Quantitative Marketing and Economics, 3, 365–392, 2005.

Solving and Testing for Regressor-Error (in)Dependence When no Instrumental Variables are Available: With New Evidence for the Effect of Education on Income

PETER EBBES*
Smeal College of Business, The Penn State University, University Park, PA 16802-3603, USA
E-mail: pebbes@psu.edu

MICHEL WEDEL
Ross School of Business at the University of Michigan, Ann Arbor, MI 48109-1234, USA

ULF BÖCKENHOLT
Faculty of Management, McGill University, Montreal, PQ, H3A 1G5, Canada

TON STEERNEMAN
Department of Economics, University of Groningen, P.O.Box 800, 9700AV, Groningen, NL

Abstract. This paper has two main contributions. Firstly, we introduce a new approach, the latent instrumental variables (LIV) method, to estimate regression coefficients consistently in a simple linear regression model where regressor-error correlations (endogeneity) are likely to be present. The LIV method utilizes a discrete latent variable model that accounts for dependencies between regressors and the error term. As a result, additional 'valid' observed instrumental variables are not required. Furthermore, we propose a specification test based on Hausman (1978) to test for these regressor-error correlations. A simulation study demonstrates that the LIV method yields consistent estimates and the proposed test-statistic has reasonable power over a wide range of regressor-error correlations and several distributions of the instruments.

Secondly, the LIV method is used to re-visit the relationship between education and income based on previously published data. Data from three studies are re-analyzed. We examine the effect of education on income, where the variable 'education' is potentially endogenous due to omitted 'ability' or other causes. In all three applications, we find an upward bias in the OLS estimates of approximately 7%. Our conclusions agree closely with recent results obtained in studies with twins that find an upward bias in OLS of about 10% (Card, 1999). We also show that for each of the three datasets the classical IV estimates for the return to education point to biases in OLS that are not consistent in terms of size and magnitude. Our conclusion is that LIV estimates are preferable to the classical IV estimates in understanding the effects of education on income.

Key words. instrumental variables, latent instruments, testing for endogeneity, mixture models, identifiability, estimating the return to education

JEL Classification: C12, C13, C21, J3, M3

*Corresponding author

1. Introduction

The standard linear regression model $y = X\beta + \epsilon$ is an important tool in (applied) statistical science to model the effect of a set explanatory variables on a dependent variable. Here $y = (y_1, \ldots, y_n)'$ is the dependent variable, $X \in \mathbb{R}^{n \times k}$ the vector of explanatory variables (regressors), β is the unknown $k \times 1$ vector of interest and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is an unobserved stochastic disturbance. An important assumption in these models is the independence of the explanatory variables X and random (error) components ϵ. In this case the regressors are said to be 'exogenous' (or determined outside the model). Failure of this assumption may lead to biased or inconsistent estimates for the parameters of interest and henceforth to faulty conclusions and erroneous decision-making (Judge et al., 1985; Bowden and Turkington, 1984; Greene, 2000). An important area where this independence assumption may not hold is in estimating the causal effect of education on earnings, see for instance Griliches (1977), Card (1999, 2001), or Uusitalo (1999). Most of the studies in question have focused on estimating a linear regression model that relates the logarithm of a measure of earnings to a measure of education and possible other explanatory variables. The effect of education on income is expected to be positive, but the simple OLS estimator cannot be used to estimate it because of omitted 'ability', in which case the OLS estimate is expected to be biased upward.

Unfortunately, in many similar situations the assumption of regressors and error independence is not satisfied. In this case the regressors are often said to be 'endogenous'. Endogeneity can arise from a number of different sources: (1) relevant omitted variables, such as omitted 'ability', (2) measurement error in the regressors, (3) the problem of self-selection, (4) simultaneity, and (5) serially correlated errors in the presence of a lagged dependent variable. Ruud (2000) shows that (2)–(5) can be viewed as a special case of (1). One way to circumvent problems of endogeneity is to find instruments, based on economic theory or intuition, and apply two-stage least squares or limited maximum likelihood estimation techniques (see e.g. Bowden and Turkington, 1984; Verbeek, 2000; Greene, 2000). Instruments are variables that mimic the troublesome regressors as well as possible but are uncorrelated with the error term. Hence, instrumental variables cannot have a direct effect on the dependent variable. However, this method suffers from at least two problems: (i) in many situations no such variables are available, and (ii) if they are available, performance of the inferential procedures critically relies on the 'validness' of these variables. Using bad quality instruments may result in estimates that are even more biased than OLS estimates (see e.g. Bound et al., 1995; Hahn and Hausman, 2003; Stock et al., 2002).

For most empirical researchers the question where to find suitable instruments is still open. Without having valid instrumental variables at hand classical instrumental variables estimation techniques cannot be relied on, and usually there is not much choice on the selection of instrumental variables. Furthermore, there is a bit of a dilemma: theory suggests that the best choice of instruments are variables that are highly correlated with the endogenous regressors. However, the more highly correlated they are, the less defensible the claim is that these variables themselves are uncorrelated with the disturbances (cf. Greene, 2000). From the applications presented in this paper it becomes clear that estimating the return to education using instrumental variable estimation is problematic.

In addition, in empirical work it is not always obvious whether it is necessary at all to search for instruments. Thus, one would like to test a priori whether $E(\epsilon \mid X) = 0$ holds.

However, as OLS always yields $X'e = 0$, it is fruitless to use the OLS estimates for that purpose. One way to test for exogeneity is through the use of instruments, and by applying a Hausman test it can be determined post-hoc whether they were actually needed (see e.g. Bowden and Turkington, 1984). This method has as drawbacks that instruments need to be available and that the performance of the test critically depends on the quality of the instruments. Besides, one may conclude that endogeneity is not a serious problem, in which case OLS is the best available estimator making the instrumental variables superfluous.

In this paper we propose a new "instrument-free" approach to solve these circular problems. This Latent Instrumental Variables (LIV) approach estimates regression parameters consistently regardless of the presence of regressor-error correlations. As this method does not rely on observable instruments, issues such as availability, validity, and weakness of the instruments can be circumvented. The proposed LIV method utilizes a latent variable model to account for dependencies between the regressors and the error. The method introduces an (unobserved) discrete binary variable to decompose x into a systematic part that is uncorrelated with ϵ and one that is possibly correlated with ϵ. We show how the approach can be used to test for regressor-error dependencies, without having instrumental variables at hand. Hence, as opposed to the classical IV approach, one can test a priori for endogeneity.[1]

This paper is organized as follows. We first introduce the LIV model and prove that the model parameters are identifiable (Section 2). In addition, we discuss several implementation issues. In Section 3 we suggest a method to test directly for regressor-error problems, which is based on a Hausman-test (Hausman, 1978). This instrument-free test can be used to assess a priori the presence of regressor-error correlations. The model estimators and test-statistic are evaluated on the basis of a simulation study (Section 4). In Section 5 we review part of the schooling literature and discuss the problems associated with classical instrumental variables estimation. As will become clear, the classical IV method has produced a less than satisfactory solution in estimating the return to education. We re-examine three empirical datasets with the LIV approach. Section 6 presents a summary of our findings. We conclude that the results presented here lend credibility to the LIV approach to solve for general regressor-error dependencies.

2. The LIV model

2.1. *The simple LIV model*

The structural form of the assumed LIV model is given by

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_1 + \epsilon_i, \\ x_i &= \pi' \tilde{z}_i + \nu_i, \end{aligned} \tag{1}$$

1 Although the idea of *latent* instruments is new, observed discrete instruments have been used before in classical IV estimation (van der Ploeg, 1997; Verbeek, 2000). Nevertheless, these dummy instruments may be weak or endogenous. Similar to LIV, methods to 'identify' instruments for unbiased estimation of regression parameters have been proposed in the measurement error literature. For instance, grouping methods have been used to construct instruments based on the method of Wald (cf. Madansky, 1959; Bowden and Turkington, 1984). Lewbel (1997) and Erickson and Whited (2002) propose method-of-moments based estimators that do not rely

with $i = 1, \ldots, n$ and π a $(m \times 1)$-vector of category means. Here, we assume a single unobserved discrete instrument $\tilde{z}_i$. The discrete instrument should have at least two categories ($m > 1$), which is in accordance with van der Ploeg's (1997) result for the standard IV model with observed discrete instruments. Below we suggest how to include additional exogenous variables into the model. We will also show that the model is robust against (under) misspecification of the true number of categories of the instruments, and to error distributions that are fatter in the middle, skewed, or that have fatter tails. It is assumed that $\tilde{z}$ is independent of the error terms (ϵ, ν), that are specified to follow a normal distribution with mean zero and variance-covariance matrix

$$\sum = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\nu} \\ \sigma_{\epsilon\nu} & \sigma_\nu^2 \end{bmatrix}.$$

As can be seen, the endogenous regressor x is split into an exogenous part and an endogenous part. The correlation between x and ϵ is captured through the covariance $\sigma_{\epsilon\nu}$. The identifiability of this covariance term is an important feature of our model. The unobserved instrument is a nuisance parameter, but can be profiled to give it an (economical) interpretation if additional data is available, and it may be compared to available observed instruments. If we had observed the instruments, $\tilde{z}$ would simply separate the sample into m groups, with known category-membership for each observation and model (1) is in form identical to a standard instrumental variables model (e.g. van der Ploeg, 1997). In fact, the LIV estimator for the regression parameters in this case is identical to the well known LIML estimator. The LIML estimator assumes normality of the error terms and is often recommended instead of 2SLS when the observed instruments are weak (Bekker, 1994; Staiger and Stock, 1997). We assume, however, that the category indicators are unknown a priori and have a multinomial distribution with parameters (t, λ), where $t = 1$ and $\lambda = (\lambda_1, \ldots, \lambda_m)'$, where $\sum_j \lambda_j = 1$.

Conditionally given the category number $j = 1, \ldots, m$, the reduced form distribution corresponding to (1) is

$$\mathcal{L}(y_i, x_i \mid \tilde{z}_i = e_j) = N_2(\mu_j, \Omega) \tag{2}$$

with mean

$$\mu_j = \begin{pmatrix} \beta_0 + \beta_1 \pi_j \\ \pi_j \end{pmatrix} \tag{3}$$

and variance-covariance matrix

$$\Omega = \begin{bmatrix} \beta_1^2 \sigma_\nu^2 + 2\beta_1 \sigma_{\epsilon\nu} + \sigma_\epsilon^2 & \beta_1 \sigma_\nu^2 + \sigma_{\epsilon\nu} \\ \beta_1 \sigma_\nu^2 + \sigma_{\epsilon\nu} & \sigma_\nu^2 \end{bmatrix}, \tag{4}$$

(*Continued.*) on observable instruments, and show that, under certain higher-order moment conditions, the constructed instrumental variables can be used to consistently estimate the regression parameters. However, the LIV model is developed for situations with general regressor-error dependencies and, hence, is more generally applicable. Furthermore, the LIV method estimates the instruments via the available data rather than constructing them on basis of moments that may or may not hold.

where e_j is the j-th column of the $m \times m$-identity matrix. If f_j denotes the normal bivariate probability density function conditionally given $\tilde{z}_i = e_j$, then the unconditional (marginal) probability density function for (y_i, x_i) can be computed as

$$f(y_i, x_i) = \sum_{j=1}^{m} \lambda_j f_j(y_i, x_i \mid \tilde{z}_i = e_j), \tag{5}$$

Thus, $f(y_i, x_i)$ is a mixture of bivariate homoscedastic normal distributions and it has expectation

$$\mu_{y,x} = \begin{pmatrix} \beta_0 + \beta_1 \sum_{j=1}^{m} \lambda_j \pi_j \\ \sum_{j=1}^{m} \lambda_j \pi_j \end{pmatrix}, \tag{6}$$

with variance-covariance

$$\Omega_{yx} = \Omega + \begin{pmatrix} \beta_1 \pi' \\ \pi' \end{pmatrix} \mathrm{Var}(\tilde{z}_i) \begin{pmatrix} \beta_1 \pi' \\ \pi' \end{pmatrix}',$$

where $\mathrm{Var}(\tilde{z}_i) = \mathrm{diag}(\lambda) - \lambda\lambda'$, $\lambda = (\lambda_1, \ldots, \lambda_m)'$. The parameters $\lambda_1, \ldots, \lambda_m$ represent the category sizes, with $\lambda_j > 0$ and $\sum_j \lambda_j = 1$. The parameters to be estimated are the regression parameters β_0 and β_1, the category means π_j, the variances σ_ϵ^2 and σ_ν^2, the covariance $\sigma_{\epsilon\nu}$, and the category sizes λ_j, for $j = 1, \ldots, m$. The parameters are identified as is shown in appendix 1.

For estimation of the parameters, assume that a sample of n IID observations (y_i, x_i) is available. The method of maximum likelihood can be used to estimate the model parameters. The likelihood function is obtained as the product of (5) across the observations. The resulting (log-)likelihood equations, however, are nonlinear and do not allow a closed-form expression. Therefore we use quasi-Newton numerical optimization routines (the BFGS-method) for the maximization of the likelihood function that are provided with the GAUSS package (Aptech, 2000). As the LIV model belongs to the class of mixture models, standard results on estimation schemes and statistical properties for these models can be applied (see Titterington et al., 1985; Redner and Walker, 1984; McLachlan and Peel, 2000, for an overview).

Contrary to traditional instrumental variables estimation, the LIV approach is not identified through the first two moments of the data, but is identified by statistical assumptions, similar in spirit to the measurement error methods proposed by e.g. Wald (1940) or Lewbel (1997). As can be seen from the results in appendix 1, the model is identified for every $m > 1$. This means that, unless the joint distribution of (y_i, x_i) is perfectly normal, a mixture of normals with more than one component can be fitted to approximate the true distribution. As long as $m > 1$, i.e. there exist at least two different group means π, the covariance $\sigma_{\epsilon\nu}$ is identifiable and the endogeneity can be accounted for. Titterington et al. (1985) put forward that mixtures of normals often approximate densities of various shapes quite well and that mixtures are used in robustness studies to investigate non-normal conditions. Hence, the LIV model can be expected to be relatively insensitive to the shape of the (joint) distribution of the data or to the (non)existence of higher order moments. Furthermore, identifiability does not break down when $\beta_1 = 0$. We show this below in Monte Carlo studies. These simulation studies indicate that the simple LIV model estimates the regression parameters consistently and is flexible in adapting to the different joint distributions of (y_i, x_i). We also show below that the results are not sensitive to misspecification of the true number of categories of the latent instrument.

2.2. *LIV: Implementation issues*

Additional exogenous regressors. Additional exogenous regressors can be included straightforwardly in both rows of $\mu_{y,x}$ in (6), so that (say) $2 \times k$ additional parameters have to be estimated. Similarly, possible available observed instrumental variables can be included in the second row of $\mu_{y,x}$. It can be shown that these additional parameters are identifiable as well (Ebbes, 2004).

Selection of the number of categories of the discrete instrument. In empirical studies, it needs to be decided how many categories the latent instrument has, i.e. how large m should be. Standard model selection methods, like AIC, CAIC, or BIC are often found to overestimate the number of groups. Naik et al. (2003) argue that information criteria like AIC are designed for selecting regressors, but not groups. The integrated classification likelihood (ICL) criterion (Biernacki et al. 2000) has also been shown to be suitable for selecting the number of components in mixture models. Since our aim is to select the number of categories for the discrete instruments, i.e. the number of groups representing the endogenous regressor best, and given the importance of not overestimating the number of components, we prefer the ICL criterion, which is more conservative than the other statistics.

The ICL criterion is a modification of BIC. Biernacki, Celeux and Govaert (2000) suggest to choose the model that maximizes the complete integrated maximum likelihood and show that the resulting ICL criterion is essentially the BIC statistic penalized by the subtraction of the mean entropy $-2\sum_i \sum_j \hat{z}_{ij} \log \hat{p}_{ij}$, where $\hat{p}_{ij}$ are the posterior probabilities that observation i comes from category j and $\hat{z}_{ij} = 1$ whenever $\hat{p}_{ij} = \max_j \hat{p}_{ij}$, and zero otherwise. It follows that if the categories are not well separated, this term has a large value and BIC is penalized more severely. If the groups found by the LIV model are not well separated, it resembles a situation in classical IV where the instruments are weak. Furthermore, overfitting in terms of the number of groups results in using a too large number of instruments which is not preferred, since degrees of freedom are lost which reduces efficiency.

Although the ICL criterion can be used in determining the number of instruments, we emphasize that several choices of m should be examined. This approach allows investigating whether the estimated regression coefficients are stable across the considered choices of m. In general, to prevent over-fitting we recommend using the extra penalization term proposed by Biernacki et al. (2000) that adjusts the BIC statistic more severely when the instruments yield posterior groupings that are fuzzy.

LIV diagnostics. We extend several diagnostics, originally proposed for the classical regression model (Fox, 1991; Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982) to the LIV case. Outliers and influential observations can be problematic because they may influence estimation results. Their presence may point out that the estimated model fails to capture important aspects of the data. Analyzing residuals can reveal important information for assessing model assumptions. Although maximum likelihood estimation is approximately valid in all but small samples, it is still important to examine carefully systematic deviations from normality in the distributions of the residuals.

Analyzing residuals. The model residuals from (1) can be examined (i) to investigate the normality assumption of the disturbances; (ii) to detect potential outliers; and (iii) to examine heteroscedasticity. In appendix 2 we derive two type of residuals: conditional residuals, and IV-type residuals. To examine (lack of) normality we propose using the IV-type residuals to compute kurtosis and skewness. Heteroscedasticity can be examined using conditional residuals (i.e. examine scatterplots of the residuals versus explanatory variables and the predicted values). To correct the estimated standard errors for heteroscedasticity, White's (1980) method can be used. Finally, outliers are identified by examining standardized versions of the above residuals and the highest values are investigated.

Analyzing influential observations and outliers. We propose to approximate the Jack-knife LIV estimate $\hat{\theta}(i)$ by a few numerical optimization steps with the maximum likelihood estimate of the complete sample as starting value (Cook and Weisberg, 1982; Belsley et al., 1980; Fahrmeir and Tutz, 1994). The measures in appendix 2 can be used to examine the influence of observation i.

3. A test-statistic to test for endogeneity

We propose to apply a Hausman test directly to test for exogeneity of the regressor (see Greene, 2000) based on the parameter estimates $\hat{\beta}_{\mathrm{LIV}}$. The null hypothesis is that both OLS and LIV estimates are consistent. The alternative hypothesis states that only LIV is consistent. The Hausman-LIV statistic is defined as

$$HLIV = (\hat{\beta}_{\mathrm{LIV}} - \hat{\beta}_{\mathrm{OLS}})' \hat{\Sigma}_{\mathrm{HLIV}}^{-1} (\hat{\beta}_{\mathrm{LIV}} - \hat{\beta}_{\mathrm{OLS}}), \tag{7}$$

where $\hat{\Sigma}_{\mathrm{HLIV}}$ is the estimated asymptotic covariance of the difference of $\hat{\beta}_{\mathrm{LIV}} - \hat{\beta}_{\mathrm{OLS}}$. Hausman shows that this difference can be computed by subtracting the estimated asymptotic OLS covariance matrix from the estimated asymptotic LIV covariance matrix. We found that the estimated asymptotic covariance matrix based on the analytical first- and second-order derivatives is more stable and gives more accurate results than a numerical approximation of the gradient or Hessian. Under the null hypothesis, the statistic follows asymptotically a $\chi^2(1)$-distribution.

This Hausman-LIV test proposed has a great practical advantage over classical IV methods. In the classical case, one would first need to find observable instruments of decent quality, after which a test to investigate whether or not the instruments were needed can be performed. If the test does not reject the null hypothesis, the instrumental variables are simply discarded since the OLS estimator is used in that case. Besides, weak and/or endogenous instruments will bias the test leading to false conclusions. Our approach circumvents this circular problem since observed instrumental variables are not required to test for endogeneity.

4. Monte Carlo experiments

This section presents the results of a Monte Carlo experiment to demonstrate that the proposed simple LIV model and Hausman-LIV test are well suited to identify and resolve regressor-error dependencies. In all cases we estimate the LIV model assuming that the latent instrument has two categories, i.e. we use $m = 2$ in estimation. We show that even if the true number of the categories of the instrument is larger than two and for various distributions of the instruments, the LIV estimates are approximately consistent and the power of the test appears to be satisfactory.

In the simulation study the data were generated as follows. The error terms (ϵ, ν) are drawn from a bivariate normal distribution with unit variances. The endogenous regressor x was constructed by varying the correlation between x and ϵ and the true number of instruments $\tilde{m}$. We considered three specifications for $\tilde{m}$: 2, 4, and 8, using equal group sizes $1/\tilde{m}$. This results in a bimodal distribution with $\tilde{m} = 2$ (bim2), and two unimodal distributions with $\tilde{m} = 4$ (unim4) and $\tilde{m} = 8$ (unim8) for the endogenous x. Furthermore we consider two other distributions for the instruments both with $\tilde{m} = 8$ support points, resulting in a bimodal distribution (bim8) and a skewed distribution (skew8). The values for $\tilde{m}$, $\sigma_{\epsilon\nu}$, and $\pi_1, \ldots, \pi_{\tilde{m}}$ are chosen such that the mean of x is zero, its variance is 2.5 and the correlation between x and ϵ is $0, 0.1, \ldots, 0.5$. Since in all simulations the endogenous regressor has mean zero, the OLS estimate of the constant is consistent and it can be used as an estimate for β_0. Hence, we omit further details on β_0 in the following. The Hausman-LIV test statistic for the regression coefficient has a $\chi^2(1)$-distribution under the null hypothesis of no regressor-error dependency. Data were generated for 1000 observations and 250 Monte Carlo replications. We use the analytical expressions for the gradient and Hessian.

Figure 1 shows the bias plots for β_1 estimated for the six different correlations between x and ϵ by, respectively, OLS and the LIV method for data generated with $\tilde{m} = 2, 4, 8$. Two observations are in order. First, increasing the degree of endogeneity decreases the amount of uncertainty in the LIV estimator. This result is expected as the proposed method is designed for situations with endogeneity. In the case of a perfectly exogenous regressor, OLS provides the 'best linear unbiased' estimator and outperforms LIV, but OLS performs worse as the correlation between x and ϵ increases. Secondly, when there are four or eight instruments in the unimodal distribution (unim4 and unim8), some efficiency is lost with the LIV approach since the model is misspecified under these conditions, as we specified $m = 2$ in all cases. Furthermore, less well separated mixture components may lead to lower efficiency (Titterington et al., 1985, Redner and Walker, 1984), and unmixing the distribution becomes more difficult for these cases. However, when the true instrument has an obvious grouped structure, as in the case of the two bimodal (bim2 and bim8) and the skewed (skew8) distribution of the instrument, it is well approximated by the assumed discrete instrument.

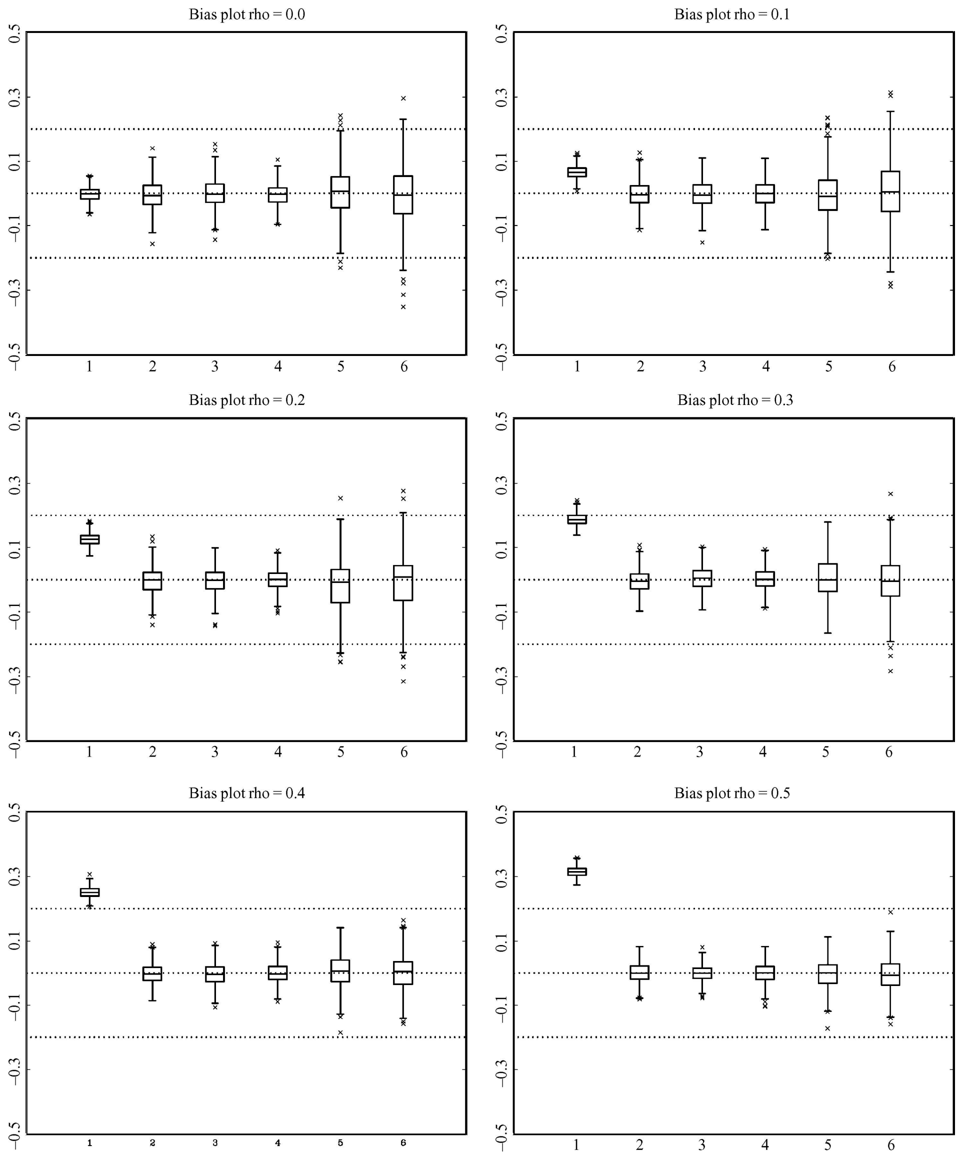

Figure 1. Bias plots β_1 for the simple LIV model, where 1: OLS, 2: bimodal $\tilde{m} = 2$ (bim2), 3: bimodal $\tilde{m} = 8$ (bim8), 4: skewed $\tilde{m} = 8$ (skew8), 5: unimodal $\tilde{m} = 4$ (unim4), and 6: unimodal $\tilde{m} = 8$ (unim8).

In these cases, the LIV model represents the true instruments quite accurately, resulting in more efficient estimates. This illustrates that the LIV model is flexible in adapting to different shapes of the true distribution of (y, x), which is a typical property of mixture models. Simultaneously, the mixture enables identifiability of the regression parameters. In all cases, the LIV estimates for the regression parameters appear to be consistent.

Table 1. Power of the Hausman-LIV test using the simple LIV model for various degrees of endogeneity, for sizes (respectively) $\alpha = 0.50, 0.05$, and 0.01.

		Distribution				
α	$\rho_{x,\epsilon}$	bim2	bim8	skew8	unim4	unim8
0.50	0	0.53	0.51	0.54	0.47	0.54
	0.1	0.92	0.86	0.96	0.71	0.59
	0.2	1.00	1.00	1.00	0.95	0.86
	0.3	1.00	1.00	1.00	0.99	0.97
	0.4	1.00	1.00	1.00	1.00	1.00
	0.5	1.00	1.00	1.00	1.00	1.00
0.05	0	0.04	0.06	0.08	0.06	0.07
	0.1	0.44	0.40	0.56	0.20	0.13
	0.2	0.95	0.97	0.99	0.52	0.42
	0.3	1.00	1.00	1.00	0.85	0.76
	0.4	1.00	1.00	1.00	1.00	0.97
	0.5	1.00	1.00	1.00	1.00	1.00
0.01	0	0.02	0.01	0.01	0.02	0.02
	0.1	0.22	0.22	0.37	0.08	0.06
	0.2	0.87	0.82	0.96	0.27	0.24
	0.3	1.00	1.00	1.00	0.68	0.61
	0.4	1.00	1.00	1.00	0.97	0.94
	0.5	1.00	1.00	1.00	1.00	1.00

Table 1 shows the results for the Hausman-LIV test. The degrees of endogeneity are presented row-wise, each entry represents the fractions of rejections of the null hypothesis. Increasing the correlation between x and ϵ increases the number of times the null hypothesis is rejected, as is to be expected. Comparing the two bimodal distributions, the test performs slightly better for bim2 in which case the number of instruments is correctly specified, although the results are very close.[2] Comparing the two unimodal distributions (unim4 and unim8) the test tends to reject the null-hypothesis somewhat too often when the true correlation is zero for the case with $\tilde{m} = 8$. As before, this is caused by efficiency loss due to misspecification, and the approximation of the distribution of x to a normal distribution. When the true instrument has a skewed distribution, the power of the Hausman-LIV test for $\rho_{x\epsilon} > 0$ is higher than for any other distribution that we investigated. But, in this case the test is also too liberal for zero regressor-error correlations. The power of the test is highest for the bimodal distributions and the skewed distribution, and the lowest power for the unimodal distributions. If the instrument has a bimodal or a skewed distribution, the two groups imposed on the endogenous x by the simple LIV model are a more adequate representation, allowing for precise LIV estimates.

We also examined the situation where the endogenous regressor x has no effect on y (i.e. $\beta_1 = 0$) and the impact of misspecified error distributions: (1) when ϵ has a skewed gamma

2 We did not report the standard deviations. These can be computed easily as $\sqrt{f(1-f)/l}$, where f are the reported fractions and l is the total number of simulation runs.

distribution, with a shape parameter equal to 4, (2) when ϵ follows a mixture of a normal and an uniform distribution, which is fatter than a normal distribution, and (3) when ϵ has a t_6 distribution, which has a higher peak than a normal distribution. All distributions were normalized to have mean 0 and variance 1, and we specified the same degrees of endogeneity as before. We used the same specification for the latent instrument as the bim8 case above. We found that the LIV results for the situation where $\beta_1 = 0$ and the distribution of the errors is normal are similar to the ones before. This confirms our results from appendix 1 that identifiability does not break down when x has no effect on y. Secondly, we found that the LIV model is robust against misspecified error distributions, yet the likelihood tends to display multiple local optima in these cases, in particular for the skewed gamma distribution for ϵ. Several issues seem worth noting. First, such local optima can be identified by running the algorithm multiple times from different starts. We found that these locally optimal solutions are often boundary solutions where the estimated correlation between ϵ and ν is close to 1, which may be used as an indicator of a problematic LIV estimate. We did not find multiple local optima when the distribution of ϵ is correctly specified, and, hence, presence of multiple optima itself may indicate a misspecified likelihood. Secondly, when their is no endogeneity, the LIV model does not take this nonnormality as evidence for the presence of endogeneity. Thirdly, and most importantly, we found that when the algorithm is started from the OLS solution, i.e. assuming that there is no endogeneity, it converges to the correct optimum in all cases and the LIV model estimates for the regression parameters are unbiased. We found that in case of misspecified distributions the LIV estimates of σ_ϵ^2 and $\sigma_{\epsilon\nu}$ have a larger root mean-squared error, but are still unbiased. Importantly, we found for β_1 that the LIV estimates are not aversely affected by a misspecified likelihood, and the results are very similar to the ones presented in Figure 1. Finally, residual checks as proposed in the appendix seem particularly relevant to identify cases in which the assumption of a normal distribution of the error is violated. These results corroborate that the LIV model is robust against misspecified error distributions, but it is advisable to start the maximum-likelihood estimation algorithms from the OLS estimates, to search for multiple local optima, and to investigate model residuals.

The results of our simulation studies for the Hausman-LIV test suggest that the test has reasonable power across a wide range of regressor-error correlations and for different kinds of distributions of the instruments. Furthermore, the proposed model test and estimation work well even if the number of instruments is misspecified. We find the test to be fairly robust under such misspecifications with a small bias towards rejecting the null hypothesis somewhat too often. These results are obtained without requiring observed instrumental variables, which is the main contribution and an important feature of the LIV method.

5. Application: The effect of education on earnings

In this section we consider OLS-, IV-, and LIV-estimation of the return to education. As will become clear, OLS-and IV-estimation is not straightforward in such applications. We provide LIV results for three empirical datasets in Section 5.3 and argue that our results are preferable to OLS and IV.

5.1. OLS estimation of returns to education

Over the past decades, much research has been conducted to investigate the return to education (Card, 1999, 2001). Most work has employed a version of the following linear regression model:

$$y_i = \beta_0 + \beta_1 S_i + X_i \beta_2 + \epsilon_i, \tag{8}$$

where y_i is the logarithm of a measure of earnings, S_i is a measure of education and X_i is a collection of other explanatory variables assumed to influence y_i. The disturbances ϵ_i represent all other influences not explicitly accounted for. If the disturbances are distributed independently of the explanatory variables S_i and X_i, the simple OLS estimator can be used to estimate β_1. However, the independence assumption may not be realistic.

Four major potential sources of bias have been identified in the literature on the relationship between education and income. Much work has focused on the issue whether the presence of a —so called—'ability bias' overstates the true causal effect of education on earnings (for instance, Angrist and Krueger, 1991; Harmon and Walker, 1995; Verbeek, 2000). 'Ability' can be seen as an omitted variable that enables (certain) individuals to obtain more income. When individuals with higher ability have chosen to obtain more education, the effect of education on income is overstated, since the effect of unobserved ability is falsely attributed to it. As such, exogenous shocks in education levels will have less effect on individual wages than what is predicted by the OLS regression model. Other sources of potential bias are error in the measurement of the education variable S_i, which may result in downward biases, heterogeneity in the regression coefficients, and optimizing behavior of individuals, that both may lead to either an upward or a downward bias in OLS (e.g. Chamberlain and Griliches, 1975; Griliches, 1977; Harmon and Walker, 1995; Card, 1999, 2001; Verbeek, 2000).

There is little agreement on the direction and magnitude of the potential bias in the OLS estimator of the return to education effect. This situation is not surprising in view of the many sources of potential regressor-error dependencies, with each of them having their own specific impact on the direction and magnitude of the bias in OLS. A further complicating factor is that these causes offset or enforce each other.

Card (1999, 2001) surveys several empirical studies on the return to education and finds regression estimates ranging from about 0.03–0.14. Quite often, the OLS estimates were not found to be statistically different from the instrumental variable estimates. As suggested above and discussed in more detail in Section 5.2, instrumental variables estimates for these kind of studies are potentially biased as well because the instruments used are possible weak and/or endogenous. Recent evidence from Twin studies suggests an upward bias in the OLS estimator of about 10–15% (cf. Card, 1999). The major advantage of using data on twins is that no observed instruments are required as the within-family estimator can be used.

5.2. IV estimation of returns to education

Given the divergent and a priori unknown sources of potential regressor-error dependencies in estimating the return to education, it is not an easy task to find appropriate instruments that alleviate regressor-error dependencies in model (8). Card (1999, 2001) gives an overview of recent studies that use instrumental variables to estimate the return to schooling. He distinguishes two sets of instrumental variables that are commonly used: (1) those that are based on institutional features of the school system and (2) those that are based on family background characteristics. We discuss them briefly.

Institutional features of the schooling system. When instrumental variables based on institutional features are used, the resulting IV estimates are approximately 30% higher than the corresponding OLS results. This finding does not agree with current beliefs in the literature about the traditional ability bias. Card (1999, 2001) provides a number of explanations. Firstly, instruments based on institutional features of the schooling system may not be truly exogenous, since a direct effect of the instruments on earnings may exist. This may hold for such instruments as 'college proximity', 'quarter-of-birth', and 'schooling reforms' (Verbeek, 2000; Bound and Jaeger, 1996; Angrist and Krueger, 1991; Card, 1999). Bound et al. (1995) show that, in finite samples, IV estimates based on weak instruments are biased in the same direction as OLS. Secondly, the downward bias in OLS can be a result of error in the measurement of education. However, the strength of this effect is doubtful in view of Card (1999) who argues that it is unlikely that measurement error alone can account for the large positive gap between IV and OLS estimates. Thirdly, factors like compulsory schooling or schooling availability are most likely to affect individuals who otherwise would have had relatively low schooling. If, because of potential heterogeneity, these individuals have higher than average marginal returns to schooling, then instruments based on these variables tend to recover the returns to education for a subset of individuals with relatively high returns to education, resulting in estimates higher than OLS. Uusitalo (1999) notes in this respect that presence of heterogeneity in the coefficient of the returns to education yields an additional error term $u_i^{\beta} S_i$. Since the instrument Z_i is correlated with S_i, it cannot be uncorrelated with the error term of the wage equation.

Family background. The second type of instrumental variables commonly used are instruments based on family background characteristics, for instance measures on education levels of family members (Card, 1999). The use of these variables as instruments is motivated by the fact that children's education tend to exhibit a high correlation with parents' education. However, he concludes that if the OLS estimator is upward-biased, one would expect an IV estimator based on family background to be even more upward-biased. He shows that when the OLS estimator is biased upward because of unobserved ability, the bias in the IV estimator is at least as large, and potentially larger, depending on the strength of the instruments and its possible direct effect on the dependent variable.

A particularly powerful approach to address regressor-error dependencies in schooling models is to use data on twins (or siblings) (Card, 1999). This approach attempts to eliminate possible omitted variable biases by assuming that some of the unobserved factors (e.g. ability or motivation) are identical within families (or twin/sibling pairs). In this case, differences of levels of schooling and education for the twins or siblings can be exploited to estimate the effect of education on wage. Card (1999) gives an overview of several studies that use twin-data. He finds that under the assumption that identical twins have identical abilities, the within-family estimator gives a consistent estimate for the average marginal returns to schooling. Furthermore, this estimator can be corrected for measurement error. Card (1999) concludes from his survey that the OLS estimator obeys a slight upward-bias of the order of 10–15%. A drawback of these methods is the (possible) lack of generalization to non-twins and the potential failure of the identical abilities assumptions for identical twins and siblings. If the assumption does not hold, twin studies might still overestimate to some extent the effect of education on earnings. In a recent study, Hertz (2003) also finds that OLS results are biased upward, based on measurement-error corrected estimators.

This review of the literature demonstrates that IV estimation has produced a less than satisfactory solution to the endogeneity problem of the schooling effect. In the following sections we present the LIV model results for three applications to examine the effect of education on income. Each of these three applications are based on previously published data. First we briefly describe the three datasets, where a more detailed description can be found in appendix 3. We then estimate model (8) with latent instrumental variables and compare these results with the traditional IV and OLS estimates. Furthermore, we investigate the strength of the available observed instruments, and conclude that the LIV results are to be preferred over IV and OLS.

5.3. LIV estimation of the returns to education

5.3.1 Data used.

NLSY data. The first dataset is a sample of 3010 men taken from the US National Longitudinal Survey of Young Men (NLSY) from 1976. This dataset is analyzed in Card (1995) and Verbeek (2000). The dataset contains several exogenous variables and one dummy instrumental variable measuring the presence of a nearby college, i.e. an instrument based institutional features of the school system.

Brabant data. The second dataset was originally sampled in 1952 from the Dutch province 'Noord-Brabant'. Thirty years later the same individuals were contacted to collect data on, among other things, educational level, income, and social background statistics. The labor market information used here is from 1983, and the dataset used contains observations on 833 men who had reached a stable labor market position. As with the NLSY dataset, several exogenous explanatory variables are available. We have two instrumental variables: measures on the educational level of the respondents' father and mother, i.e. family background characteristics (see also Hartog, 1988, for a more detailed description of the data).

PSID data. The third dataset contains data on 424 working, married white women between the ages 30 and 60 in 1975, and comes from the University of Michigan Panel Study of Income Dynamics (PSID), analyzed in Wooldridge (2002) and Mroz (1987). The labor market information is from 1975. This dataset has several exogenous variables. The available instruments are also family background variables: the respondents' fathers and mothers level of education. For more details on the datasets and the used regressors and instruments, we refer to Appendix 3.

The three datasets differ on various key aspects (sample sizes, region, sex of respondents, year of labor market information), which makes direct comparison of the estimated

Table 2. Results of OLS, IV and LIV for the schooling coefficient for the three datasets.

$\hat{\beta}_1$	OLS	IV	LIV2	LIV3	LIV4	LIV5
NLSY	0.074	0.133	0.050	0.065	0.068	0.069
	(0.0035)	(0.0518)	(0.0099)	(0.0041)	(0.0040)	(0.0040)
Brabant	0.043	0.056	0.040	0.042	0.040	-
	(0.0044)	(0.0075)	(0.0051)	(0.0049)	(0.0049)	
PSID	0.102	0.073	0.134	0.099	0.099	0.096
	(0.0139)	(0.0321)	(0.0282)	(0.0160)	(0.0153)	(0.0142)

regression coefficients superfluous. However, we compute the relative bias in OLS with respect to the LIV and IV estimates, which, as will become clear, straightforwardly allows us to compare the results across the three applications. The application of LIV with its assumption of discrete levels of the latent variable may well correspond to the existence of discrete levels of schooling, underlying the measured education variables, that are free of measurement error and that represent the levels of education that one would obtain regardless of ability, but is not predicated on that. The latent instrument relates to factors that affect the choice of education but do not directly affect income, such as e.g. cost-, parental-, or school-characteristics. Alternatively, as one reviewer to this study pointed out, LIV can be interpreted as identifying 'latent twins' and using an analogue of the twin estimator, conceptually.

5.3.2. LIV results. We estimate the LIV model with $m = 2, \ldots, 5$ and with the inclusion of extra exogenous variables. We emphasize that the LIV model does *not* require the availability of instrumental variables, and the results in this section are obtained *without* using the available observed instruments mentioned above. We also present here the results for the standard OLS estimator, the IV estimator, and LIV model fit diagnostics, but postpone a detailed discussion of the IV results until Section 5.3.3.

Estimated coefficients. In Table 2 we present the results for the estimated schooling coefficients for the datasets using OLS, IV, and LIV. It can be seen that for all specifications for m in the LIV model, the resulting estimate for the schooling coefficient is below the OLS estimate, indicating a small upward bias in the OLS estimate. On the other hand, the direction of the bias for the IV results using the observed instruments is not the same, and we discuss this in more detail below. The only downward bias found by LIV is for the PSID data when $m = 2$. This can be expected if a dummy variable exists which is identical or nearly identical to the unobserved instrument. In this case, there is a situation of (almost) perfect multicollinearity in the second stage of the LIV model and the parameters are only nearly identified. This also explains why the results for $m = 2$ have larger standard deviations than from what could have been expected and why relative large improvements in model fit occur for $m > 2$. In these applications several dummy regressors are present (see Appendix 3). In addition, the PSID data is the smallest dataset we have and the likelihood may be less smooth in this case. For the Brabant data the maximized value of the likelihood is degenerate[3] at $m = 5$ and no estimate for LIV5 is given in Table 2. Here the LIV method

3 The reason why the likelihood degenerates is that two categories of the LIV estimator are identical, so that in effect LIV recovers $m = 4$ here too.

Table 3. Computed values for the ICL criterion. Boldface values indicate the minimum value (row-wise). BIC and AIC3 are provided for comparison.

		$m = 2$	$m = 3$	$m = 4$	$m = 5$
NLSY	ICL	6942.75	5703.59	**5611.37**	5995.09
	BIC	5832.06	5404.04	5309.55	**5291.73**
	AIC3	5751.91	5313.86	5209.36	**5181.52**
Brabant	ICL	**1931.07**	1974.67	1990.93	2004.44*
	BIC	1867.02	1837.67	**1835.73**	1849.18*
	AIC3	1799.97	1763.17	**1753.77**	1759.77*
PSID	ICL	1199.23	1042.04	1020.93	**914.55**
	BIC	1164.49	1023.99	1005.42	**905.26**
	AIC3	1103.49	956.89	932.22	**825.97**

*Degenerate solution, where in fact the $m = 4$ solution is recovered, since two of the estimated class means are identical.

Table 4. Relative biases with respect to OLS and results for Hausman-test (P-values) to test for endogeneity (based on the Hessian matrix).

Data	Estimator	% Δ	P-values
NLSY	IV	−79.9	0.220
	Opt. LIV $m = 4$	7.9	0.002
	Opt. LIv $m = 5$	6.5	0.007
Brabant	IV	−30.1	0.034
	Opt. LIV $m = 2$	7.0	0.325
	Opt. LIV $m = 4$	7.0	0.105
PSID	IV	27.8	0.326
	Opt. LIV $m = 5$	5.5	0.040

indicates that the number of instruments (number of categories) should not be too large. Overall it can be seen from Table 2 that the LIV results are relatively stable for different choices of m. We consider choosing the 'best' m next.

Choosing the number of categories of the latent instrument. As argued before, we choose among the different values for m by looking at the ICL criterion, and for comparison and validity we also present AIC3 and BIC in Table 3. For the NLSY data the ICL statistic yields a minimum at $m = 4$. For the Brabant dataset ICL yields $m = 2$ and $m = 5$ for the PSID data. Importantly, we find that the estimated regression coefficients and the estimates for the schooling equation are not very different for the values of the selection criteria.[4] As we will show this result also holds for testing for (absence of) endogeneity. In the following we will only consider the optimal LIV results.

Testing for endogeneity using the Hausman-LIV test. Table 4 shows the results for the relative bias[5] in the estimated regression coefficient for schooling with respect to OLS for

4 The estimated schooling effect is given in Table 2. The results for the effects of the exogenous regressors, the π's, the λ's, and variance components, are presented in appendix 5B of Ebbes (2004).

5 We computed this percentage as $100 \times (1 - \hat{\beta}_1^{\text{LIV}}/\hat{\beta}_1^{\text{OLS}})$ and $100 \times (1 - \hat{\beta}_1^{\text{IV}}/\hat{\beta}_1^{\text{OLS}})$.

the IV and optimal LIV results. Furthermore, the test results for testing for absence of endogeneity are presented. We present the results for IV (2SLS) as well, but discuss the IV estimates and the used instruments in more detail later on. The test-statistics for LIV are computed *without* using the observed instrumental variables. The Hausman-test is based on comparing the complete vectors $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{\text{LIV}}$.[6]

Overall, we find that the differences between LIV and OLS are not large, which is also indicated by the Hausman-test (presented in the last column of Table 4). The optimal LIV solutions for the NLSY data and the PSID data indicate a significant upward bias in OLS, but for the Brabant data the estimated value for β_1 by LIV is not significantly different from OLS. Here, the classical IV estimator indicates a significant downward bias in OLS.

Before discussing the classical IV results in more detail, we first examine various diagnostics for the above presented LIV estimates, where we only report the results for the LIV model indicated by the (preferred) ICL-criteria. We note that for the Brabant data the R^2 measure for the strength of the LIV instruments is substantially better for $m = 4$ than for $m = 2$, which is discussed in Section 5.3.3.

Diagnostics: Outliers, influential observations, normality and heteroscedasticity. We examined the various diagnostics presented in Section 2.2 to investigate the fit of the (optimal) LIV model and to identify potential outliers and influential observations.

For the NLSY data ($n = 3010$), residual checks did not reveal heteroscedasticity, and residuals had a skewness of -0.28 and a kurtosis of 3.5. All standardized residuals (in absolute value) were smaller than 4.5. Examining the outliers and influential observations diagnostics did not identify highly unusual data.

For the PSID data ($n = 424$) there is evidence of weak heteroscedasticity for the variable 'experience', but this effect is rather small. The residuals are slightly skewed (-0.26) and are leptokurtic (kurtosis is 5.1). One observation was identified as an influential observation, but no outliers are present. When this observation is removed results and conclusions do not change, and all standardized residuals are smaller than 4 (in absolute value).

As for the PSID data, the results for the Brabant data ($n = 833$) indicate slight evidence of weak heteroscedasticity, here for the dummy variable whether the father is self employed at the age of 12. For this dataset, the residuals are more skewed (skewness is -1.25) and more leptokurtic (kurtosis is 12.7). Examination for potential outliers and influential data identifies three observations that clearly do not 'fit' the rest of the data. We re-estimated the model without these three observations, and found that the estimates and test statistics are not affected. The Hausman-statistic to test for endogeneity (see Table 4) for the $m = 4$ solution now becomes 3.47, which is significant at $\alpha = 0.10$. After omission of these outlying data, the residuals are less skewed and leptokurtic. All but four of the absolute values of the standardized residuals are smaller than 4.5, with a maximum of 5.9.

5.3.3 Relative biases and comparison with classical IV. It can be seen from the relative percentage bias in OLS with respect to the optimal LIV and IV estimates in Table 4 (the

6 We used the regression-based form of the Hausman test for the classical IV (2SLS) estimator, which is easier to compute in this case and equivalent to the original form of the Hausman test (e.g. Wooldridge, 2002). See also Ebbes (2004) for an alternate account and for the estimates of the complete vector of regression coefficients.

column indicated by $\%\Delta$), that the LIV method reveals an upward bias of OLS ranging from 5.5–8%. When traditional IV is used, the conclusions are very different for the three studies, ranging from an $\approx$ 80% downward bias to an $\approx$30% upward bias in OLS. For the NLSY data, the IV estimate for the return to schooling, based on a dummy for college proximity, is about 80% higher than OLS and equal to (approximately) 0.13 (0.052). For the Brabant data, we find that the IV estimate is 0.056 (0.008), which is also substantially higher ($\approx$30%) than OLS. Here the instruments are the levels of education of the respondents' parents. Using the same set of instruments, we find for the PSID data an upward bias of $\approx$30% in the OLS estimate. It can be seen that in all cases the IV estimate has a standard deviation that is substantially higher than OLS. The instability of the 2SLS results and the high standard deviations may be a result of weak and/or endogenous instruments.

Strength of the available observed instruments. In applying classical IV estimation it is recommended to report the R^2 or F-statistic from the regression of the endogenous regressors on the instrumental variables, i.e. on the instruments and the other explanatory variables (Bound et al., 1995). When the instruments explain only a small part of the variation of the endogenous regressors, the instruments are weak and using the IV results in this case is not recommendable. Instruments can be computed as a byproduct of the LIV results by computing a posteriori category membership using Bayes theorem. Subsequently, these estimates can be 'treated' as observed instruments.

The third column of Table 5 reports the difference in R^2 of the regression of schooling on the explanatory variables and the available observed instruments, or, in case of LIV, the optimal LIV instruments, and the R^2 of the regression of schooling on the explanatory variables only. Hence, a large increase in R^2 indicates that the instruments explain a substantial amount of the variance in the endogenous schooling variable. It can be seen that in particular for the the NLSY data the observed instrument 'Nearc' appears to be weak. The family background instruments (Brabant and PSID data) explain a larger part of the variance in schooling, in particular for the PSID dataset. However, the increase in R^2 is in all cases substantially larger when using the optimal LIV-instruments. It follows that the optimal LIV instruments do a much better job in explaining the variance of x than the available observed instruments, because they are estimated from the available data. These findings explain the

Table 5. Results strength of observed versus predicted LIV instruments. Instruments NLSY: 'Nearc', Brabant and PSID: 'FatherEd' and MotherEd', respectively (based on the Hessian matrix).

Data	Method	ΔR^2
NLSY	Obs IV	0.0029
	LIV4 IV	0.7503
	LIV5 IV	0.7976
Brabant	Obs IV	0.0922
	LIV2 IV	0.3906
	LIV4 IV	0.5247
PSID	Obs IV	0.1658
	LIV5 IV	0.8312

loss of efficiency in the 2SLS estimates for the regression coefficients in Table 2, where the IV estimated standard deviations are (0.052), (0.008), and (0.032) and, respectively, 14.8, 1.7, and 2.3 times higher than the OLS standard deviations. Not surprisingly, the estimated standard deviations for (the optimal) LIV estimates are only 1.14, 1.16, and 1.02 times the OLS estimated standard deviations.

These results illustrate the difficulties associated with classical IV estimation in these applications. The conclusions for the three datasets with respect to the magnitude and sign of the bias in the estimated OLS coefficient for schooling differ highly, even with a similar set of instruments. Card's (1999) reasoning that the IVs used are potential endogenous may explain part of this variability. He argues that instruments based on family background characteristics are likely to be endogenous. Furthermore, the best available empirical evidence from studies on identical twins suggest a small upward bias on the order of 10–15% in the OLS estimator (cf. Card, 1999), which is not supported by the standard IV estimates from the three datasets analyzed here. Our estimates have the same order of magnitude found in the twin studies but do not fully recover the 10% difference. A reason for this result might be that estimating the model by simple OLS yields in general only a modest fit (the OLS results presented in Table 2 have R^2's of respectively 29, 23 and 17%), i.e. the regressors do not explain a large part of the variance in wage. The fact that LIV indicates a smaller positive bias might also indicate that a part of the positive ability bias is offset by negative biases due to e.g. measurement error or heterogeneity, which is expected to be less in the twin studies. Further, in the twin studies there may still be a limited amount of unobserved ability if the abilities of twins and siblings differ.

6. Conclusions

Searching for valid instruments is a long-standing problem in estimating IV models that account for regressor-error problems in the social sciences. In addition, the identification of regressor-error correlation has been impossible without such valid instruments. Our proposed instrument-free approach presents a practical solution to this circular problem: it can be used to estimate regression parameters and test for regressor-error correlations without the necessity of first finding 'valid' instruments.

We proved that the LIV model is identifiable through the likelihood and, hence, usual results on consistency of maximum likelihood estimation can be applied. A potential limitation of the model is that it is predicated on distributional assumptions of the x and y variables. However, the Monte Carlo studies show that the model yields unbiased results for several types of distributions for x, and it is robust to misspecification of the error distribution on ϵ and to the particular choice of m. The Hausman-LIV test detects departures from independence of regressor and model error with a reasonable power across a wide range of regressor-error correlations. Importantly, the test, as well as the estimates for β, are rather insensitive to the true number of instruments and the distribution of the true instruments. In the case of severe model violations, the Hausman-LIV test becomes too strict in rejecting the null hypothesis when it is true. As a result, in applications of this test researchers may search unnecessarily for manifest instruments in a small fraction of cases. We feel that this is a small price to pay in view of the simplicity and ease of implementing the proposed test.

However, of course, the LIV method is not without caveats. Firstly, one may encounter situations in which the model is severely misspecified, in particular with respect to the distribution of ϵ where it may yield biased estimates. This is one reason why we consider the analysis of local optima, residuals (even from OLS as a first indication), outliers, influential cases, model fit, and diagnostic tests of endogeneity as important to address the former question. We illustrated this approach in our empirical analysis of the education effects on earnings. In addition, the simulation studies suggest that moderate misspecifications of the distribution of ϵ are not a severe problem. Secondly, a latent instrument may not be identifiable from the data, which happens if the true data-generating process is $x_i = c + \nu_i$, where c is a constant. In this case x is normally distributed and the LIV model is unidentified. For other error distributions identifiability still seems to be possible, but future research is needed at this point. A situation where x had an unimodel, symmetric distribution, and was thus approximately normally distributed, was investigated in the simulation study and we found that in this case the LIV model yields unbiased but less efficient results, reflecting the near unidentifiability. The distribution of x can be investigated a priori for that purpose. Thirdly, the assumption that the latent levels of the endogenous regressor are uncorrelated with the error term of the dependent variable are not testable and they hold by assumption as in the classical instrumental variables framework. This assumption implies that the endogenous regressor can be split in an endogenous and exogenous part. The idea to decompose the endogenous regressor to correct for regressor-error biases is used in other studies as well, e.g. Manchanda et al. (2004), Van Dijk et al. (2004), Bronnenberg and Mahajan (2001), or Chamberlain and Griliches (1975).[7] It may be important to stress that we view these latent instruments as auxiliary parameters that one needs to get unbiased estimates, but that one may not be primarily interested in themselves.

The studies of Card (1999, 2001) clearly indicate the difficulties associated with applying standard IV estimation to estimating the returns to education. The results in this area are often found to be counterintuitive and different across studies. Furthermore, in many instances it can be questioned whether the instruments used were 'valid'. Unfortunately, in general it is not possible to test for the validity of a specific instrument. We show that the LIV method can be successfully applied to solve these problems. The OLS estimates are found to be biased upward by about 7%. Equally important, the instruments that have been used in the literature seem to be often inadequate, and produce results that are both more biased than the OLS results and have much lower efficiency.

The advantage of the LIV approach is that no observable instruments are needed. Furthermore, once estimates have been obtained, endogeneity can be tested for. For the NSLY and the PSID dataset we find significant evidence of an 'ability' bias. Furthermore the standard errors of the estimates are much smaller than the standard errors for classical IV, and not much larger than for OLS. Because of the relative large number of observations in the NLSY data, it is to be expected that the power of the Hausman test is larger. In using LIV to test for endogeneity it is recommended to use datasets of substantial size to ensure a reasonable power. Using the proposed diagnostics, we do not find any evidence that the LIV models used for the three applications here do not fit the data well.

7 We thank an anonymous reviewer for pointing this out to us.

The relative size and magnitude of the bias in the OLS estimator that was found is somewhat smaller, but still close to the numbers reported in Card (1999) for the Twin studies: 6–8% for all three datasets. The results in this paper are convergent and add credibility to the LIV approach. We conclude that the LIV model presents a solution to the circular problem of searching for valid instruments in empirical studies. The model and Hausman-LIV test are fairly simple and easy to implement, and the results in this paper illustrate its usefulness across a wide variety of problems.

Appendix 1: Identification

The parameters π and σ_v^2 are identified using first and second order moments, but the parameters $\beta_0, \beta_1, \sigma_\epsilon^2$, and $\sigma_{\epsilon v}$ are not identified in this case. However, these parameters become identified by considering finite mixtures. Let

$$\mathcal{F} = \{F(x, \theta), \theta \in \Theta, x \in \mathbb{R}^d\}$$

be the class of d-dimensional distribution functions from which mixtures are to be formed. Here Θ will be a Borel measurable set in $\mathbb{R}^q$ and θ is formed from the elements of the mean (3) and variance (4). For the simple LIV model above, $d = 2, q = 5$, and $F(\cdot, \theta)$ is a bivariate normal c.d.f..

The class of finite mixtures $\mathcal{H}$ generated by $\mathcal{F}$ is defined as

$$\mathcal{H} = \left\{H(x) : H(x) = \sum_{j=1}^{m} \psi_j F(x, \theta_j), \psi_j > 0, \right.$$
$$\left. \sum_{j=1}^{m} \psi_j = 1, F(x, \theta_j) \in \mathcal{F}, \ \forall j, m = 1, 2, \ldots; x \in \mathbb{R}^d \right\}. \quad \text{(A1.1)}$$

So, $\mathcal{H}$ is the convex hull of $\mathcal{F}$. For the sake of simplicity, we will use some abbreviations: $F(x, \theta_j)$ will be written as $F_i(x)$ or just F_j and the (corresponding) mixture as $H = \sum_{j=1}^{m} \psi_j F_j$. We use definition 1 for the identifiability of mixtures in $\mathcal{H}$. Here we are interested in pure mixtures ($\psi_j > 0$) of order $m = 1, 2, \ldots$.

Definition 1. Suppose H and H' are any two members of $\mathcal{H}$, given by

$$H = \sum_{j=1}^{m} \psi_j F_j, \quad H' = \sum_{j=1}^{m'} \psi'_j F'_j. \quad \text{(A1.2)}$$

$\mathcal{H}$ is *identifiable* when $H \equiv H'$ if and only if $m = m'$, and the order of summation can be chosen such that $\psi_j = \psi'_j$, $F_j = F'_j$, $j = 1, \ldots, m$.

Stated differently, $H \in \mathcal{H}$ is identifiable, if there is a unique solution (up to a permutation of subscripts) of the identity defining H in (A1.2). Several theorems on the identifiability of finite mixtures are available and linear independence of the members of $\mathcal{F}$ is the key to answering the question (cf. Titterington et al., 1985). Core papers on this issue are Teicher (1963) and Yakowitz and Spragins (1968).

In several studies it is proved that certain families $\mathcal{F}$ of d-dimensional c.d.f.'s, for instance Gaussian c.d.f.'s, generate identifiable finite mixtures. As we show below, these results do not carry over directly to the LIV model and we have to prove identifiability in two steps. Similarly, related work by Hennig (2000) on identifiability of mixtures of regressions cannot be extended straightforwardly because of structural differences between his and our framework, and model assumptions (in the context of Hennig at least $\sigma_{\epsilon\nu} \neq 0$ is required whereas $\sigma_{\epsilon\nu} = 0$ is also of interest in our model).

Let

$$\begin{aligned}\mathcal{F}_{\beta,\Sigma} = \{F \mid F &\text{ is a bivariate normal c.d.f. on } \mathbb{R}^2 \text{ of the pair } (y_i, x_i)\\ &\text{with mean and variance } (\mu(\beta,\pi), \Omega(\beta,\Sigma)), \pi \in \mathbb{R}\}\end{aligned} \tag{A1.3}$$

where $\mu(\beta,\pi) = (\beta_0 + \beta_1\pi, \pi)'$ and $\Omega(\beta,\Sigma) = \Omega$ as in (4), be the class of general LIV models, β and Σ are known, and $\mathcal{H}_{\beta,\Sigma}$ the set of all pure finite mixtures of order m of the class $\mathcal{F}_{\beta,\Sigma}$ We will consider general $m > 1$. We apply standard results of identifiability of finite mixtures to establish identifiability of the class $\mathcal{H}_{\beta,\Sigma}$ in terms of the π's and the mixture probabilities. However, the identifiability of the parameters β and Σ does not follow immediately. In fact, we are not seeking for the identifiability of $\mathcal{H}_{\beta,\Sigma}$ but for the identifiability of the larger class

$$\mathcal{G} = \bigcup_{\beta,\Sigma} \mathcal{H}_{\beta,\Sigma}. \tag{A1.4}$$

In the following, we first prove the identifiability of the class $\mathcal{H}_{\beta,\Sigma}$. Subsequently, we use the identifiability of $\mathcal{H}_{\beta,\Sigma}$ to prove identifiability of $\mathcal{G}$ the class of LIV models. Identifiability of $\mathcal{G}$ is equivalent with identifiability of the parameters in the general LIV model.

Proof of identifiability of $\mathcal{H}_{\beta,\Sigma}$. Let $F_{j,X}$ be the marginal distribution function of F_j for X. More specifically, $F_{j,x}(x) = \lim_{y\to\infty} F_j(y,x)$. From (1) and (3) it can be seen that X has mean π_j, and variance ω_{22}. Here, all $F_{j,X}$, $j = 1, 2, \ldots$, are normal distribution functions with different location parameters but with the same variances.

Since we assume for the moment that β and Σ are known, identifiability of the class $\mathcal{H}_{\beta,\Sigma}$ is established if there is a unique solution of $F(y,x) = \sum_{j=1}^{m} a_j F_j(y,x)$ in terms of a_j and π_j for $m = 1, 2, \ldots$. Hence, we only have to look at the marginal distribution of X, since this distribution contains all the relevant parameters. The c.d.f. F is a finite mixture of $N(\pi, \sigma_\nu^2)$, distribution functions with $\pi \in \mathbb{R}$ and σ_ν^2 is fixed for the moment. According to proposition 1 of Teicher (1963) or proposition 2 of Yakowitz and Spragins (1968) F is identifiable. It follows[8] that there is a unique solution in terms of m, a_j and π_j for $j = 1, 2, \ldots, m$ for any $F \in \mathcal{H}_{\beta,\Sigma}$.

8 It can be seen from the following that as long as x can be unmixed, the parameters of the LIV model are identifiable. A mixture of normal distributions can be used for that purpose, but this is not the only family that yields identifiable mixtures (Redner and Walker, 1984; Al-Hussaini and Ahmad, 1981; Yakowitz and Spragins, 1968), and the normality assumption is not crucial. We already observed in Section 4 that the maximum-likelihood method is quite robust for moderate deviations of normality in the LIV model. We plan to report on identifiability results for the LIV model with more general distributed error terms in the future.

In the preceding we assumed β and Σ known, which will not be the case in general. But the previous result can be used to proof identifiability of the larger class $\mathcal{G} = \bigcup_{\beta,\Sigma} \mathcal{H}_{\beta,\Sigma}$ which is the union across all possible values of β, Σ. If a distribution from this class has a unique solution in terms of its unknown parameters, than the parameters of the general LIV model are identified (including the relative class sizes).

Proof of identifiability of $\mathcal{G}$. In the following we prove that $\mathcal{G}$ is also identified, i.e. we prove the following theorem.

Theorem 1. *Assume that $m \geq 2$. $\mathcal{H}_{\beta,\Sigma}$ is identified for all β and Σ positive semi-definite $\Leftrightarrow$ $\mathcal{G}$ is identified.*

Proof: ($\Rightarrow$) Let $F, G \in \mathcal{G}$ such that $F \equiv G$, where

$$F = \sum_{i=1}^{m} a_i F_i \in \mathcal{H}_{\beta,\Sigma}$$
$$G = \sum_{j=1}^{k} b_j G_j \in \mathcal{H}_{\delta,\Psi},$$

and $a_1, \ldots, a_m$ and $b_1, \ldots, b_k$ are the positive mixing proportions, and the distributions $F_1, \ldots, F_m \in \mathcal{F}_{\beta,\Sigma}$ and $G_1, \ldots, G_k \in \mathcal{F}_{\delta,\Psi}$ are different in terms of their means and variances, i.e.

$$\begin{aligned} &F_i \text{ is the c.d.f. of } N_2(\mu(\beta, \pi_i), \Omega(\beta, \Sigma)) \\ &G_i \text{ is the c.d.f. of } N_2(\mu(\delta, \gamma_i), \Omega(\delta, \Psi)). \end{aligned} \tag{A1.5}$$

We need to show that $F \equiv G$ implies $m = k$, $a_i = b_i$ and $F_i = G_i$ modulo permutation (definition 1). By definition 1, $\mathcal{G}$ is identified if $F \equiv G$ implies that $m = k$, $a_i = b_i$, and $F_i = G_i$ eventually after relabeling (and vice versa) for $i = 1, \ldots, k$.

$F \equiv G$ implies that

$$\sum_{i=1}^{m} a_i F_i = \sum_{j=1}^{k} b_j G_j. \tag{A1.6}$$

Both F and G have unique representations (up to a permutation of indices) in terms of m, a_i, and π_i, respectively, k, b_j, and γ_j because $\mathcal{H}_{\beta,\Sigma}$ and $\mathcal{H}_{\delta,\Psi}$ are assumed to be identified ($\Rightarrow$). In (A1.6) we have a finite mixture of bivariate normal distribution functions. According to proposition 2 of Yakowitz and Spragins (1968) such mixtures are identifiable, hence we must have $m = k$, $a_i = b_i$, and $F_i = G_i$ (eventually after relabeling). But $F_i = G_i$ implies that $\mu(\beta, \pi_i) = \mu(\delta, \gamma_i)$ and $\Omega(\beta, \Sigma) = \Omega(\delta, \Psi)$. Thus,

$$\beta_0 + \beta_1 \pi_i = \delta_0 + \delta_1 \gamma_i \tag{A1.7}$$
$$\pi_i = \gamma_i \tag{A1.8}$$
$$\beta_1^2 \sigma_\nu^2 + \sigma_\epsilon^2 + 2\beta_1 \sigma_{\epsilon\nu} = \delta_1^2 \psi_\nu^2 + \psi_\epsilon^2 + 2\delta_1 \psi_{\epsilon\nu} \tag{A1.9}$$

$$\beta_1\sigma_\nu^2 + \sigma_{\epsilon\nu} = \delta_1\psi_\nu^2 + \psi_{\epsilon\nu} \tag{A1.10}$$

$$\sigma_\nu^2 = \psi_\nu^2 \tag{A1.11}$$

for $i = 1, \ldots, m$. Since the $F_i = G_i$ are different for $i = 1, \ldots, m$, we have $\pi_i \neq \pi_j$ and $\gamma_i \neq \gamma_j$ for all $i \neq j$. Using this and $m \geq 2$, it follows from (A1.7) and (A1.8) that $\beta_0 = \delta_0$ and $\beta_1 = \delta_1$. Subsequently, from (A1.9)–(A1.11), we have $\sigma_\epsilon^2 = \psi_\epsilon^2$, $\sigma_\nu^2 = \psi_\nu^2$ and $\sigma_{\epsilon\nu} = \psi_{\epsilon\nu}$. So, $F \in \mathcal{G}$ has an unique representation and $\mathcal{G}$ is thus identified. The reverse of the proof ($\Leftarrow$) follows immediately (i.e. a subset of an identified set must be identified as well).

To conclude, from theorem 1 it follows that if $m \geq 2$ and all the group means π_j, $j = 1, \ldots, m$, are different, the parameters of the LIV model in (1) with normally distributed errors are identified, including the mixture probabilities.

Appendix 2: LIV residuals and measures for outliers and influential observations

Conditional residuals. One way to examine residuals is to look at the conditional distribution of y given x in the LIV model (for the sake of notational simplicity we omit other exogenous regressors here). Conditional on category j, we obtain from (1): $(y_i \mid x_i, j) \sim N(\mu_{y\mid x,j}, \sigma^2_{y\mid x,j})$, where the conditional mean of $y_i|x_i, j$ is

$$E(y_i \mid x_i, j) = (\beta_0 - \frac{\sigma_{\epsilon\nu}}{\sigma_\nu^2}\pi_j) + (\beta_1 + \frac{\sigma_{\epsilon\nu}}{\sigma_\nu^2})x_i \tag{A2.1}$$

$$= \beta_0 + \beta_1 x_i + \frac{\sigma_{\epsilon\nu}}{\sigma_\nu^2}(x_i - \pi_j), \tag{A2.2}$$

with $x_i - \pi_j = \nu_i$, and $\mathrm{var}(y_i \mid x_i, j) = \sigma_\epsilon^2 - \frac{\sigma_{\epsilon\nu}}{\sigma_\nu^2}$ The residual $e_i = y_i - \hat{y}_i$ with $\hat{y}_i = \hat{y}_{y_i|x_i j,}$ is equal to $e_i = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i - (\hat{\sigma}_{\epsilon\nu}/\hat{\sigma}_\nu^2)\hat{\nu}_i + \epsilon_i$ with $\hat{\nu}_i = x_i - \hat{x}_i = (\pi_j - \hat{\pi}_j) + \nu_i$. Because of the presence of ν_i, normality of ϵ_i, alone cannot be examined via this type of residuals. Since a priori it is not know to which category/group individual i belongs, we use the a posteriori group probabilities $\tilde{p}_{ij}$, and compute the residuals as $e_i = y_i - \hat{y}_i$ where $\hat{y}_i = \sum_{j=1}^m \tilde{p}_{ij}\hat{\mu}_{y_i|x_i,j}$.

IV-type residuals. In classical IV estimation with observed instruments, the residual is estimated as $y - X\hat{\beta}$ and not as $y - \hat{X}\hat{\beta}$, see e.g. Greene (2000) or Pagan (1984). Using the latter residuals results in an incorrect estimate of the standard errors. Applying this result to LIV gives $e_i = y_i - \hat{y}_i = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i + \epsilon_i$ where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, and the estimated values are the maximum likelihood estimates from LIV. Note that there is no 'direct' effect of ν_i Furthermore, there is no need in using the a posteriori group memberships. Unfortunately, we found this type of residual to be misleading in detecting heteroscedasticity (see Ebbes, 2004, for more details).

Outliers and influential observations. Once the model is estimated, we propose to use the following measures to determine the influence of observation i on:

1. the likelihood, measured by the likelihood distance $LD(i) = 2[LL(\hat{\theta}) - LL(\hat{\theta}(i))]$ where $LL(\theta)$ denotes the value of log-likelihood for the complete sample in point θ and $\hat{\theta}(i)$ is the maximized value of the log-likelihood leaving observation i out of the sample.

Table 6. Descriptive statistics NLSY dataset ($n = 3010$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant	–	–
schooling (β_1)	Years of schooling in 1976	13.26	2.68
experience (β_2)	Potential experience	8.86	4.14
black (β_3)	Equals 1 if black	0.23	0.42
smsa (β_4)	Equals 1 if lived in metropolitan area in 1976	0.71	0.45
south (β_5)	Equals 1 if lived in south in 1976	0.40	0.49
Dependent			
log wage	Logarithm of hourly wage	6.26	0.44
Instruments			
Nearc	Grew up near a 4 year college	0.68	0.47

2. the estimated parameters, measured by Cook's distance $CD(i) = (\hat{\theta} - \hat{\theta}(i))' H(\hat{\theta})(\hat{\theta} - \hat{\theta}(i))$, where $H(\hat{\theta})$ is Hessian evaluated at $\hat{\theta}$.
3. the estimated covariance matrix, measured by $COVRATIO1(i) = \det[V(\hat{\theta}(i))]/\det[V(\hat{\theta})]$ where $V(\theta)$ denotes the estimated variance covariance matrix for θ
4. the estimated covariance matrix of (ϵ, ν), given by $COVRATIO2(i) = \det[\hat{\Omega}(i)]/\det[\hat{\Omega}]$, where Ω is given in (4). Because Ω is essential in correcting for endogeneity, this measure may point towards observations having a large effect on the relation between ϵ and ν.

Our experience based on simulation studies is that the four measures mention above, together with an examination of the residuals, can be fruitfully applied to detect outliers and influential observations. We propose examining the ranking of the largest values of $LD(i)$ or $CD(i)$, and of $|COVRATIO1(i) - 1|$ and $|COVRATIO2(i) - 1|$, where large jumps between subsequent observations indicate potential influential or outlying observations.

Appendix 3: Descriptive statistics used datasets

NLSY data. The total sample consists of 3010 men taken from the National Longitudinal Survey of Young Men (see Verbeek, 2000; Card, 1995).[9] In this survey, a group of individuals in the age of 14–24 years is followed since 1966. The labour market information used is from 1976. In this year, the individuals had on average a little more than 13 years of schooling, with a maximum of 18 years. The average working experience was about 8.86 years (in 1976 those men aged 24–34) with an average hourly wage rate of \$5.77. The variables used can be found in Table 6. We used the values centered around the mean for schooling in estimation.

Brabant data. The initial dataset used in this paper consisted of 839 observations, but we deleted 5 observations with very low wages (log hourly wages < 0). Another observation with an extreme large reported wage was also removed ($> 9 \times$ IQR from median). This

9 http://www.econ.kuleuven.ac.be/GME/.

Table 7. Descriptive statistics Brabant dataset ($n = 833$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant	–	–
schooling (β_1)	Years of schooling after age 12	4.35	4.00
experience (β_2)	Potential experience	25.52	4.19
nr. children (β_3)	Number of children present at age 12	4.91	2.68
av. mark (β_4)	Average school mark in final year of primary education	5.62	1.42
anti-social (β_5)	Equals 1 comes from antisocial background	0.10	0.29
self (β_6)	Equals 1 if father is self employed at age 12	0.31	0.46
Dependent			
log wage	Logarithm of hourly wage	2.70	0.42
Instruments			
Father Ed.	Education level father	2.35	0.70
Mother Ed.	Education level mother (levels: 1–6, higher categories = higher education)	2.22	0.54

Table 8. Descriptive statistics PSID dataset ($n = 424$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant		
schooling (β_1)	Years of schooling	12.66	2.29
experience (β_2)	Actual labor market experience	13.09	8.05
kidslt6 (β_3)	Number of children younger than 6	0.14	0.39
kidsgr6 (β_4)	Number of children older than 6	1.34	1.32
unempl (β_5)	Unemployment rate in county of residence	8.54	3.04
city (β_6)	Equals 1 if lives in SMSA	0.64	0.48
nwincome (β_7)	Family income less total income wife/1000	18.99	10.62
Dependent			
log wage	Logarithm of hourly wage	1.22	0.67
Instruments			
Father Ed.	Years of schooling father	8.80	3.57
Mother Ed.	Years of schooling mother	9.24	3.37

data was collected in 1983 in the Netherlands' southern province of Noord-Brabant. At that time the average age of the men in the sample was about 43. This cohort was confronted with compulsonary schooling until 12 years of age. The schooling measure used is the number of post-compulsonary years of schooling; on average 4.35 years. The average hourly wage was Dfl. 16.72 and the individuals had, on average, 25 years of work experience at the time of the survey. See Table 7 for more information on the variables used. As before, we centered the schooling variable around the mean.

PSID data. As for the Brabant dataset, we removed a few observations prior to data analysis: four observations had an obvious lower (log) wage ($\ll -1$) than the rest and were not used for estimation. This data come from the University of Michigan Panel

Study of Income Dynamics (PSID),[10] obtained in 1976 (also used in Mroz, 1987). The sample consists of working married white women, who were aged in between 30 and 60 in 1975. They earned of average \$4.18 per hour. The women reported an average 12.7 years of schooling and a little over 13 years of labor market experience. For a detailed description of the used variables, see Table 8, where for estimation, the schooling variable was mean-centered.

Acknowledgments

The research of Peter Ebbes was supported in part by the 'Prince Bernhard Cultural Foundation' and the 'Netherlands Organization for Scientific Research'. Ulf Böckenholt was supported in parts by grants from the Social Sciences and Humanities Research Council of Canada and the Canadian Foundation of Innovation. This study was completed when the first author was a PhD student at the University of Groningen (NL) and visiting at the Ross School of Business at the University of Michigan.

References

Al-Hussaini, E.K. and K.E.-D. Ahmad. (1981). "On the Identifiability of Finite Mixtures of Distributions." IEEE Transactions on Information Theory 27, 664–668.

Angrist, J.D. and A.B. Krueger. (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" The Quarterly Journal of Economics 56, 979–1014.

Bekker, P.A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators." Econometrica 62, 657–681.

Belsley, D.A., E. Kuh, and R.E. Welsch. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons, Inc.

Bound, J. and D.A. Jaeger. (1996). "On the Validity of Season of Birth as an Instrument in Wage Equations A Comment on Angrist and Krueger's Does Compulsory School Attendance Affect Schooling and Earnings?" Technical Report 5835, NBER.

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, 443–450.

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

Bronnenberg, B.J. and V. Mahajan. (2001). "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables." Marketing Science 20, 284-299.

Card, D. (1995). "Using Geographical Variation in College Proximity to Estimate the Return to Schooling." In L.N. Christofides, E. Grant, and R. Swidinsky (eds.), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, Toronto: University of Toronto Press, pp. 201–222.

Card, D. (1999). "The Causal Effect of Education on Earnings." In Ashenfelter, O.C. and D. Card (eds.), Handbook of Labor Economics volume 3A, North-Holland: Elsevier Science B.V., pp. 1801–1863

Card, D. (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." Econometrica 69, 1127–1160.

Chamberlain, G. and Z. Griliches. (1975). "Unobservables with a Variance-Components Structure: Ability, Schooling, and the Economic Success of Brothers." International Economic Review 16, 422–449.

Cook, R.D. and S. Weisberg. (1982). Residuals and Influence in Regression. New York: Chapman and Hall.

Dijk, van, A., H.J. van Heerde, P.S.H. Leeflang, and D.R. Wittink. (2004). "Similarity-Based Spatial Methods for Estimating Shelf Space Elasticities From Correlational Data." Quantitative Marketing and Economics 2, 257–277.

10 http://mitpress.mit.edu/Wooldridge-EconAnalysis.

Ebbes, P. (2004). "Latent Instrumental Variables: A New Approach to Solve for Endogeneity." PhD thesis, SOM Research School, University of Groningen (http://irs.ub.rug.nl/ppn/270945334).

Erickson, T. and T.M. Whited. (2002). "Two-Step GMM Estimation of the Errors-in-Variables Model Using High-Order Moments." Econometric Theory 18, 776–799.

Fahrmeir, L. and G. Tutz. (1994). Multivariate Statistical Modelling Based on Generalized Linear Models. New York: Springer-Verlag.

Fox, J. (1991). Regression Diagnostics. London: Sage Publications, inc.

Greene, W.H. (2000). Econometric Analysis. New Jersey: Prentice-Hall, Inc., Upper Saddle River.

Griliches, Z. (1977). "Estimating the Returns to Schooling: Some Econometric Problems." Econometrica 45, 1–22.

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

Harmon, C. and I. Walker. (1995). "Estimates of the Economic Return to Schooling for the United Kingdom." American Economic Review 85, 1278–1286.

Hartog, J. (1988). "An Ordered Response Model for Allocation and Earnings." Kyklos 41, 113–141.

Hausman, J.A. (1978). "Specification Tests for Econometrics." Econometrica 46, 1251–1271.

Hennig, C. (2000). "Identifiability of Models for Clusterwise Linear Regression." Journal of Classification 17, 273–296.

Hertz, T. (2003). "Upward Bias in the Estimated Returns to Education: Evidence from South Africa." The American Economic Review 93, 1354–1368.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T.-C. Lee. (1985). The Theory and Practice of Econometrics. New York: John Wiley & Sons Inc.

Madansky, A. (1959). "The Fitting of Straight Lines When Both Variables are Subject to Error." Journal of the American Statistical Association 54, 173–205.

Manchanda, P., P.E. Rossi, and P.K. Chintagunta. (2004). "Response Modeling with Non-Random Marketing Mix Variables." Journal of Marketing Research 41, 467–478.

McLachlan, G.J. and D. Peel. (2000). Finite Mixture Models. New York: John Wiley & Sons, Inc.

Mroz, T.A. (1987). "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." Econometrica 55, 765–799.

Pagan, A. (1984). "Econometric Issues in the Analysis of Regressions with Generated Regressors." International Economic Review 25, 221–247.

Ploeg, van der, J. (1997). "Instrumental Variable Estimation and Group-Asymptotics." PhD thesis, SOM Research School, University of Groningen (http://irs.ub.rug.nl/ppn/157854507).

Redner, R.A. and H.F. Walker. (1984). "Mixture Densities, Maximum Likelihood and the EM Algorithm." SIAM Review 26, 195–239.

Ruud, P.A. (2000). An Introduction to Classical Econometric Theory. New York: Oxford University Press.

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, 518–529.

Teicher, H. (1963). "Identifiability of Finite Mixtures." The Annals of Mathematical Statistics 34, 1265–1269.

Titterington, D.M., A.F.M. Smith, and U.E. Makov. (1985). Statistical Analysis of Finite Mixture Distributions. Chichester: John Wiley & Sons Ltd.

Uusitalo, R. (1999). "Essays in Economics of Education." PhD thesis, University of Helsinki.

Verbeek, M. (2000). A Guide to Modern Econometrics. Chichester: John Wiley & Sons Ltd.

Wald, A. (1940). "The Fitting of Straight Lines if Both Variables are Subject to Error." The Annals of Mathematical Statistics 11, 284–300.

White, H. (1980). "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." Econometrica 48, 817–838.

Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge: Massachusetts Institute of Technology.

Yakowitz, S.J. and J.D. Spragins. (1968). "On the Identifiability of Finite Mixtures." The Annals of Mathematical Statistics 39, 209–214.