

## University of Groningen

### Latent instrumental variables

Ebbes, P.

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2004

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Ebbes, P. (2004). *Latent instrumental variables: a new approach to solve for endogeneity*. s.n.

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Chapter 8

# Discussion

The primary objective of this thesis is to develop a new method, the latent instrumental variables (LIV) method, to solve and test for regressor-error dependencies in linear models. The traditional instrumental variables (IV) method is limited in its use because it requires the availability of instruments of decent quality. In many situations such instruments are not available. Besides, in applications where instruments are available, the performance of inferential procedures critically depends on the quality of such variables, and results have to be interpreted with caution. The proposed LIV method allows for consistent estimation in the presence of regressor-error dependencies and testing for such dependencies without having observed instrumental variables at hand. In this chapter we present the conclusions of our findings. Table 8.1 gives an overview of the main topics and findings of the chapters. Furthermore, we provide a discussion of the LIV model and suggest steps for further research.

### 8.1 Summary and conclusions

An important assumption in the linear regression model is independence of the regressors and the error term. In chapter 2 we presented five situations in which this assumption is questionable: (i) relevant omitted variables, (ii) measurement error, (iii) self-selection, (iv) simultaneous equation models, and (v) lagged dependent variables and autocorrelation. In many empirical applications one or more of these situations may apply and standard estimation

Ch.	Subject	Model	Main findings
2	Literature review instrumental variables (IV) method	–	<ul style="list-style-type: none"> <li>• Bias OLS in presence of <math>X\epsilon</math>-dependency</li> <li>• Possible caveats with classical IV estimation</li> </ul>
3	Simple LIV model and tests for $X\epsilon$ -dependency	Linear model, one endogenous $x$ , an <i>un</i> observed discrete instrument with $m > 1$ categories	<ul style="list-style-type: none"> <li>• Sim. studies for wide range of settings, regr. par. estimated consistently, proposed tests powerful detecting <math>X\epsilon</math>-dependency</li> <li>• Results insensitive for misspecification of <math>m</math></li> <li>• Identification proof</li> </ul>
4	Tests for instrument weakness and endogeneity, and implementation issues	Extension of model Ch.3, add exogenous regressors and observed IVs	<ul style="list-style-type: none"> <li>• Sim. studies: proposed tests powerful to detect bad quality IVs</li> <li>• LIV model robust against misspecifying likelihood</li> <li>• Diagnostics to choose <math>m</math> and examine outliers/infl. observations</li> </ul>
5	Estimating the return to education	Application of model Ch.4	<ul style="list-style-type: none"> <li>• Results for three empirical datasets</li> <li>• OLS estimate for schooling biased upward (<math>\approx 7\%</math>)</li> <li>• Tests indicate bad quality of available observed IVs</li> </ul>
6	Multilevel models and random-effects (RE) regressor-dependencies	Several random intercept models	<ul style="list-style-type: none"> <li>• Tests for RE regressor-dependencies are reviewed</li> <li>• Sim. studies: bad performance tests and estimation methods in certain situations</li> <li>• Recommendations are made to target these situations</li> </ul>
7	Hierarchical models and endogeneity	Hierarchical models with Dirichlet prior process for <i>un</i> observed instrument	<ul style="list-style-type: none"> <li>• Alleviating discreteness assumption latent instrument</li> <li>• Simulation studies promising</li> <li>• Work in progress</li> </ul>

**Table 8.1:** Overview thesis-chapters and main findings.

procedures for the linear regression model are known to give biased and inconsistent results. Important examples are, for instance, estimating the effect of marketing mix variables in sales response models and estimating the return to education on income. Studies in marketing and industrial economics (Berry, Levinsohn and Pakes, 1995, or Besanko, Gupta and Jain, 1998) find that the estimated price response parameter in choice models is biased towards zero when endogeneity of prices is ignored. Managerial decisions based on price response measures that are not corrected for endogeneity are likely to have underestimated the effect of a price change on sales or market share. Similarly, policy makers that rely on the OLS estimates for the return to education (Card, 1999) find themselves over-ambitious because the true effect of education on wages can be expected to be lower. Hence, ignoring endogeneity leads to false conclusions and erroneous decision making.

The ‘classical’ instrumental variables (IV) method can be used to estimate models where regressor-error dependencies may be present. This method assumes that an additional set of instruments is available that can be used to separate the endogenous regressors into an exogenous part and an endogenous part. If the instruments are of good quality, then the IV estimates are known to be consistent. However, the literature review given in chapter 2 points out two problems with classical IV estimation: (i) instruments need to be available, and (ii) performance of the IV method critically relies on the quality of the instruments used. Despite (ii), these variables are often chosen on basis of ad-hoc arguments or convenience, as in many empirical applications instruments are not readily available. Several studies in econometrics have proposed solutions to the problem of weak instruments (Stock, Wright and Yogo, 2002, and Hahn and Hausman, 2003). The results from these studies present a toolbox with methods and tests to improve on classical IV inference in presence of weak instruments. Most of these studies, however, do not address instrument endogeneity and are conditional on the availability of a set of instrumental variables.

For empirical problems the question how and where to find instruments is still open. The latent instrumental variables (LIV) method presented in chapter 3

addresses this issue right at the heart. We propose a new method that does *not* require the availability of observed instrumental variables. We prove that the LIV model parameters can be identified through the likelihood and we illustrate the method on synthetic data. The simulation studies show that the LIV model gives consistent results for the regression parameters and the proposed test to test for regressor-error dependencies has a reasonable power across a wide variety of settings. These results are obtained *without* having observed instrumental variables at hand. In addition, the LIV model gives identical results to classical IV estimation for a measurement error application where a laboratory dummy instrumental variable is available. Furthermore, we show that the LIV results are rather insensitive to misspecification of the true number of categories of the discrete instrument. These results are important for empirical researchers because our ‘instrument-free’ approach does not require the necessity of first finding good quality instrumental variables when regressor-error dependencies are suspected.

In chapter 4 we extend the simple LIV model by allowing for additional exogenous regressors and possible available instruments. Furthermore, we discuss several implementation issues that complete an LIV analysis. The results of the identification proof for the more general LIV model suggest two procedures to investigate the validity of instrumental variables: (i) a test for a zero effect of the instrument on the endogenous variable (i.e. whether the instrument is ‘weak’), (ii) a test for a direct effect of the instrument on the dependent variable (i.e. whether the instrument is exogenous). Our synthetic data results show that the proposed procedures have a reasonable power to detect ‘bad’ quality instruments. Furthermore, our results indicate that the LIV estimates for the regression parameters are rather insensitive to misspecification of the true distribution of the error terms. This can be expected, since the LIV model belongs to the class of mixture models, that are known to be flexible in adapting to a broad range of distributions.

The literature review in chapter 5 illustrates the difficulties in estimating the return to education on income due to potential ability bias and the lack of good

quality instrumental variables. The LIV results for three empirical datasets indicate an upward ability bias of approximately 7%. This number is close to recent results from twin studies (Card, 1999). On the contrary, the classical IV results are highly unstable, inconsistent with the traditional ability bias criticism, and suffer from large standard deviations. We investigate the quality of the available instrumental variables in the three datasets and compare them with the ‘optimal’ LIV instruments. We find in two of the three applications that the available instruments are weak and/or exogenous. In all cases the optimal LIV instruments are found to be much stronger and, hence, the LIV results are more efficient than the classical IV results. The results that we find are convergent and lend credibility to the usefulness of the LIV method in empirical settings.

Chapters 6 and 7 consider endogeneity problems in multilevel models. In many applications data has an hierarchical structure, which introduces additional error terms and possible endogeneity-relations in the model. The model we consider in chapter 6 has two levels, and endogeneity may arise at the individual-specific level (level-one) or at the group level (level-two). In this chapter we review previous literature on estimating random intercept models in presence of regressor-error dependencies. Traditional methods (fixed-effects estimation, the Hausman-Taylor approach, Mundlak’s approach) to solve for level-two dependencies are shown to be limited in their use in presence of level-one dependencies. Our results reveal that even small violations of level-one independence may lead to fallacious conclusions in applying these traditional methods. Besides, we provide evidence that the problem of weak instruments also applies to multilevel applications, in particular to multilevel methods that solve for level-one dependencies, but also to the Hausman-Taylor approach to address level-two dependencies. We argue that much work needs to be done before problems of endogeneity in multilevel models can be adequately addressed and we present a list of open problems.

In chapter 7 we address two issues. Firstly, we present a solution for two multilevel models discussed in chapter 6 that may suffer from regressor error-

dependencies: the standard (random-intercept) multilevel model and the random coefficient regression model with individual-level covariates to explain part of the heterogeneity-variance. Furthermore, we suggest how our results may improve on the standard Hausman-Taylor approach. Secondly, we propose a nonparametric Bayesian method to alleviate the discreteness assumption of the unobserved instrument. The model can be estimated using Markov Chain Monte Carlo methods. The advantage of a Bayesian approach is that it provides a general framework that can be extended easily to incorporate more general models (e.g. choice models or models with several endogenous variables). Besides, a Bayesian analysis facilitates exact small sample inference. By assuming that the unobserved instrument has a Dirichlet prior process, the unobserved distribution of the instrument can adapt to any distribution. As opposed to the LIV model, it is not necessary to specify the number of support points of the mixture distribution since the model estimates the distribution from the available data. We present several simulation studies and show that the results are promising, yet several issues are still open for future research.

## 8.2 Limitations and future research

There are several issues concerning the LIV method that we did not address in this thesis. We will discuss the following issues in more detail below:

- **Methodological (technical) issues**
  - Large sample results
  - Identification in more general settings
  - Testing for a discrete instrument
  - Relation with classical IV estimation
  
- **Substantive issues**
  - Extensions to more than one endogenous variable
  - Choice models and more general GLM
  - Self-selection problems
  - Comparison to Lewbel's approach and heterogenous LIV
  - Straightforward testing for endogeneity

– Generalizing the unobserved instrument

These issues mostly apply to the standard LIV model introduced in chapters 3 and 4. A discussion on and steps for further research for the Bayesian approach in chapter 7 was given in section 7.5.

### 8.2.1 Methodological (technical) issues

*Large sample results.* Two technical issues that we did not address in this thesis are the consistency and the asymptotic distribution, that approximates the finite sample distribution, of the LIV estimator. The simulation studies presented in this thesis indicate that the LIV estimates are consistent, but we have not yet proven this.

The LIV estimates are maximum-likelihood (ML) estimates and consistency can be examined using basic results from maximum likelihood theory (e.g. Ferguson, 1996). Redner and Walker (1984), and Titterington, Smith and Makov (1985) summarize large sample results for ML estimation in mixture models, the class to which the LIV model belongs. They find that asymptotic theory for mixtures is not always straightforward because of possible singularities in the likelihood surface. Besides, the likelihood may be unbounded. However, Titterington, Smith and Makov (1985) state that the regularity conditions for consistency and asymptotic normality are satisfied in many well known and commonly occurring cases.

It may be more interesting, however, to investigate whether the regression parameter  $\beta$ , which is not a mixing parameter in the LIV model, can be estimated consistently by maximum likelihood when the model fitted has fewer components than the actual model. In other words, can consistency be proven for  $m = 2$ , regardless of whether the true value for  $m$  is larger than two. The simulation studies in section 3.5 suggest a positive answer to this question. Besides, if one has a set of strong instruments at hand, then adding a few additional instruments does not change the asymptotic results in a classical IV framework.



Two recent articles (Cheng and Liu, 2001, and Zhu and Zhang, 2004) establish asymptotic theory for comparing nested mixture models in which case the distribution is represented by a subset in the parameter space. Their results suggest that under certain regularity conditions the ML estimator converges to an arbitrary point in this subset, and quantities of interest such as means or variances may be estimated consistently even though the distribution is not uniquely represented. These results are supported by our simulation results in section 3.5. Andrews (1999) considers asymptotic theory for extremum estimators (e.g. ML) when a parameter is on the boundary. His results are interesting because he establish conditions under which the asymptotic distribution of a subvector of the parameter is not affected by the true values of another sub-vector being on a boundary of a parameter space. For instance, he shows that for a random coefficient model, the quasi-ML estimator for the regression coefficients are asymptotically normal whether or not some of the random coefficient variances are zero. His theory appears to be very general and may be applicable to the LIV model. The conditions he establishes, however, may be difficult to verify.

The LIV model in reduced form is quite similar to measurement error models, although standard measurement error models assume zero covariance between the errors. As mentioned in section 2.4, the grouping results of Wald (1940) and Madansky (1959) are similar in thought to the grouping idea of the LIV model. Wald and Madansky assume that a grouping of the data into two groups exists, or can be constructed. Once a ‘valid’ grouping is available, a line can be drawn, because it is determined by two points. This line is estimated consistently under certain conditions (e.g. Neymann and Scott, 1951). Madansky also considers another grouping method from an ANOVA point of view, where  $k_i$  observations for  $X_i$ ,  $i = 1, \dots, k$ , are available. He shows that the within mean square error and between mean square error can be used to obtain a consistent estimate for  $\beta$  when the grouping is independent of the model error<sup>1</sup>, hence, consistency is independent of  $k$ . The LIV model does not assume prior

---

<sup>1</sup>See also his discussion on the Housner-Brennan estimate (p. 189 - p. 191).

existence of such a grouping and uses mixture methodologies to classify the sample into groups. The results that we found using synthetic data also suggest that consistency of the LIV estimate for  $\beta$  does not depend on the number of categories chosen for the discrete instrument.

Another model closely related to the LIV model is a measurement error model considered by Kiefer and Wolfowitz (1956), who prove the consistency of the ML estimator in the presence of infinitely many incidental parameters. The model considered is

$$\begin{aligned} X_{i1} &= \alpha_i + u_i \\ X_{i2} &= \theta_{01} + \theta_{20}\alpha_i + v_i, \end{aligned} \tag{8.1}$$

where  $(v_i, u_i)$  have a bivariate normal distribution with mean zero and a covariance matrix consisting of the elements  $\{d_{11}, d_{12}, d_{22}\}$ . They find that the maximum-likelihood estimates for  $(\theta_1, \theta_2)$  are strongly consistent, given that  $d_{11}$ ,  $d_{22}$ , and  $d_{11}d_{22} - d_{12}^2$  are bounded away from zero. Reiersøl (1950) proves for normally distributed errors that  $\theta_1$  and  $\theta_2$  are nonidentifiable if and only if  $X_1, X_2$  are constants or normally distributed (cf. Madansky, 1959, p. 180). Something similar was observed in chapter 7 using the nonparametric Bayesian LIV model. Furthermore, the mixture approach for measurement error models advocated by Carroll, Roeder and Wasserman (1999), and their discussion, may be applicable to our framework as well.

Although we have not proven consistency of the maximum-likelihood estimates for the LIV model introduced in chapters 3 and 4, the simulation studies presented in this thesis suggest that they are. Furthermore, the articles cited above consider similar models, and provide intuition for the simulation results found, and a possible starting point to formally prove consistency and asymptotic normality of  $\hat{\beta}_n^{\text{LIV}}$ .

*Identification in more general settings.* Identification of all LIV model parameters was proven in chapters 3 and 4 assuming a bivariate normal distribution

for the error terms  $(\epsilon, \nu)$ . Although a mixture of normals can adapt to a broad class of distributions (Kim, Menzefricke and Feinberg, 2004), it is desirable to generalize the LIV model to allow for non-normally distributed error terms. In some applications, for instance, the normality assumption may be too restrictive and a more robust or general specification (e.g.  $t$ , gamma, logistic, or Gumbel distributions) may be desirable. We found in subsection 4.5.2 that the LIV model appears to be fairly robust against misspecified errors, although in case of severe misspecification of the error distribution of the regression equation this may present a problem. In such a case, a more robust distribution for the errors may circumvent this.

*The existence of a discrete instrument.* Identification of the LIV model requires the existence of a discrete instrument with at least two categories. Subsequently, a likelihood-framework can be used to estimate the regression parameters. Two important questions that were not considered in this thesis are: (i) is it possible to test for the existence of a discrete instrument, and (ii) what happens if the category means  $\pi$  in (3.1) for  $k = 2$  are not very distinct, i.e.  $\|\pi_2 - \pi_1\|$  is small?

Recent studies (Cheng and Liu, 2003, and Zhu and Zhang, 2004) have developed tests to test for a simpler mixture model versus a full mixture model, i.e. tests of the form  $H_0 : \lambda(1-\lambda)\|\pi_1 - \pi_2\| = 0$  versus  $H_1 : \lambda(1-\lambda)\|\pi_1 - \pi_2\| \neq 0$ . These tests may be applicable to the LIV model to investigate the assumption of the existence of a discrete instrument. However, given that mixture models are often used to approximate continuous distributions, we feel that the discreteness assumption, which does not imply that  $x$  is discrete, is not limiting in most empirical applications. Besides, many classical IV studies rely on discrete instruments.

The second question is an important issue in the mixture model literature and is closely related to the information matrix and the Mahalanobis distance between mixture components. It is known that if the mixture components do not separate well, large sample sizes may be required to obtain precise maximum-

likelihood estimates (e.g. Redner and Walker, 1984, or Titterington, Smith, and Makov, 1985). Something similar was observed in subsection 3.5.3 where we found for synthetic data that using  $m > 2$  in the LIV model, increases the occurrence of degenerate solutions. This is not much of an issue in most applications since the latent category instrument is a ‘nuisance’ parameter rather than of theoretical interest. However, estimation may be problematic if the true distribution of the unobserved instrument consists of only two groups that are not well separated. In this case the model is weakly identified and this issue is related to (i). The distribution of the latent instrument is now very close to a normal distribution, or a constant. Deriving the actual information matrix may give some insights in these issues. Furthermore, increasing the sample size and EM-algorithm estimation may improve estimation results in such situations.

*Relation with classical IV estimation.* The basic LIV model does not assume the existence of observed instrumental variables, and identification is established through the likelihood. The classical IV approach assumes the existence of good quality instrumental variables and the model parameters can be identified via the first two moments or via the likelihood. Although we argued and showed in both synthetic and real data examples that the LIV model results are rather insensitive to the different choices for  $m$ , to different shapes of the distribution of  $x$ , or to a modest misspecification of the likelihood, researchers who have been using the traditional instrumental variables approach (i.e. identification via theory and observed data) may be skeptical in adopting the latent instrumental variables approach. In this research we have not explicitly pursued the relation with classical instrumental variables, because the main goal is to formulate a new method that does not require such instruments (an exception is the study in section 3.6). However, in order to introduce the LIV method to more traditional IV users, we feel that future research should emphasize the relation between LIV and classical IV. This can be done in one or more of the following four ways.

Firstly, as was shown before, the LIV estimates can be used to obtain an a posteriori clustering of the data using Bayes’ rule, which gives the ‘optimal’ LIV

instrument  $\tilde{Z}$ , a  $n \times m$  matrix. This instrument matrix can be used to compute a 2SLS estimate for the regression parameters. In a simulation study the following questions can be investigated: (1) are the 2SLS estimates for  $\beta$  using  $\tilde{Z}$  similar to the LIV estimates, (2) is the optimal LIV instrument  $\tilde{Z}$  uncorrelated with  $\epsilon$ , (3) what is the  $R^2$  of a regression of  $x$  on  $\tilde{Z}$  compared to the  $R^2$  of a regression of  $x$  on the true (discrete)  $Z$ , and (4) what is the relation between  $x$  and  $\hat{x} = \tilde{Z}\hat{\pi}_{LIV}$ . For the simulation results presented in section 3.5 we find that the 2SLS estimate, based on LIV instruments, yields approximately similar results (means and standard deviations) to the maximum likelihood (LIV) estimate of  $\beta$  (in most cases the values are exactly identical, but for the unimodel case with eight instruments there are small differences). We also examined the correlation between  $\hat{x}$  and  $\epsilon$ , and the correlation between  $\hat{x}$  and  $x$ . We found, on average, that the correlation between  $\hat{x}$  and the true errors is approximately zero, while the correlation between  $\hat{x}$  and  $x$  was found to be much larger than zero. Although these preliminary findings suggest that the LIV predicted instruments are possibly ‘optimal’, because they are not correlated with  $\epsilon$  and are of considerable strength, future research is needed to give more conclusive results.

Secondly, in empirical applications the LIV instruments  $\tilde{Z}$  can be profiled using (additional) observed data. The results in section 3.6, for instance, illustrate that the predicted LIV instrument is identical to the laboratory temperature effect. We have not yet found interpretations for the predicted instruments for ‘schooling’ in chapter 5. However, if an instrument can be given a sensible interpretation, it may inspire confidence in the results found, or even point out new theories that can be used in subsequent studies to obtain instrumental variables.

Thirdly, another empirical validation of the LIV model for schooling applications (chapter 5) can be obtained using twin or sibling data. In twin or sibling studies the schooling parameter is estimated using a fixed-effects estimator because unobserved ‘ability’ cancels out within families (see also section 5.3.3 and chapter 6). Ideally, both methods should give similar results. In addition,

the predictive validity of the estimated LIV model can be examined using the transformed ‘within-family’ data, since differences in years of schooling of twins or siblings is exogenous, because the effect of omitted ability is eliminated. However, to assess predictive validity, the schooling variable has to be measured without error, which is questionable, see recent results on twin studies (e.g. Bonjour et al., 2003, Hertz, 2003, Isacsson, 2004).

Finally, it is interesting to investigate in what situations the LIV model can be used to improve efficiency in standard IV models if ‘valid’ observed instruments are available. Since IV estimates often suffer from large standard deviations, addition of an unobserved discrete instrument may improve on efficiency. Furthermore, the more traditional IV users are now still identifying the model through a priori formed theories or reasoning. The simulation study in section 4.4 indirectly addresses this issue and we found that combining observed instruments with a latent discrete instrument may be beneficial.

### 8.2.2 Substantive issues

*Extensions to more than one endogenous variable.* Although one right-hand side endogenous variable is the most commonly occurring situation (cf. Hanh and Hausman, 2003), applications may suffer from two or more endogenous regressors. For instance, marketing managers not only set prices based on unobserved information, but also other marketing mix variables like advertising or shelf-space location (Chintagunta, Kadiyali, and Vilcassim, 2003, Manchanda, Rossi, and Chintagunta, 2004). Furthermore, in estimating the return to schooling it is common to include measures for experience and squared experience that are constructed from ‘years of schooling’, and hence also endogenous (Verbeek, 2000).

The nonparametric Bayes approach in chapter 7 is applicable to problems with more than one endogenous variable. The standard LIV model in (3.1) can be extended to (say)  $l$  endogenous variables by taking for  $x_i$  a  $(l \times 1)$ -vector and extending the variance-covariance matrix  $\Sigma$  to a  $(l + 1) \times (l + 1)$  matrix. Hence, the more general LIV model is a mixture of  $(l + 1)$ -dimensional mul-

tivariate normal distributions. The identification proof has to be modified and we suspect that a discrete instrument with at least two categories has to exist for each endogenous variable. Consequently, the resulting mixture LIV model has  $m \geq 2l$  categories. Simulation studies and theoretical results need to be obtained prior to applying the outlined approach to empirical applications.

*Choice models and more general linear models (GLM).* The models considered in this thesis are simple linear models. However, for many applications the linearity assumption is too restrictive whereas endogeneity may be present. For instance, most studies cited in subsection 2.1.4 and section 2.3 (methods that model demand, cost, and competition) are choice models. An interesting and important extension of the simple LIV model is a generalization to this class of models.

Observed choices can be modeled using a random utility framework. It is assumed that the alternative with the highest utility is chosen. Let  $y_j$  denote the (unobserved) utility derived from choosing alternative  $j = 1, \dots, m$ , and let  $c$  be the observed choice. Then  $c = j$  if  $y_j = \max_{l=1, \dots, m} y_l$ . The model for the unobserved utility is just a standard linear model. If the errors are assumed to have a normal distribution and one of the explanatory variables is endogenous, then model (3.1) can be augmented with the maximum utility framework to obtain a ‘LIV-probit’ model. Furthermore, the LIV approach can be applied to the type of problems and the linearization of choice models introduced by Berry (1994) and Berry, Levinsohn, and Pakes (1995), that has recently generated a stream of subsequent research.

However, extending endogeneity issues to general nonlinear models is not straightforward. Dubé and Chintagunta (2003) argue that “Characterizing [endogeneity] bias is not straightforward in the context of non-linear models [...] it is unclear how strong the correlation between prices and [the errors] must be to generate statistical bias. [...] It is also unclear how the endogeneity bias will manifest itself in the estimates”. Cramer (2004) considers omitted variables bias in discrete models. He observes that “Even if the omitted variable

is orthogonal to the other regressors, its effect shows up in the variance of the disturbance. Since the slope coefficients of discrete models are scaled by the standard deviation, [...] the remaining coefficients are depressed towards zero". Furthermore, he finds that the omitted variables bias may be larger because of a misspecification of the disturbances. Mullahy (1997) considers dependence of covariates and unobservables in count data models. He observes that the standard assumption of separable additivity of the unobservables from the parametric structural model does generally not hold. Hence, even certain nonlinear IV estimators (e.g. Bowden and Turkington, 1984) may not be consistent. He proposes an alternative approach based on transforming the basic model that may be more appropriate to use. Foster (1997) also notes that traditional instrumental variables estimation does not simply extend to non-linear models. He proposes a non-linear two stage least squares estimator for a logit model, but the comments made by Mullahy (1997) may still apply. See also Blundell and Powell (2001a,b) for a more detailed discussion. From this discussion it becomes clear that extending the LIV approach to general nonlinear model is of great importance, yet nontrivial because researchers do not agree on how to model endogeneity in such models.

*Self-selection problems.* As discussed before, self-selection issues arise when an individual tends to select itself in a certain state (treated vs. non-treated, internet user vs. non-user) in a non-random way. A simple self-selection model is given by  $y_i = \beta_0 + \beta_1 d_i + \epsilon_i$ , where  $d_i$  is zero or one, depending on the 'state' of individual  $i$ . This model is similar to (3.1) with a single discrete endogenous regressor. The LIV model, however, assumes that the endogenous regressor is a continuous variable. But, using similar arguments as above for choice models, the LIV approach can possibly be extended by incorporating a probit model for  $x_i$  to handle self-selection problems.

*Comparison to Lewbel's approach and heterogeneous LIV.* As mentioned in section 2.3, Lewbel's approach (Lewbel, 1997, Erickson and Whited, 2002) is in spirit similar to the LIV approach in the sense that Lewbel's approach also does not require the availability of observed instrumental variables. In-



stead, Lewbel proposes to construct instruments from the available data based on higher-order moment restrictions. Subsequently, 2SLS or GMM estimates can be computed to estimate the regression parameters. The identifying conditions for Lewbel's approach are not similar to the conditions for the LIV model (see also appendix 6C). Hence, it is interesting to compare the performance of the LIV- and Lewbel estimates for  $\beta$  under the different identifying conditions using synthetic data.

For instance, identification for the Lewbel estimator requires that the distribution of the unobserved instrument is non-symmetric. The LIV model, however, is not restricted to non-symmetric distributions, as was shown in (e.g.) section 3.5. Secondly, as opposed to the LIV model, Lewbel's approach requires  $\beta_1$  in (3.1) to be nonzero, and situations where it is close to zero are weakly identified. On the other hand, the LIV model assumes the existence of a discrete unobserved instrument. If, for instance, the true distribution of the instrument is a skewed gamma distribution, Lewbel's method can be used, whereas the LIV model is 'technically' not identified, because all observations belong to the same group ( $m = 1$ ). However, as stated before, mixture models are generally used to approximate continuous distributions, a property that also extends to the LIV model. This was illustrated for the nonparametric Bayes model in chapter 7, and the standard LIV model was estimated for a situation where the latent instrument had a skewed gamma distribution. The LIV model in chapters 3 and 4 assumes that the mixture components for  $x$  have equal variances. This assumption may be too restrictive to approximate general continuous bivariate densities of  $(y, x)$ . Hence, an interesting development is to extend the LIV model to the class of heterogeneous mixture models where the variance  $\sigma_v^2$  in (3.2) can be different for each group  $j = 1, \dots, m$ . This model may be very robust in adapting to any distribution.

We emphasize that Lewbel's method for measurement error models has not yet been extended to models with general regressor-error dependencies. The results presented in appendix 6C for a general multilevel model have, to the best of our knowledge, not appeared in the literature before.

*Straightforward testing for endogeneity without having instruments.* We proposed two tests to test for endogeneity in standard linear models without having observed instruments at hand: a Hausman test (section 3.4) and a Wald test (section 4.6). Both were shown to have a reasonable power to detect an endogenous regressor. Another asymptotical equivalent test is a Lagrange-multiplier test (e.g. Greene, 2000). The potential advantage of this test is that it operates under the restricted model, i.e. when  $\sigma_{\epsilon_v} = 0$  ( $x$  is endogenous). As such, the model parameters  $\beta$  and  $\sigma_{\epsilon}^2$  can be estimated by OLS in a standard statistical package, and estimates for the group means  $\pi$ , the group sizes  $\lambda$ , and  $\sigma_v^2$ , can be obtained using standard software for mixture models. Subsequently, the estimated values can be substituted in the gradient vector (evaluated at the restricted parameter vector), which should give a vector of zeros, at least within the range of sample variability, if the restrictions are valid.

The only complicated step is to evaluate the score vector, that is based on the first-order derivatives in appendix 3B. However, once these derivatives are programmed, this test is potentially easy to apply, because it does not require the availability of observed instrumental variables, and may serve as a standard diagnostic tool to investigate endogeneity in linear regression estimation.

*Generalizing the unobserved instrument.* Finally, an interesting empirical question is whether the exogenous part (i.e. the unobserved discrete instrument) of the endogenous regressor can be profiled and given an interpretation. We elaborated on this before and suggested to examine the posterior classifications. Alternatively, one can investigate this formally by using a concomitant mixture model (Wedel and Kamakura, 2000) in which case the prior group sizes  $\lambda$  are made dependent on individual level covariates, i.e.

$$\lambda_{j|i} = \frac{\exp(\gamma_{0j} + v_i' \gamma_j)}{\sum_{l=1}^k \exp(\gamma_{0l} + v_i' \gamma_l)}, \quad (8.2)$$

for  $j = 1, \dots, k$ . The parameter  $\gamma_j$  represents the effect of the concomitant variables  $v_i$  on the prior probabilities  $\lambda_j$ . As such, each observation has its

own prior probability  $\lambda_{ji}$  of belonging to the  $j$ -th group of the discrete instrument. This generalizes the standard LIV model where the observations have the same prior probabilities  $\lambda_j$ . An important question is to investigate under which conditions inclusion of concomitant variables yields improved results. For instance, if the  $v_i$  are observed instrumental variables, this approach may give more efficient results than classical IV estimation and simple LIV estimation.

Furthermore, it is interesting to investigate whether a generalization of the prior distribution of the latent instrument can identify patterns of endogeneity. For instance, Dubé and Chintagunta (2003) observe for the results obtained by Yang, Chen and Allenby (2003), that the pattern of endogeneity is most pronounced at the lower price levels. In other applications similar observations can possibly be made and the pattern of endogeneity may depend on certain covariates.

In summary, we believe that the LIV method is a powerful approach to address endogeneity issues, it is simple to implement, and it presents an avenue for further research and future applications that can shed light on the issues raised in this discussion.