Latent Instrumental Variables: A Critical Review

Irene Hueter[*+]

Working Paper No. 46

July 20, 2016

Abstract

This paper considers the estimation problem in linear regression when endogeneity is present, that is, when explanatory variables are correlated with the random error, and also addresses the question of a priori testing for potential endogeneity. We provide a survey of the latent instrumental variables (LIV) approach proposed by Ebbes (2004) and Ebbes et al. (2004, 2005, 2009) and examine its performance compared to the methods of ordinary least squares (OLS) and IV regression. The distinctive feature of Ebbes' approach is that no observed instruments are required. Instead 'optimal' instruments are estimated from data and allow for endogeneity testing. Importantly, this Hausman-LIV test is a simple tool that can be used to test for potential endogeneity in regression analysis and indicate when LIV regression is more appropriate and should be performed instead of OLS regression. The LIV models considered comprise the standard one where the latent variable is discrete with at least two fixed categories and two interesting extensions, multilevel models where a nonparametric Bayes algorithm completely determines the LIV's distribution from data. This paper suggests that while Ebbes' new method

[*] Columbia University, Department of Statistics, 1255 Amsterdam Avenue, New York, NY 10027, USA, email: irene.hueter@columbia.edu.

[+] This research was partially supported by INET grant #INO15-00036.

is a distinct contribution, its formulation is problematic in certain important respects. Specifically the various publications of Ebbes and collaborators employ three distinct and inequivalent statistical concepts exchangeably, treating all as one and the same. We clarify this and then discuss estimation of returns of education in income based on data from three studies that Ebbes (2004) revisited, where 'education' is potentially endogenous due to omitted 'ability.' While the OLS estimate exhibits a slight upwards bias of 7%, 8%, and 6%, respectively, relative to the LIV estimate for the three studies, IV estimation leads to an enormous bias of 93%, 40%, and -24% when there is no consensus about the direction of the bias. This provides one instance among many well known applications where IVs introduced more substantial biases to the estimated causal effects than OLS, even though IVs were pioneered to overcome the endogeneity problem. In a second example we scrutinize the results of Ferguson et al. (2015) on the estimated effect of campaign expenditures on the proportions of Democratic and Republican votes in US House and Senate elections between 1980 and 2014, where 'campaign money' is potentially endogenous in view of omitted variables such as 'a candidate's popularity.' A nonparametric Bayesian spatial LIV regression model was adopted to incorporate identified spatial autocorrelation and account for endogeneity. The relative bias of the spatial regression estimate as compared to the spatial LIV estimate ranges between -17% to 18% for the House and between -25% to 7% for the Senate.

JEL Codes: No. C01, C8, C36.

Keywords: Endogeneity, instrumental variables, latent instrumental variables, omitted variables, regression, returns to education, return to campaign money in Congressional elections.

1 Introduction

In applied statistics, social science, and medicine the instrumental variable (IV) method has been a near-miraculous statistical technique until recently as widely acclaimed in econometrics as the linear regression model $Y = X\beta + \varepsilon$ on which it is built. A crucial assumption to the vigor of the ordinary least squares (OLS) approach in estimating the regression parameters β is that the random explanatory variables X and random error component ε are uncorrelated. X is then said to be *exogenous.* Otherwise when X is correlated with ε, X is said to be *endogenous.* In the latter event the OLS estimator for β may be biased, inconsistent, and inefficient, and thus lose all features that make it preferable over other estimators. IVs were pioneered to overcome this troublesome issue in a variety of contexts, such as the effect of a change in the price of a product for estimating the elasticity of supply and demand, education or training's relation to income, annual income on consumption, medical treatment or intervention on health, and policing on crime.

Wright [37] had developed the notion of IVs to estimate the elasticity of flaxseed demand and supply in 1928, although, as Angrist and Krueger [3] point out, this econometric advance went unnoticed in the literature until the rediscovery of the method in the 1940s. But the IV method carries its own assumptions that often are challenging to meet and has its own pitfalls. In many applications, IVs notoriously introduced more substantial biases to the estimates of causal effects than OLS estimation did and incidentally did not provide the remedies that they were supposed to deliver. The standard IV model augments the regression model with an equation $X = Z\Pi + V$ that decomposes X into an exogenous variable Z and an endogenous noise variable V, where Z is uncorrelated with the random error term ε and has no direct effect on the regressor Y yet has effect Π on the explanatory variable X, and thus explains part of the variability of X. The variable Z is said to be an *instrumental variable.*

Endogeneity is known to arise in the following situations that feature: (1) omitted relevant variables, (2) measurement error in the explanatory variables, (3) self-selection, (4) simultaneity, and (5) serially correlated errors injected by lagged explanatory variables. Ebbes [9] notes that Ruud [30] showed that scenarios (2)-(5) can be regarded as special cases of (1). In estimating the causal effect of education on earnings, for example, a typical omitted variable that appears to be relevant is ability. Individuals with higher ability may be more marketable and successful at earning higher wages. Yet, at the same time, they may have acquired a higher degree of education, as measured in terms of number of years, say.

Unobserved ability will thus affect both education and earnings and introduce dependence between the explanatory variable 'education' and the model error term.

The possible sources of endogeneity indicated in (1)-(5) and their consequences are described in detail in Section 3 of this paper along with the shortcomings of the OLS method in the presence of endogeneity and the benefits and pitfalls of IV regression. Section 4 reviews the standard latent instrumental variables (LIV) model where the latent variable is discrete with at least two fixed distinct categories and examines a simulation study that Ebbes [9] conducted. We also discuss the estimation of returns to education in income based on data from three studies that Ebbes [9] revisited, where 'education' is potentially endogenous due to omitted 'ability.' In Section 5 we consider two interesting extensions, two multilevel models where a nonparametric Bayes algorithm completely determines the LIV's distribution from data and look at a simulation study that Ebbes [9] performed. Section 6 contains a second data example, where we scrutinize the results of Ferguson, Jorgensen, and Chen [17] for the estimated effect of total campaign expenditures on the proportions of the Democratic and Republican votes in the US House and Senate elections between 1980 and 2014. Here 'campaign money' is potentially endogenous in view of omitted variables such as 'a candidate's popularity.' A nonparametric Bayesian spatial LIV regression model was adopted for the data analyses in order to incorporate identified spatial autocorrelation and accommodate endogeneity.

2 How to Avoid Endogeneity: Zero Correlation, $\mathbf{E}(\varepsilon|X) = 0$, or Independence of X and ε ?

2.1 Endogeneity definition

First, however, it is necessary to look more closely at the concept of endogenity as put forward in Ebbes [9] and Ebbes et al. [12, 10, 11]. It is unsettling that three distinct and inequivalent statistical concepts closely connected with endogeneity are used totally exchangeably without differentiation and treated as one and the same concept throughout [9, 12, 10, 11]. These three concepts are (i) linear association or correlation of two random variables X and ε, (ii) independence of X and ε, and (iii) the requirement that $\mathbf{E}(\varepsilon|X) = 0$. However, these three requirements or concepts are not equivalent, as we will discuss below. While one particular requirement can imply another among the three, the converse of this step generally is false, that is, the second requirement does not imply the first requirement.

For example, $\mathbf{E}(\varepsilon|X) = 0$ implies zero correlation between X and ε, if one also assumes $\mathbf{E}(\varepsilon) = 0$, but not vice versa, that is, zero correlation does not imply $\mathbf{E}(\varepsilon|X) = 0$. Moreover, independence implies $\mathbf{E}(\varepsilon|X) = 0$, if one also assumes $\mathbf{E}(\varepsilon) = 0$, but not vice versa. The question emerges as to which of the three situations the LIV approach is supposed to address: Correlation, $\mathbf{E}(\varepsilon|X) \neq 0$, or dependence of X and ε?

Having treated each of these concepts as one single concept in [9, 12, 10, 11] makes their presentation of the LIV approach not only confusing, it also leads to multiple recurring incorrect conclusions about when biases emerge, properties of estimators are violated, and the performance of algorithms is hampered. For example, since dependence does not imply (nonzero) correlation, the condition of zero correlation may not be violated even when X and ε are dependent.

Let us be more specific on the issues at stake and briefly look at some instances in [9, 12, 10, 11] where one definition was mistaken for the other. It is stated in [9], page 1, and in slightly rewritten form in [12], 'Abstract', that

'... we propose a new method to estimate regression coefficients in linear regression models where regressor-error correlations are likely to be present. ... the Latent Instrumental Variables (LIV) method utilizes a discrete latent variable model that accounts for dependencies between regressors and the error term.'

These initial lines have set the stage for either correlations or dependencies. Further down, at the bottom of page 1 going into page 2, one continues to read about these, namely,

'We focus on a situation where the regressor is random and possibly correlated with the disturbance, in which case it is not 'exogenous' but 'endogenous.'... The standard inferential methods are invalid if regressor-error dependencies exist.'

A close look, however, reveals that while Ebbes, Wedel et al. [12] just defined 'endogeneity' as regressor-error correlations in the 'Abstract,' within the first ten lines of the 'Introduction,' right underneath the 'Abstract,' they change their mind to define 'non-endogeneity' or 'exogeneity' as independence of the explanatory variables X from the random (error) components.

'An important assumption in these models is the independence of the explanatory variables X and random (error) components ε. In this case the regressors are said to be 'exogenous' ...'

Or perhaps, dependence was simply mistaken for correlation; equivalently, independence was mistaken for zero correlation. Another paragraph further down, Ebbes et al. [12] do indeed confirm that the latter definition for endogeneity relating to dependence between regressor and error will be relied on going forward:

'Unfortunately, in many similar situations the assumption of regressors and error independence is not satisfied. In this case the regressors are often said to be 'endogenous'...'

Four years later, Ebbes et al. [11] change back. They base the definition of endogenous regressor upon correlation:

'These studies show that when regressors and errors are correlated, and regressors are said to be endogenous, traditional inferential techniques for multilevel models yield biased and inconsistent estimates of the model parameters. ... We view this extension of the LIV method as critical because the problem of regressor-error dependencies in multilevel models is even more complicated than for one-level regression models.'

Now, on page 15, [9], there is a subsection entitled *'Bias in OLS when $E(\varepsilon|X) \neq 0$,'* where a bias does not necessarily exist since $E(\varepsilon|X) \neq 0$ does not imply that $\mathbf{E}(\hat{\beta}^{\text{OLS}} - \beta) \neq 0$. Of course, it is possible that $\mathbf{E}(\hat{\beta}^{\text{OLS}} - \beta) \neq 0$ but it is not forced by the condition $E(\varepsilon|X) \neq 0$ and we may well have $\mathbf{E}(\hat{\beta}^{\text{OLS}} - \beta) = 0$. Ebbes et al. [10], 'Introduction', reach similar conclusions:

'When the assumption of independence of regressors and errors does not hold (i.e. when $E(\varepsilon|X) = 0$) it follows immediately that the OLS estimator is biased.'

Furthermore on page 35 in [9] we are falsely reassured that

'In applying the classical linear regression model, the assumption that $E(\varepsilon|X) = 0$ ... may not hold. As a consequence, the OLS estimator is biased.'

Not far below, we arrive at a statement that says that zero correlation between ε and Z is equivalent to $E(\varepsilon|Z) = 0$:

'Instruments mimic the troublesome regressors but are uncorrelated with the error term (i.e. $E(\varepsilon|Z) = 0$...).'

But, as we already remarked and we will prove below, zero correlation is not equivalent to $E(\varepsilon|Z) = 0$. Another instance where these two definitions are exchangeably referred to happens on the same page:

'This method has as drawbacks that instruments need to be available and that once they are available, they may be weak and/or correlate with the error (i.e., $E(\varepsilon|Z) = 0$ may not hold).'

On page 36, Ebbes [9] continues to focus on zero correlation between the error term and explanatory variable:

'...the 'latent instrumental variable (LIV) method'... introduces an (unobserved) discrete binary variable to decompose x into a systematic part that is uncorrelated with ε and one that is possibly correlated with ε.'

Upon a brief look back at page 35, [9], the condition $E(\varepsilon|X) = 0$ gains importance in view of the following assertion:

'... Thus, one would like to test a priori whether $E(\varepsilon|X) = 0$ holds.'

The same statement emerges in [12], page 366. Next, on page 37, [9], apparently the definitions of dependence and correlations are freely exchanged, and in fact, they were traded at once for $E(\varepsilon|X) = 0$, which was suggested as a claim to be tested:

'Furthermore, we suggest a method which is based on a Hausman-test (Hausman, 1978) to test directly for regressor-error dependencies ... This instrument-free test can be used to assess a priori the presence of regressor-error correlations...'

It will become evident in Section 4 that the exogenous instrumental variable Z is uncorrelated with the error term ε. Yet, for the purpose of answering the question as to which requirements the LIV approach is intended to assure among (i) zero correlation, (ii) $\mathbf{E}(\varepsilon|Z) = 0$, and (iii) independence of the exogenous variable Z and ε, we present a few observations next.

2.2 From dependence to correlation: A few observations

Throughout this subsection we assume that $\mathbf{E}(\varepsilon) = 0$, unless otherwise stated. This assumption is commonly imposed on the random error term in the regression equation. Note that, since $\mathbf{E}(\varepsilon) = 0$, the correlation $\rho(\varepsilon, X) = 0$ if and only if $\mathbf{E}(\varepsilon X) = 0$ by the definition of the correlation $\rho(\varepsilon, X)$ between X and ε. In other words, X and ε are correlated, that is, $\rho(\varepsilon, X) \neq 0$ if and only if $\mathbf{E}(\varepsilon X) \neq 0$.

Observation 1. If X and ε are independent random variables, then $\mathbf{E}(\varepsilon|X) = \mathbf{E}(\varepsilon) = 0$. Equivalently, if $\mathbf{E}(\varepsilon|X) \neq 0$, then X and ε are not independent, thus, dependent.

Observation 2. If $\mathbf{E}(\varepsilon|X) = 0$, then $\mathbf{E}(\varepsilon X) = \mathbf{E}[X\,\mathbf{E}(\varepsilon|X)] = \mathbf{E}[X \cdot 0] = \mathbf{E}(0) = 0$, that is, X and ε are uncorrelated. Equivalently, if X and ε are correlated, then $\mathbf{E}(\varepsilon|X) \neq 0$.

Observation 3. Zero correlation $\rho(\varepsilon, X) = 0$ does not imply that $\mathbf{E}(\varepsilon|X) = 0$. Equivalently, $\mathbf{E}(\varepsilon|X) \neq 0$ does not imply that X and ε are correlated.

This is highlighted by the following counterexample. Suppose that $\mathbf{E}(\varepsilon|X) = C$ for some real constant $C \neq 0$ and let X have zero mean. We obtain $\mathbf{E}(\varepsilon X) = \mathbf{E}[X\,\mathbf{E}(\varepsilon|X)] = \mathbf{E}(X\,C) = C\,\mathbf{E}(X) = C{\cdot}0 = 0$. Thus, X and ε have zero correlation even though $\mathbf{E}(\varepsilon|X) \neq 0$.

Observation 4. $\mathbf{E}(\varepsilon|X) = 0$ does not imply that X and ε are independent. Equivalently, if X and ε are dependent, it does not follow that $\mathbf{E}(\varepsilon|X) \neq 0$.

This may be seen in the following example. Assume that X and Y are two independent random variables that come from a normal distribution $\mathrm{N}(0, 1)$. Furthermore, assume that ε is defined, conditional on X, namely, $\varepsilon|X = Y$ if $X > 0$ and $\varepsilon|X = Y^3$ if $X \leq 0$. We see that $\mathbf{E}(\varepsilon|X) = \mathbf{E}(Y)$ when a positive value of X is observed, while $\mathbf{E}(\varepsilon|X) = \mathbf{E}(Y^3)$ when a nonpositive value of X is observed. Hence, $\mathbf{E}(\varepsilon|X) = 0$ because $\mathbf{E}(Y) = 0$ and $\mathbf{E}(Y^3) = 0$. However, the value of ε obviously depends on the value of X, thus, ε depends on X. Consequently, ε and X are not independent.

Observation 5. Zero correlation $\rho(\varepsilon, X) = 0$ does not imply that X and ε are independent. Equivalently, if X and ε are dependent, it does not follow that X and ε are correlated.

The following example sheds light on this observation. Assume that ε is distributed $\sim \mathrm{N}(0, 1)$ and let $X = \varepsilon^2$. We verify that $\mathbf{E}(\varepsilon X) = \mathbf{E}(\varepsilon^3) = 0$. Thus, X and ε are not correlated. But it is immediately clear that X and ε are dependent, since X is a function of ε.

Remark. It is noteworthy that there is a unique instance when zero correlation and independence are equivalent definitions and concepts, namely, when (X, ε) is a bivariate normal random variable. Unfortunately, this is exactly one of the situations that is excluded as possible scenario in the LIV approach proposed in [9]. This particular scenario corresponds to the case where the latent instrumental variable equals a constant with probability 1, and thus, the discrete distribution of the latent variable has only $m = 1$ category. It is mentioned in [9] that in this case the LIV model is underidentified and the parameters are nonidentifiable. Consequently, X could not come from a normal distribution. Thus, for all practical purposes, we can assume that (X, ε) does not have a bivariate normal distribution. Hence, independence of X and ε is distinctively different from zero correlation between X and ε.

2.3 OLS and the Gauss Markov theorem

While the question remains open as to whether the LIV approach is supposed to handle situations when X and ε are correlated, $\mathbf{E}(\varepsilon|X) \neq 0$, or X and ε are not independent, we are inclined to think that $\mathbf{E}(\varepsilon|X) \neq 0$ may be sufficiently relevant that it should not be ignored. This view is informed by the following well known results about OLS estimators. In the linear regression model $Y = X\beta + \varepsilon$ with non-random matrix X, where ε is a vector of uncorrelated errors with mean zero and the same finite variance σ_ε^2 (that is, the errors are homoscedastic), the OLS estimator for β given by $\hat{\beta}^{\mathrm{OLS}} = (X'X)^{-1}X'Y$ is unbiased for β, consistent, and efficient (e.g. see [26]). Here efficient means that $\hat{\beta}^{\mathrm{OLS}}$ has the smallest variance among all linear unbiased estimators, and thus is the best linear unbiased estimator (BLUE) for β. This result is known as the Gauss-Markov theorem, named after Carl Friedrich Gauss and Andrey Markov. The errors do not need to be independent and identically distributed nor do they need to be normal. However, $\hat{\beta}^{\mathrm{OLS}}$ is only best possible within this specific class of unbiased estimators. There are smaller variance estimators that are biased. Furthermore, the mean square error $S^2 = (Y - X\hat{\beta}^{\mathrm{OLS}})'(Y - X\hat{\beta}^{\mathrm{OLS}})/(n-p)$, defined as the squared error term corrected for the degrees of freedom $n - p$ (n = number of observations, p = number of model parameters) is an unbiased estimator for the error variance σ_ε^2. The generalized least squares (GLS) or Aitken estimator extends the Gauss-Markov theorem to the case when the error vector has a non-scalar covariance matrix. This estimator is also a BLUE (Aitken [1], 1934).

When X is random, an appropriate framework in econometrics, the assumptions of the Gauss-Markov theorem are stated conditional on X. It is assumed that $\mathbf{E}(\varepsilon \mid X) = 0$ and the variance-covariance matrix is stated conditional on X. The OLS estimator $\hat{\beta}^{\mathrm{OLS}}$ is unbiased, consistent, and efficient.

3 Instrumental Variables

3.1 Sources of endogeneity

Next we highlight five reasons for the pitfalls that create endogeneity. For references to the literature, we refer the reader to [9].

(1) Omitted Variables. Endogeneity rooted in *omitted relevant variables* occurs in marketing models used in research studies where the response variable is consumer behavior and the explanatory variables include price of a product but omit unobserved local market information or product characteristics such as coupon availability, local competition, and taste changes that impact both, the price and consumer behavior. In the omitted variables model with conditional expected response

$$\mathbf{E}(Y|(X,\omega)) = X\beta + \omega\gamma$$

that Judge [25] examined, where ω denotes the *latent* or unobserved variables and X the observed variables, OLS estimation provides the estimator $\hat{\beta}^{\mathrm{OLS}} = (X'X)^{-1}X'Y$ and introduces the expected bias $\mathbf{E}(\hat{\beta}^{\mathrm{OLS}} - \beta \,|(X,\omega)) = \Pi\gamma$ with $\Pi = (X'X)^{-1}X'\omega$. When X is known but without knowledge of ω, the best linear estimator of β that minimizes the mean square error is given by the conditional expectation $\mathbf{E}(Y \mid X) = X\beta + \mathbf{E}(\omega| \, X)\gamma$. However, this estimation is biased because $\mathbf{E}(Y \mid X)$ is different from $X\beta$, unless $\gamma = 0$, that is, the omitted explanatory variables are irrelevant, or $\mathbf{E}(\omega| \, X) = 0$.

(2) Measurement Error. One is confronted with *measurement error* in the explanatory variables X when the observed variables differ from those that are specified to be of interest. Examples for questionable variable representations are IQ substituted for intelligence or ability, net or gross household income for household wealth, list price for transaction price that depends on vendors and geographic regions, instances of incorrectly aggregated and pooled measures from different data sources such as GDP, price inflation, unemployment rate, and work productivity, and questionnaire items and rating scales that collect assessments of perceptions, attitudes, and beliefs for which physical measures are not available. Biases in estimating β_1 via OLS are potentially reinforced when these assessments are additionally exposed to interviewer biases and 'halo effects.' Similarly in the model to explain income, it is evident that 'years of schooling completed' neglects an individual's on-the-job training or evening and weekend studies. Both, measurement error and omitted variables were recognized as sources of position endogeneity in paid search data, where one is interested in assessing the purchase conversion performance of individual keywords in paid search advertising (see Rutz et al. [31]). That measurement error leads to correlation between the explanatory and noise variables, and thus, endogeneity is easily illustrated in the simple linear regression model on the variable of interest $\tilde{X}$ given by

$$Y = \beta_0 + \beta_1\tilde{X} + \varepsilon$$

with $\beta_1 \neq 0$, where $X = \tilde{X} + \nu$ is observed instead of $\tilde{X}$. Assume expectations $\mathbf{E}(\varepsilon) = \mathbf{E}(\nu) = 0$, variances $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2 > 0$ and $\mathrm{Var}(\nu) = \sigma_\nu^2 > 0$, and zero covariances $\mathbf{E}(\varepsilon\,\nu) = \mathbf{E}(\tilde{X}\varepsilon) = \mathbf{E}(\tilde{X}\nu) = 0$. Consequently, $\mathbf{E}(X\,\varepsilon) = 0$. Rewrite Y as a regression on the observed variable X given by

$$Y = \beta_0 + \beta_1 X + \tilde{\varepsilon}$$

with corrected error component $\tilde{\varepsilon} = \varepsilon - \beta_1\nu$. We see that $\mathbf{E}(\tilde{\varepsilon}) = 0$. If X were exogenous, by definition X and $\tilde{\varepsilon}$ would be uncorrelated, equivalently, $\mathbf{E}(X\,\tilde{\varepsilon}) = 0$. However, $\mathbf{E}(X\tilde{\varepsilon}) = \mathbf{E}[X(\varepsilon - \beta_1\nu)] = -\beta_1\,\mathbf{E}(X\nu) = -\beta_1\,\mathbf{E}[(\tilde{X} + \nu)\nu] = -\beta_1\,\mathbf{E}(\nu^2) = -\beta_1\,\sigma_\nu^2 > 0$, which is a contradiction. We conclude that X and $\tilde{\varepsilon}$ are correlated, thus, X is not exogenous.

(3) Self-Selection. It is inevitable that certain individuals are more frequent internet consumers than others or are more likely to attend to activities streamlined through the internet, click on web pages, purchase products online, and fill in online feedback or survey forms. Unobserved personal characteristics, employment type and other circumstances may play a role. Analyzing research data to estimate the causal effect on purchase power within such subpopulations that are formed by group membership defined by web usage, a treatment, or another intervention can lead to misguided and ill-informed decisions, for example, when investigating quantities of products that need to be purchased and stored for online sale. Some causal questions can be tackled with randomized experiments. Yet often in practical situations, randomized trials are difficult to implement. But even in randomized medical trials where a key interest lies in estimating the treatment effect, not everyone is treated as intended. For instance, non-compliance issues may happen because of treatment related adverse events. Lower compliance may relate to a weaker treatment response [13]. At the same time, control subjects at times benefit from experimental interventions or concomitant therapies and medications. *Self-selection* models can be regarded as special cases of omitted variables models, where the omitted variable is a binary indicator that is correlated with the model error term. How self-selection may lead to endogeneity can be seen in the simple self-selection model with the equation being expressed at the observation level for $i = 1, \dots, n$ as

$$Y_i = X_i' \beta + S_i X_i' \delta + \varepsilon_i,$$

where the dummy variable $S_i = 1$ if $i \in \mathcal{S}$ and $S_i = 0$ if $i \notin \mathcal{S}$ indicates membership of individual i to a set $\mathcal{S}$ of states that represent treatment or another intervention. When $\mathbf{E}(\varepsilon_i | S_i) \neq 0$, then S_i and ε_i possibly are correlated. The reason consists in the observation that $\mathbf{E}(\varepsilon_i S_i) = \mathbf{E}[S_i \mathbf{E}(\varepsilon_i | S_i)] = \mathbf{P}(S_i = 1)\mathbf{E}(\varepsilon_i | S_i = 1)$, which is $\neq 0$ for $\mathbf{P}(S_i = 1) > 0$, unless $\mathbf{E}(\varepsilon_i | S_i = 1) = 0$. Hence, $\mathbf{E}(\varepsilon_i | S_i = 1) \neq 0$ implies that S_i and ε_i are correlated, while $\mathbf{E}(\varepsilon_i | S_i = 1) = 0$ implies that S_i and ε_i are uncorrelated. In the event of nonzero correlation, endogeneity is present and OLS estimation fails in this regression in the sense that the estimator for β may be biased, inconsistent, and inefficient.

(4) Simultaneity. When the value of an explanatory variable such as product price is determined at the same time as that of the response variable such as demand and supply, endogeneity will emerge. In this instance, one equates two interrelated equations that both depend on the price in order to determine the equilibrium price. Consequently, the model error term once again may be correlated with the price, and thus, the explanatory variable is endogenous. Another example arises when individuals make simultaneous or intertwined choices about labor market participation and embarking on additional educational opportunities.

(5) Serially Correlated Errors. If lagged response variables show up in the explanatory variables and we effectively deal with a time series model, the exogeneity assumption is violated so that OLS estimation is no longer suitable. For example, promotional activities at time t may depend on promotional activities at time $t-1$. The frequency of catalogue mailings may be based on the chance of customers' purchasing a product, other customers' behaviors in the past, and passed sales. Consider the model

$$Y_t = X_t' \beta_1 + Y_{t-1} \beta_2 + \varepsilon_t \quad \text{and} \quad \varepsilon_t = \phi \, \varepsilon_{t-1} + \nu_t,$$

where $|\beta_2| < 1$, $\phi \neq 0$, and $|\phi| < 1$. Assume that, for all t, we have $\mathbf{E}(\varepsilon_t) = \mathbf{E}(\nu_t) = 0$, $\mathbf{E}(X_t \, \varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 > 0$, and ν_t is uncorrelated with Y_s for $s \leq t-1$. Furthermore, assume weak stationarity of $\{Y_t\}_{t>0}$ and the existence of second moments. Y_t and X_t may represent the sales and promotional activities, respectively, at time t. Observe that $\varepsilon_t \, Y_{t-1} = \phi \, \varepsilon_{t-1} \, Y_{t-1} + \nu_t \, Y_{t-1}$ and $\mathbf{E}(\varepsilon_t \, Y_{t-1}) = \phi \, \mathbf{E}(\varepsilon_{t-1} \, Y_{t-1})$ because ν_t is uncorrelated with Y_{t-1}. Since $\mathbf{E}(X_t \, \varepsilon_t) = 0$, we see that $\mathbf{E}(\varepsilon_t \, Y_t) = \beta_2 \, \mathbf{E}(\varepsilon_t \, Y_{t-1}) + \mathrm{Var}(\varepsilon_t) = \beta_2 \, \phi \, \mathbf{E}(\varepsilon_{t-1} \, Y_{t-1}) + \sigma_\varepsilon^2$. Next, thanks to the stationarity assumption, we obtain $\mathbf{E}(\varepsilon_t \, Y_t) = \mathbf{E}(\varepsilon_{t-1} \, Y_{t-1})$ and

$$\mathbf{E}(\varepsilon_{t-1} \, Y_{t-1}) = \mathbf{E}(\varepsilon_t \, Y_t) = \sigma_\varepsilon^2/(1 - \beta_2 \, \phi) \quad \text{and} \quad \mathbf{E}(\varepsilon_t \, Y_{t-1}) = \phi \, \sigma_\varepsilon^2/(1 - \beta_2 \, \phi)$$

for all t, where we again rely on $\mathbf{E}(\varepsilon_t \, Y_{t-1}) = \phi \, \mathbf{E}(\varepsilon_{t-1} \, Y_{t-1})$. Now, if ε_t and Y_{t-1} were uncorrelated, this would imply that $\mathbf{E}(\varepsilon_t \, Y_{t-1}) = 0$. However, this contradicts $\mathbf{E}(\varepsilon_t \, Y_{t-1}) = \phi \, \sigma_\varepsilon^2/(1 - \beta_2 \, \phi) \neq 0$ and $\mathbf{E}(\varepsilon_t \, Y_t) = \sigma_\varepsilon^2/(1 - \beta_2 \, \phi) \neq 0$. We conclude that ε_t and Y_{t-1} are correlated and $\mathbf{E}(\varepsilon_t \,|\, Y_{t-1}) \neq 0$. Therefore, the exogeneity assumption is violated in this regression model.

3.2 OLS method's shortcomings in the face of endogeneity

In the situations that meet with endogeneity and the condition $\mathbf{E}(\varepsilon \,|\, X) \neq 0$, as we elaborated in the previous subsection, OLS estimation often falls short. When exogeneity indeed is violated and $E(\varepsilon X) \neq 0$ and $\mathbf{E}(\varepsilon \,|\, X) \neq 0$, the OLS estimator

$$\hat{\beta}^{\mathrm{OLS}} = \beta + (X'X)^{-1} X' \varepsilon \tag{3.1}$$

ceases to be appealing, because it no longer possesses basic key features expected of good estimators for model parameters such as unbiasedness, consistency, and efficiency. First, when $\mathbf{E}(\varepsilon \mid X) \neq 0$, then $\mathbf{E}(\hat{\beta}^{\mathrm{OLS}} - \beta \mid X) = \mathbf{E}[(X'X)^{-1}X'\varepsilon) \mid X] = (X'X)^{-1}X'\,\mathbf{E}(\varepsilon \mid X) \neq 0$. Second, since $\mathbf{E}(\hat{\beta}^{\mathrm{OLS}} - \beta) = \mathbf{E}[(X'X)^{-1}X'\,\mathbf{E}(\varepsilon \mid X)] = \mathbf{E}[(X'X)^{-1}X'\varepsilon)]$, it follows that $\hat{\beta}^{\mathrm{OLS}}$ is biased, unless $\mathbf{E}[(X'X)^{-1}X'\,\mathbf{E}(\varepsilon \mid X)] = 0$. Furthermore, the mean square error $S^2 = (Y - X\hat{\beta}^{\mathrm{OLS}})'(Y - X\hat{\beta}^{\mathrm{OLS}})/(n-p)$, defined as the squared error term corrected for the degrees of freedom $n-p$ (n = number of observations, p = number of model parameters), that is routinely engaged as an unbiased estimator for the error variance σ_ε^2 may not be an unbiased estimator either, unless $\mathbf{E}[(X'X)^{-1}X'\varepsilon] = 0$. When σ_ε^2 is underestimated, the model's accuracy is overestimated. Since the standard error of β is proportional to S, underestimating σ_ε^2 also means underestimating the standard error of $\hat{\beta}^{\mathrm{OLS}}$ and overestimating the precision of $\hat{\beta}^{\mathrm{OLS}}$.

In addition, the OLS estimator is inconsistent when exogeneity is compromised, that is, $\hat{\beta}^{\mathrm{OLS}}$ does not converge to β with probability 1 as the sample size increases without bound (e.g. see [33]). This signifies that increasing the sample size will not rectify the problems of biases in the parameter and error variance estimates but the biases may be equally large or larger for a larger number of observations.

Moreover, biases in estimating β also affect the efficiency in the estimation. The OLS estimator is efficient under the assumption that the errors have finite variance and are *homoscedastic,* which means that $\mathbf{E}(\varepsilon^2 \mid X_i)$ does not depend on i for $i = 1, 2, \ldots, n$. This is assured by the Gauss-Markov theorem, which states that $\hat{\beta}^{\mathrm{OLS}}$ provides the BLUE for β in the sense that $\hat{\beta}^{\mathrm{OLS}}$ has minimal variance among all linear unbiased estimators. Since there are biased estimators with smaller variance, the requirement of unbiasedness is necessary and cannot be relaxed. Let us add that, when $\mathbf{E}(\varepsilon \mid X) = 0$ but the errors are correlated or *heteroscedastic,* thus, the conditional covariance matrix of the errors is expressed by $\mathrm{Var}(\varepsilon \mid X) = \Omega$, GLS estimation (Aitken [1]) produces an estimator that is unbiased, consistent, efficient, and asymptotically normal. The GLS estimator is given by

$$\hat{\beta}^{\mathrm{GLS}} = (X'\,\Omega^{-1}\,X)^{-1}\,X'\,\Omega^{-1}\,Y. \tag{3.2}$$

These properties hold because GLS estimation can be regarded as OLS estimation after a transformation has been applied to the variables X that standardizes the errors and removes the correlation between them, thus, created a perfect setting for the OLS method to perform.

3.3 IV regression model

In instances of endogeneity, the IV model expressed by the two simultaneous equations $Y = X\beta + \varepsilon$ and $X = Z\,\Pi + V$, where the observable instruments $Z \in \mathbf{R}^{n\times q}$ are uncorrelated with ε but explain a portion of the variability of the endogenous variables $X \in \mathbf{R}^{n\times k}$, was spearheaded to overcome potential problems that the OLS technique can create. Here Z is exogenous and also comprises the exogenous variables in X, whereas V is endogenous and accounts for the correlation with ε. It is assumed that $q \geq k$ and $\operatorname{rank}(Z) = q < n$ in order to assure identifiability of the model parameters. Since the IV model turns out to be a special case of a simultaneous equation model (SEM), estimation techniques in SEMs can be adopted to estimate β. The 2-stage least squares (2SLS) estimator and the limited information maximum likelihood (LIML) estimator are among the most widely used. 2SLS is available in many software packages. We remark that, while the instrument Z is observed, Π and $Z\,\Pi$ are not observable and an estimate for Π is arrived at by regressing X on Z. The IV estimator is the GLS estimator given by

$$\hat{\beta}^{\mathrm{IV}} = (X'\,P_Z\,X)^{-1} X'\,P_Z\,Y \tag{3.3}$$

with $P_Z = Z(Z'Z)^{-1}Z'$, which is the usual projection matrix in regressing X on Z to obtain an estimate for Π and is the projection operator onto the space spanned by the columns of Z. Hence, we gather from (3.2) that P_Z plays the role of $\Omega^{-1} = [\mathrm{Var}(\varepsilon \mid X)]^{-1}$ in the usual GLS estimation of β, which minimizes the squared Mahalanobis length $(Y - X\beta)'\,\Omega^{-1}\,(Y - X\beta)$ of the residual vector.

An important point to take note of is that, while 2SLS and other IV estimators are consistent, generally $\mathbf{E}(\hat{\beta}^{\mathrm{IV}} - \beta) = \mathbf{E}[(X'P_ZX)^{-1}X'\,P_Z\,\varepsilon] \neq 0$ and $\hat{\beta}^{\mathrm{IV}}$ is not unbiased. Unbiasedness of an estimator $\hat{\beta}$ means that $\mathbf{E}(\hat{\beta}) = \beta$ for any fixed sample size, in other words, in average the unknown parameter β is estimated precisely for any sample size. However, the 2SLS IV estimator $\hat{\beta}^{\mathrm{IV}}$ has an expected bias of $\mathbf{E}(\hat{\beta}^{\mathrm{IV}} - \beta) = \mathbf{E}[(X'P_ZX)^{-1}X'\,P_Z\,\varepsilon] \neq 0$ for any finite samples. This is true for exogenous IVs. For invalid IVs, the bias can be substantially greater, as we shall discuss below. This IV estimator's bias under exogenous Z is in the same direction as the bias of the OLS estimator $\hat{\beta}^{\mathrm{OLS}}$ in the presence of endogeneity, which possesses expected bias $\mathbf{E}(\hat{\beta}^{\mathrm{OLS}} - \beta) = \mathbf{E}[(X'X)^{-1}X'\varepsilon] \neq 0$.

Consistency means that, with increasing sample size, the error in the estimation of β is as small as possible with probability 1. Thus, while for small or medium sample sizes, the IV estimator is expected to estimate β with a bias or error, the bias will decrease with growing sample size. Therefore, researchers who adopt the IV approach should rely on large samples and keep in mind that standard errors that software packages routinely post are only approximate standard errors.

3.4 Benefits and pitfalls of IV regression

From an application point of view, the IV approach essentially is nearly as simple as an ordinary linear regression analysis, even though a few of the latter's favorable estimator properties are lost. But the IV method seems to present the advantage of avoiding the troubling shortcomings of OLS estimation when that method meets with endogeneity. It thus is not surprising that, during the last eight to nine decades, the IV method has become a flowering and celebrated one-size-fits-all method in economics and empirical applications for modeling linear associations and causal relations in the presence of endogeneity. Yet it is revealing to reflect about the questions of where and how to find suitable and relevant IVs in practical situations and how to confirm their suitability and relevance post hoc. While additional variables that can be incorporated as IVs in the model indeed may be uncorrelated with the model error term or explain a substantial amount of the variability, the real difficulty is to ensure both requirements are satisfied simultaneously. Should these requirements not be valid at the same time, one needs to be aware of two other avenues that wreak havoc with the IV regression's success, which have frequently been witnessed in empirical studies. On the one hand, instrumental variables that are strongly correlated with the endogenous variables may also be correlated with the model error term. On the other hand, IVs that are uncorrelated with the model error component may only be weakly correlated with the endogenous variables and be poor IVs. The occurrence of such flawed IV candidates is well reviewed and documented in the literature [3, 9]. Finding relevant IVs that are exogenous reportedly is challenging [3]. However one does not have to go as far to encounter another issue and possibly walk into another dead end. It is not always obvious at the outset that IVs are necessary and IV estimation should be favored over OLS estimation, especially, since IV estimates are biased while OLS estimates are unbiased. Suspected endogeneity may not be present after all. Endogeneity can be statistically tested for in the framework of IVs. Yet, the pitfall of this option is a circular problem, because the endogeneity test relies on IVs. Thus, strikingly, IVs need to be available a priori to test whether they are indeed required. Since IVs often are hard to find, evidently, the burdensome line of pursuit to first address the challenge to identify IVs that are relevant and exogenous and hence fulfill the two key conditions and then possibly conclude in a second step that no IVs are needed because of absent endogeneity in the explanatory variables seems to be more trouble than most situations warrant. It is difficult to imagine that one would embark on such efforts. On the other hand if the IVs are weak, it should be obvious that these may be misleading and misinform conclusions. Let us elaborate on these hidden traps that are (1) endogenous instruments, (2) weak instruments, and (3) endogeneity testing based on flawed IVs.

Trap 1: Non-Exogenous Instruments. One of the two requirements of a valid instrument is exogeneity. It is apparent that exogenous instruments need to avoid all sources of endogeneity which are (1) omitted relevant variables, (2) measurement error in the explanatory variables, (3) self-selection, (4) simultaneity, and (5) serially correlated errors caused by lagged explanatory variables. A typical example for issue (5) would be when lagged prices or promotional variables are used as instruments in marketing response models. Employing IVs for price from other markets is problematic when promotional or advertising activities are synchronized across markets or price shocks synchronize globally, which exemplifies the issue in (4). In each of situations (1)-(5), the primary problem arises from an instrument that is correlated with omitted relevant variables, thus, with the error term, which is equivalent to issue (1) above. As Angrist and Krueger [3] put it, 'seemingly appropriate instruments can turn out to be correlated with omitted variables on closer examination...' They illuminate this point in an example of weather in Brazil that may well shift the supply curve for coffee, and thus, be a plausible instrument to estimate the effect of price on demand. In addition, the Brazilian weather might possibly shift the demand curve for coffee in the exchange, where coffee futures are traded, if traders use weather data to adjust holdings in anticipation of price increases that may not even materialize. Angrist and Krueger [3] go on to remark that 'especially worrisome is the possibility that an association between the IVs and omitted variables can lead to a bias in the resulting estimates that is much greater than the bias in OLS estimates....' In order to illustrate this phenomenon and its magnitude, in Section 4 below, we look to some concrete numbers in three data set examples examined in [9]. High biases of the IV estimates compared to the OLS estimates were confirmed [9] by applying the LIV method that he proposed that should be adopted instead, which we will describe in detail in Section 4, as well. For return to schooling, he reported that the IV estimate is about 80% higher than the OLS estimate in the NLSY data, based on an indicator variable for proximity, about 30% higher for the Brabant data based on instruments that are education levels of respondents' parents, and about 30% higher for the PSID data.

Yet, other troubling instances commonly arise where the IVs only affect a subgroup of the study sample or target population. Thus, the IV approach provides an estimate for the causal effect of the instrument only for the particular subgroup of individuals whose behavior is changed by the instrument. For example, the quarter-of-birth instrument changed the level of schooling of some individuals (see Angrist and Krueger [2] for a discussion). Alternatively, IV regression estimates the effect of the instrument for those individuals whose behavior would be changed by the instrument if it were assigned to them entirely at random in a trial (Imbens and Angrist [24]). When an IV merely partially explains the variation in an explanatory variable due to this sort of heterogeneity, it should be obvious that the interpretation of the IV estimates is problematic at best. Consider as an example the subgroup of patients randomized to active treatment in a clinical trial. An IV that has an indirect effect on the efficacy for patients in the active group is non-compliance to treatment. Typically, the same IV's effect is de-emphasized in the control group. Here the treatment effect is supposed to be estimated of the subjects who will take the study treatment that was assigned (see [13]). In trials with substantial noncompliance, an IV estimate sometimes is obtained, as well, in order to consistently estimate the average treatment effect in those subjects who comply with the treatment assignment. Yet the IV estimate is not always consistent, which was shown in [32].

Let us return to the effect of schooling on later earnings. IVs that are suitable for some individuals may have no effect on individuals who pursue higher education and academic degrees with a higher chance. In turn, these longer-term students will likely move up into a higher income class when they graduate from additional years of education. But the same IVs may affect individuals' return on schooling in terms of later income who have a higher chance to quit school as early as possible [3]. In sum, in all these situations where the exogeneity property must be questioned, the IV method is at risk of creating biases in the estimates that possibly exceed any bias induced by OLS estimation by orders of magnitudes.

Trap 2: Weak Instruments. In addition to the exogeneity feature, a valid instrument needs to meet the requirement to be correlated with the explanatory variables and display a high degree of explanatory power. Stock, Wright and Yogo [34] express that 'Finding exogenous instruments is hard work, and the features that make an IV plausibly exogenous [...] can also work to make the instrument weak.' An instrument is termed 'weak' if it is poorly correlated with the endogenous explanatory variables X. It is important to note that the statistical properties of IV estimators and the deduced inferential methods and results are sensitive to the choice and validity of the instruments, even for large sample sizes. An obvious implication is that when the instruments are not valid, the statistical inferential methods are not applied successfully and are invalid, too. Nor are the deduced estimates, confidence intervals, hypothesis tests, and other inferential statistics robust in situations of deviations from valid instruments and statistical properties are futile. It is inevitable that different researchers who engage different weak instruments end up with drastically different IV estimates and divergent conclusions. These unpleasant consequences are simi-lar to those caused by another lurking pitfall, non-identifiability, which is yet another key concept that researchers who rely on regression analysis need to be cautious about. When IVs do not exhibit a high degree of predictive power for the endogenous variables (or the number of instruments is large), a variety of weaknesses of the instruments emerge, even in the presence of exogeneity. Again the IV estimates may possibly be more biased than OLS estimates, inconsistent, and exhibit lower precision than anticipated. They thus lose all properties expected of estimators and become worthless. Consequently, confidence intervals and hypothesis tests based on IV estimates become unreliable, inaccurate, and meaningless. Aside from the nonzero finite sample bias $\mathbf{E}[(X'P_Z X)^{-1}X'\,P_Z\,\varepsilon]$ of $\hat{\beta}^{\mathrm{IV}}$, weak instruments bring about another flaw in the estimation, which is a poorly performing asymptotic approximation of the sampling distribution of IV estimators to the true but unknown value of β. This finite sample bias increases with the number of less relevant instruments that are added to the model. Bound, Jaeger and Baker [5] and Hahn and Hausman [21] argue that the bias is inversely related to the F-statistic of the first stage regression of X onto Z. These two observations combined imply that the partial R^2 and F statistics of the first stage regression give an indication of the quality of the IV estimates, and for that reason, should be routinely reported.

It is worthwhile to examine the magnitude of the relative inconsistency of the IV estimator compared to the OLS estimator and the finite sample bias of $\hat{\beta}^{\mathrm{IV}}$ when the instruments are weak and not exogenous. In this case, the finite sample bias is substantially more magnified. For this purpose, consider the simplest IV model with one explanatory variable and one instrument. Let $\rho_{Z,\varepsilon}$, $\rho_{X,\varepsilon}$, and $\rho_{X,Z}$ denote the respective correlation between ε and Z, ε and X, and between X and Z. Notice that exogeneity of Z is equivalent to $\rho_{Z,\varepsilon} = 0$. Endogeneity of X is equivalent to $\rho_{X,\varepsilon} \neq 0$. A weak instrument Z will have a relatively small value of $|\rho_{X,Z}|$. Bound, Jaeger and Baker [5] demonstrate that, when Z is not exogenous, that is, $\rho_{Z,\varepsilon} \neq 0$, the relative inconsistency of the IV estimator compared to the OLS estimator is equal to

$$\frac{\hat{\beta}^{\mathrm{IV}} - \beta}{\hat{\beta}^{\mathrm{OLS}} - \beta} = \frac{\rho_{Z,\varepsilon}}{\rho_{X,\varepsilon} \cdot \rho_{X,Z}} \,.$$

Thus, when the instrument Z is weak, which is expressed in terms of a small value of $\rho_{X,Z}$, this relative inconsistency blows up, possibly, to the extent that it renders the inconsistency of the IV estimator larger than the one of the OLS estimator. This occurs if $|\rho_{Z,\varepsilon}/(\rho_{X,\varepsilon}\, \rho_{X,Z})| > 1$. In conclusion, when the instruments are weak, even in large samples, the large sample asymptotic approximations are poor, which invalidates the classical results that are vital for statistical inference such as the construction of confidence intervals and hypothesis testing of the overall model or parameters (Nelson and Startz [29]).

Trap 3: Endogeneity Testing Based on IVs. In the presence of endogeneity, it would be beneficial to test whether externally available IVs are exogenous and provide explanatory power for the endogenous X. Hahn and Hausman [20] developed a validity test for IVs that simultaneously addresses exogeneity and strength. Rejecting the null hypothesis indicates either a lack of the assumed exogeneity or weakness of the instruments. The initial exogeneity test procedure of Hausman [23] obviously relies on external instruments, since it depends on the difference between the OLS and IV estimates. However, amid the dilemmas one possibly encounters with IV regression, it would be advantageous to test potential regressor error correlation prior to having to find instruments. This would avoid the search for suitable instruments that frequently is problematic in practice, for the sake of only coming to the conclusion later on that there is no need for the instruments because OLS estimation performs equally well. One possible outcome of performing the Hausman test for endogeneity is the conclusion that previously identified instruments are not needed. An additional complication with this circular problem that the Hausman test creates is the failure to reject the suitability of the OLS estimator simply because of the bias that weak instruments introduced in the test statistic (see Hahn and Hausman [21]).

4 Latent Instrumental Variables Approach

The latent instrumental variables method that Ebbes [9] presented attempts to solve these circular problems. This is an instrument-free approach that utilizes a latent variable to estimate regression parameters when endogeneity is present. The endogenous explanatory variable is decomposed into an exogenous part and an endogenous error term, where the exogenous term is an unobserved discrete variable and the model parameters are identified and estimated via maximum likelihood methods. Hence observed IVs are not required. Interestingly, 'optimal' IVs are estimated from the data and endogeneity can be tested as well. Mixture modeling techniques [27] estimate a grouping associated with the discrete latent variable simultaneously with the other parameters.

4.1 Model

Suppose we fit an LIV model to n available independent and identically distributed (i.i.d.) observations $(Y_1, X_1)', \ldots, (Y_n, X_n)'$. The LIV model is assumed to have the form

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ X_i &= \pi' Z_i + \nu_i, \end{aligned} \tag{4.4}$$

where $\pi = (\pi_1, \pi_2, \ldots, \pi_m)'$ for $m \geq 2$, the Z_i are unobserved categorical variables that are assumed to be uncorrelated with the errors (ε, ν) and follow a multinomial $(1, \lambda)$ distribution with jth group mean $\lambda_j > 0$ and $\sum_{j=1}^{m} \lambda_j = 1$, and the errors (ε, ν) are assumed to obey the normal distribution $F = N(\mathbf{0}, \Sigma)$ with variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\nu} \\ \sigma_{\varepsilon\nu} & \sigma_v^2 \end{bmatrix}. \tag{4.5}$$

Hence, the correlation $\rho(X_i, \varepsilon_i)$ between X_i and ε_i is fully captured by the correlation $\rho(\nu_i, \varepsilon_i)$ between ν_i and ε_i, whereas $\rho(Z_i, \varepsilon_i) = 0$. Let us investigate the implications of this structure on the unconditional and conditional distributions of $(Y_i, X_i)'$, given the group membership of Z_i, in the case when $m = 2$. Denote $e_1 = (1, 0)'$ and $e_2 = (0, 1)'$. The conditional distribution $\mathcal{L}((Y_i, X_i)' \,|\, Z_i = e_j) = N(\mu_j, \Omega)$, given that Z_i belongs to the jth group, has expectation

$$\mu_j = \mathbf{E}\left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \,|\, Z_i = e_j \right) = \begin{pmatrix} \beta_0 + \beta_1 \pi_j \\ \pi_j \end{pmatrix} \tag{4.6}$$

and variance-covariance matrix

$$\Omega = \begin{bmatrix} \beta_1^2 \sigma_v^2 + 2\beta_1 \sigma_{\varepsilon\nu} + \sigma_\varepsilon^2 & \beta_1^2 \sigma_v^2 + \beta_1 \sigma_{\varepsilon\nu} \\ \beta_1^2 \sigma_v^2 + \beta_1 \sigma_{\varepsilon\nu} & \sigma_v^2 \end{bmatrix}. \tag{4.7}$$

The marginal probability density function (pdf) of $(Y_i, X_i)'$ is the pdf of a mixture [27] of two bivariate normals

$$f(y_i, x_i) = \lambda f_1(y_i, x_i) + (1 - \lambda) f_2(y_i, x_i) \tag{4.8}$$

with mean

$$\mu_{y,x} \quad = \quad \begin{pmatrix} \beta_0 + \beta_1[\lambda\pi_1 + (1-\lambda)\pi_2] \\ \lambda\pi_1 + (1-\lambda)\pi_2 \end{pmatrix} \tag{4.9}$$

and variance-covariance matrix

$$\Omega_{y,x} \quad = \quad \Omega + \lambda(1-\lambda)(\pi_1 - \pi_2)^2(\beta_1, 1)'(\beta_1, 1), \tag{4.10}$$

where f_j denotes the bivariate normal pdf, given $Z_i = e_j$. The parameters $\beta_0, \beta_1, \Sigma, \pi_1, \pi_2$, and λ that need to be estimated turn out to be identifiable in this model. Specifically, Ebbes [9] showed that, if $m \geq 2$ is fixed and all group means π_j are different, all parameters including the mixture probabilities can be uniquely determined (up to a possible permutation of the group labels) for the LIV model with normally distributed errors.

Importantly, however, a normal distribution for the unobserved instrument results in an underidentified model where the parameters are unidentifiable. Apparently the same situation emerges when $m = 1$ and the instrument equals a constant. A speculation in [9] has it that if the error distribution of ε is assumed to be non-normal, a normal distribution for the latent instrument may be feasible.

Ebbes [9] mentions that incorporating analytical expressions of the gradient vector and Hessian matrix in numerical optimization techniques such as the quasi-Newton routines, for instance, the BFGS method, drastically increases the algorithmic convergence speed and provides more stable results than numerically approximating the gradient and Hessian.

Equipped with the LIV estimate $\hat{\beta}_{\text{LIV}}$ for β, Ebbes [9] proposed a Hausman exogeneity test of the explanatory variables X that relies on the Hausman-LIV test statistic

$$H_{\text{LIV}} = (\hat{\beta}_{\text{LIV}} - \hat{\beta}_{\text{OLS}})' \, \hat{\Sigma}^{-1}_{\text{H}_{\text{LIV}}} (\hat{\beta}_{\text{LIV}} - \hat{\beta}_{\text{OLS}}), \tag{4.11}$$

where $\hat{\beta}_{\text{OLS}}$ is the OLS estimate for β and $\hat{\Sigma}_{\text{H}_{\text{LIV}}}$ is the estimated asymptotic covariance of the difference $\hat{\beta}_{\text{LIV}} - \hat{\beta}_{\text{OLS}}$. The null hypothesis is that both the OLS and LIV estimates are consistent, while the alternative hypothesis asserts that only $\hat{\beta}_{\text{LIV}}$ is consistent. Under the null hypothesis, H_{LIV} has an asymptotic χ^2_1 distribution. Importantly, since this LIV variant of the Hausman test does not require any observed IVs, it eliminates the circular problem that the IV approach poses and avoids potential traps that weak IVs inject in the test procedure.

4.2 Simulation study

Beyond the solution to the identifiability issue that was covered in the last subsection, a couple of obvious questions emerge in the LIV approach. Whether the LIV estimates are unbiased and consistent is not addressed from a mathematical point of view in [9] but the performed Monte Carlo simulation in various scenarios establishes that potential endogeneity issues are adequately resolved. In contrast with OLS, the LIV estimates are approximately unbiased, the parameters of interest can be recovered effectively, and the statistical power of the Hausman-LIV test is judged to be satisfactory. Thus, the test is sufficiently powerful to suitably identify endogeneity. Importantly, the exact choice of the number of categories of the unobserved discrete instrument is not sensitive to the final outcomes when the LIV method is applied. Subsequent sensitivity analysis for the simple LIV model, based on the specifications that the unobserved discrete instrument has three or four categories, reveals no significant impact on the estimation results nor on the statistical power estimates of the Hausman-LIV test.

In the Monte Carlo (=MC) simulations that explored the consistency of the model parameters and accuracy of the estimation, Ebbes [9] considers various scenarios for the true number of instruments $\tilde{m} \in \{2, 4, 8\}$, the distribution of group allocation π, and correlation $\rho \in \{0.0, 0.1, 0.2, \ldots, 0.5\}$ of the endogenous variable X with ε. It is assumed that $\mu_X = 0$ and $\sigma_X^2 = 2.5$ in all simulations. The resulting distribution includes unimodal and bimodal distributions and a skewed distribution. Note that the constant can be estimated in an unbiased way via the OLS method because $\mu_X = 0$. Data were generated for $n = 1000$ observations and 250 Monte Carlo replications. We focus on the model parameters β_1 and σ_ε that apparently are of principal interest. The estimation of the remaining parameters is discussed [9], as well. In the presented bias plots for β_1 and σ_ε, the simple LIV model with $m = 2$ is contrasted with the OLS estimate $\hat{\beta}^{\text{OLS}} = \beta + (X'X)^{-1}X'\varepsilon$. Only one OLS bias plot is shown, which may indicate that the OLS estimate was not sensitive to the different scenarios that the MC simulations highlighted.

Bias Plots for Estimates of β_1. A first observation in Figure 3.1, [9], is that the OLS estimate only performs well in the situation of perfect exogeneity, in which it is known to provide the 'best linear unbiased' estimator in terms of smallest possible variance among all unbiased linear estimators. However, in all other situations of nonzero correlation ρ, the OLS estimate carries a substantial bias across the majority of MC replications. The magnitude of bias begins to be increasingly and outstandingly high for $\rho \geq 0.3$. Moreover, since the predicted uncertainty of the OLS estimator remains small across various degrees of endogeneity, this estimator is extremely misleading not only because the estimate's bias is strikingly large but it also gives the false impression of being highly precise at the same time that it is highly inaccurate. Its distribution is narrowly centered around a blatantly wrong estimate. A second observation is that, with $n = 1000$ data points, the simple LIV estimate provides an asymptotically unbiased estimator across all scenarios including with no, small, or high correlation ρ, where a higher degree of endogeneity decreases the amount of uncertainty of the estimate. The precision of the LIV estimator is reasonably high for none or small correlation ρ and is high for medium to high correlation ρ. Furthermore, when the true instrument has an obvious cluster structure, it is well approximated by the assumed discrete instrument. When there are four or eight instruments in the assumed unimodal distribution for the group allocation and the model is mis-specified, some efficiency is lost with the simple LIV approach, since the model approaches the vicinity of the non-identifiability region. Lower information in the mixture models translates into lower efficiency of the LIV estimate. When the true number of categories is larger, it is expected that the distribution of X becomes closer to a normal distribution. We know that if the unobserved instrument has a normal distribution, this results in an unidentified model. This particular phenomenon is more emphasized for none or minor endogeneity. Yet, this only is a minor concern at best, since in that case one can resort to the OLS approach.

Bias Plots for Estimates of σ_ε^2. We gather from Figure 3.2, [9], that the LIV estimates appear to be asymptotically unbiased in all cases. For the skewed and two bimodal distributions for the group allocation, the simple LIV estimation results in high precision even without endogeneity. When there are four or eight instruments in the assumed unimodal distribution, the distribution of the LIV estimate is slightly positively skewed across scenarios. With increasing endogeneity, the variance of the estimates marginally increases, which is expected since Y and $\mathbf{E}(Y|X)$ tend to be further apart. Similarly as we remarked for β_1, the OLS estimator for $\rho > 0.2$ experiences a substantial bias, which is a downward bias, that becomes significantly more emphasized with increasing endogeneity. Again the OLS estimator turns out to be utterly misleading, as it falsely indicates high precision, which at best would be high precision around a glaringly wrong estimate.

Estimated Power of Hausman-LIV Test. For the distributions that were selected for the MC simulation study [9], the test exhibits reasonably good power for a wide range of levels of statistical significance and is robust under model mis-specification with a minor tendency towards rejecting the null hypothesis too often. The power of the test is the lowest for the unimodal distributions and highest for the bimodal distributions and skewed distribution, which reflects on the simple LIV model adequately representing the different clusters. Yet when endogeneity is lacking, the Hausman-LIV test too frequently rejects the null hypothesis for the skewed distribution and unimodal distributions with $\tilde{m} = 2$ or $\tilde{m} = 4$. The latter instance is caused by lacking information in the mixture models.

Choice of Number of Clusters. Making the right choice for the value of m is not crucial in successfully applying the Hausman-LIV test. Sensitivity analyses in [9] that rely on $m = 3$ and $m = 4$ instead of $m = 2$ demonstrate that the conclusions relating to the statistical power and estimation of β_1 and σ_ε remain intact.

4.3 Estimating the returns to education: Application of LIV method

Whether time spent in education completely or partially explains who lands the top earning jobs in the market at a given time and the impact of education on income in general are topics of wide public interest and attention. It is apparent that other factors than education tenure, such as ability, may predict income in subsequent years and would need to be taken into account to avoid biases in the statistical inference. For this reason and by means of various comparisons, the OLS estimate based on linear regression of income on years of education has been believed to be biased. This example seems to be a quintessential arena to resort to instrumental variables as remedy – we will mention another example, as well, in Section 6. Strikingly, though, when indeed other explanatory variables are included in the regression and IV regression is adopted rather than OLS regression because of suspected endogeneity, the injected bias in the estimated explanatory power of education for income often has been reported to be stupendous. In order to put the biases by OLS and IV regression on scale, Ebbes [9] applied the LIV approach to three empirical datasets to investigate the return of education on income. Consider the following linear regression

$$Y_i = \beta_0 + \beta_1 S_i + X_i \beta_2 + \varepsilon_i, \tag{4.12}$$

where Y_i denotes the logarithm of a measure of earnings, S_i a measure of education and X_i is a collection of other explanatory variables that are assumed to influence Y_i. The coefficient β_1 measures the effect of education on income when adjustments are incorporated for other variables X_i and is expected to be positive. The error term ε_i absorbs the remaining factors that may influence Y_i that are not captured in this equation.

We gather from the findings in [9] that the results based on the LIV approach are more consistent than those based on the classical IV approach. There are strong indications that the observed instruments that have been engaged are either weak or endogenous in two of the three applications. Judging from the LIV estimates, the OLS estimate of β_1 exhibits a moderate upward bias of around 7%, which supports the ability bias hypothesis. However, relying on IV estimates to gauge bias had previously indicated a bias of the OLS estimate that stunningly ranges from -80% to 30% for these three data applications. This lucidly illuminates the crucial major concern that if, for a particular research problem, one uses different sets of instrumental variables that are not all suitable because of issues with potential weakness or endogeneity, one is gravely misled about the magnitude and direction of bias in estimating model parameters. Bias produced by OLS estimation can be dramatically over- or underestimated, while there are no tools to double-check whether the IV approach was applied suitably or with minor or serious flaws. The resulting extremely slim agreement in the direction and magnitude of the potential bias is not at all surprising since the multitude of bias sources have their own specific impact on direction and magnitude and may offset or reinforce each other in ways that cannot be assessed.

Violation of the zero correlation assumption of education, the variable X, with the error term ε has been attributed in the literature to the pitfalls of omitting ability, measurement error, and heterogeneity among others. As an example, we illustrate the implied 'ability' bias when one presumes that ability enables some individuals to secure higher income than they would without that ability. When individuals with higher ability tend to acquire more education, the ability bias is further reinforced. If education is the only explanatory variable besides education tenure, equation (4.12) becomes $Y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + \varepsilon_i$, where the coefficient β_2 associated with ability represents the effect of ability on income. It is expected that $\beta_2 > 0$ because in average higher ability and more years in education should benefit income. When ability is unobserved, the OLS estimator for the effect of education on income is falsely exaggerated by approximately $\beta_2\, \sigma_{\mathrm{SA}}/\sigma^2_{\mathrm{S}}$, where σ^2_{S} denotes the variance of S_i and σ_{SA} the covariance of ability and the earnings quantity. Whenever the linear association between ability and earnings is positive, this causes an upward bias. However, this upward bias can be offset by a downward bias of the OLS estimate, for instance, if the measure for education is imperfect or other other explanatory variables are omitted. Griliches [19] argued that, if 'years of schooling' does not fully capture total education, a large downward bias may be triggered but the latter may itself be offset by other omitted relevant variables.

In an overview of studies that employ IVs to estimate the returns to education, Card

[7, 8] distinguishes between IVs based on (i) institutional features of the school system and (ii) family background characteristics. In the former case, the produced IV estimates are approximately 30% higher than the OLS estimates, which is in contradiction with the current belief about ability bias in the literature (see Card [7, 8] for possible explanations). In the latter case, when, for example, measures on education levels of family members are incorporated in the model, Card [7] shows that the IV estimator is at least as biased upwards as the OLS estimator if the OLS estimator is biased, and the IV bias might exceed the OLS bias. Card's [7] approach that uses twins or siblings data exploits information on the twins or siblings and attempts to eliminate biases by correcting for omitted variables or measurement error. He concludes that the OLS estimator exhibits a slight upward bias of around $10\% - 15\%$. A possible pitfall of this avenue is that these results may not readily generalize to non-twins and the assumption of identical abilities of identical twins or siblings may be questionable and violated.

Next we describe the three empirical datasets to gauge the biases that were encountered in the results from OLS or IV regression to examine the return of education on income by means of the LIV method.

NLSY Data. The US National Longitudinal Survey of Young Men (NLSY) data set with observations drawn in 1976 on 3010 men aged between $24 - 34$ years followed since 1966 that contains several supposedly exogenous variables and one dummy IV indicator for the presence of a nearby college was analyzed in [6] and [35].

Brabant Data. This data set was originally sampled from the Dutch province 'Noord-Brabant' in 1952 and contains observations on 833 men who were contacted 30 years later when expected to hold a stable labor market position. Available variables are educational level, income, and social background, several exogenous explanatory variables, and quantities on the educational level of the respondents' father and mother (see [22] for a more detailed description of the data).

PSID Data. The University of Michigan Panel Study of Income Dynamics (PSID) data set which contains observations on 424 working married white women between the ages of 30 and 60 in 1975 was analyzed by Wooldridge [36] and Mroz [28]. Available variables include several exogenous explanatory variables and the educational level of the respondents' father, mother, and husband.

Results. We compare the results between the LIV, OLS, and IV methods presented in [9] on the estimated schooling effect on income for these three datasets. Of vital interest are the biases of the OLS and IV estimates, when the LIV estimates are used to gauge the bias. It turns out, while for each of the data sets, the OLS estimate slightly over-estimates the returns to education, the IV estimate substantially over-estimates the return to schooling for the NLSY and Brabant data sets but noticeably under-estimates the return for the PSID data set. It is important to note that, while the IV approach utilized the available observed instruments mentioned above, the LIV approach was applied without them. When detailing the LIV estimates, we turn to the optimal group size m_* for the latent variable identified based on BIC and AIC3. This optimal cluster size is $m_* = 5, 4, 5$ for NLSY, Brabant, and PSID, respectively. The LIV results are fairly stable across different choices of m.

Table 1: LIV, OLS, and IV Estimates and Bias of OLS and IV Estimates Relative to LIV Estimate for NLSY, Brabant, and PSID Data Sets

Data Set	$\hat{\beta}_1^{\text{LIV}}$ (SE)	$\hat{\beta}_1^{\text{OLS}}$ (SE)	$\hat{\beta}_1^{\text{IV}}$ (SE)	Bias (%) of $\hat{\beta}_1^{\text{OLS}}$	Bias (%) of $\hat{\beta}_1^{\text{IV}}$
NLSY	0.069 (0.004)	0.074 (0.004)	0.133 (0.052)	7.25	92.75
Brabant	0.040 (0.005)	0.043 (0.004)	0.056 (0.008)	7.5	40
PSID	0.096 (0.014)	0.102 (0.014)	0.073 (0.032)	6.25	−23.96

For NLSY, Brabant, and PSID, the respective estimate $\hat{\beta}_1^{\text{LIV}}$ (SE) is given by 0.069 (0.0040), 0.040 (0.0049), and 0.096 (0.0142), the estimate $\hat{\beta}_1^{\text{OLS}}$ (standard error) is 0.074 (0.0035), 0.043 (0.0044), and 0.102 (0.0139), and the estimate $\hat{\beta}_1^{\text{IV}}$ (standard error) is 0.133 (0.0518), 0.056 (0.0075), and 0.073 (0.0321). We observe that the OLS estimates are above the LIV estimates in all cases, while the IV estimates exceed the OLS and LIV estimates for the NLSY and Brabant data but lie below the OLS and LIV estimates for the PSID data. Hence, the LIV approach leads to disagreement about the direction of the OLS bias across the three data sets. Let us examine the relative errors of the OLS and IV estimates compared to the LIV estimate, derived as $100\%(\hat{\beta}_1^{\text{OLS}}/\hat{\beta}_1^{\text{LIV}} - 1)$ and $100\%(\hat{\beta}_1^{\text{IV}}/\hat{\beta}_1^{\text{LIV}} - 1)$, respectively. As we gather from Table 1, the OLS estimate indicates a slight upwards bias of 7.25%, 7.5%, and 6.25% relative to the LIV estimate for NLSY, Brabant, and PSID, whereas the IV estimate indicates an enormous or substantial bias of 92.75%, 40%, and −23.96% relative to the LIV estimate when there also is no consensus about the direction among the three situations.

Testing for Endogeneity. We recollect from (4.11) that the Hausman-LIV and Hausman test statistics rely on the difference between the estimators $\hat{\beta}_1^{\text{OLS}}$ and $\hat{\beta}_1^{\text{LIV}}$ and between the estimators $\hat{\beta}_1^{\text{OLS}}$ and $\hat{\beta}_1^{\text{IV}}$, respectively. Judging from the Hausman-LIV test, endogeneity is clearly expressed for the NLSY and PSID data with test statistic values of 7.18 and

4.20, but to a lesser extent for the Brabant data with a test statistic value of 2.63. In contrast, if the IV estimate is adopted, the Hausman test is misleading and erroneously strongly indicates endogeneity only for the Brabant data with a test statistic value of 4.35, but to a smaller degree for the NLSY and PSID data with test statistic values of 1.31 and 0.95. However, it should be the other way around by virtue of the endogeneity test via LIV estimates.

4.4 Residuals, outliers, and influential observations in LIV models

Outliers and influential observations can reveal deviations from model assumptions and impact estimates, even though the ML estimation remains approximately valid except for small samples. Inspecting residuals helps to detect potential departures from model assumptions such as non-normality, marked asymmetry, bimodality, heavy tails, and heteroscedasticity. Since residuals are estimates of the model error term, they are expected to mimic the features of the disturbance term under a particular assumed distribution. If we compute the a posteriori category membership by means of the LIV method, we obtain estimated latent instruments without using instrumental variables. These LIVs can be incorporated in the regression equation of the endogenous variables in order to compute the R^2 statistics (see [5]) that serves as diagnostics in the IV estimation to identify weak instruments that may be omitted due to lack of relevance. In most empirical work it is sensible to assume bivariate normal residuals. Yet if the data were generated from a sharply different distribution, the MLE should take this into account.

Defining LIV residuals requires caution. We restrict the discussion to conditional residuals. Another option that Ebbes [9] mentions consists in IV-type residuals. Since the explanatory variables X are random, it is sensible to condition on X, when viewing residuals. Recall that we assumed that the complete LIV model is a bivariate mixture model. We recollect the LIV model equations $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and $X_i = \pi' Z_i + \nu_i$ for $i = 1, \ldots, n$. Conditional on $X_i = x_i$ and the j-th class $Z_i = j$, the distribution of Y_i is $N(\mu_{Y_i | X_i = x_i, Z_i = j}, \sigma^2_{Y|X,j})$ for each $1 \le i \le n$, where

$$\begin{aligned} \mu_{Y_i|X_i=x_i,Z_i=j} &= \mathbf{E}(Y_i \,|\, X_i = x_i,\, Z_i = j) = (\beta_0 - \frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}\,\pi_j) + (\beta_1 + \frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2})\, x_i \\ &= \beta_0 + \beta_1 x_i + \frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}\, v_i \end{aligned} \tag{4.13}$$

with $v_i = x_i - \pi_j$ and $\mathrm{Var}(Y_i \,|\, X_i,\, Z_i = j) = \sigma^2_{Y|X,j} = \sigma^2_\varepsilon - \sigma^2_{\varepsilon\nu}/\sigma^2_\nu$. Here $\sigma_{\varepsilon\nu}$ denotes the covariance between ε and ν and captures the linear association between ε and X, which is non-zero unless X is exogenous. The conditional mean $\mu_{Y_i|X_i=x_i,Z_i=j}$ is estimated by $\hat{\mu}_{Y_i|X_i=x_i,Z_i=j} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_i + (\hat{x}_i - \hat{\pi}_j)\,\hat{\sigma}_{\varepsilon\nu}/\hat{\sigma}^2_\nu$, where the unknown model parameters and latent variable are replaced by the LIV estimates. Since the instrumental variable category is unknown a priori, we estimate it by $\sum_{j=1}^m \tilde{p}_{ij}\,\hat{\pi}_j$ for the ith observation, where $\tilde{p}_{ij}$ denotes the posterior probability of the ith observation to belong to group j. Consequently, the prediction of Y_i, conditional on $X_i = x_i$, is given by

$$\hat{Y}_i|\{X = x_i\} = \sum_{j=1}^m \tilde{p}_{ij}\,\hat{\mu}_{Y_i|X_i=x_i,Z_i=j} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_i + (\hat{x}_i - \sum_{j=1}^m \tilde{p}_{ij}\,\hat{\pi}_j)\,\hat{\sigma}_{\varepsilon\nu}/\hat{\sigma}^2_\nu.$$

Hence, the ith conditional LIV residual, given $X_i = x_i$, is

$$e_i|\{X_i = x_i\} = (Y_i - \hat{Y}_i)|\{X_i = x_i\} = \beta_0 - \hat{\beta}_0 + (\beta_1 - \hat{\beta}_1)\hat{x}_i - (\hat{x}_i - \sum_{j=1}^m \tilde{p}_{ij}\,\hat{\pi}_j)\,\hat{\sigma}_{\varepsilon\nu}/\hat{\sigma}^2_\nu.$$

Curiously, if we treat X with mean μ_X and variance σ^2_X as given in $Y = \beta_0 + \beta_1 X + \varepsilon$ with $\sigma_{X\varepsilon} \neq 0$, then the probability limit of the OLS estimator for β_0 equals $\beta_0 - \mu_X \cdot \sigma_{X\varepsilon}/\sigma^2_X$ and for β_1 equals $\beta_1 + \sigma_{X\varepsilon}/\sigma^2_X$, which are both inconsistent estimates unless $\sigma_{X\varepsilon} = 0$.

The conditional LIV residuals are helpful to detect heteroscedasticity via scatterplots of the residuals versus the explanatory variables and the predicted values. While heteroscedasticity does not affect the consistency of the estimated regression parameters, it entails a loss in efficiency. Ebbes [9] remarks that the IV-type residuals can be employed to compute kurtosis and skewness and detect departure from normality for the error distribution. Importantly, in MC simulations of samples of size $n = 1000$ carried out in [9] with a misspecified error distribution being either the χ^2_1 distribution or t_3 distribution, the LIV method was found to be fairly robust against misspecification, unlike the OLS estimator when substantial endogeneity exists. It is noted, though, that the maximum likelihood LIV estimator may not be fully efficient anymore and more efficient estimators may exist. For such misspecified error distributions, the LIV approach can face a dilemma if the numerical optimization algorithm mistakenly mixes on a skewed error distribution instead of that of the latent variable.

We conclude this section by noting that similar tools are available as in OLS regression to identify outliers and influential observations (the reader is referred to [9]).

5 Bayesian LIV Approach

Interesting extensions of the standard LIV model are investigated in [9] from a Bayesian point of view. The multinomial distribution with a fixed number of categories, which had to be estimated for the latent instrument, is relaxed to a general distribution G, and endogeneity is examined in two commonly used multilevel models that possibly carry more than one endogenous variable. In the nonparametric Bayes LIV approach, the 'best' distribution of the latent instrument is completely determined and estimated from the data. In the presence of endogeneity, evidently multi-level models become more complex and traditional methods such as fixed-effects estimation and random-effects estimation were shown to be limited in multiple ways [10].

We add a few comments on the properties of Bayesian estimators pertaining to features that are preferable in the repeated sampling or frequentist approach. Bayesian estimates are reasonable if they are unbiased in the limit for large samples but otherwise this concept is and cannot be a criterion in the Bayesian estimation (see [18], Chapter 8, for further details). In this case, the estimate is asymptotically unbiased. Yet a Bayesian estimate generated by an algorithm that converges to the true value of the unknown estimate with increasing sample size, and thus is consistent, is appealing. This happens, for example, under common regularity conditions of the maximum likelihood function when the posterior distribution tends to be concentrated around the mode with increasing sample size and the posterior mode is consistent for a parameter that is estimated. Gelman et al. [18] point out that the posterior mode, median, and mean are consistent and asymptotically unbiased when the true distribution is included in the family of models relied on for the model fit and under mild regularity conditions. In addition, under mild regularity conditions, the center of the posterior distribution in terms of the mode, median, or mean (as appropriate) is asymptotically efficient, that is, there is no other estimator that estimates the unknown parameter with a lower mean squared error.

5.1 A simple multilevel model with LIVs

It is noted [9] that, when additional explanatory variables and random terms are added to a simple multilevel model, the Dirichlet process for the latent variable and the MCMC estimation are not affected, as the MCMC estimation is performed conditional on all other parameters and observations. A simple multi-level model with allowed possible endogeneity is given by

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij} \qquad (5.14) \\ X_{ij} &= \theta_{ij} + \nu_{ij} \end{aligned}$$

for $i = 1, \ldots, n$, $j = 1, \ldots m_i$, and the total number of observations $N = \sum_{i=1}^{n} m_i$, where X_{ij} is a covariate for the ith individual at level j, and θ_{ij} is an unobserved instrument. Assume that the error terms ε_{ij} have zero mean and variance σ_ε^2 and are independently and identically distributed across and within individuals. The variable X_{ij} is endogenous, when the correlation $\rho(X_{ij}, \varepsilon_{ij}) \neq 0$, equivalently, the covariance $\mathbf{E}(X_{ij}\,\varepsilon_{ij}) \neq 0$. In the latter case, the endogenous explanatory variable X_{ij} is decomposed into an exogenous part θ_{ij} and an endogenous part ν_{ij} so that ν_{ij} and ε_{ij} have non-zero correlation. Moreover, it is assumed that ε_{ij} and ν_{ij} have zero mean and variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\nu} \\ \sigma_{\varepsilon\nu} & \sigma_v^2 \end{bmatrix}. \tag{5.15}$$

Unlike in the LIV regression model in (4.4), where the latent instrument is assumed to be a discrete variable with m possible categories, a Dirichlet process is assumed for the distribution of θ_{ij} (Ferguson [16], Antoniak [4], Escobar [14], Escobar and West [15]). Specifically, the θ_{ij} are i.i.d. according to G, where G is not in parametric form but rather has a Dirichlet process prior $\mathcal{DP}(\alpha, G_0)$ with precision or dispersion parameter α and initial prior distribution G_0 for the location parameter of $\mathcal{DP}$. The Dirichlet prior puts a probability measure (or distribution) on the space of all possible probability distributions for G. While for large values of α, G tends to be concentrated close to G_0, for small values of α, the probability mass of G has a high probability to be concentrated on a few distinct atoms. The support of the Dirichlet process is the class of all distribution functions. This nonparametric model offers the advantage to fit a distribution G to the data that can take on a multitude of different shapes, including distributions that are skewed, heavy-tailed, or multimodal or that have shoulders, regardless of the shape of G_0. For G_0, we may pick the normal distribution with mean μ_g and variance σ_g^2. Moreover, for α we may pick a gamma distribution, for $\beta = (\beta_0, \beta_1)$ a normal distribution, and for Σ an inverted two-dimensional Wishart distribution as respective prior distribution. In addition, the error terms are assumed to be i.i.d. and normally distributed. MCMC results can be employed to approximate the nonparametric Bayes LIV model (for more details, see [9]).

5.2 Endogenous subject-level covariates and random coefficients

In random coefficients models with endogenous subject-level covariates, standard estimation techniques such as hierarchical Bayes can lead to severely biased estimates, as we will see below. Endogeneity may be injected when relevant covariates that are correlated with included covariates are omitted, or when some of the covariates are measured with error. We assume that a set of individual level covariates are available to explain part of the variance of the random coefficients. We study the following standard linear two-level model [9] with random coefficients

$$\begin{aligned} Y_{ij} &= \beta_0 + X_{ij}'\,\beta_i + \varepsilon_{ij} \\ \beta_i &= \gamma_0 + \gamma\, Z_i + \nu_i \\ Z_{1i} &= \theta_i + \alpha\, Z_{2i} + \xi_i \end{aligned} \tag{5.16}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, where X_{ij} is a vector of explanatory variables, β_i is a k-vector, $Z_i = (Z_{1i}', Z_{2i}')'$ are covariates for the i-th individual, with Z_{1i} being possibly endogenous and correlated with ν_i (but uncorrelated with ε_{ij}) and Z_{2i} being exogenous, θ_i is a vector of unobserved instruments, and ξ_i an error term vector with zero mean. The variable Z_{1i} is endogenous, when the correlation $\rho(Z_{1i}, \nu_i) \neq 0$. For the latent instrument θ_i, a Dirichlet process is assumed with prior distribution G_0.

5.3 Simulation study for nonparametric Bayesian LIV approach

Ebbes [9] compares the performance of the nonparametric Bayesian LIV estimation algorithm of these two models to the standard LIV regression and the OLS method in a simulation study for three different choices of the distribution of the latent instrument: a discrete distribution with two categories, a gamma distribution, and a t_6 distribution with 6 degrees of freedom. In the latter case, which is a heavy-tailed distribution, the LIV model is weakly identified, since it is not identified for an exact normal distribution. In each of the three cases, it is assumed that the latent variables have mean zero and equal variance 1.5. The sample size is chosen as $n = 1000$ for the bimodal and gamma distributions and as $n = 10,000$ for the t_6 distribution. Additionally, $\beta_0 = 1$, $\beta_1 = 2$, $\sigma_\varepsilon^2 = \sigma_\nu^2 = 1$, and $\sigma_{\varepsilon\nu} = 0, 0.36$, and$= 0.79$, respectively, for no, moderate, and severe endogeneity. For the nonparametric Bayes model, the posterior means were computed for β_1, σ_ε^2, $\sigma_{\varepsilon\nu}$, and k,which represents the number of clusters or different values of the latent variable. Then the means and standard deviations over 2000 saved MCMC iterations were obtained for these variables and tabulated in Tables 7.1 through 7.5, [9].

Let us begin by considering the simple multi-level model. Unless the model is exogenous, when the OLS approach is best, the OLS results are biased and misleading in the same sense we observed earlier. For the bimodal distribution and gamma distribution with scale parameter 0.5 and shape 0.58, the simple nonparametric Bayes model furnishes approximately unbiased results in all cases. In the presence of endogeneity, the results are similar for the Bayesian and classical LIV methods in the bimodal case when the classical LIV model is correctly specified with two clusters. Yet for the gamma case, when $\sigma_{\varepsilon\nu} > 0$, the results of the Bayesian LIV estimation are superior to those of the classical LIV estimation, which is best for $k = 3$. It thus is apparent that the nonparametric Bayes model can adapt more easily to a situation where the true distribution of the instrument is continuous, and thus, is more flexible. It is noteworthy that k is overestimated for the bimodal case, but gets closer to the true cluster size of 2 as $\sigma_{\varepsilon\nu}$ increases. As for the classical LIV model, in the nonparametric Bayesian model, endogeneity can be tested for by testing whether $\sigma_{\varepsilon\nu} \neq 0$, based on the fraction of MCMC samples with $\sigma_{\varepsilon\nu} > 0$ versus those with $\sigma_{\varepsilon\nu} < 0$ and posterior p-values can be relied on. For the t_6 distribution, Ebbes [9] reports that a sample size of 1000 was insufficient to estimate the model because the MCMC chain did not converge. To get around this nonidentifiability situation and underidentified model, a total sample size of $10,000$ was utilized in the simulations, which is rather enormous. It turns out that the standard LIV model remained unidentified and hence behaved quite sensitively to a distribution of the instrument that is somewhat close to normal. However, the nonparametric Bayes LIV model yields approximate unbiased results with relatively large standard deviations. These indicate that the model is weakly identified.

Additionally for the random coefficients model, similar findings suggest that the nonparametric Bayesian LIV approach can be successfully employed to estimate the model parameters in the presence of endogenous covariates. In contrast, the estimates obtained from the standard hierarchical Bayes model are highly biased. The standard hierarchical Bayes model also substantially underestimates the degree of heterogeneity in the regression coefficients.

6 Campaign Money's Return in Congressional Elections via LIV Regression

6.1 Election data and Bayesian spatial LIV model

Ferguson, Jorgensen and Chen [17] examined the effect of total campaign expenditures on the proportions of the Democratic and Republican votes in US House and Senate Elections between 1980 and 2014. They pooled all spending by and on behalf of candidates and considered relative differences in total campaign outlays and proportional vote differentials. The focus was on investigating conventional claims that political money is of limited importance in predicting Congressional voting outcomes. To the contrary, their findings vividly demonstrated that '... in three widely spaced years – 1980, when Congress functioned very differently than it does today, 1996, and 2012 – the relation between major party candidates' shares of the two party vote and their proportionate share of total campaign expenditures were strongly linear – more or less straight lines, in fact. The relationship was strong for the Senate and almost absurdly tight for the House.' They went on to show that these findings of a strong linear relationship and the observation that the relative total campaign expenditures by major political parties effectively predict the proportion of votes they win are valid across all election years that the data analyses comprise, namely, between 1980 and 2012 for the House and between 1980 and 2014 for the Senate.

Ferguson et al. [17] were cognizant of the apparent problem of omitted relevant variables such as popularity of candidates, and thus, endogeneity in estimating the relative effects of campaign money and spending in elections and politics. But they also recognized the troubling pitfalls that surround instrumental variables in regression analyses, as we discussed earlier, and that useful IVs are elusive in such politics and money questions. Thus, instead of adopting IVs and IV regression for the analyses of campaign money's return, which would require leaning heavily on thin reeds, they employed the LIV regression approach where no IVs are required. Moran's test identified the presence of spatial autocorrelation in the data of the majority of the Senate elections and virtually all House contests. For that reason, they took care to achieve higher accuracy in the LIV estimates and deployed a nonparametric Bayesian spatial LIV regression model. The spatial dependence is incorporated in a random intercept term of the primary regression equation, and thus has the effect of a shift in the response variable, which here is the percentage of the votes of the Democratic candidates' share. The random intercept variables are chosen independently from the two error terms in the LIV regression and their distribution is conditional on the values of the neighbors.

Figure 3 in [17] presents graphs of LIV regression analysis results for the Senate elections in every election between 1980 and 2014 together with estimated explained variation and goodness-of-fit. The analyses confirm the strong linear relationship between total campaign expenditures by major political parties and the proportion of votes they win in the Senate in each election year. Importantly, relative total campaign expenditures by major political parties effectively predict the proportion of votes they win.

The effect of relative total campaign expenditures on the major party candidates' shares of the two party vote was estimated in [17] using the Bayesian spatial LIV estimate (=median of posterior distribution) $\hat{\beta}_1^{\mathrm{BSLIV}}$ together with 95% credible intervals for every election year for the House and Senate.

Figures 1-2 in [17] display the LIV regression results, together with the estimated model fit in terms of variability explained, for the House elections in 2012 and every other election year between 1980 and 2012, respectively. At the bottom left, when Democrats spend no money, they essentially get no percentage of the votes; at the top right, when all the money spent is allocated to the Democrats, they garner all the ballots, calculated as proportions of totals for the major parties. The results, which are visualized in these graphs, establish a strong linear relationship between Democratic candidates' shares of total two party spending in House elections and the percentage of major party votes they won in each election year.

Figure 1: The estimated coefficients for the return of the 'relative difference of the total campaign money' in terms of the 'difference of the major party candidates' proportional shares of the two party vote' is plotted over the 18 election years from 1980 to 2014 (from left to right) for the House (purple solid line) and Senate (black broken line).

Figure 2: The estimated bias of the spatial regression estimate relative to the Bayesian spatial LIV estimate for the House elections between 1980 and 2012 (purple solid line) and the Senate elections between 1980 and 2014 (black broken line).

Figure 1 – Effect of Campaign Money on Proportional Vote for Democrats in Elections

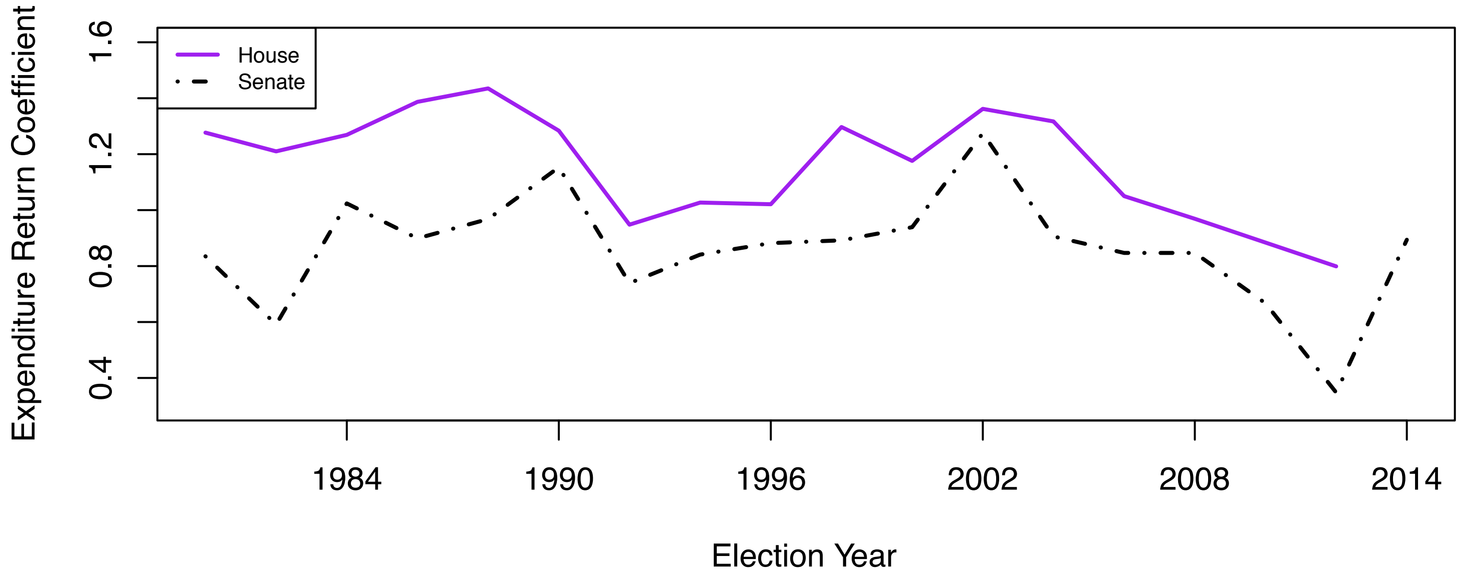

Figure 2 – Relative Bias of Spatial Regression Estimate Compared to BSLIV

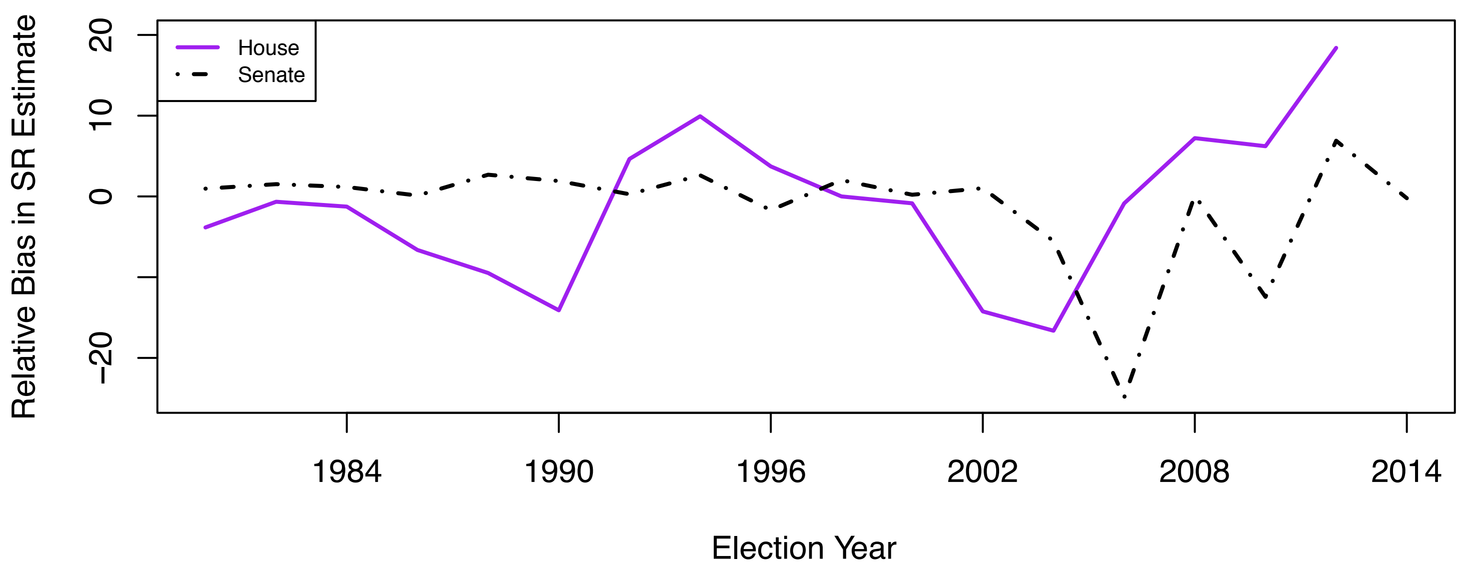

Figure 1 above visualizes an analogue, which is inflated by a certain factor that depends on the year, of the 'one dollar one vote' phenomenon and displays the estimated coefficients for the return of the 'relative difference of the total campaign money' in terms of the 'difference of the major party candidates' proportional shares of the two party vote' over the 18 election years from 1980 to 2014 (from left to right) for the House and Senate. For example, if the Democrats gathered 64% of the votes, the value used for the response variable in the LIV regression was 0.28. More abstractly, if v is the share of the Democrats' votes, the response variable in the LIV regression is $2v - 1$.

While no exogeneity test in terms of the deviance of the Bayesian spatial LIV estimator and the spatial regression estimator is available that could have been used to a priori test for endogeneity when possible spatial autocorrelation accompanies possible endogeneity, Ferguson et al. [17] provide both estimates, the Bayesian spatial LIV estimate $\hat{\beta}_1^{\mathrm{BSLIV}}$ and the spatial regression (SR) estimate $\hat{\beta}_1^{\mathrm{SR}}$ together with the standard error (SE) for each election year for comparative purposes. This allows us to estimate the bias of the spatial regression estimate $\hat{\beta}_1^{\mathrm{SR}}$ relative to the LIV estimate $\hat{\beta}_1^{\mathrm{BSLIV}}$. Similarly as for the estimated return to schooling in the income example, we inspect the relative error or bias of the SR estimate compared to the BSLIV estimate, derived as $100\%(\hat{\beta}_1^{\mathrm{SR}}/\hat{\beta}_1^{\mathrm{BSLIV}} - 1)$, for the estimated return of campaign money in the vote differential in the House and Senate elections in every year. We note that in the nonparametric Bayesian model, endogeneity can be tested for by testing whether $\sigma_{\varepsilon\nu} \neq 0$, by estimating posterior p-values from the proportion of MCMC samples with $\sigma_{\varepsilon\nu} > 0$ versus those with $\sigma_{\varepsilon\nu} < 0$, as explained in Section 5.

For the House elections the sample sizes were between 424 and 435, which is close in order of magnitude to the sample size of 1000 used for the MCMC simulations highlighted in Section 5. For the House, we gather from Figure 1 that the values of $\hat{\beta}_1^{\mathrm{BSLIV}}$ range from 0.80 (in 2012) to 1.44 (in 1988) with all values being > 1 before 2006 except for 1992 and being < 1 since 2006. Note that the return to campaign expenditure is the highest, when the value of $\hat{\beta}_1^{\mathrm{BSLIV}}$ is the largest. In various elections, the bias of $\hat{\beta}_1^{\mathrm{SR}}$ appears to be unsubstantial. Furthermore, the spatial regression over-estimates the return to campaign money in some elections, whereas it underestimates the return in other elections. The relative bias of the spatial regression estimate as compared to the spatial LIV estimate, displayed in Figure 2, ranges between -17% in 2004 to 18% in 2012 and is below 10% in absolute value in all but four instances (in 1990, 2002, 2004, and 2012). We suspect moderate to substantial endogeneity is present in those instances when the spatial regression estimate markedly differs from the spatial LIV estimate, since the the regression estimates were found to be biased under endogenous conditions in [9], as detailed above in subsection 4.3.

For the Senate elections, the sample sizes were between 32 and 36. Without doubt, these sample sizes are overwhelmingly different and smaller than the sample size of 1000 adopted for the simulations summarized in Section 5. The convergence properties of the MCMC runs applied to the elections data sets are not discussed in [17]. For the Senate elections, we notice from Figure 1 that the values of $\hat{\beta}_1^{\mathrm{BSLIV}}$ range from 0.35 (in 2012) to 1.28 (in 2002) and only three values are > 1, namely, in 1984, 1990, and 2002 when the return to campaign expenditure was the greatest. Similarly as seen for the House elections, the direction of the bias is not consistent across elections. A number of biases of $\hat{\beta}_1^{\mathrm{SR}}$ appear to be minor. The relative bias of the spatial regression estimate as compared to the spatial LIV estimate, depicted in Figure 2, ranges between -25% in 2006 to 7% in 2012 and is below 5% in absolute value in all but four instances (in 2004, 2006, 2010, and 2012).

6.2 Estimation of posterior distribution of latent variable

Before we conclude, in the hope that we may shed light on the role and functioning of the latent variable in the LIV regression, we present estimates of the latent variable in the elections example. The latent instrumental variable can represent a ‘candidate’s popularity.’ In the Bayesian LIV approach assumed, the parameters of the posterior distribution of the latent variable are estimated from data (subsection 7.1, Ebbes [9]). As prior distribution of the latent variable, Ferguson et al. [17] assumed a mixture of normal distributions, where the number of mixing components was to be determined post hoc. A mixture of normal distributions is estimated as the posterior distribution of the latent variable by using the WinBugs software. Let us look at the 1980, 1986, and 2012 House election data in [17]. The fitted models are normal mixtures $\sum_{i=1}^{4} p_i \, \mathrm{N}(\mu_i, \sigma_i)$ with 4 normal component distributions in each year and associated class weights (p_1, p_2, p_3, p_4) with $\sum_{i=1}^{4} p_i = 1$. Generated simulations in R from the estimated posterior distribution of the instrumental variable are depicted in the form of a histogram as estimated distribution by year in an appendix, ‘Appendix I: A Note on Methods,’ that appears in a forthcoming revision of [17]. The estimated distributions are shown for the election years 1980, 1986, and 2012 with the same sample sizes of 382, 361, and 384 as for the House data. The respective estimated normal mixture in each year is given by

$$0.01 \cdot \mathrm{N}(0.13, 0.03) + 0.03 \cdot \mathrm{N}(0.26, 0.03) + 0.08 \cdot \mathrm{N}(0.38, 0.03) + 0.88 \cdot \mathrm{N}(0.49, .02),$$

$$0.02 \cdot \mathrm{N}(0.14, 0.03) + 0.03 \cdot \mathrm{N}(0.27, 0.03) + 0.07 \cdot \mathrm{N}(0.39, 0.03) + 0.88 \cdot \mathrm{N}(0.51, 0.02),$$

$$0.01 \cdot \mathrm{N}(0.13, 0.03) + 0.02 \cdot \mathrm{N}(0.25, 0.04) + 0.16 \cdot \mathrm{N}(0.36, 0.04) + 0.81 \cdot \mathrm{N}(0.43, 0.02).$$

It is worthwhile mentioning that the component normal distributions and the cluster weights essentially remain the same across these years. Glancing at the posterior distribution of a candidate's popularity, we see a skewed distribution, where the rightmost normal component carries most of the observations and the components to the left represent less frequent and outlying observations.

7 Conclusions

In summary, Ebbes' method [9], which is instrument-free and uses latent instrumental variables to estimate linear regression parameters, when endogeneity is present, that is, when explanatory variables are correlated with the random error, appears to be a powerful tool and has great potential for applications. This approach also solves some of the circular problems and dilemmas with endogeneity testing and other pitfalls that have often been encountered with IV regression. The LIV approach removes expected bias that frequently is introduced in the estimation methods of ordinary least squares and IV regression, when the random error has nonzero correlation with some of the explanatory variables. Thus, if only such correlation exists with no other dependencies between the explanatory variables and error term, then the LIV approach adequately performs and exhibits favorable estimator properties. We noted that, while Ebbes' novel method is a distinct contribution, its formulation is problematic in certain important aspects. Specifically the various publications of Ebbes and collaborators [9, 12, 10, 11] employ three distinct and inequivalent statistical concepts exchangeably, treating all as one and the same. These three concepts are closely interwoven with endogeneity, as we explained. We clarified the issues that emerge and its implications. Nevertheless, in various simulations and data examples that Ebbes [9] examined in detail, the approach performed well and was successfully applied.

Importantly, since no observed instruments are required and 'optimal' instruments are estimated from data, the Hausman-LIV test provides a simple tool to a priori test for potential endogeneity in regression analysis and indicate when LIV regression is more appropriate and should be performed instead of OLS regression. Consequently, the LIV method fully addresses the question of a priori testing for potential endogeneity in linear regression analyses. It thus offers what could be an add-on feature built into linear regression in any software program to avoid biases in the regression estimates due to correlation that is present between explanatory variables and the random error.

In this paper we surveyed settings where the identifiability problem was completely settled. Even though whether the LIV estimates are unbiased and consistent is not addressed from a mathematical point of view in [9], the performed Monte Carlo simulations in various scenarios establish that possible endogeneity issues are adequately resolved. In contrast with OLS, the LIV estimates are approximately unbiased, the parameters of interest can be recovered effectively, and the statistical power of the Hausman-LIV test, a modified version of the Hausman test is judged to be satisfactory. We discussed LIV estimation of returns to education in income based on data from three studies that Ebbes [9] revisited, where 'education' is potentially endogenous due to omitted 'ability.' This example supports earlier well-known findings in the literature that the IV method indeed can lead to enormous biases, while the OLS estimation only injects relatively 'modest' biases in comparison. In a second example we closely looked at the results of Ferguson, Jorgensen and Chen [17] on the estimated effect of campaign expenditures on the proportions of Democratic and Republican votes in US House and Senate elections between 1980 and 2014, where 'campaign money' is potentially endogenous in view of omitted variables such as 'a candidate's popularity.' The relative bias we summarized of the spatial regression estimates as compared to the spatial LIV estimates indicates that endogeneity is present in the data, in addition to the identified presence of spatial autocorrelation. This provides another vigorous example where the LIV approach was successfully and adequately applied and bears its fruits.

In conclusion, numerous questions, including some mathematical and statistical ones relating to the performance of the LIV approach, remain open for future research. At the same time, it is hoped that this approach will gain wider visibility, applicability, and recognition for its merits and that its applicability will be further tested in all fields that routinely resort to linear regression approaches.

Acknowledgment. I am grateful to Professor Thomas Ferguson for piquing my interest in this topic, fruitful conversations, his encouragement, support and useful comments. I am also grateful to Professor Andrew Gelman for helpful comments.

References

[1] Aitken, A. C. (1934). On least-squares and linear combinations of observations. *Proceedings Royal Soc. of Edinburgh* **55,** 42–48.

[2] Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly J. of Economics* **56,** 979–1014.

[3] Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. of Economic Perspectives* **15(4),** 69–85.

[4] Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *Ann. Statist.* **2,** 1152–1174.

[5] Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Soc.* **90,** 443–450.

[6] Card, D. (1995). Using geographical variation in college proximity to estimate the return to schooling. In Christofides, L. N., Grant, E. and Swidinsky, R. (editors), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp,* University of Toronto Press, Toronto, 201–222.

[7] Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. C. and Card, D. (editors), *Handbook of Labor Economics,* Vol. **3A**, Elsevier Science B.V., North-Holland, 1801–1863.

[8] Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* **69,** 1127–1160.

[9] Ebbes, P. (2004). Latent Instrumental Variables – A New Approach to Solve for Endogeneity. Ph.D. thesis, University of Groningen Economics and Business.

[10] Ebbes, P., Böckenholt, U. and Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica* **58,** 161–178.

[11] Ebbes, P., Böckenholt, U., Wedel, M. and Nam, H. (2009). A Bayesian modeling framework to account for regressor-error dependencies in multilevel models. Technical Report.

[12] Ebbes, P., Wedel, M., Böckenholt, U. and Steerneman, A. G. M. (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. *Quantit. Marketing and Economics* **3,** 365–392.

[13] Efron, B. and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *J. Amer. Statist. Soc.* **86,** 9–17.

[14] Escobar, M. D. (1994). Estimating normal means with a dirichlet process prior. *J. Amer. Statist. Soc.* **89,** 268–277.

[15] Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Soc.* **90,** 577–588.

[16] Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Ann. Statist.* **1,** 209–230.

[17] Ferguson, T., Jorgensen, P. and Chen, J. (2015). How money drives US congressional elections: More evidence. *Institute for New Economic Thinking,* http://ineteconomics.org/uploads/papers/How-Money-Drives-US-Congressional-Elections-More-Evidence.pdf .

[18] GELMAN, A., CARLIN, J. B., STERN, H. S. AND RUBIN. D. B. (2004). *Bayesian Data Analysis,* 2nd edition. Chapman & Hall/CRC, Boca Raton, FL.

[19] GRILICHES, Z. (1977). Estimating the returns to schooling: Some econometric problems *Econometrica* **45,** 1–22.

[20] HAHN, J. AND HAUSMAN, J. A. (2002). A new specification test for the validity of instrumental variables. *Econometrica* **70,** 63–189.

[21] HAHN, J. AND HAUSMAN, J. A. (2003). Weak instruments: Diagnosis and cures in empirical econometrics. *Recent Adv. in Econom. Methodology* **93,** 118–125.

[22] HARTOG, J. (1988). An ordered response model for allocation and earnings. *Kyklos* **41,** 113–141.

[23] HAUSMAN, J. A. (1978). Specification tests for econometrics. *Econometrica* **46,** 1251–1271.

[24] IMBENS, G. W. AND ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62,** 467–475.

[25] JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., LÜTKEPOHL, H. AND LEE, T.-C. (1985). *The Theory and Practice of Econometrics.* John Wiley & Sons, New York.

[26] KUTNER, M. K., NACHTSHEIM, C. J. AND NETER, J. (2004). *Applied Linear Regression Models,* 4th edition. McGraw-Hill Irwin, New York.

[27] MACLACHLAN, G. AND PEEL, D. (2000). *Finite Mixture Models.* John Wiley & Sons, New York.

[28] MROZ, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* **55,** 765–799.

[29] NELSON, C. R. AND STARTZ, R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* **58,** 967–976.

[30] RUUD, P. A. (2000). *An Introduction to Classical Econometric Theory.* Oxford University Press, New York.

[31] RUTZ, O. J., BUCKLIN, R. E. AND SONNIER, G. P. (2012). A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *J. Marketing Res.* **XLIX,** 306–319.

[32] SCOSYREV, E. (2013). Identification of causal effects using instrumental variables in randomized trials with stochastic compliance. *Biometrical J.* **55,** 97–113.

[33] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

[34] STOCK, J. H., WRIGHT, J. H., AND YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *J. of Business & Economic Statistics* **20,** 518–529.

[35] VERBEEK, M. (2000). *A Guide to Modern Econometrics.* John Wiley & Sons Ltd, Chichester.

[36] WOOLDRIDGE, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Report. Massachusetts Institute of Technology, Cambridge.

[37] WRIGHT, P.G. (1928). *The Tariff on Animal and Vegetable Oils.* Macmillan.