1997).[6] The average first-stage F statistics of our 2SLS by dropping the weakest instruments reaches 7 for direct regression and 10 for reverse regression, which means the estimates have corrected at least 80% of the inconsistency of OLS.

4.2.2 Direct regressions versus reverse regressions

Direct OLS and reverse OLS Proposition 1 and 2 provide us a guidance to discuss the inconsistency and efficiency of OLS. Let us for instance consider the case of $\beta = 1$. The direct OLS is more biased than the reverse OLS when $\sigma_u^2 > \sigma_v^2$, otherwise, the reverse OLS is more biased. When $\sigma_u^2 = \sigma_v^2$, the direct OLS is more biased than the reverse OLS as long as $\beta > 1$. Since both factors which affect the bias, namely, the true value of β and the ratio of σ_u^2 and σ_v^2, are not observed, it is hard to say which estimator should be preferred in term of consistency. However, by assuming that the latent variable x^* and the error terms are normally distributed, Corollary 1 points out that the variance of the direct OLS estimator is always smaller than the variance of the reverse OLS estimator.

Direct 2SLS and reverse 2SLS Under the assumption that the instrument set used is valid, Propositions 3 and 4 point out that both direct and reverse 2SLS regression estimate consistently the true value of β with the same efficiency. However, there are three interesting points which emerge from examination of the empirical results.

First, the direct 2SLS estimates differ from the reverse 2SLS estimates due to finite-sample bias. According to Propositions 3 and 4, the difference converges asymptotically to zero when the F-test rejects the hypothesis of weak instruments. I found some empirical (finite sample) support for this result, as the differences between the direct and reverse 2SLS estimates are often smaller than those obtained with OLS.

Second, the F statistics of the reverse 2SLS (the F statistic of the regression of y on Z) are generally larger than those of the direct 2SLS (the F statistic of the regression of x on Z). In order to explain this phenomenon, recall the model structure, $\tilde{y} = \beta\tilde{x}^* + v - \beta u$. The instrument set Z is assumed to be exogenous w.r.t. u and v, and correlated with the latent variables x^*. Then by construction the F statistic of the reverse regression is larger when $\beta > 1$. The difference between the F statistics of the direct and reverse 2SLS regression increases with the degree of returns to scale. Therefore, the comparison of F statistics can be viewed as a sort of indicator of increasing returns to scale.[7] In our 2SLS estimation results the reverse regressions have a larger first-stage F statistic and generally the estimated returns to scale are larger than 1. For example, both direct and reverse 2SLS indicate that the "Chem & Allied Prod" sector exhibits strong increasing returns to scale, 1.498 and 2.249, the first-stage F statistic of direct regression is 4.78, much smaller than its reverse counterpart 10.91.

Third, the empirical evidence and the Monte Carlo simulation (see next section) seem to show that in a finite sample, the reverse 2SLS estimator is less efficient than the direct 2SLS estimator when instruments are weak. In the "Misc. Manufacturing" sector for instance, the reverse 2SLS estimate for the returns to scale is 5.362 (with $F = 1.75$) whereas the direct 2SLS estimate is 0.436 (with $F = 1.46$). Even more exaggerated results are given by Hall (1990), the reverse 2SLS estimator suggests increasing returns to scale of 139.47 for "Chem & Allied Prod" sector and 33.53 for "Food & Kindred Prod" sector. Bartelsman (1995) also found the same results by using Monte Carlo simulations. An intuitive explanation is that the 2SLS estimator tends to behave like the inconsistent OLS estimator when the set of instruments is weak, and the 2SLS regression produces

[6]For instance, the average first-stage F statistics of 2SLS estimation is 3, in the empirical study of Basu and Fernald (1997).

[7]However, this claim holds only if the errors are relatively small.

a much less precise estimation result. Since Corollary 1 pointed out that the direct OLS is always more efficient that the reverse OLS, therefore, we can think that the reverse 2SLS is also less efficient than the direct 2SLS. However, this suggestion needs to be confirmed by working out formally the finite-sample properties of the direct and reverse weak 2SLS estimators.

So far, I summarize here the theoretic results and the empirical evidences presented in this paper. 1) The direct and reverse OLS estimators are biased in opposite directions. 2) The direct OLS is asymptotically more efficient than the reverse OLS. 3) The direct and reverse 2SLS are consistent and asymptotically equivalent when instruments are not weak. 4) The direct 2SLS seems to be more precise than its reverse counterpart when instruments are weak. 5) The reverse 2SLS regression has larger first-stage F statistics than the direct 2SLS. In the next section all these points will be illustrated by Monte Carlo simulations.

5 Monte Carlo Simulation

Informally, Monte Carlo simulation is essentially a way to generate many artificial finite samples from a parametrized model. The statistics of interest are calculated on each of these samples and analyzed. Here, the experiment is designed to confirm the results of our analytical development.

Data generating process By using a pseudo-random number generator, I simulate five series of sample size $T = 50$ of normally distributed exogenous variables, which are denoted by z_1,..., z_4 and e. The first four variables are considered as instruments. The series of latent variables, x^* and y^* are built as

$$x^* = \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + e \;;$$

$$y^* = \alpha + \beta x^* \;.$$

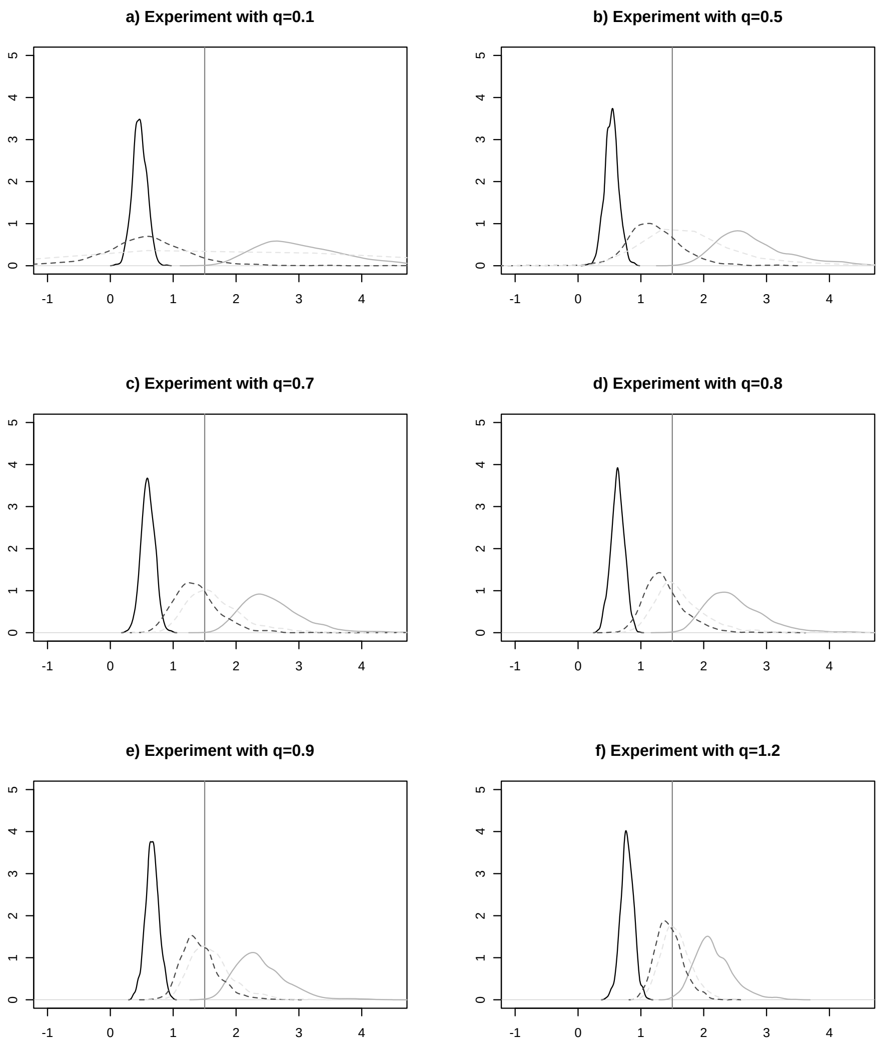
The true value of α and β are respectively 0.001 and 1.5. Without loss of generality, I restrict $\pi_1 = \pi_2 = \pi_3 = \pi_4 \equiv q$. The term e follows a zero-mean normal distribution with standard error equals 4. The coefficient q adjust the correlation between latent variables and instruments, hence the strength of the instrument set. The observable variables x and y are additively affected by u and v as in (2). The error terms u and v follow a zero mean normal distribution with the same standard error $\sigma_u = \sigma_v = 2$.

Results The statistics of interest, including the mean of OLS and 2SLS estimates, the corresponding standard error, the average first-stage F statistics and the ratio of average bias are reported in Table 3 for the direct and reverse regressions. Figure 1 provides the empirical distributions of the estimates generated by six Monte Carlo experiments. Each experiment, of one thousand replications, corresponds to the model with different quality of instruments.

Table 3: Monte Carlo results

q	0.1	0.5	0.7	0.8	9	1.2
Direct Regression						
$Mean(F)$	1.07	2.61	4.19	5.11	6.31	10.27
$Mean(\hat{\beta}_{ols})$	0.75(0.14)	0.85(0.15)	0.90(0.14)	0.93(0.14)	0.97(0.13)	1.06(0.13)
$Mean(\hat{\beta}_{2sls})$	0.80(0.75)	1.27(0.42)	1.36(0.37)	1.40(0.30)	1.41(0.28)	1.45(0.20)
$\frac{Bias_{2sls}}{Bias_{ols}}$	94%	35%	23%	17%	16%	11%
Reverse Regression						
$Mean(F)$	1.12	3.22	5.38	6.67	8.15	13.94
$Mean(\hat{\beta}_{ols})$	2.27(0.51)	2.08(0.37)	2.00(0.34)	1.94(0.31)	1.92(0.28)	1.80(0.22)
$Mean(\hat{\beta}_{2sls})$	3.28 (26.03)	1.86(1.17)	1.57(1.97)	1.64(0.46)	1.60(0.36)	1.56(0.23)
$\frac{Bias_{2sls}}{Bias_{ols}}$	231%	62%	14%	31%	24%	18%

Figure 1: Simulated distribution of estimators



Six figures illustrate the simulated distributions of direct and reverse OLS/IV estimators of β with different value of q. Simulated distributions of the direct (black) / reverse (gray) OLS and 2SLS are respectively depicted by a continuous and dashed line. The vertical straight line represents the true $\beta = 1.5$.

The parameter q in Table 3, varies from 0.1 to 1.2, and as a direct consequence the F statistic increases from 1.07 to 10.27 for direct 2SLS and from 1.12 to 13.94 for reverse 2SLS. We notice that the F statistics of reverse 2SLS are larger than those of direct regressions. If we run a series of experiments with $\beta < 1$, the opposite results can be found. The simulated distribution of estimators in Figure 1, show that neither direct OLS nor reverse OLS estimators can approximate the true value of β. The direct OLS estimators, however, are more precise than their reverse counterpart. Given the parametrization of this DGP, the reverse OLS estimates are less biased than the direct OLS estimates (which illustrates Corollary 1). The distributions of the 2SLS estimator in Figure 1 shows that the finite-sample bias of 2SLS estimator is essentially dependent on the strength of the instruments. When q is high, both direct and reverse 2SLS estimates are close to the true value of β. For instance, for $q = 1.2$ the direct and reverse 2SLS estimates are respectively 1.45 and 1.56. They have corrected about 89% and 82% of the bias of OLS estimations. On the other hand, the 2SLS estimator produces a particularly bad result when instruments are weak. For example, when $q = 0.1$ the instruments are too weak to reduce the bias (F equals 1.07 and 1.12). In the reverse regression, the 2SLS estimator is even worse than OLS (the relative bias of 2SLS is 231%). Therefore, I believe that the weak instrument problem is the main reason for explaining the Hall's implausible results. Finally, the main message which we learned from this series of Monte Carlo simulations, is that the relative bias and the standard errors of 2SLS estimators decrease progressively with the quality of instruments which is characterized by first-stage F statistics.

6 Conclusion

Following the theoretical developments, the empirical outcomes of this paper provide evidence of increasing returns to scale and relatively small technical progress effects. Compared with prior studies, including Hall (1990), Bartelsman (1995), Basu and Fernald (1997) and Diewert and Fox (2008), this paper contains a more complete econometric analysis and more convincing empirical results on the estimation of returns to scale.

In this paper, a comparative study between the OLS and 2SLS estimators for the direct and reverse regression models is conducted within the EIV framework. Our findings suggest that the 2SLS is clearly a better estimation strategy, because both direct and reverse 2SLS are consistent and they are asymptotically equivalent. The gap between direct and reverse 2SLS estimates are generally smaller than those of OLS estimates. Both empirical evidences and Monte Carlo simulations show that the reliability of 2SLS estimates is highly dependent on the strength of the instruments. This finding implies that the weak instrument test needs to be systematically included in empirical studies. The next step of research on this topic may be oriented towards exploring richer databases, such as the four-digit NAICS manufacturing industries database from BLS or the six-digit NAICS database from NBER-CES, and using the panel data analysis. A further purpose is to generalize these findings to nonlinear specifications of the production technology along the lines of Kumbhakar and Tsionas (2011).

References

[1] R.J. Adcock, “A problem in least squares,” The Analyst, vol. 5, 1878, pp. 53–54.

[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley New York, 1958.

[3] E.J. Bartelsman, “Of empty boxes: Returns to scale revisited,” Economics Letters, vol. 49, Jul. 1995, pp. 59-67.

[4] S. Basu and J.G. Fernald, “Returns to Scale in U.S. Production: Estimates and Implications,” The Journal of Political Economy, vol. 105, Apr. 1997, pp. 249-283.

[5] A.C. Cameron and P.K. Trivedi, *Microeconometrics: methods and applications*, Cambridge Univ Pr, 2005.

[6] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, *Measurement error in nonlinear models: a modern perspective*, CRC Press, 2006.

[7] H. Cramer, *Mathematical Models of Statistics*, Princeton University Press, 1946.

[8] W.E. Diewert, "Exact and superlative index numbers," Journal of Econometrics, vol. 4, May. 1976, pp. 115-145.

[9] W.E. Diewert and K.J. Fox, "On the estimation of returns to scale, technical progress and monopolistic markups," Journal of Econometrics, vol. 145, Jul. 2008, pp. 174-193.

[10] S.G. Donald and W.K. Newey, "Choosing the Number of Instruments," Econometrica, vol. 69, 2001, pp. 1161-1191.

[11] E.F. Drion, "Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error," Indagationes Math, vol. 13, 1951, pp. 256260.

[12] R. Farmer and J. Gou "Real Business Cycles and the Animal Spirits Hypothesis," Journal of Economic Theory. vol 63(1), 1994, pp. 42-72.

[13] W.A. Fuller, *Measurement error models*, Wiley Series in Probability and Mathematical Statistics, 1987.

[14] J.W. Gillard, "An Historical Overview of Linear Regression with Errors in both variables," Cardiff University School of Mathematics Technical Report, 2006.

[15] G.H. Golub and C.F. Van Loan, "An analysis of the total least squares problem," SIAM Journal on Numerical Analysis, vol. 17, 1980, pp. 883–893.

[16] A.S. Goldberger, "Reverse Regression and Salary Discrimination," The Journal of Human Resources, vol. 19, 1984, pp. 293-318.

[17] J. Hahn and J. Hausman, "A New Specification Test for the Validity of Instrumental Variables," Econometrica, vol. 70, 2002, pp. 163-189.

[18] R.E. Hall, "The Relation Between Price and Marginal Cost in U.S. Industry," The Journal of Political Economy, vol. 96, Oct. 1988, pp. 921-947.

[19] R.E. Hall, "Invariance Properties of Solow's Productivity Residual," In: Diamond, p. (Ed), Growth, Productivity, Employment. MIT Press, Cambridge MA, 1990.

[20] C. Jones, "Growth and ideas," NBER working paper. 2004.

[21] G.G. Judge, W.E. Griffiths, R.C. Hill and T.Lee, *Theory and practice of econometrics,* John Wiley and Sons, New York, 1980, pp.509-531.

[22] M.G. Kendall and A. Stuart, *The advanced Theory of Statistics - vol. 2*, Hafner Publishing, New York, 1973, pp. 395–397.

[23] B. Koebel, and F. Laisney, "The aggregate Le Chatelier Samuelson principle with Cournot competition", ZEW Discussion Paper No 10-009. 2010.

[24] S.C. Kumbhakar and E.G. Tsionas, "Stochastic error specification in primal and dual production systems," Journal of Applied Econometrics, vol. 26(2), 2011, pp. 270-297.

[25] A. Madansky, "The fitting of straight lines when both variables are subject to error," Journal of the American Statistical Association, vol. 54, 1959, pp. 173–205.

[26] Y. Mundlak, "Production Function Estimation: Reviving the Primal," Econometrica, vol. 64, 1996, pp. 431-38.

[27] A. Pakes, "On the asymptotic bias of Wald-type estimators of a straight line when both variables are subject to error," International Economic Review, 1982, pp. 491497.

[28] K. Pearson, "Researches on the Mode of Distribution of the Constants of Samples Taken at Random from a Bivariate Normal Population," Proceedings of the Royal Society of London. Series A, vol. 112, Aug. 1926, pp. 1-14.

[29] D.H. Richardson and D. Wu, "Least Squares and Grouping Method Estimators in the Errors in Variables Model," Journal of the American Statistical Association, vol. 65, Jun. 1970, pp. 724-748.

[30] P.A. Samuelson, "A Note on Alternative Regressions," Econometrica, vol. 10(1), 1942, pp.80 - 83.

[31] J. Shea, "Instrument Relevance in Multivariate Linear Models: A Simple Measure," Review of Economics and Statistics, vol. 79, May. 1997, pp. 348-352.

[32] D. Staiger and J.H. Stock, "Instrumental Variables Regression with Weak Instruments," Econometrica, vol. 65, May. 1997, pp. 557-586.

[33] L.A. Stefanski, "Measurement Error Models.," Journal of the American Statistical Association, vol. 95, 2000.

[34] J.H. Stock and M. Yogo, "Testing for Weak Instruments in Linear IV Regression," SSRN eLibrary, Nov. 2002a.

[35] J.H. Stock and M. Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," Journal of Business and Economic Statistics, vol. 20, 2002b, pp. 518-529.

[36] A. Wald, "The Fitting of Straight Lines if Both Variables are Subject to Error," The Annals of Mathematical Statistics, vol. 11, 1940, pp. 284-300.

[37] T.J. Wansbeek and E. Meijer, *Measurement error and latent variables in econometrics*, Elsevier, New York, 2000.

Appendix

Appendix A: Statistical Properties of the direct and reverse OLS estimators

A.1 Proof of Proposition 1

The density function of the direct OLS estimator, $\hat{\beta}_{ols} = s_{xy}/s_{xx}$, see Cramer (1946):

$$f_{\hat{\beta}_{ols}}(\beta) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{h_{20}^{\frac{T-2}{2}}(h_{20}\beta^2 - 2h_{11}\beta + h_{02})^{T/2}} .$$

We perform a change of variable.

$$S = \frac{h_{20}\sqrt{T-1}}{\sqrt{H}}(\hat{\beta}_{ols} - \beta_{21}) ,$$

where $\beta_{21} \equiv h_{11}/h_{20}$. Thus, the density function of this new variable is

$$\begin{aligned} f_S(s) &= \frac{\sqrt{H}}{h_{20}\sqrt{T-1}} \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{h_{20}^{\frac{T-2}{2}}(h_{20}\beta^2 - 2h_{11}\beta + h_{02})^{T/2}} \\ &= \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi(T-1)}\Gamma(\frac{T-1}{2})}[\frac{h_{20}}{H}(h_{20}\beta^2 - 2h_{11}\beta + h_{02})]^{-T/2} \\ &= \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi(T-1)}\Gamma(\frac{T-1}{2})}[(\frac{s^2}{(T-1)} + 2\frac{h_{20}\sqrt{H}\beta_{21}s}{H\sqrt{T-1}} + \frac{h_{20}^2\beta_{21}^2}{H}) \\ &\qquad -2(\frac{h_{11}\sqrt{H}}{H\sqrt{T-1}}s + \frac{h_{11}h_{20}}{H}\beta_{21}) + \frac{h_{20}h_{02}}{H}]^{-T/2} \end{aligned}$$

where $\beta_{21} = h_{11}/h_{20}$ and $H = h_{20}h_{02} - h_{11}^2$. Therefore,

$$f_S(s) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi(T-1)}\Gamma(\frac{T-1}{2})}[1 + \frac{s^2}{(T-1)}]^{-T/2} .$$

It implies that S is distributed as a Student's distribution with degree of freedom $T-1$. Hence, the variance of S is $\frac{T-1}{T-3}$, and the variance of $\hat{\beta}_{ols}$ is

$$V(\hat{\beta}_{ols}) = \frac{H}{h_{20}^2(T-3)} .$$

Replacing H and h_{20} by their expressions, then we find

$$V(\hat{\beta}_{ols}) = \frac{1}{T-3}[\frac{\sigma_v^2}{\sigma^2 + \sigma_u^2} + \frac{\beta^2\sigma^2\sigma_u^2}{(\sigma^2 + \sigma_u^2)^2}] .$$

A.2 Proof of Proposition 2

By permuting the indexes of the density function of the direct OLS estimator, we obtain the density function of $\hat{\rho}_{ols} \equiv s_{xy}/s_{yy} = 1/\hat{\beta}_{rols}$.

$$f_{\hat{\rho}_{ols}}(\rho) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{h_{02}^{\frac{T-2}{2}}(h_{02}\rho^2 - 2h_{11}\rho + h_{02})^{T/2}} .$$

The reverse OLS estimator is the inverse of $\hat{\rho}_{ols}$, its density function is obtained from f_s after a change in variables $\beta = 1/\rho$.

$$f_{\hat{\beta}_{rols}}(\beta) = \frac{\Gamma(\frac{T}{2})}{\sqrt{\pi}\Gamma(\frac{T-1}{2})} \cdot \frac{H^{\frac{T-1}{2}}}{\beta^2 h_{02}^{\frac{T-2}{2}} (h_{02}\beta^{-2} - 2h_{11}\beta^{-1} + h_{02})^{T/2}} \, .$$

This change of variable makes the density function even more nonlinear and hard to solve. So that, in order to obtain the variance I prefer to use directly the density function of $\hat{\rho}_{ols}$. Similarly, define a variable

$$P = \frac{h_{02}\sqrt{T-1}}{\sqrt{H}}(\hat{\rho}_{ols} - \beta_{12}) \, ,$$

where $\beta_{12} \equiv h_{11}/h_{02}$, we can show that P follows a Student distribution of degree of freedom $T-1$. Therefore the variance of $\hat{\rho}_{ols}$ is

$$V(\hat{\rho}_{ols}) = \frac{1}{T-3}[\frac{\sigma_u^2}{\beta^2\sigma^2 + \sigma_v^2} + \frac{\sigma^2\sigma_v^2}{(\beta^2\sigma^2 + \sigma_v^2)^2}] \, .$$

By using a first order Taylor expansion of $1/\hat{\rho}_{ols}$ around $E(\hat{\rho}_{ols})$ and assume that $E(\hat{\rho}_{ols}) \neq 0$, the variance of the reverse OLS estimator can be approximated as

$$\begin{aligned} V(\hat{\beta}_{rols}) &\simeq \frac{1}{E(\hat{\rho}_{ols})^4} V(\hat{\rho}_{ols}) \\ &= \frac{1}{T-3}[\frac{(\beta^2\sigma^2 + \sigma_v^2)^2(\beta\sigma^2\sigma_u^2 + \sigma_v^2\sigma^2 + \sigma_v^2\sigma_u^2)}{\beta^4\sigma^8}] \end{aligned}$$

A.3 Proof of Corollary 1

The variance of the direct and reverse OLS estimators can be respectively rewritten as

$$V(\hat{\beta}_{ols}) = \frac{1}{T-3}[\frac{\beta^2\sigma^2\sigma_u^2 + \sigma_v^2\sigma^2 + \sigma_v^2\sigma_u^2}{(\sigma^2 + \sigma_u^2)^2}]$$

and

$$V(\hat{\beta}_{rols}) = \frac{1}{T-3}[\frac{(\beta^2\sigma^2 + \sigma_v^2)^2(\beta^2\sigma^2\sigma_u^2 + \sigma_v^2\sigma^2 + \sigma_v^2\sigma_u^2)}{\beta^4\sigma^8}] \, .$$

Then

$$V(\hat{\beta}_{rols}) - V(\hat{\beta}_{ols}) = \frac{(\beta^2\sigma^2\sigma_u^2 + \sigma_v^2\sigma^2 + \sigma_v^2\sigma_u^2)}{T-3} \cdot \triangle \, ,$$

where

$$\begin{aligned} \triangle &= \frac{(\beta^2\sigma^2 + \sigma_v^2)^2}{\beta^4\sigma^8} - \frac{1}{(\sigma^2 + \sigma_u^2)^2} \\ &= \frac{(\beta^2\sigma^2 + \sigma_v^2)^2(\sigma^2 + \sigma_u^2)^2 - \beta^4\sigma^8}{\beta^4\sigma^8(\sigma^2 + \sigma_u^2)^2} > 0 \, . \end{aligned}$$

Therefore, the variance of the reverse OLS regression estimator is always larger than the variance of the direct OLS regression estimator.

Appendix B: Asymptotic Properties of the direct and reverse 2SLS estimators

B.1 Proof of Proposition 3

The direct 2SLS estimator of our model can be written as

$$\hat{\beta}_{2sls} = [(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1}(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{y}_t) \,.$$

Since the observable output index is $\tilde{y}_t = \beta \tilde{x}_t + \varepsilon_t$, we can rewrite the estimator as

$$\hat{\beta}_{2sls} = \beta + [(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1}(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \varepsilon_t) \,.$$

Under Assumption C.2 and the law of large numbers,

$$(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{x}_t)$$

is non singular and Assumption C.1 implies that

$$\text{plim}[(\sum_{t=1}^{T} \tilde{x}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \varepsilon_t)] = 0 \,.$$

Thus, we have $\text{plim}[\hat{\beta}_{2sls}] = \beta + A^{-1}0 = \beta$. We can write

$$\sqrt{T}(\hat{\beta}_{2sls} - \beta) = [(T^{-1}\textstyle\sum_{t=1}^{T} \tilde{x}_t' Z_t)(T^{-1}\sum_{t=1}^{T} Z_t' Z_t)^{-1}(T^{-1}\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1}$$
$$(T^{-1}\textstyle\sum_{t=1}^{T} \tilde{x}_t' Z_t)(T^{-1}\sum_{t=1}^{T} Z_t' Z_t)^{-1}(T^{-1/2}\sum_{t=1}^{T} Z_t' \varepsilon_t) \,.$$

Under Assumption C.2,

$$[(T^{-1}\sum_{t=1}^{T} \tilde{x}_t' Z_t)(T^{-1}\sum_{t=1}^{T} Z_t' Z_t)^{-1}(T^{-1}\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1} - A^{-1} = o_p(1) \,.$$

The central limit theorem implies that

$$(T^{-1}\sum_{t=1}^{T} \tilde{x}_t' Z_t)(T^{-1}\sum_{t=1}^{T} Z_t' Z_t)^{-1}(T^{-1/2}\sum_{t=1}^{T} Z_t' \varepsilon_t) \sim N(0, B) \,,$$

where $B \equiv E(\tilde{x}'Z)E(Z'Z)^{-1}E(\varepsilon^2 Z'Z)E(Z'Z)^{-1}E(Z'\tilde{x})$. Therefore,

$$\sqrt{T}(\hat{\beta}_{2sls} - \beta) = A^{-1}(T^{-1}\sum_{t=1}^{T} \tilde{x}_t' Z_t)(T^{-1}\sum_{t=1}^{T} Z_t' Z_t)^{-1}(T^{-1/2}\sum_{t=1}^{T} Z_t' \varepsilon_t) + o_p(1)$$

and

$$\sqrt{T}(\hat{\beta}_{2sls} - \beta) \sim N(0, A^{-1}BA^{-1}) \,.$$

Assumption C.3, namely the homoskedasticity assumption allows to simplifying the form of 2SLS asymptotic variance to $\sigma_\varepsilon^2 A^{-1}$.

B.2 Proof of Proposition 4

The reverse 2SLS estimator is

$$\hat{\beta}_{r2sls} = [(\sum_{t=1}^{T} \tilde{y}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1}(\sum_{t=1}^{T} \tilde{y}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{y}_t)$$

$$= \beta + [(\sum_{t=1}^{T} \tilde{y}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \tilde{x}_t)]^{-1}(\sum_{t=1}^{T} \tilde{y}_t' Z_t)(\sum_{t=1}^{T} Z_t' Z_t)^{-1}(\sum_{t=1}^{T} Z_t' \varepsilon_t) \, .$$

Similarly to Proposition 3, this estimator is consistent under Assumption C.1, C.2 and C.3. $\text{plim}[\hat{\beta}_{r2sls}] = \beta + C^{-1}0 = \beta$, where $C \equiv [E(\tilde{y}'Z)E(Z'Z)^{-1}E(Z'\tilde{x})]$. Again, by using the central limit theorem, we have

$$\sqrt{T}(\hat{\beta}_{r2sls} - \beta) \sim N(0, C^{-1}DC^{-1}) \, ,$$

where $D \equiv E(\tilde{y}'Z)E(Z'Z)^{-1}E(\varepsilon^2 Z'Z)E(Z'Z)^{-1}E(Z'\tilde{y})$. Then, using Assumption D.3,

$$Avar[\sqrt{T}(\hat{\beta}_{r2sls} - \beta)] = \sigma_\varepsilon^2 [E(\tilde{x}'Z)E(Z'Z)^{-1}E(Z'\tilde{x})]^{-1} = \sigma_\varepsilon^2 A^{-1} \, .$$

Documents de travail du BETA

2011–**01** *La création de rentes : une approche par les compétences et capacités dynamiques*
Thierry BURGER-HELMCHEN, Laurence FRANK, janvier 2011.

2011–**02** *Le Crowdsourcing : Typologie et enjeux d'une externalisation vers la foule.*
Claude GUITTARD, Eric SCHENK, janvier 2011.

2011–**03** *Allocation of fixed costs : characterization of the (dual) weighted Shapley value*
Pierre DEHEZ, janvier 2011.

2011–**04** *Data games: sharing public goods with exclusion (2nd version)*
Pierre DEHEZ, Daniela TELLONE, janvier 2011.

2011–**05** *Règle du taux d'intérêt et politique d'assouplissement quantitatif avec un rôle pour la monnaie*
Meixing DAI, janvier 2011.

2011–**06** *Ambiguity and Optimal Technological Choice: Does the Liability Regime Matter?*
Julien JACOB, février 2011.

2011–**07** *Politique budgétaire et discipline budgétaire renforcée dans une union monétaire*
Irem ZEYNELOGLU, mars 2011.

2011–**08** *L'évolution de la segmentation du marché du travail en France : 1973-2007.*
Magali JAOUL-GRAMMARE, avril 2011.

2011–**09** *The scarring effect of unemployment in ten European countries : an analysis based on the ECHP.*
Olivia EKERT-JAFFE, Isabelle TERRAZ, avril 2011.

2011–**10** *Almost common value auctions: more equilibria.*
Gisèle UMBHAUER, juin 2011.

2011–**11** *Increasing returns to scale in U.S. manufacturing industries: evidence from direct and reverse regression*
Xi CHEN, juin 2011.

La présente liste ne comprend que les Documents de Travail publiés à partir du 1er janvier 2011. La liste complète peut être donnée sur demande.
This list contains the Working Paper writen after January 2011, 1rst. The complet list is available upon request.