JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **24**: 446–468 (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1058

FRUGAL IV ALTERNATIVES TO IDENTIFY THE PARAMETER FOR AN ENDOGENOUS REGRESSOR

PETER EBBES,[a]* MICHEL WEDEL[b] AND ULF BÖCKENHOLT[c]

[a] *Smeal College of Business, Pennsylvania State University, University Park, PA, USA*
[b] *Robert H. Smith School of Business, University of Maryland, College Park, MD, USA*
[c] *Faculty of Management, McGill University, Montreal, Quebec, Canada*

SUMMARY

A review of the econometric literature on instrumental variables (IV) estimation shows that the performance of traditional IV estimation relies critically on the quality of the instruments. We discuss three different approaches that do not require the availability of observed instrumental variables: the 'Higher Moments' (HM) estimator, the 'Identification trough Heteroscedasticity' (IH) estimator, and the 'Latent Instrumental Variable' (LIV) approach. These methods attempt to identify the regression parameters not through observed instruments but by using other information that enables identifiability. The performance of these methods is illustrated on simulated and empirical data. Copyright © 2009 John Wiley & Sons, Ltd.

1. INTRODUCTION

The claim that regressors, such as price or other marketing mix variables, are exogenous is often not tenable, due, for example, to the omission of 'key' aspects, feedback loops, or measurement error. For instance, store managers set the marketing mix (price or advertising variables) based on market information (e.g., competition, coupon availability, word-of-mouth effects) or product characteristics unknown to the researcher. If this unobserved information affects consumer behavior, then regressors, usually price, are correlated with the error term in a typical marketing model (Villas-Boas and Winer, 1999; Chintagunta, 2001; Nevo, 2001; Petrin and Train, 2002). Although price endogeneity can be caused by a price-setting firm taking unobserved product attributes into account while interacting with competition, it can also be a result of the mechanics of consumers' optimization problems or measurement error in price (Nevo, 2000; Sudhir, 2001). These causes may enforce or offset each other to an extent that depends on the empirical context. Wansbeek and Wedel (1999) put forward that the exogeneity assumption of regressors, including price, is a shortcoming of standard market response models.

It is well understood that standard inferential methods are invalid in such cases. For instance, the ordinary least squares (OLS) estimator is biased when $E(\in |X) \neq 0$ and it loses its attractiveness as an estimator. This bias generally does not reduce when the sample size gets larger, and the OLS estimates are inconsistent. Instrumental variables (IV) methods were developed to overcome these problems, at least in large samples, and have a long history in econometrics (Bowden and Turkington, 1984; Greene, 2000; White, 2001). Instruments are variables that mimic the endogenous regressors as well as possible, but are uncorrelated with the error term. Once 'valid'

* Correspondence to: Peter Ebbes, Smeal College of Business, Pennsylvania State University, University Park, PA 16802, USA. E-mail: pebbes@psu.edu

instruments are available, the regression parameters can be consistently estimated via, for instance, two-stage least squares. The quality of the inference from the classical instrumental variables approach, however, strongly depends on the instruments used. This has been a topic of investigation in the econometric literature for many years (Bound *et al*., 1995; Stock *et al*., 2002; Hahn and Hausman, 2003). If instruments are weak, not truly exogenous, or both, the standard IV estimation and inferential procedures are inaccurate and produce results that are potentially worse than obtained when simply ignoring the endogeneity problem. Furthermore, there is 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 is the claim that these variables themselves are uncorrelated with the disturbances (cf. Greene, 2000).

The main purpose of this study is twofold. Firstly, we briefly summarize the econometric literature on IV estimation. We cover most of the recent literature and discuss the problems surrounding weak and endogenous instruments. We point to several methods that have been proposed to deal with weak instruments. Most of this work depends on the availability of exogenous instruments. Apart from the problems surrounding invalid instruments, instruments may not be available to the researcher. This review motivates the second purpose of this study. We discuss three estimation methods that do not rely on observed IV: the 'Higher Moments' (HM) approach (Erickson and Whited, 2002; Lewbel, 1997), the 'Identification through Heteroscedasticity' estimator (Rigobon, 2003; Hogan and Rigobon, 2003), and the Latent Instrumental Variables (LIV) method (Ebbes *et al*., 2005), and we illustrate these on simulated and real data, in an attempt to show the pros and cons of these methods under various conditions and in realistic settings. We then provide a discussion summary of the conditions under which 'frugal' IV alternatives are most useful for dealing with endogeneity.

The paper is organized as follows. In the next section we present a literature review of classical IV estimation. In Section 3 we present the above-mentioned three methods that do not rely on traditional instruments to solve for regressor-error dependencies. In Section 4 we compare the properties of these methods using simulated data. The simulation study is motivated from a real data application on sales and prices, where price is potentially endogenous. A second empirical illustration is given in Section 5, and in Section 6 we summarize our major findings and provide a discussion of the results found.

2. THE CLASSICAL IV APPROACH

The IV method assumes that a set of instruments is available that are uncorrelated with the model error term and explain part of the variability in the endogenous regressors. The IV model with one endogenous regressor x_i is given by

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + x_{2i}\beta_2 + \in_i \\ x_i &= \gamma_0 + x_{2i}\gamma_2 + \theta_i + \nu_i \end{aligned} \tag{1}$$

for $i = 1, \ldots, n$, where y_i is the dependent variable, x_i is the endogenous regressor, and x_{2i} is a $1 \times k_2$ vector of exogenous regressors, which are assumed to appear in the equation for x_i as well (Wooldridge, 2002). The random errors $\in_i$ and ν_i have mean zero and variances $\sigma^2_{\in}$ and σ^2_{ν}, respectively. The covariance between $\in_i$ and ν_i is unequal to zero to model the dependency between x_i and $\in_i$. Further, $\theta_i \equiv z_i\gamma_3$ represents the effect of the instruments z_i, where z_i is a

DOI: 10.1002/jae

$1 \times g_2$ vector. The IV approach assumes that at least one instrument is available to measure θ_i in the above model. In order for an instrument to be valid, it must hold that $E(\in_i | \theta_i) = 0$ and $\gamma_3 \neq 0$. This implies that the instruments cannot have a direct effect on y_i and are excluded from the y_i equation. In this case the instruments z_i are exogenous. Hence the endogenous regressor is split up into an exogenous part and an endogenous part.

The most common IV estimators for the regression parameters are the 2SLS estimator (or a method of moments estimator) and the limited information maximum likelihood (LIML) estimator. These are consistent and approximately normally distributed for large n if (i) $\text{plim}(1/n)Z' \in = 0$, and (ii) both $\text{plim}(1/n)Z'Z$ and $\text{plim}(1/n)Z'X$ exist and have full column rank, where X is the $n \times k$ matrix containing the endogenous and exogenous regressors (including a constant), and Z an $n \times g$ matrix containing the observed instruments and the exogenous regressors. Provided that the instruments are not too weak, the asymptotic properties of the 2SLS and LIML estimator are the same (Davidson and MacKinnon, 1993; Kleibergen and Zivot, 2003).

2.1. Considerations when Using Instrumental Variables

The problem in empirical applications is how to find 'valid' instruments. Often there are no clear guidelines and theory may be lacking, so that instruments are not easy to obtain. The condition $E(\in_i | \theta_i) = 0$ requires that there is no direct association between the instruments and the dependent variable, which is debatable in many empirical situations (e.g., Bound *et al.*, 1995; Card, 1999, 2001; Wooldridge, 2002). For instance, lagged prices or promotional variables are often used as instruments in marketing response models, but the validity of these variables is questionable when reference prices exist and are formed historically (cf. Bronnenberg and Mahajan, 2001). Yang *et al.* (2003) note that lagged prices may not be appropriate, for reasons such as forward buying and stockpiling. Besides, treating lagged variables as 'exogenous' creates a potential source of endogeneity itself (Arellano, 2002). Nevo (2001) used price data from other markets as instruments for price, but notes that these instruments are invalid when common (national) demand shocks occur, or when advertising or promotion activities are coordinated across markets. This is likely when the same manufacturer or retailer is active in several markets. Although cost drivers may be potential instrumental variables for price, Nevo (2000) concludes that these are rarely observed, while proxies for cost usually do not exhibit sufficient variation.

Weak instruments

Exogeneity of instruments is only one of the two criteria for an instrument to be valid and available instruments may be weak in the sense that they are poorly correlated with the endogenous regressors. The econometric literature has shown that the presence of weak instruments reduces the precision of the estimates, and may lead to biases that are potentially larger than OLS, even for large samples (Bound *et al.*, 1995). Furthermore, standard asymptotic approximations break down (Staiger and Stock, 1997; Bound *et al.*, 1995; Hahn and Hausman, 2002; Kleibergen and Zivot, 2003). As a consequence, standard hypothesis tests and confidence intervals are unreliable and sensitive to the choice and validity of the instruments. Weak instruments may arise when the instruments do not have a high degree of explanatory power for the endogenous regressors or when the number of instruments is large (cf. Hahn and Hausman, 2002, 2003). However, features that make an IV plausibly exogenous can also work to make the instrument weak (cf. Stock *et al.*, 2002).

DOI: 10.1002/jae

In the presence of weak instruments there are three potential pitfalls.[1] Firstly, in finite samples the IV estimator (2SLS) is biased in the same direction as OLS, which often goes unnoted in empirical studies. This bias arises because coefficients of the first-stage regression are not observed. Hahn and Hausman (2002, 2003) and Buse (1992), among others, show that this finite sample bias is a function of the number of instruments, which suggests that augmenting the set of instruments may even increase the bias in the estimator, in particular when adding less important or weak instruments. Sargan (1958) concluded that if the first few instrumental variables are well chosen there is usually no improvement, and possibly a deterioration, in the estimates when the number of IV is further increased. Bound *et al.* (1995) and Hahn and Hausman (2003) show that the bias is inversely related to the F-statistic of the regression of the endogenous explanatory variable on the instruments.

Secondly, in the presence of weak instruments the inconsistency of the IV estimator, relative to OLS, can be very large when the instrument is correlated with $\in$. Bound *et al.* (1995) show that this relative inconsistency of IV to OLS is equal to (for simplicity it is assumed that $k = g = 1$)

$$\frac{\text{plim }\hat{\beta}_1^{\text{IV}} - \beta_1}{\text{plim }\hat{\beta}_1^{\text{OLS}} - \beta_1} = \frac{\rho_{z,\in}/\rho_{x,\in}}{\rho_{x,z}}$$

where $\rho_{x,z}$ is the correlation between x and z, the other terms being defined similarly. When the instrument is weak, $\rho_{x,z} \to 0$, and even a small correlation between z and $\in$ can produce a large relative inconsistency in the IV estimator.

Thirdly, if the instruments are weak, then classical (first-order) asymptotic approximations are poor (e.g., Nelson and Startz, 1990). It is known that in the presence of weak instruments the classical asymptotic variance matrix is large, and the asymptotic distribution of β is dispersed and a very poor approximation to the finite sample density function (which is bimodal, fat tailed and concentrated closer to the probability limit of least squares than the true value). Hence, inferential procedures based on classical asymptotic results are unreliable in this case.

In sum, the finite sample properties and bias of IV estimators are dependent on the choice of instruments and the number of instruments.[2] The outcome of several recent studies suggests routinely reporting F-statistics (which should be at least 10 for 2SLS inference to be reliable) and R^2 measures of the first-stage regression to measure the strength of the instruments used (Stock *et al.*, 2002; Bound *et al.*, 1995). Staiger and Stock (1997) develop a data-based measure for the relative bias, where large values should alert the researcher to potential problems of correlations between the instruments and the error. Bowden and Turkington (1984) and Verbeek (2000) (among others) present tests for instrument admissibility when $g > k$ (overidentification). This test does not address the potential weakness of the instruments and Hahn and Hausman (2003) argue that this test rejects too often when weak instruments are present. Hahn and Hausman (2002) develop a test for the validity of instrumental variables that jointly addresses exogeneity and strength. Finally, Donald and Newey (2001) propose a mean-squared error criterion that can be minimized to decide among several instrumental variables and find that this method of choosing instruments may improve the finite sample performance of IV estimators.

[1] See also Stock *et al.* (2002) and Hahn and Hausman (2003) for an extensive summary.

[2] Large sample theory, however, indicates that an IV estimator with one more (valid) instrument is at least as efficient (Davidson and MacKinnon, 1993).

DOI: 10.1002/jae

2.2. IV-Based Solutions to the Weak Instrument Problem

Several studies have presented improved asymptotic approximations to finite-sample distributions and inferential procedures, which provide better alternatives in the presence of weak instruments (e.g., Staiger and Stock, 1997; Bekker, 1994). Generally it is concluded that LIML is to be preferred over the standard IV estimator, but that it may not completely solve the problem (Hahn and Hausman, 2003). Besides, several fully robust methods are developed to test hypothesis and construct confidence sets for β_1 that have approximately the correct size and coverage rates when instruments are weak (e.g., Anderson and Rubin, 1949; Kleibergen, 2002).[3] However, these methods do not provide point estimates for β_1 and may be difficult to compute. Second-order unbiased estimates, such as LIML or Nagar estimators, are often suggested as robust alternatives. These estimators have no finite sample moments, which may present a problem in empirical situations (Hahn and Hausman, 2003). Other alternatives are Jackknife Instrumental Variables (Angrist *et al*., 1999), Fuller-*k* Estimator (Fuller, 1977), or bias-adjusted 2SLS (Donald and Newey, 2001). Stock *et al*. (2002) find that these partially robust estimators provide relatively reliable alternatives to 2SLS in applications with weak instruments. However, based on Monte Carlo evidence Hahn and Hausman (2003) recommend caution using 'no moment' estimators and argue that 'instrument pessimism seems overstated for 2SLS'.

Most of the previously presented results are conditional on instrument exogeneity. Failure of the exogeneity restriction, in particular in combination with weak instruments, leads to additional complications and situations in which OLS may do better than the above suggested remedies to weak instruments. In the next section we consider three alternative methods that do not rely on observed instrumental variables. Hence, these methods may present an alternative to a situation where no observed instruments of decent quality are available to the researcher.

3. THREE FRUGAL ALTERNATIVES TO CLASSICAL IV

We discuss three methods that can potentially be used as an alternative to classical IV when no or only weak instruments are available: (1) the HM method (Erickson and Whited, 2002; Lewbel, 1997); (2) the IH estimator (Rigobon, 2003; Hogan and Rigobon, 2003); and (3) the LIV method (Ebbes *et al*., 2005). We use the structural model given in (1), in which we are primarily interested in the effect of x and x_2 on y. The methods discussed next assume that no exogenous variables are available that can measure θ_i, in contrast to the classical IV model.

3.1. Higher Moments (HM) estimators

Erickson and Whited (2002), Lewbel (1997), and Dagenais and Dagenais (1997) show that for measurement error models instruments can be constructed from available data by exploiting higher-order moments. Consistent estimation requires, among other things, that measurement and structural equation errors are independent and have moments of every order, but no further assumptions have to be made about distributional forms. Although proposed for measurement error models, this approach may also be applicable to more general regressor-error dependency models as in (1) (e.g., Ebbes *et al*., 2004). To show this, we first partial out the correctly measured

[3] See also Stock *et al*. (2002) for power comparisons for several tests under different conditions.

DOI: 10.1002/jae

exogenous regressors x_{2i}, including the constant (Erickson and Whited, 2002). The reduced form of (1) is given by

$$y_i = \beta_1\theta_i + x_{2i}\tilde{\beta}_2 + \beta_1\nu_i + \in_i \tag{2}$$

$$x_i = x_{2i}\gamma_2 + \theta_i + \nu_i \tag{3}$$

where $\tilde{\beta}_2 = \beta_2 + \beta_1\gamma_2$. It is assumed that the latent factor θ_i is independent of $\in_i$ and ν_i, but no measures need to be available for it. The population regression of x_i on x_{2i} is $\mu_x = [\mathrm{E}(x'_{2i}x_{2i})]^{-1}\mathrm{E}(x'_{2i}x_i)$. Denote the regression of θ_i on x_{2i} by $\mu_\theta = [\mathrm{E}(x'_{2i}x_{2i})]^{-1}\mathrm{E}(x'_{2i}\theta_i) = \mu_x - \gamma_2$. It is assumed that x_{2i} is exogenous and free of measurement error. Subtracting $x_{2i}\mu_x$ from both sides of (3) gives

$$\tilde{x}_i = x_i - x_{2i}\mu_x = \theta_i + x_{2i}(\gamma_2 - \mu_x) + \nu_i = \tilde{\theta}_i + \nu_i \tag{4}$$

where $\tilde{\theta}_i = \theta_i - x_{2i}\mu_\theta$. By construction, the residuals $\tilde{\theta}_i$ from the regression θ_i on x_{2i} have expectation zero.[4] Similarly, $\mu_y = [\mathrm{E}(x'_{2i}x_{2i})]^{-1}\mathrm{E}(x'_{2i}y_i) = \beta_1\mu_\theta + \tilde{\beta}_2$. Subtracting $x_{2i}\mu_y$ from both sides of (2) yields

$$\tilde{y}_i = y_i - x_{2i}\mu_y = \beta_1\theta_i + x_{2i}(\tilde{\beta}_2 - \mu_y) + \beta_1\nu_i + \in_i = \beta_1\tilde{\theta}_i + \beta_1\nu_i + \in_i \tag{5}$$

Note that for the reduced form model in (4) and (5) it holds that $\mathrm{E}(\tilde{\theta}_i) = \mathrm{E}(\in_i) = \mathrm{E}(\nu_i) = 0, \mathrm{E}(\in_i \nu_i) = \sigma_{\in\nu}$, and $\tilde{\theta}_i$ is independent of $\in_i$ and ν_i.

Following Erickson and Whited (2002) and Lewbel (1997), the following instruments can potentially be used to identify β_1: (i) $z_{1i} = \tilde{x}_i\tilde{y}_i, z_{2i} = \tilde{x}_i^2$, and $z_{3i} = \tilde{y}_i^2$. In order to be valid instruments, they need to satisfy the following three conditions: (a) $\mathrm{E}z_{ji} \in_i = 0$, (b) $\mathrm{E}z_{ji}\nu_i = 0$, and (c) $\mathrm{E}z_{ji}\tilde{x}_i \neq 0$, for $j = 1, 2, 3$. We prove in the Appendix that these conditions hold for the proposed instruments z_{ji}, $j = 1, 2, 3$.

It follows that three conditions are important to form the proposed IVs: (1) symmetry of the error distribution (e.g. normality); (2) non-symmetry of the distribution of $\tilde{\theta}_i$; and (3) the magnitude of β_1. In addition, $\tilde{\theta}_i$ has to be independent of $\in_i$ and ν_i. The performance of the method depends on the skewness of the endogenous regressor, and the method may yield weak instruments z_{ji}. In addition, higher-order moments are sensitive to outliers and robust estimates for the moments may be needed. The proposed instruments will be most useful in large cross-section datasets (Lewbel, 1997).

An estimate of β_1 can be obtained in two steps (e.g., Erickson and Whited, 2002). First, an estimate for the population means μ_y and μ_x can be obtained from $\hat{\mu}_y = \left[\sum_i x'_{2i}x_{2i}\right]^{-1}\left(\sum_i x'_{2i}y_i\right)$ and $\hat{\mu}_x = \left[\sum_i x'_{2i}x_{2i}\right]^{-1}\left(\sum_i x'_{2i}x_i\right)$, respectively. Subsequently, these results can be substituted in the expressions for $\tilde{y}_i = y_i - x_{2i}\mu_y$ and $\tilde{x}_i = x_i - x_{2i}\mu_x$, and the resulting $\hat{\tilde{y}}_i$ and $\hat{\tilde{x}}_i$ are used to form the instruments $\hat{z}_{ji}$, $j = 1, 2, 3$. A standard IV regression of $\hat{\tilde{y}}_i$ on $\hat{\tilde{x}}_i$ or a generalized method of moments (GMM) approach can then be computed to estimate β_1. An estimate for β_2 is obtained from the relations $\mu_y = \beta_1\mu_\theta + \tilde{\beta}_2$, where $\mu_\theta = \mu_x - \gamma_2$ and $\tilde{\beta}_2 = \beta_2 + \beta_1\gamma_2$, hence $\mu_y = \beta_1(\mu_x - \gamma_2) + (\beta_2 + \beta_1\gamma_2) = \beta_1\mu_x + \beta_2$. The full asymptotic covariance matrix for (β_1, β_2) can be obtained using the Delta method. Hence, the HM approach may provide an easy way to

[4] In general, a population regression v_i on w_i implies the model $v_i = w_i\mu + \xi_i$, where ξ_i has mean zero by construction. Hence, $\hat{\mu} = [\mathrm{E}(w'_iw_i)]^{-1}\mathrm{E}(w'_iv_i)$ and $\mathrm{E}(\hat{\xi}_i) = \mathrm{E}(v_i - w_i\hat{\mu}) = 0$, by using the law of iterated expectations (on w_i).

DOI: 10.1002/jae

obtain instruments from the available data in general regressor-error dependencies models, can be used together with or in the absence of traditional observed instruments, and can be extended with higher-order moments or functions of the exogenous regressors x_{2i} (Erickson and Whited, 2002).

3.2. The 'Identification through Heteroscedasticity' (IH) Estimator

Hogan and Rigobon (2003) and Rigobon (2003) also propose an identification strategy using higher-order moments based on heteroscedasticity of the structural shocks. For example, in estimating price effects in market response models, this heteroscedasticity may naturally arise because of (unobserved) differences in price-setting behavior of firms across regions or time. Hence, in contrast to classical IV, one does not need a variable that shifts the price variable and it is not necessary to observe the source of this variation.

Using model (1), the IH estimator in its simplest form can be explained as follows. The 'residuals' of the reduced form regression in (2) and (3) are given by

$$u_i^y = \beta_1\theta_i + \beta_1\nu_i + \in_i \tag{6}$$

$$u_i^x = \theta_i + \nu_i \tag{7}$$

with corresponding variance–covariance matrix

$$\Omega = \begin{bmatrix} \beta_1^2(\sigma_\theta^2 + \sigma_\nu^2) + \sigma_\in^2 + \sigma_{\in\nu} & \beta_1(\sigma_\theta^2 + \sigma_\nu^2) + \sigma_{\in\nu} \\ \beta_1(\sigma_\theta^2 + \sigma_\nu^2) + \sigma_{\in\nu} & \sigma_\theta^2 + \sigma_\nu^2 \end{bmatrix} \tag{8}$$

Hogan and Rigobon (2003) assume that $\in_i$ and ν_i are independent, but instead include an omitted variable and measurement error directly to model the dependency between x_i and the model error. Identifiability requires that the variances of these common shocks, and β_1, are constant across regions. We model this dependency by specifying a non-zero covariance between $\in_i$ and ν_i that is constant across regions. Assuming that there exists m regions or groups that have different variances for θ_i, it follows that

$$\Sigma_j = \Omega_j - \Omega = \begin{bmatrix} \beta_1^2(\sigma_{\theta,j}^2 - \sigma_\theta^2) & \\ \beta_1(\sigma_{\theta,j}^2 - \sigma_\theta^2) & (\sigma_{\theta,j}^2 - \sigma_\theta^2) \end{bmatrix} \tag{9}$$

for $j = 1, \ldots, m$. Hence, for each group j, β_1 can be solved from $\sigma_{22,j}\beta_1 - \sigma_{21,j} = 0$, where $\sigma_{vw,j}$ is the vwth element of Σ_j. Hogan and Rigobon (2003) propose to treat these equations as moment conditions, and GMM techniques can be used to estimate β_1. They suggest letting the weighting matrix be determined by the group sizes. The matrix Σ_j, $j = 1, \ldots, m$ can be estimated as the difference between the variance–covariance matrix of the reduced form residuals in group j and the complete sample. Once β_1 is obtained, an estimate for β_2 can be computed from the reduced form estimates, and sampling simulation-based methods or Jackknife approaches can be used to construct the standard errors.

The IH estimator can be extended in various ways to accommodate more general situations, such as accounting for heterogeneity or feedback loops (Rigobon, 2003; Hogan and Rigobon, 2003). In addition, the average levels of the common shocks may vary across groups, and Hogan and Rigobon (2003) suggest including the region or group dummies in the set of regressors in the reduced forms to account for that.

The IH estimator is a general method that can be used when no realistic exclusion restrictions are available to identify the IV model, but some information is available about the error distribution. The exact cause of the heteroscedasticity is not of direct concern and does not need to be observed. In order to compute the IH estimator, however, one needs to observe group indicators. Rigobon (2003) and Hogan and Rigonbon (2003) emphasize that it must have an economic foundation, but provide evidence that the model is fairly robust against changes and misspecification in the sample split variable. Standard tests can be applied to the reduced form residuals to investigate heteroscedasticity.

3.3. The Latent Instrumental Variables (LIV) Method

Ebbes *et al.* (2005) propose to approximate the unobserved θ_i by a latent discrete variable with m levels (or groups). To be more specific, $\theta_i \equiv \tilde{z}_i \pi$, where $\tilde{z}_i$ is a $1 \times m$ vector that is unobserved and discrete, and π is a vector of group means. They show that the resulting distribution of (y, x) belongs to the class of normal mixture models, under the assumption that the errors have a bivariate normal distribution. Here, θ_i can be seen analogous to a factor in a factor model, and is assumed to be independent of $\in_i$ and ν_i. They prove that all the model parameters are identifiable as long as there exist at least two groups with different means π_j. Their simulation studies show that the results are insensitive to the misspecification of the number of levels of the latent discrete variable, and that the estimation results can be used to test for endogeneity.

Here we extend the LIV method by allowing for different group variances $\sigma_{\nu,j}$. As Rigobon (2003) and Hogan and Rigobon (2003) observe, unobserved group differences may induce heteroscedasticity in the distribution of the endogenous regressors. This heteroscedasticity may yield more efficient estimates when appropriately accounted for. The reduced form of the heterogenous LIV model, conditional on group $j = 1, \ldots, m$, is given by

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 \pi_j + x_{2i}(\beta_2 + \beta_1 \gamma_2) + \in_i + \beta_1 \nu_{ij} \\ x_{ij} &= x_{2i}\gamma_2 + \pi_j + \nu_{ij} \end{aligned} \tag{10}$$

where $\text{var}(\nu_{ij}) = \sigma^2_{\nu,j}$. The covariance is assumed to be equal across the groups. The conditional mean of (y_i, x_i) is

$$\begin{aligned} \mu^y_{i|j} &= \beta_0 + \beta_1 \pi_j + x_{2i}(\beta_2 + \gamma_2 \beta_1) \\ \mu^x_{i|j} &= \pi_j + x_{2i}\gamma_2 \end{aligned} \tag{11}$$

and the conditional variance is

$$\Omega_j = \begin{bmatrix} \beta_1^2 \sigma^2_{\nu,j} + 2\beta_1 \sigma_{\in\nu} + \sigma^2_{\in} & \beta_1 \sigma^2_{\nu,j} + \sigma_{\in\nu} \\ \beta_1 \sigma^2_{\nu,j} + \sigma_{\in\nu} & \sigma^2_{\nu,j} \end{bmatrix} \tag{12}$$

$j = 1, \ldots, m$. The group sizes are denoted by λ_j, $\sum_{j=1}^{m} \lambda_j = 1$. The heterogeneous LIV model also belongs to the class of normal mixture models, assuming normality of the error terms. The likelihood can be obtained from (10) and maximized using numerical optimization routines.

It can be seen that, similar to the classical IV and the HM models, the LIV model assumes that the endogenous regressor can be separated into an exogenous part and an endogenous part.

DOI: 10.1002/jae

However, Ebbes *et al*. (2005) propose to model the exogenous part as an unobserved discrete variable and the correlation between x_i and $\in_i$ that causes the endogeneity is captured through the covariance $\sigma_{\in\nu}$. Hence the latent instrument is uncorrelated with the error component and correlated with x_i for the assumed model by construction. The parameters are identified through the likelihood. As a consequence, observed instrumental variables are not required to estimate the regression parameters, conditional upon the identifying restrictions.[5] The LIV model, however, depends on the adequacy of the underlying assumptions, i.e., the existence of discrete instruments, as well as on the distribution assumptions on the error terms in (10). Although it has been shown to be somewhat robust against departures of that assumption, the method breaks down in particular if the distribution of the unobserved instruments and x_i is normal. Further, the method works well in particular in cases where the variance of the error term $\sigma^2_{\nu,j}$ is smaller than the variance of θ_i.

3.4. On Differences and Similarities among the Three Methods

Although similar in spirit, the methods are different on the following key points. The LIV method is a likelihood-based approach, whereas both the HM and the IH methods are based on moments. The likelihood-based approach can be adapted easily to more general situations, for instance a choice-based framework, and allows for precise statements about the probability of the observed data (Yang *et al*., 2003). A likelihood approach has desirable optimality properties and can be expected to be more efficient than method-of-moments estimation. LIV results are likely to be less sensitive to outliers than methods based on higher-order moments. On the other hand, a method-of-moments-based approach does not require the specification of specific distributional forms, but requires only the correct specification of the moments. Thus, the LIV model may be more sensitive to non-normal errors than moment-based estimators, although the HM estimator for model (1) does require a symmetric error distribution.

The identifiability conditions of the HM estimator and the LIV estimator in model (1) are somewhat similar. An important condition for the HM estimator is $E(\theta_i^3) \neq 0$. Hence, the performance of this method is better if the endogenous regressors are more skewed (assuming symmetrical error distributions). Ebbes *et al*. (2005) observe that the performance of the (homogeneous) LIV model is worse when the distribution of x is more symmetric and tends to a normal distribution. In this case, the group means π_j are closer, and identification and estimation become more problematic. This occurs in general in mixture models when the mixture components are poorly separated (Redner and Walker, 1984). Large samples may be required to overcome this problem. The heterogeneous LIV model proposed in this paper, however, can be applied even when $\pi_j = 0$, $j = 1, \ldots, m$, under the condition that there exists at least two sufficiently different $\sigma_{\nu,j}$s. This condition is similar to the one used in the IH estimator.

In order to compute the IH estimator additional information has to be available to define the group indicator. As Rigobon (2003) and Hogan and Rigobon (2003) note, this should be done ideally on the basis of economic theory. Such knowledge may not always be available. The LIV and HM estimator do not require such additional information, but are as a consequence more

[5] The LIV approach has some features that are similar to the grouping method developed by Wald (Wald, 1940; Madansky, 1959; Bowden and Turkington, 1984). It is assumed that the data can be divided into two groups according to statistical criteria. A line can then be fitted which is a consistent estimate of the true line if the grouping criteria are satisfied. The LIV model does not require the grouping of the data to be observed but estimates such a grouping simultaneously with the other parameters using mixture modeling techniques.

DOI: 10.1002/jae

Table I. OLS, IV, and frugal IV estimates of the price effect in the ice cream data

Method	OLS	IH	HM	LIVhom	LIVhet
	−0.13	−0.28	−0.45	−0.21	−0.18
	(0.11)	(0.34)	(0.31)	(0.12)	(0.22)

sensitive to functional form assumptions. As for LIV and HM, the IH estimator does not rely on exclusion restrictions of observed data, and the region or group dummy variable is included in the reduced form equations. The IH estimator can be applied to simultaneous equations to estimate a full system of structural equations. Ruud (2000, pp. 491–499) shows that measurement error and simultaneity can be viewed as a special case of an omitted variables problem. The frugal IV methods are formulated as an omitted variables model to estimate a single equation and are not proposed to estimate a full system. We show in our simulation study that the frugal IV methods perform reasonably well in estimating a single equation from a larger simultaneous system.

4. A SALES-PRICE ILLUSTRATION AND A SIMULATION STUDY

4.1. Analyzing Endogenous Price

The dataset we use to illustrate the frugal IV methods is described in Hand *et al*. (1993). It contains measures on (aggregate) ice cream consumption (pints per capita) over 30 4-week periods. All variables were standardized to have mean zero and variance one before the analyses. It is thought that the variables affecting ice cream consumption may be price, weekly family income, and temperature (average). The OLS price coefficient is not significantly different from zero. For a number of reasons, one may expect that price and the model error are not independent. Firstly, the aggregation of the variables may have led to measurement error. Secondly, if managers set prices based on factors that are observed by consumers but are not included in the regression model, price is endogenous. For instance, a lower price of competitive categories such as soft drinks or beer may induce lower ice cream sales, and a manager may respond by lowering the price of ice creams, and vice versa. Hence, part of the effect of competitive category prices is falsely attributed to the effect of ice cream prices on sales and we expect that price is biased upward towards zero.

We apply the above approaches to this dataset to explore endogeneity. The HM-IVs are based on higher moment transformations of the sales and price data. Since the observed price is rather symmetric and the R^2 of the first-stage regression of the HM-IVs on price is 19%, we conclude that the HM-IVs are weakly informative. For the IH estimator we need to define a grouping variable. Although we have no strong economic theory (Rigobon, 2003) to guide us in this example, we constructed a variable as follows. We argue that the variance of price fluctuates more in periods when the temperature is lower, for instance, because a drop in temperature makes ice cream vendors compete more on price when sales slow down. We therefore define a dummy variable indicating when the temperature is above and below the average temperature in the sample. We find that for the warmer and colder period the variance of price is 0.83 and 1.22, respectively. Although a 50% increase, this difference is not statistically significant, possibly because the sample at hand is small, which may make the IH split variable somewhat weak. Finally, we estimate the two LIV methods (with homogeneous and heterogeneous error variances, respectively), with $m = 2$. The estimates of the price coefficients are presented in Table I.

The OLS estimate for price is −0.13 (0.11), and the R^2 of this regression is 0.72. The other two regressors (income and temperature) explain most of the variance. The effect of price is not significant, which may be because it is endogenous and biased towards zero, or because of the small sample. It can be seen that all frugal IV models indicate an upward bias in OLS, where the HM and IH estimators indicate the largest and the heterogeneous LIV model the lowest bias. However, as indicated above, the HM and IH instruments may be weak, which is corroborated by relative large standard deviations (0.34 and 0.31) for the estimated price coefficients of −0.28 and −0.45, respectively. The homogeneous LIV model indicates a bias which is larger than for the heterogeneous LIV model. For the homogeneous LIV model, which has one parameter less to estimate, the price coefficient has a lower estimated standard deviation. In order to compare the results with a classical IV estimator, we included lagged price as an instrument[6] in the two methods that use mean shifts information for identification (LIV and HM). The estimated price coefficients are −0.29 (0.24), −0.28 (0.12), and −0.19 (0.18) for the homogeneous LIV model, the heterogeneous LIV model, and the HM-IV model, respectively.[7] The LIV and IH methods now yield comparable estimates. When the HM-IVs are combined with the observed IV, the estimated price coefficient is reduced substantially in magnitude. The R^2 of the first-stage regression increases from 19% to 37% and this set of IVs is more informative. The classical IV estimate for price is −0.07 (0.25), and indicates that price coefficient has a downward bias. Its large standard deviation suggests that the lagged price instrument is weak, which is confirmed by its low R^2 of the first-stage regression of 22%. This shows that combining a 'classical' instrument with frugal IV methods may be beneficial, in particular when the instruments are weak.

The application reveals that frugal IV methods should also be judiciously applied. Although all three methods show that the OLS estimates are biased toward zero, they do not agree on the magnitude of the bias. Careful consideration of the model specification is required, including assumptions on the form and homoscedasticity of the error distributions, possible autocorrelation of the errors for this application (Kadiyala, 1970; Verbeek, 2000), inclusion of exogenous variables in the second-stage regression, and inclusion of additional observed IVs in the first-stage regression. In addition, possibly due to the small sample size, weakly identified symmetric latent instruments, or both, none of the estimated price coefficients was significant.

This empirical illustration motivates a simulation study to investigate the performance of these methods in a realistic setup. Firstly, we are interested in investigating the comparative performance of the three frugal IV approaches for different scenarios when at least one of the methods is not or only weakly identified. Secondly, we investigate sample size effects on the performance of the methods. We use the above ice cream parameter values and data as input in the design of the simulation study. The specific simulation design is discussed next.

4.2. Design

From Section 3 it follows that identification of the three methods is largely determined by the distribution of θ_i (for the HM and LIV approaches) and the heteroscedasticity of ν_i (for the IH approach). Hence, to obtain insight into the performance of the frugal IV alternatives, we primarily

[6] Lagged price may not be a valid instrument for reasons mentioned in Section 2, but other instruments are not available for this dataset.

[7] For the LIV models there is evidence of multiple optima of the log-likelihood equation. We used the optima that resulted from starting the algorithm from the OLS results, as recommended by Ebbes *et al*. (2005).

DOI: 10.1002/jae

consider different specifications for these components, while keeping the other parameters in model (1) fixed.

We use model (1) with $\beta_0 = 0$ and $\beta_1 = -0.28$ (the average of the four frugal IV estimates in Table I). In addition, we included income and temperature as regressors (x_{2i}), and simulated seasonality in these variables similar to that in the observed data, i.e. $\text{Inc}_i = 0.06 + 0.88\text{Inc}_{i-1} + 0.47\omega_{1i}$, and $\text{Temp}_i = 0.06 + 0.89\text{Temp}_{i-1} + 0.52\omega_{2i}$, where $\omega_{ji} \sim N(0, 1)$, $j = 1, 2$. We set $\beta_2 = (0.31, 0.86)'$, and $\gamma_2 = (-0.063, -0.304)'$, unless stated differently. In all cases we take $\in_i \sim N(0, 0.312)$, unless otherwise specified. We keep the covariance between x_i and $\in_i$ fixed to 0.14 such that the OLS results for the simulated data are, on average, equal to the estimated price coefficient of the real data (−0.13). For each of the conditions we generated 500 datasets. The LIV and IH models are estimated with two groups. The simulation conditions are presented in Table II.

We focus on eight different conditions for the distribution of price (x_i). Different distributions of x_i are obtained by specifying different values for the parameters of the x equation in model (1). In all cases, the parameters were chosen such that the mean and variances of the simulated price are approximately equal to the mean and variance of the observed price, and only the shape of its distribution differs. Each of the cases in Table II corresponds to a situation where at least one of the methods is weakly identified.

In *case 1* we specify a data generating process (DGP) for price that is similar to the observed price. The following parameter values are obtained from the estimated price (x) equation using

Table II. Overview conditions simulation study

Case	Description	Identified
1	θ_i discrete (two groups), different means, and more or less symmetric. Heteroscedastic variances	Classical IV LIVhom and LIVhet IH HM—weakly
2	θ_i discrete (two groups), different means, and symmetric. Homoscedastic variances	Classical IV LIVhom and LIVhet
3	θ_i discrete (two groups), equal means, and symmetric. Heteroscedastic variances	LIVhet IH
4	θ_i continuous and skewed	Classical IV LIVhom and LIVhet HM
5	θ_i discrete (four groups), different means, and skewed. Heteroscedastic variances (three different variances across the four groups)	Classical IV LIVhom and LIVhet HM IH—weakly
6	θ_i continuous and symmetric. $\in_i$ misspecified and strongly skewed. Heteroscedastic variances (two groups)	Classical IV IH LIVhet—weakly
7	θ_i constant. $\in_i$ misspecified, mixture of normals. Heteroscedastic variances (two groups)	IH HM—weakly LIVhet—weakly
8	θ_i as in case 1. $\in_i$ and ν_i independent normals. y_i has a positive effect on x_i	Classical IV LIVhom and LIVhet IH HM—weakly

DOI: 10.1002/jae

the real price data in (1) only. We take $\theta_i = z_i\pi$, where $z_{i1} = 1$ with probability 0.3, $z_{i2} = 1$ with probability 0.7, and zero otherwise, and $\pi = (-1.21, 0.53)'$. The variances of the two groups are 0.067 and 0.440, respectively. Hence, $x_i \sim 0.3 \times N(-1.21, 0.067) + 0.7 \times N(0.53, 0.440)$ (we here omit the effect of x_{2i} for simplicity of notation, as well as in the following). In this case, the classical IV, the LIV methods, and the IH method are identified. The HM method is weakly identified because the data distribution of price is not very strongly skewed.

Case 2 assumes a distribution for price that is symmetric with homoscedastic variances. We specify the parameters for the price equation such that the mean and variance of x_i are approximately equal to the mean and variance in the observed data. We take $\pi = (-0.797, 0.797)'$ with probability 0.5 each, and ν_i is homoscedastic and is sampled from $N(0, 0.328)$. In this case, the classical IV and the LIV methods are identified. The HM method is not identified because θ_i is symmetric, and the IH estimator is not identified because the errors are homoscedastic across the groups.

Case 3 takes $\pi = (0, 0)'$ and ν_i is heteroscedastic and sampled from $N(0, \sigma^2_{\nu,j})$ if $z_{ij} = 1$, $j = 1, 2$, with probability 0.5, where $\sigma^2_{\nu,1} = 0.482$ and $\sigma^2_{\nu,2} = 1.446$. As before, the parameters are chosen such that the mean and variance are approximately equal to the observed data; only the shape of the price distribution differs. Hence, $x_i \sim 0.5 \times N(0, \sigma^2_{\nu,1}) + 0.5 \times N(0, \sigma^2_{\nu,2})$, and the IH and heterogeneous LIV approaches are identified. The IV estimator is not identified because the instrument is weak and the HM model lacks identification because $E(\theta_i^3) = 0$.

For *case 4* we assumed that $\theta_i \sim$ Gamma(1) and $\nu_i \sim N(0, 0.328)$. Since we do not observe a grouping variable, the IH estimator is not identified, whereas the others are.

In *case 5*, $\theta_i = z_i\pi$ is discrete with four levels. Here $z_{ij} = 1$ with probability λ_j, and zero otherwise, for $j = 1, \ldots, 4$. The means $\pi = (\pi_1, \ldots, \pi_4)$ and probabilities are chosen such that x_i has a skewed distribution. The groups have different variances. To be more specific, $x_i \sim 0.375N(-0.8, 0.35) + 0.375N(0, 0.35) + 0.2N(0.85, 0.2) + 0.05N(2.275, 0.4)$. Because the models are estimated with two groups, the observed group variable for IH is misspecified and the IH method is weakly identified. This is discussed in more detail in the results section. The LIV approaches are also misspecified with respect to the true number of groups, but still identifiable. The HM estimator is identified because θ_i is skewed.

Cases 6 and *7* consider situations where $\in_i$ is misspecified. For *case 6*, we take $\in_i \sim$ Gamma(2), normalized to have mean zero and variance 0.312, $\nu_i \sim 0.5N(0, 0.244) + 0.5N(0, 1.38)$, and $\theta_i = 0.48z_i$, where $z_i \sim N(0, 1)$. These parameter choices yield simulated price and sales data that have approximately similar moments to the real data. The classical IV and the IH estimators are identified, and the heterogeneous LIV model is only weakly identified because of the severe misspecification of the true DGP. The homogeneous LIV model and the HM model are not identified, because θ_i (and x_i) has a normal distribution.

Similarly, in *case 7* we assume that $\in_i$ has a mixture distribution and we take it to be similar to x_i in *case 1*, but normalized to have mean zero and variance 0.312 and without x_{2i}. We take θ_i to be a constant and assume that ν_i has the following structure: $\nu_i \sim 0.3N(0, 0.264) + 0.7N(0, 1.264)$, with the true grouping identical to the grouping of $\in_i$. For both errors, the observations are classified similarly and this classification variable is used for the IH and classical IV estimator. For classical IV, this instrument is weak and not exogenous. The LIV models are strongly misspecified in this case. The performance of the HM method depends on the final skeweness in x_i and the symmetry of the errors. Similar to *case 6*, because the true classifications are available and the errors exhibit heteroscedasticity, we expect the IH estimator to be most robust for this case.

DOI: 10.1002/jae

Finally, for *case 8* we assume that the data are generated with a feedback loop as follows: $y_i + 0.28x_i = 0.31x_{2i} + 0.86x_{3i} + \in_i; x_i - 0.4y_i = \theta_i + \nu_i$, where we assume that $\in_i$ has a normal distribution with mean 0 and variance 0.312. We assume that $\in_i$ and ν_i are independent. The latent instrument is specified similarly as in *case 1* and the simulated data has approximately similar moments as the real data. An example of this set of equations arises in a pricing context, where y_i would denote sales, x_i prices, and x_{2i} and x_{3i} are exogenous variables, such as temperature. As in *case 1*, here all methods are identified to estimate the effect of x_i on y_i except for the HM approach, which is weakly identified because of a close-to-symmetric price distribution. Note that all methods are misspecified in that they do not accommodate the feedback loop explicitly.[8]

For each of the above eight cases we considered a situation where $n = 60$, $n = 150$, and $n = 500$ to investigate the effect of sample size on the performance of the frugal IV methods. In our simulation study the true instrument θ_i is observed without error, which allows us to investigate the performance of the three frugal IV alternatives relative to the classical IV estimator as if we were observing the true instrument. Similarly, the IH method assumes the existence of an observable grouping variable that corresponds to the heteroscedasticity, which is perfectly observed for six of the eight situations we consider.

4.3. Results

We compare the results of the three estimators presented in the previous section and the homogeneous LIV (LIVhom) model with the standard IV estimator. The latter assumes that an instrumental variable is available. In all cases OLS is biased with a mean bias of approximately 0.14 and mean square error of approximately 0.15. We first discuss the results for $n = 500$.

The results[9] for $n = 500$ are given in Table III. In *case 1* four of the five estimators are identified. It can be seen that information in both mean shift and heteroscedastic variances in price is more informative than the heteroscedasticity across the groups alone and the LIV and IV methods outperform the IH method. The HM method is weakly identified because the distribution of price is fairly symmetric and it has a bias in the direction of the OLS estimate (see also Section 2). The R^2 of the first-stage regression using the constructed instruments to compute the HM estimator is very small. The LIV methods give similar results to classical IV because the mixture components are well separated and the sample size is relatively large. The heterogeneous LIV (LIVhet) method uses both information in the means and variances and has the lowest root mean squared error (RMSE).

In *case 2*, the HM and IH estimators are not identified, which explains their poor performance. We found that the Bruesch–Pagan method to investigate whether the reduced form residuals are homoscedastic works reasonably well, and can serve as a guide to determine whether the IH estimator can be applied. As for the previous case, the R^2 of the first-stage regression is small. Both LIV models are identified and perform well and β_1 is estimated approximately unbiased.

[8] As stated before, the IH estimator can be adapted to estimate the full system (Rigobon, 2003).

[9] The LIV estimates were obtained using numerical optimization of the log-likelihood equation. The starting values used were in general the true parameter values. We eliminated the results that yield degenerate LIV estimates, as indicated by the eigenvalues of the 'outer product of the gradient' (OPG) matrix, which occurred mostly for smaller sample sizes and for cases 6 and 7, in which the LIV models are strongly misspecified. Degenerate solutions may indicate identification or starting value problems. For the moment-based estimators identification problems generally result in extreme estimates and large standard deviations.

DOI: 10.1002/jae

Table III. Simulation results for the price effect ($n = 500$). See Table II for a description of cases 1–8

Case		IV	LIVhom	LIVhet	HM	IH
1	Mean bias	0.002	0.002	0.003	0.129	0.004
	RMSE	0.028	0.030	0.026	0.304	0.061
2	Mean bias	0.000	−0.002	−0.001	0.167	0.391
	RMSE	0.026	0.035	0.036	0.340	1.362
3	Mean bias	0.043	—	0.000	0.087	0.002
	RMSE	1.613	—	0.106	0.293	0.045
4	Mean bias	0.000	0.002	0.001	0.003	—
	RMSE	0.028	0.052	0.047	0.045	—
5	Mean bias	0.001	0.002	0.003	0.010	0.196
	RMSE	0.028	0.074	0.074	0.068	1.140
6	Mean bias	0.000	—	0.210	1.491	0.000
	RMSE	0.046	—	0.652	1.550	0.046
7	Mean bias	0.104	−0.661	0.506	0.048	0.052
	RMSE	1.334	0.699	1.183	0.074	0.055
8	Mean bias	−0.001	0.000	0.000	0.140	0.003
	RMSE	0.031	0.034	0.030	0.344	0.075

Similar to *case 1*, the loss of information with respect to a situation with the true observable group dummies z_i to compute the classical IV estimator is in this case rather small.

For *case 3*, the observed instrument z_i is weak, as $\pi_1 = \pi_2 = 0$, and the classical IV estimator is not identified. Similarly, the homogeneous LIV model and the HM estimator are not identified. The numerical optimization procedure for the homogeneous LIV model converged either to solutions with a non-positive definite information matrix or to boundary solutions, and the R^2 of the first-stage regression for the HM method is very small. Because of a presence of a heteroscedastic error, both the IH and the heterogeneous LIV model are identified. It can be seen that the parameter β_1 is estimated approximately unbiasedly by the IH and the LIVhet approaches. The RMSE for the LIVhet model is about twice as large as for the IH estimator, which is caused by the fact that the LIVhet model does not use the true group split variable, which is the case for the IH estimator. If the difference between the variances of the two groups is large, and if the sample is large, the loss of information of ignoring the true group indicators is not disadvantageous to the heterogeneous LIV model.

When the true distribution of θ_i is a skewed gamma distribution (*case 4*), both the LIV and the HM estimators work very well, where the HM estimator has a slightly lower RMSE, and the regression coefficient is estimated approximately unbiasedly. The RMSE for the LIV and HM estimators are larger than for a situation where the true instrument is observed, but this difference is small because θ_i is strongly skewed. In this case, the IH estimator cannot be computed since no observable group indicator variable is available as θ_i is continuous.

In *case 5* we assume that the true instrument is discrete with four categories. The group means and sizes are specified such that θ_i is skewed (the average third-order moment of θ_i was smaller than in *case 4*). The LIV models are misspecified with respect to the true number of categories, since they are estimated with $m = 2$ groups. To compare the effect of misspecification of the split variable for the IH estimator, we constructed a split variable (with two groups) from the true split variable (with four groups) by taking groups 1 and 2 as one group, and 3 and 4 as one group. It turned out that for this grouping the IH estimator is weakly identified, as indicated by the Breusch–Pagan tests. This is also obvious from the results in Table III. In addition, the IH

DOI: 10.1002/jae

estimator does not use unobserved information present in shifts in first moments, as opposed to the LIV models. It can be seen that the LIV results are not very sensitive to the misspecification because the estimates are unbiased and the RMSE is small and not much larger than the results for the benchmark classical IV approach with the true instrument. The HM estimator also performs well, because the distribution of x is skewed. It has a somewhat smaller RMSE than the LIV models.

In *case 6* we used a skewed gamma distribution for $\in$. In addition, the instrument θ follows a normal distribution. We included heteroscedasticity. It can be seen that both the classical IV and the IH estimator perform well in this situation, because these estimators use the additional data (instrument and group indicator) and both are moment-based estimators and less sensitive to misspecified error distributions. The LIV estimator is biased upward and has a small sample distribution that is bimodal (this property of LIV was also observed by Ebbes *et al*., 2005). Numerical optimization of the homogeneous LIV model yielded solutions where the information matrix is not positive definite (as in *case 3*). The HM estimator performs worst, because $E(\theta_i^3) = 0$ and $E(\in_i^3) \neq 0$.

In *case 7* the errors $\in_i$ have a mixture distribution. The same classification is used to generate the error distribution for ν_i. The observed simulated price x_i is somewhat skewed through the dependence of $\in_i$ and ν_i. The observed true classifications that are used for classical IV are endogenous and somewhat weak. It can be seen that the classical IV estimator is biased in the same direction as OLS with large standard deviations, in concordance with the results in Section 2. The LIV models are strongly misspecified. We find that the LIVhom falsely attributes part of the error distribution to the price coefficient, the estimate of which is biased. Similar to *cases 3* and *6*, the information matrix in the maximized value of the log-likelihood equation was often not positive definite. The LIVhet results are biased as well, yielding a multimodal sampling distribution as in the previous case. Because the price variable is somewhat skewed, the HM-IVs are not very weak. However, the error distributions are not symmetric and the HM-IVs are not truly exogenous and a small upward bias remains (in particular, we find that z_{2i} correlates with the true errors). Finally, the IH estimator is most robust and has the smallest RMSE with a small upward bias. The results from *cases 6* and *7* show that the LIV and the HM approaches are more sensitive to distributional assumptions, which necessitates the investigation of the OLS, LIV, or HM residuals (Ebbes *et al*., 2005). For the previous two cases, the OLS residuals, although biased in the presence of endogeneity, clearly indicated a non-normal error distribution. The HM and, especially IH methods, being moment estimators, do relatively well under these conditions.

Finally, the LIV methods and the IH estimator work well for *case 8*. The HM estimator is only weakly identified as the variable in question (x_i) is not very skewed under this DGP. The frugal IV methods presented here are formulated in terms of an omitted variable model, and, as these simulation results suggest, can handle endogeneity cases that arise from simultaneity, under appropriate identification restrictions (as presented and discussed above). Simultaneity can be seen as a special omitted variable problem (Ruud, 2000). We note that, unlike the IH method, the LIV and HM methods in their current forms are not designed to estimate the full system; that is, the effect of y_i on x_i is not explicitly estimated here. Yet, the effects of x_i, x_{2i}, and x_{3i} on y_i are estimated approximately unbiasedly.

The results for $n = 150$ and $n = 60$ are presented in Tables IV and V, respectively. All estimators suffer from smaller sample sizes. In particular, for $n = 60$ the results of the frugal IV methods are the least stable with large RMSEs. The homogeneous LIV model tends to perform better in terms of lower RMSE than the heterogeneous LIV model for the cases where it is

DOI: 10.1002/jae

identifiable. In fact, the heterogeneous LIV model is most sensitive to small sample sizes. The exception is *case 7*, for which the LIVhet model has a smaller bias and RMSE for smaller sample sizes. Here the increased sample size works against the model as the multiple modes in the log-likelihood arising from misspecified errors become more pronounced, increasing bias and RMSE. Furthermore, the IH estimator outperforms the LIVhet model when both are identified on small samples. Hence, observing a valid split variable yields substantial improvements, in particular for smaller sample sizes. The HM estimator tends to have a small bias in the same direction as the OLS estimate. This result is possibly caused by the fact that identification proofs are large sample results, but in smaller samples the constructed instruments may not be truly exogenous. Overall, the classical IV estimator, which uses a perfect instrument, suffers the least from small sample sizes. Of the frugal IV alternatives, the LIVhom approach tends to be more stable in small sample sizes than the others. We found for this simulation study that the OLS estimates for price are generally not significant for $n = 60$ (as for the real data in Table I), but the price effect is significant when estimated with classical IV. This suggests that price endogeneity may make price insignificant.

5. AN ILLUSTRATIVE EXAMPLE: A SIMPLE MEASUREMENT ERROR MODEL

We now present a second empirical example in which a laboratory instrument is observed (Madansky, 1959). The dataset contains a random sample of 50 measures on the yield strength of artillery shells (x) and on the hardness of the shells (y). The shells were manufactured from steel at two different temperatures, which constitutes a 'natural' data-grouping criterion (25 observations in each group) that can serve as an instrumental variable. According to Madansky, this constitutes an important, useful, and valid application of the method of grouping. Hence this dataset is of particular interest for comparing the LIV methods to the classical IV methods because of the presence of a discrete laboratory instrument.

We present the results from the IV method that uses the laboratory instrument and the LIV methods in Table VI and we estimate model (1) without other explanatory variables. The IV

Table IV. Simulation results for the price effect ($n = 150$). See Table II for a description of cases 1–8

Case		IV	LIVhom	LIVhet	HM	IH
1	Mean bias	−0.002	0.002	0.024	0.124	0.018
	RMSE	0.051	0.062	0.139	0.307	0.117
2	Mean bias	0.000	0.005	0.001	0.175	0.432
	RMSE	0.051	0.068	0.082	0.338	1.241
3	Mean bias	0.089	—	0.012	0.080	−0.002
	RMSE	1.425	—	0.221	0.247	0.085
4	Mean bias	0.001	−0.002	0.003	0.014	—
	RMSE	0.054	0.099	0.092	0.082	—
5	Mean bias	0.001	−0.004	0.040	0.030	0.157
	RMSE	0.050	0.156	0.199	0.112	0.942
6	Mean bias	0.002	—	0.342	0.868	0.010
	RMSE	0.084	—	0.776	0.982	0.079
7	Mean bias	0.201	−0.662	−0.037	0.054	0.051
	RMSE	1.584	0.798	0.638	0.109	0.063
8	Mean bias	−0.001	0.001	0.001	0.130	0.012
	RMSE	0.056	0.064	0.059	0.356	0.137

DOI: 10.1002/jae

Table V. Simulation results for the price effect ($n = 60$). See Table II for a description of cases 1–8

Case		IV	LIVhom	LIVhet	HM	IH
1	Mean bias	0.000	0.013	0.095	0.148	0.051
	RMSE	0.080	0.143	0.277	0.346	0.190
2	Mean bias	0.007	0.016	0.015	0.160	0.455
	RMSE	0.084	0.154	0.149	0.343	1.107
3	Mean bias	0.155	—	0.042	0.117	−0.004
	RMSE	1.590	—	0.288	0.265	0.152
4	Mean bias	0.003	0.002	−0.023	0.034	—
	RMSE	0.085	0.190	0.181	0.135	—
5	Mean bias	0.016	0.013	0.064	0.071	0.122
	RMSE	0.087	0.228	0.288	0.196	0.766
6	Mean bias	−0.013	—	0.152	0.501	0.003
	RMSE	0.195	—	0.581	0.665	0.131
7	Mean bias	−0.044	−0.436	−0.044	0.063	0.046
	RMSE	1.604	0.588	0.434	0.143	0.074
8	Mean bias	−0.001	0.006	0.001	0.147	0.033
	RMSE	0.090	0.122	0.108	0.359	0.231

estimate with the temperature dummy as natural instrument for the effect of yield strength on hardness is $\hat{\beta}_{\mathrm{IV}} = 3.204$. The LIVhom estimate (with $m = 2$) is also equal to 3.204. Note that both methods assume homoscedasticity across the groups, while the latter estimate is obtained without using the observed instrument. The LIVhet results are close to the LIVhom results.

We compare the a posteriori instrument classes found by the LIV models to the 'natural' dummy instrument (high temperature–low temperature). The LIV classifications and the observed classification are identical; i.e., the observed instrument is recovered exactly by the LIV methods (Table VII). All posterior probabilities for the two categories of the latent instrument are either zero or one (up to three decimals). The LIV methods, which do not rely on observed instruments, give results that are identical to those of IV in this situation where a 'natural laboratory' instrument is observed.

When applying the HM estimator, we found that it is only weakly identified for this application. The distribution of $\tilde{x}$ appears symmetric and a boxplot of the observed values $\tilde{x}^3$ is symmetric

Table VI. OLS, IV, and LIV results for the Madansky measurement error data ($n = 50$)

Method	OLS	IV	LIVhom	LIVhet
$\hat{\beta}_1$	3.288	3.204	3.204	3.243
	(0.400)	(0.453)	(0.445)	(0.434)

Table VII. Predicted instrument versus observed instrument (temperature) for the LIV models

		Observed instrument		Total
		Low temp.	High temp.	
Predicted instrument LIV	Group 1	25	0	25
	Group 2	0	25	25
	Total	25	25	50

DOI: 10.1002/jae

around zero. In addition, the R^2 of the first stage regression of the yield strength on the constructed instruments is only 8.4%. The HM estimate for β_1 yields 5.658 (4.926) and is far off, as can be expected. The IH estimator relies on the variances in the high- and low-temperature groups for the yield strength being different. We illustrate the IH estimator by including the temperature effect in both equations (1), as per Hogan and Rigobon (2003), and the observed laboratory instrument is not used as a standard instrument. Now, the IH estimator is identified solely through heteroscedasticity of the reduced form across the temperature groups. The reduced form residuals are somewhat heteroscedastic, but not very strongly so. The IH estimate for β_1 is 4.551 (1.886) and the OLS estimate for β_1 in the case when the temperature effect is also included as explanatory variable for the hardness is 4.491 (1.306). In both cases, the temperature effect on y_i is not significant. The IH estimator indicates a downward bias in OLS, as can be expected from the classical measurement error model, but the standard deviation is large, indicating weak identifiability.

The estimates reported here fall within the range of estimates reported by Madansky (1959). The small size of the dataset does not allow for a precise estimation of the relation between yield strength and hardness of artillery shells for either of the methods, and the standard errors are large, as we found in the previous section.[10] Nevertheless, the equivalence between LIV and IV with the laboratory instrument is remarkable. A 'natural' instrument will rarely be available in empirical studies in economics or marketing, however.

6. DISCUSSION AND CONCLUSIONS

The classical instrumental variables method may be limited in practice if instrumental variables of reasonable quality are not available, since the inferential results are sensitive to the quality of the instruments used. We have discussed and illustrated three other methods that attempt to solve that problem: (1) the HM approach (Erickson and Whited, 2002; Lewbel, 1997); (2) the IH approach (Hogan and Rigobon, 2003; Rigobon, 2003); and (3) the LIV approach (Ebbes *et al*., 2005). These methods identify the model parameters in the presence of an endogenous regressor through statistical assumptions and functional form. We illustrated the three methods on simulated data and two empirical datasets and compared the results to the classical IV approach. Our simulation experiment was motivated from a dataset on sales and prices. Here price is potentially endogenous and the OLS estimate is biased towards zero. Although the frugal IV alternatives all indicate a positive bias in the price coefficient, they do not agree much on the magnitude of this bias. In addition, standard errors are large and price is not significant. Our simulation study was designed to provide additional insight. Under favorable model conditions, all methods work well. They are approximately unbiased and the information loss compared to a situation where the true instrument is observed is rather small. Our simulation studies demonstrate that the HM estimator is sensitive to the skewness of the distribution of the latent factor θ_i, and it only uses higher moment information in the data. However, it works quite well when the data are more skewed (see also Erickson and Whited, 2002; Lewbel, 1997).

The IH estimator performs well in most cases, but requires the availability of an observable group variable. We saw from our simulation studies that complications arise with this method when such a variable cannot be observed or is misspecified. Fortunately, these situations can be detected by investigating the reduced form residuals. Our ice cream application, and accompanying

[10] The reported standard deviations are computed using a Jackknife 'leave-one-out' method.

DOI: 10.1002/jae

simulations, suggest that the small sample size impairs identifiability of the IH estimator. Hogan and Rigobon (2003) and Rigobon (2003) applied the IH method to estimate the returns to education and to estimate the relation across emerging-markets' sovereign bonds yields using fairly large samples, respectively.

The homogeneous and heterogeneous LIV estimators do not require any grouping information, and they perform fairly well in the cases considered in the simulation study. However, as any likelihood-based method, the LIV estimators are more sensitive to distributional and functional form assumptions than the IH approach. We found that in the cases where the true error distribution was non-normal, checks of the OLS residuals indicated non-normality of the error distribution, suggesting that these are good diagnostics for the situations considered here. Ebbes (2004) and Ebbes *et al*. (2005), however, suggest that the normality assumption of the errors is not crucial in LIV models and that more general specifications of those models including non-normal errors may be developed. We find that the LIV approach identifies the discrete laboratory instrument perfectly in the Madansky data. This supports the validity of LIV in empirical applications, where a discrete instrument can be assumed to be a reasonable approximation to the DGP. The heterogeneous LIV model proposed here is more generally applicable than the homogeneous LIV model because it exploits information in the heteroscedasticity of the errors. Drawbacks of the heterogeneous LIV model are, however, that it is more sensitive to small sample sizes, and that the likelihood may become unbounded, which then complicates estimation and asymptotic analysis. However, this limitation may be resolved by restrictions on the variances (Titterington *et al*., 1985). We find that the performance of all three non-IV approaches is much better for larger sample sizes, but the heterogeneous LIV model is most sensitive to small sample sizes.

In some applications additional information, such as the nature of the data-generating process, the suspected cause of endogeneity, or fixed-effects approaches, suggest suitable instruments (e.g., Wooldridge, 2002). In addition, methods to address price endogeneity have recently appeared, including methods that jointly model demand, cost and competition (e.g., Berry *et al*., 1995; Yang *et al*., 2003; Sudhir, 2001), and spatial econometrics approaches (Bronnenberg and Mahajan, 2001; Van Dijk *et al*., 2004). Such methods can be considered as alternatives for, or in addition to, the three frugal methods investigated in this article. We believe that when applied sensibly the frugal IV approaches are powerful methods to identify regression parameters in the presence of an endogenous regressor in economics and marketing.

APPENDIX

Proof Validity z_{1i}

Condition (a) holds, because

$$\begin{aligned} \mathrm{E}z_{1i} \in_i &= \mathrm{E}(\tilde{\theta}_i + \nu_i)(\beta_1\tilde{\theta}_i + \beta_1\nu_i + \in_i) \in_i \\ &= \mathrm{E}(\beta_1\tilde{\theta}_i^2 + \beta_1\nu_i\tilde{\theta}_i + \tilde{\theta}_i \in_i + \beta_1\tilde{\theta}_i\nu_i + \beta_1\nu_i^2 + \nu_i \in_i) \in_i \\ &= \mathrm{E}(\beta_1\tilde{\theta}_i^2 \in_i + \beta_1\nu_i\tilde{\theta}_i \in_i + \tilde{\theta}_i \in_i^2 + \beta_1\tilde{\theta}_i\nu_i \in_i + \beta_1\nu_i^2 \in_i + \nu_i \in_i^2) \end{aligned} \quad (13)$$

DOI: 10.1002/jae

which is zero by using the rule of iterated expectations, when $\in_i$ and ν_i have a symmetric distribution in the origin[11] such that $E(\nu_i^3) = E(\in_i^3) = 0$. Similarly, we can check condition (b), which holds under the assumption that $E(\nu_i^3) = 0$. Finally, for condition (c) we obtain

$$\begin{aligned} Ez_{1i}\tilde{x}_i &= E(\tilde{\theta}_i + \nu_i)^2(\beta_1\tilde{\theta}_i + \beta_1\nu_i + \in_i) \\ &= E(\tilde{\theta}_i^2 + 2\nu_i\tilde{\theta}_i + \nu_i^2)(\beta_1\tilde{\theta}_i + \beta_1\nu_i + \in_i) \\ &= E(\beta_1\tilde{\theta}_i^3 + \beta_1\nu_i\tilde{\theta}_i^2 + \tilde{\theta}_i^2 \in_i + 2\beta_1\nu_i\tilde{\theta}_i^2 + 2\beta_1\nu_i^2\tilde{\theta}_i + 2\nu_i\tilde{\theta}_i \in_i + \beta_1\nu_i^2\tilde{\theta}_i + \beta_1\nu_i^3 + \nu_i^2 \in_i) \end{aligned} \tag{14}$$

which is equal to $\beta_1 E(\tilde{\theta}_i^3)$ under the assumption that the distribution of ν_i is symmetric. Hence, this instrument is valid when $\beta_1 \neq 0$ and $\tilde{\theta}_i$ is not symmetric.

Proof Validity z_{2i}

For $z_{2i} = \tilde{x}_i^2$ we find that conditions (a) and (b) hold using similar assumptions and steps as above. For condition (c) we obtain

$$\begin{aligned} Ez_{2i}\tilde{x}_i &= E(\tilde{x}_i^3) \\ &= E(\tilde{\theta}_i^2 + 2\nu_i\tilde{\theta}_i + \nu_i^2)(\tilde{\theta}_i + \nu_i) \\ &= E(\tilde{\theta}_i^3 + \nu_i\tilde{\theta}_i^2 + 2\tilde{\theta}_i^2\nu_i + 2\tilde{\theta}_i\nu_i^2 + \nu_i^2\tilde{\theta}_i + \nu_i^3) \end{aligned} \tag{15}$$

which is equal to $E(\tilde{\theta}_i^3)$ and is non-zero when $\tilde{\theta}_i$ has a non-symmetrical distribution.

Proof validity z_{3i}

Similar to z_{1i} and z_{2i}, it can be shown that $z_{3i} = \tilde{y}_i^2$ satisfies (a) and (b) when the error distribution is symmetric. The strength of the instrument is given by

$$\begin{aligned} E(z_{3i}\tilde{x}_i) &= E(\tilde{y}_i^2\tilde{x}_i) \\ &= E(\beta_1\tilde{\theta}_i + \beta_1\nu_i + \in_i)^2(\tilde{\theta}_i + \nu_i) \\ &= E(\beta_1^2\tilde{\theta}_i^2 + \beta_1^2\tilde{\theta}_i\nu_i + \beta_1\tilde{\theta}_i \in_i + \beta_1^2\nu_i^2 + \beta_1^2\tilde{\theta}_i\nu_i + \beta_1\nu_i \in_i \\ &\quad + \beta_1\tilde{\theta}_i \in_i + \beta_1\nu_i \in_i + \in_i^2)(\tilde{\theta}_i + \nu_i) \end{aligned}$$

which is equal to $\beta_1^2 E(\tilde{\theta}_i^3)$. Hence, the instrument is weak when $\beta_1 = 0$ or $E(\tilde{\theta}_i^3) = 0$.

REFERENCES

Anderson TW, Rubin H. 1949. Estimators on the parameters of a single equation in a complete set of stochastic equations. *Annals of Mathematical Statistics* **21**: 570–582.

[11] If (a) (X, Y) and $(-X, -Y)$ have the same distribution, with (b) expectation zero, variances σ_X^2 and σ_Y^2, covariance σ_{XY}, and moments up to the third order, then (i) $EX^3 = -EX^3$, hence $EX^3 = 0$ (same for Y), (ii) $EXY^2 = E(-X)(-Y)^2 = -EXY^2$. Hence, $EXY^2 = 0$. Alternatively, omitting condition (a), we can find a random variable Z independent of Y, such that $X = \frac{\sigma_{XY}}{\sigma_Y^2}Y + Z$. Then, $EXY^2 = E\left(\left[\frac{\sigma_{XY}}{\sigma_Y^2}Y + Z\right]Y^2\right) = E\left(\frac{\sigma_{XY}}{\sigma_Y^2}Y^3\right) = 0$ if Y is symmetric or $\sigma_{XY} = 0$. The authors thank Ton Steerneman and Duncan Fong for pointing this out.

DOI: 10.1002/jae

Angrist JD, Imbens GW, Krueger AB. 1999. Jacknife instrumental variables estimation. *Journal of Applied Econometrics* **14**: 57–67.
Arellano M. 2002. Sargan's instrumental variables estimation and the generalized method of moments. *Journal of Business and Economic Statistics* **20**: 450–459.
Bekker PA. 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* **62**: 657–681.
Berry S, Levinsohn J, Pakes A. 1995. Automobile prices in market equilibrium. *Econometrica* **63**: 841–890.
Bound J, Jaeger DA, Baker RM. 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 RJ, Turkington DA. 1984. *Instrumental Variables*. Cambridge University Press: New York.
Bronnenberg BJ, Mahajan V. 2001. Unobserved retailer behavior in multimarket data: joint spatial dependence in market shares and promotion variables. *Marketing Science* **20**: 284–299.
Buse A. 1992. The bias of instrumental variables estimators. *Econometrica* **60**: 173–180.
Card D. 1999. The causal effect of education on earnings. In *Handbook of Labor Economics*, Vol. 3A, Ashenfelter OC, Card D (eds). Elsevier Science: Amsterdam; 1801–1863.
Card D. 2001. Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica* **69**: 1127–1160.
Chintagunta PK. 2001. Endogeneity and heterogeneity in a probit demand model: estimation using aggregate data. *Marketing Science* **20**: 442–456.
Dagenais MG, Dagenais DL. 1997. Higher moment estimators for linear regression models with errors in variables. *Journal of Econometrics* **76**: 193–221.
Davidson R, MacKinnon JG. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: New York.
Donald SG, Newey WK. 2001. Choosing the number of instruments. *Econometrica* **69**: 1161–1191.
Ebbes P. 2004. Latent instrumental variables: a new approach to solve for endogeneity. PhD thesis, SOM Research School, University of Groningen.
Ebbes P, Böckenholt U, Wedel M. 2004. Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica* **58**: 161–178.
Ebbes P, Wedel M, Böckenholt U, Steerneman AGM. 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. *Quantitative Marketing and Economics* **3**: 365–392.
Erickson T, Whited TM. 2002. Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory* **18**: 776–799.
Fuller W. 1977. Some properties of a modification of the limited information estimator. *Econometrica* **45**: 939–953.
Greene WH. 2000. *Econometric Analysis*. Prentice-Hall: Upper Saddle River, NJ.
Hahn J, Hausman J. 2002. A new specification test for the validity of instrumental variables. *Econometrica* **70**: 163–189.
Hahn J, Hausman J. 2003. Weak instrumens: diagnosis and cures in empirical econometrics. *Recent Advances in Econometric Methodology* **93**: 118–125.
Hand DJ, Daly F, Lunn AD, McConway KJ, Ostrowski E (eds). 1993. *Handbook of Small Data Sets*. Chapman & Hall/CRC: London.
Hogan V, Rigobon R. 2003. Using unobserved supply shocks to estimate the returns to educations. Technical report, University College Dublin.
Kadiyala KR. 1970. Testing for the indepence of regression disturbances. *Econometrica* **38**: 97–117.
Kleibergen F. 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* **70**: 1781–1803.
Kleibergen F, Zivot E. 2003. Bayesian and classical approaches to instrumental variables regression. *Journal of Econometrics* **114**: 29–72.
Lewbel A. 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* **65**: 1201–1213.
Madansky A. 1959. The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association* **54**: 173–205.

Nelson CR, Startz R. 1990. Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* **58**: 967–976.

Nevo A. 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy* **9**: 513–548.

Nevo A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**: 307–342.

Petrin A, Train K. 2002. Omitted product attributes in discrete choice models. Working paper, University of Berkeley.

Redner RA, Walker HF. 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**: 195–239.

Rigobon R. 2003. Identification through heteroskedasticity. *Review of Economics and Statistics* **85**: 777–792.

Ruud PA. 2000. *An Introduction to Classical Econometric Theory*. Oxford University Press: New York.

Sargan JD. 1958. The estimation of economic relationships using instrumental variables. *Econometrica* **26**: 393–415.

Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**: 557–586.

Stock JH, Wright JH, Yogo M. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **20**: 518–529.

Sudhir K. 2001. Competitive pricing behavior in the auto market: a structural analysis. *Marketing Science* **20**: 42–60.

Titterington DM, Smith AFM, Makov UE. 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley: Chichester.

Van Dijk A, Van Heerde HJ, Leeflang PSH, Wittink DR. 2004. Similarity-based spatial methods for estimating shelf space elasticities from correlational data. *Quantitative Marketing and Economics* **2**: 257–277.

Verbeek M. 2000. *A Guide to Modern Econometrics*. Wiley: Chichester.

Villas-Boas JM, Winer RS. 1999. Endogeneity in brand choice models. *Management Science* **45**: 1324–1338.

Wald A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* **11**: 284–300.

Wansbeek T, Wedel M. 1999. Marketing and econometrics: editors' introduction. *Journal of Econometrics* **89**: 1–14.

White H. 2001. *Asymptotic Theory for Econometricians*. Academic Press: New York.

Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. Massachusetts Institute of Technology: Cambridge, MA.

Yang S, Chen Y, Allenby GM. 2003. Bayesian analysis of simultaneous demand and supply. *Quantitative Marketing and Economics* **1**: 251–275.

DOI: 10.1002/jae