Latent instrumental variables
Ebbes, P.

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

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2004

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 2

Instrumental variables: a survey

2.1 Introduction and bias in OLS

The standard linear regression model $y = X\beta + \epsilon$ is an important tool in (applied) statistical science to model the effect of a set of explanatory variables on a dependent variable. Here $y = (y_1, ..., y_n)'$ denotes the $n \times 1$ vector of observations on the dependent variable, $X \in \mathbb{R}^{n\times k}$ denotes the $n \times k$ matrix of observations on the explanatory variables (regressors), $\beta$ is the unknown $k \times 1$ vector of regression parameters and $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ is an unobserved stochastic disturbance. Because of identifiability it is assumed that rank $X = k < n$. Although the standard linear regression model is frequently used in cross-sectional applications, in many situations data has an hierarchical structure (see also chapter 6). For instance, when it is investigated how workplace characteristics affect a worker’s productivity, both workers and firms are units in the analysis. Similarly, hierarchical data arise in the context of panel data, when multiple observations are available on the ‘objects’ under study. This type of data is modeled through multilevel models, panel data models, or hierarchical linear models, which generalize the standard linear regression model (Judge et al., 1985, Wooldridge, 2002, Snijders and Bosker, 1999, Bryk and Raudenbush, 1992, or Greene, 2000).
An important assumption in these models is the independence of the explanatory variables \((X)\) and the random error components \((\epsilon)\). In this case the regressors are said to be ‘exogenous’ and are assumed to be determined outside the model. Failure of this assumption may lead to biased or inconsistent estimates for the parameters of interest and therefore to wrong conclusions and erroneous decision-making.

Unfortunately, in many situations the assumption of regressors and error independence is not satisfied. In this case the regressors are often said to be ‘endogenous’. Endogeneity can arise from a number of different sources: (1) relevant omitted variables, (2) measurement error in the regressors, (3) the problem of self-selection, (4) simultaneity, and (5) serially correlated errors in the presence of a lagged dependent variable in the set of regressors. Ruud (2000) shows that the possibilities (2)-(5) can be viewed as a special case of (1). A similar argument is put forward by Wooldridge (2002), who notes that the distinction among the possible causes of regressor-error correlation is not always clear. Card (1999), for instance, argues that measurement error in the education variable\(^1\), on the one hand, results in a downward bias of the effect of education on income, whereas omitted ability bias may, on the other hand, results in a positive bias in OLS. Similarly, Nevo (2000) states that price endogeneity can be generated by a price-setting firm taking unobserved product attributes into account, or can be a result of the mechanics of consumer’s optimization problem. These causes may enforce or offset each other to an extent that depends on the empirical context. In the following subsections we briefly illustrate the previously mentioned causes and provide some references to empirical studies in labor economics, marketing, and industrial economics that are confronted with these problems.

2.1.1 Relevant omitted explanatory variables

Card (1999, 2001) and Uusitalo (1999), among others, consider the estimation of the causal effect of education on earnings, where ability is a typical omit-
ted variable. Individuals with a higher ability are potentially more successful on the labor market by earning higher wages, whereas these individuals may acquire more education. As such, unobserved ability affects both education and earnings, causing a dependency between the regressor ‘education’ and the model error term (see also chapter 5).

Marketing modelers are often faced with omitted variables. Wansbeek and Wedel (1999) put forward that the exogeneity assumption of regressors, including price, is a shortcoming of standard market response models. Shugan (2004) observes an increasing focus of reviewers on endogeneity. The lack of exogeneity of regressors due to the omission of ‘key’ aspects in marketing models is gaining more interest in marketing research studies. As store managers set the marketing mix variables (e.g. price or advertising variables), their decision is based on (local) market information or product characteristics unknown to the researcher. This unobserved information may affect consumer behavior, which induces a correlation between the error term and the regressors, usually price, in a typical marketing model. Examples of unobserved local market information are competition, word-of-mouth-effects, taste changes, local market shares, or coupon availability. For recent omitted variables studies in marketing, see e.g. Villas-Boas and Winer (1999), Chintagunta (2001), Nevo (2001), Petrin and Train (2002), or Vilcassim and Chintagunta (1995).

An omitted variable model is given by (Judge et al., 1985).

\[ E(y_i|x_i, w_i) = x_i'\beta + w_i'\gamma, \]  

(2.1)

where the \( w_i \)'s are the latent or unobserved variables. Conditioning on the observable \( x_i \) but omitting \( w_i \), gives

\[ E(y_i|x_i) = x_i'\beta + E(w_i'|x_i)\gamma, \]  

(2.2)

\footnote{Models that relate sales to marketing mix variables.}
which is unequal to $x_i'\beta$ whenever: (i) $E(w_i'x_i) \neq 0$ (i.e. when omitted and included regressors are not orthogonal) and (ii) $\gamma \neq 0$ (i.e. when the omitted regressors are relevant). The resulting bias in the OLS estimator for $\beta$ is equal to $E(\hat{\beta}_{OLS} - \beta) = \Pi \gamma$ of which magnitude and size depends on $\Pi = (X'X)^{-1}X'W$ and $\gamma$. As can be seen, all estimated coefficients in $\beta$ are affected by the omission of relevant explanatory variables.

### 2.1.2 Measurement error

Measurement error in regressors arises when the variables specified in the regression model are not similar to the observed measure. This may arise, for instance, due to method- or instrument-error, the absence of a ‘physical’ measure for the true construct, like IQ, ability, perceptions, ‘total price’ versus ‘money price’, or incorrectly aggregated and combined measures from different data-sources, like GDP, price inflation or productivity of employees. When the regressors used do not conform to the variables included in regression models, it is unlikely that they are independent of the random components.

As stated before, a good measure of education that corresponds to the qualities that employers are willing to pay for, needs to be available when estimating the effect of education on income. It is common practice to use ‘years of schooling completed’ as a measure for ‘total education’. Apart from errors due to recall or recording errors in ‘years of schooling completed’, it can be questioned whether this measure fully represents education levels, because individuals may, for instance, educate themselves with evening courses or on-the-job training. Besides, as most studies on labor economics rely on household interview data, all of the variables are subject to some error (Griliches, 1977). Even if the errors are small, their effect may be magnified if more variables are added in an attempt to control for e.g. omitted ability bias\(^3\) (Card, 1999, 2001, or Griliches, 1977).

Nevo (2000) and Sudhir (2001) argue that the measure for price used in es-

---

\(^3\)See also subsection 5.2.2.
2.1 Introduction and bias in OLS

timating aggregate logit demand in inferring competition may be measured with error. The price variable used in these studies is often ‘list price’ or an aggregated price measure, whereas the model specification assumes that all consumers face the same product characteristics. However, if consumers face different prices in different stores, regions, or weeks, depending on the data, the price measure used exhibits measurement error. Instead, it would be ideal to estimate the model with transaction prices (cf. Sudhir, 2001). Bagozzi, Yi and Nassen (1999) explore measurement error in marketing research data. For instance, questionnaire items or rating scales that are used to measure perceptions, beliefs, attitudes, judgements, or other theoretical constructs are likely to reflect measurement error because of the absence of physical measures corresponding to these variables. Besides, marketing research data may be subject to method errors like halo effects, interviewer effects, or social desirability distortions. Their findings suggest that measurement error in marketing data may be large and needs to be corrected for in empirical applications to improve decision making and inferences.

We illustrate the problem of measurement error in the simple bivariate case. Consider

\[
y_i = \beta_0 + \beta_1 x_i + \epsilon_i,
\]

where \( x_i \) is the ‘true’ unobserved construct. Instead, \( x_i \) is observed and \( x_i = \chi_i + \nu_i \), with \( E(\epsilon_i) = E(\nu_i) = 0, E(\epsilon_i^2) = \sigma_\epsilon^2 > 0, E(\nu_i^2) = \sigma_\nu^2 > 0, \) and \( E(\epsilon_i \nu_i) = E(\chi_i \epsilon_i) = E(\chi_i \nu_i) = 0. \) These two equations can be combined, giving \( y_i = \beta_0 + \beta_1 x_i + \tilde{\epsilon}_i, \) with \( \tilde{\epsilon}_i = \epsilon_i - \beta \nu_i. \) The OLS estimator for \( \beta_1 \) is biased towards zero as \( E(\tilde{\epsilon}_i x_i) = -\beta_1 \sigma_\nu^2 \neq 0, \) which implies that \( E(\tilde{\epsilon}_i | x_i) \neq 0. \) For more details, see e.g. Plat (1988), Wansbeek and Meijer (2000), or Carroll, Ruppert and Stefanski (1995).

\( ^4 \)A problem that arises in data collection when there is carry-over from one judgement to another (source: www.marketingpower.com).
2.1.3 Self-selection

The problem of self-selection arises when individuals tend to select themselves in a certain state, like union vs. non-union member (Vella and Verbeek, 1998), or treated vs. not treated (Angrist, Imbens and Rubin, 1996), on the basis of economic, or other, usually unknown, arguments. For instance, Angrist (1990) considers the effect of Vietnam veteran status on civilian income to investigate whether these veterans ought to be compensated by the US government for their possible loss of personal income caused by serving the army. However, civilian earnings are not easily compared by Vietnam veteran status simply because certain individuals with fewer civilian opportunities are more likely to enlist than others, and such individuals would have earned less income regardless of serving the army.

Hamilton and Nickerson (2003) give an overview of endogenous decision-making in strategic management, where managers often make strategic organizational choices between several competing strategies not ‘randomly’ but based on expectations and experience. Similarly, data collected on the internet may suffer from self-selection. Certain individuals are more likely to be on the internet and are therefore more likely to fill-in the web-survey, to click on the web pages or to purchase products online. If these unobserved individual characteristics also influence web behavior, preferences or perceptions\(^5\), then part of the effect of these latent characteristics is falsely attributed to internet usage. One could argue that these individuals would have reacted differently regardless of their frequency of being on the internet. These issues are important, for instance, when investigating purchase quantities decisions in online stores versus brick-and-mortar (“offline”) stores, whether or not to buy a certain product category, or whether or not to buy a certain brand, given category and shopping environment.

To illustrate, a simple self-selection model is given by

\(^5\)Or our phenomenon under study.
where I and II denote a certain state (e.g. treated vs. non treated, or web-user versus non web-user). More compactly,

\[ y_i = x'_i \beta + d_i x'_i \delta + \epsilon_i, \]

with \( d_i = 1 \) if \( i \in I \) and \( d_i = 0 \) otherwise. From this representation it can be seen that \( d_i \) is a dummy regressor and standard estimation fails when \( E(\epsilon_i | d_i) \neq 0 \). This assumption is possibly violated for the examples given in the previous paragraph. For more details on self-selection problems, see e.g. Vella (1998), Wooldridge (2002), or Bowden and Turkington (1984).

### 2.1.4 The simultaneous equation model

Ordinary (or hierarchical) regression analysis will not be appropriate when the right-hand variables are simultaneously determined along with the dependent variables. However, it is often hard to rule out such feedback loops. Examples are an economic agent making choices regarding education or labor market participation (Card, 1999, 2001) or the price setting behavior of firms while interacting with competition. Several studies consider simultaneity in prices and demand for markets with differentiated products, given a structure for competition. The price-setting behavior of firms due to e.g. unobserved product characteristics like coupon availability, national advertising, shelf-space (al)location, and other retail environment characteristics, or competitor’s (re)actions causes endogeneity. Berry’s (1994) work in dealing with price endogeneity in aggregated models while using instrumental variables has been widely applied and adapted. For instance, Nevo (2001) estimates a structural demand-supply model for the ready-to-eat cereal industry; Besanko, Gupta and Jain (1998) consider a scanner-data application for the two categories yoghurt and catsup; Berry, Levinsohn and Pakes (1995) and Sudhir (2001) develop a market equilibrium model with competitive pricing for an automobile market to investigate automobile pricing and competition. Using a more simple model for demand, Gasmi, Laffont and Vuong (1992) model collusive behavior on
price and advertising in a soft-drink market.

A simple supply and demand model for a product or good is given by

\[ y_d^t = (x_d^t)' \beta_d + \gamma^d p_t + \epsilon_d^t \]
\[ y_s^t = (x_s^t)' \beta_s + \gamma^s p_t + \epsilon_s^t, \]

where variables in \( x_d^t \) are factors that affect the demand or behavior of consumers, whereas the variables \( x_s^t \) only influence the behavior of producers. The price \( p_t \) is determined such that \( y_d^t = y_s^t = y^t \). When the demand equation \( y_d^t = (x_d^t)' \beta + \gamma^d p_t + \epsilon_d^t \) is estimated, it cannot be assumed that \( E(\epsilon_d^t | p_t) = 0 \), because price is simultaneously determined with the demanded quantity, i.e. unobserved positive shocks in demand or competitor (re)actions shift the demand curve upward, implying a higher equilibrium price (ceteris paribus) (van der Ploeg, 1997, or Asher, 1983). In this case, OLS cannot be used to estimate the parameters of the demand equation. For more technical details, see e.g. Judge et al. (1985) or Davidson and MacKinnon (1993).

### 2.1.5 Lagged dependent variables

The presence of lagged dependent variables in the set of regressors violates the exogeneity assumption when serial correlation is present. It is well known that OLS estimation should not be used see, for instance, White (2001). Consider

\[ y_t = x_t' \beta_1 + y_{t-1} \beta_2 + \epsilon_t \]
\[ \epsilon_t = \phi \epsilon_{t-1} + v_t, \]

where e.g. \( y_t \) are the sales at time \( t \), \( x_t \) are promotional activities at time \( t \), and \( y_{t-1} \) is included to represent lagged effects of promotional activities held in the past. Suppose that the \( v_t \) are i.i.d., \( |\phi| < 1, |\beta_2| < 1 \), \( E(v_t) = 0 \), and \( v_t \) is independent of \( y_t \) and \( x_t \), and assume that all second order moments exist. Now \( \epsilon_t y_{t-1} = \phi \epsilon_{t-1} y_{t-1} + v_t y_{t-1} \), so that \( E(\epsilon_t y_{t-1}) = \phi E(\epsilon_{t-1} y_{t-1}) \).
Furthermore, \( E(\epsilon_t y_t) = x_t' \beta_1 E(\epsilon_t) + \beta_2 E(y_{t-1} \epsilon_t) + \text{var}(\epsilon_t) \). By using the stationarity of \( \epsilon_t \) it follows that

\[
E(y_{t-1} \epsilon_t) = \frac{\phi}{1 - \phi \beta_2} \text{var}(\epsilon_t),
\]

and \( E(\epsilon_t | y_{t-1}) \neq 0 \), unless \( \phi = 0 \). Davidson and MacKinnon (1993) (p. 681) make a stronger statement and argue that the OLS estimator is biased in all models when lagged dependent variables are present (yet consistent when \( \phi = 0 \)).

In certain situations explanatory variables may ‘act’ as lagged dependent variables, which can easily be overseen. This is illustrated by Gönlü, Kim and Shi (2000), who examine the effect of sending out catalogues on the probability to buy products from that catalogue. The mailing variable and other customer shaped promotional activities, are often functions of passed sales, which implicitly introduces problems of the nature described above.

### 2.1.6 Bias in OLS when \( E(\epsilon | X) \neq 0 \)

From the preceding subsections it can be concluded that regressor-error dependencies may exist for many different applications. It follows immediately that the OLS estimator, given by \( \hat{\beta}_{n,\text{OLS}} = \beta + (X'X)^{-1}X'\epsilon \), where \( E(\epsilon) = 0 \), is biased when \( E(\epsilon | X) \neq 0 \), and it loses its attractiveness as an estimator. Similarly, in absence of heteroscedasticity and autocorrelation, the usual –for degrees of freedom corrected– estimator for the error variance that is based on the OLS residuals, is unbiased when \( E(\epsilon | X) = 0 \), see e.g. Verbeek (2000) (p. 19). Otherwise it can be expected that the true value is underestimated, since, on average, conditioning reduces the variance of the variable subject to the conditioning (cf. Greene, 2000) (p. 81).

Unfortunately, the bias in the OLS estimates does not reduce when the sample size gets larger. More specifically, the OLS estimates are inconsistent, and \( \text{plim}(\hat{\beta}_{n,\text{OLS}}) \neq \beta \) and \( \text{plim}(\hat{\sigma}_{n,\text{OLS}}^2) < \sigma^2 \), but this inconsistency can be reduced,
at least in large samples, by using instrumental variables (White, 2001, or Ferguson, 1996). The instrumental variables (IV) approach is discussed next.

### 2.2 The IV approach

The instrumental variable (IV) method assumes that a set of variables $Z$, called instrumental variables, is available. These instruments should be uncorrelated with the error term $\epsilon$, i.e. $E(\epsilon|Z) = 0$, and explain part of the variability in the endogenous regressors. This implies that the instruments $Z$ cannot have a direct effect on $y$ (the instruments $Z$ are ‘exogenous’). The standard IV regression model is obtained by augmenting the standard linear regression model with a model for the endogenous regressors and the instruments, namely

$$
\begin{align*}
y &= X\beta + \epsilon \\
X &= Z\Pi + V
\end{align*}
$$

where $y$, $X$, and $\beta$ are defined as before, $Z$ is an $n \times q$ matrix containing the instrumental variables, and $V$ is an $n \times k$ matrix containing the error terms. The matrix $\Pi$ represents the effect of the instruments on the endogenous regressors. The exogenous variables in $X$ are assumed to appear in $Z$ as well and should not be omitted (Wooldridge, 2002). It is assumed for identifiability that $q \geq k$ and rank $Z = q < n$. The correlation between $X$ and $\epsilon$, i.e. the degree of endogeneity, arises because of nonzero covariances between $\epsilon$ and $V$. The errors are assumed to have mean zero. It can be seen from (2.4) that the endogenous regressors are ‘split’ into an exogenous part and an endogenous part. This IV model is a special case of a simultaneous equation model (SEM), which is well-known in econometrics. The most common estimators for $\beta$ are the 2SLS estimator (or a method of moments estimator) and the limited information maximum likelihood (LIML) estimator, which is in fact the maximum likelihood estimator of (2.4). 2SLS is most frequently used because of its availability in many standard computing packages.

Once instruments are available, the IV estimator is given by
\[ \hat{\beta}_{IV}^n = (X'P_ZX)^{-1}X'P_Zy, \quad (2.5) \]

where \( P_Z = Z(Z'Z)^{-1}Z' \), and is consistent and approximately normally distributed for large \( n \) when (i) plim \( (1/n)Z'\epsilon = 0 \), and (ii) both plim \( (1/n)Z'Z \) and plim \( (1/n)Z'X \) exist and have full column rank. Unbiasedness of the IV estimator is discussed in the next subsection. One often relies on large-sample analysis in examining this estimator because its expected value does not exist when the number of instruments equals the number of explanatory variables (cf. Wooldridge, 2002, p.101). Standard inferential procedures can be employed to learn about the model parameters or to test hypotheses (Bowden and Turkington, 1984, or White, 2001). The maximum likelihood (LIML) estimator can be computed with a little more effort and, provided that the instruments are not too weak, the asymptotic properties of the 2SLS and LIML estimator are the same (Davidson and MacKinnon, 1993, van der Ploeg, 1997, Kleibergen and Zivot, 2003).

### 2.2.1 Considerations when using Instrumental Variables

The problem in empirical applications is how or where to find ‘valid’ instruments. In general, there are no clear guidelines, and instruments may not be easy to obtain. Besides, it can be very expensive to obtain additional data. As such, instruments are often chosen by ad hoc arguments or even by availability, resulting in potential invalid instruments. The condition \( \mathbb{E}(\epsilon|Z) = 0 \) requires that there is no direct association between the instruments and the dependent variable, which is debatable in many empirical situations.

Wooldridge (2002, p.88), for instance, discusses the (in)validity of the draft lottery number instrument used in Angrist (1990) to estimate the effect of Vietnam veteran status on personal income. Although the draft lottery number appears to be random, individuals who are more likely to get drafted may chose to obtain more education to increase the chance of obtaining a draft postponement or employers may be more willing to invest in educating and training individuals who are unlikely to be drafted. Bound, Jaeger and Baker (1995)
question the exogeneity of the quarter of birth instruments used by Angrist and Krueger (1991) who estimate the effect of schooling on income. They present evidence that a weak correlation between quarter of birth and wages, independently of the effect of quarter of birth on education, exists that is sufficiently strong to have an effect on the IV results. Card (1999, 2001) provides more extensive summaries of debates on the validity of family background variables, like parental education, and institutional features of the schooling system variables, like the presence of a nearby college, as instruments for the endogenous regressor schooling, see also chapter 5. In estimating demand, lagged prices or promotional variables are often used as instruments in marketing response models, but these are not valid, for instance, when reference prices exist\(^6\) and are historically formed (cf. Bronnenberg and Mahajan, 2001). Yang, Chen and Allenby (2003) note that lagged prices may not be appropriate due to reasons as forward buying and stockpiling. Besides, treating lagged variables as ‘exogenous’ is a potential source of endogeneity itself (see also Arellano, 2002 (p.455)). Nevo (2001) used price data from other markets as instrumental variables for price, but notes that these instruments are invalid when common (national) demand shocks occur, or when advertising or promotion activities are coordinated across markets. This is more likely when the same manufacturer or retailer is active in several markets. Although cost drivers may be potential instrumental variables for price, Nevo (2000) (p. 546) concludes that these are rarely observed, while proxies for cost usually do not exhibit sufficient variation.

Exogeneity of instruments is only one of the two criteria for an instrument to be valid, in addition, available instruments may be weak in the sense that they are poorly correlated with the endogenous regressors. Stock, Wright and Yogo (2002) state: “Empirical researchers often confront weak instruments. Finding exogenous instruments is hard work, and the features that make an IV

\(^6\)The reference price is the ‘expected price’ of a product. Several studies have found asymmetric effects when the perceived price differs from the reference price. The effect of the reference price on demand depends (among other things) on the convenience during the buying process, on the familiarity of the brand, and on the type of store the product is bought (Leeffang, 1994) (in Dutch).
plausible exogenous [...] can also work to make the instrument weak”. Unfortunately, statistical properties of IV estimators and inferential procedures based on these, turn out to be sensitive to the choice and validity of the instruments, even for large sample sizes. Consequently, researchers who study the same substantive question but use different instruments may end up with another conclusion. In the following we will review some recent results on the problem of weak instruments that appeared in the econometric literature. Most of the following discussion is developed for the linear (non-hierarchical) regression model for cross sectional studies, see Stock, Wright and Yogo (2002), and Hahn and Hausman (2003) for more details7.

**Weak instruments**

Recent results in the econometric literature has shown that the presence of weak instruments does not only reduce the precision of the estimates, but may also lead to biased and inconsistent estimates that are potentially larger than OLS. Furthermore, standard asymptotic approximations break down (Staiger and Stock, 1997, Bound, Jaeger and Baker, 1995, Hahn and Hausman, 2002 or Kleibergen and Zivot, 2003). As a consequence, standard hypothesis tests and confidence intervals are unreliable. Weak instruments may arise when the instruments do not have a high degree of explanatory power for the endogenous regressors or when the number of instruments is large (cf. Hahn and Hausman, 2002, 2003). In the following we discuss three potential pitfalls with IV estimation in the presence of weak instruments: (1) the finite sample bias of 2SLS, (2) situations where the instruments are potentially correlated with \( \epsilon \), i.e. they are not exogenous, and (3) the poor asymptotic approximation to the sampling distribution of IV estimators.

In finite samples the IV estimator (2SLS) is biased in the same direction as OLS. This fact is often unnoted in empirical studies. Even when \( E(\epsilon|Z) = 0 \),

---

7In a survey article on Sargan’s work on instrumental variables estimation, Arellano (2002) observes that “Many of the themes [on instrumental variables estimation] that appeared [...] in the econometrics literature of the 1980s and 1990s were presented in a surprisingly mature way in Sargan’s 1958 and 1959 articles”. 
\[ \hat{\beta}^{IV}_n = \beta + (X'P_Z X)^{-1} X' P_Z \epsilon \]
is, in general, biased as \( E (X' P_Z X)^{-1} X' P_Z \epsilon \neq 0 \).
This bias arises because coefficients \( \Pi \) in (2.4) are not observed. If we had observed \( Z \Pi \), an OLS regression of \( y \) on \( Z \Pi \) would be unbiased. Instead, an estimate of \( \Pi \) has to be obtained from a regression of \( X \) on \( Z \). Hahn and Hausman (2002, 2003), Buse (1992), Stock, Wright and Yogo (2002), or Bound, Jaeger and Baker (1995) show that this finite sample bias\(^8\) is a function of (among other things) the number of instruments, which suggests that augmenting the set of instruments increases the bias in the estimator. However, as Buse (1992) shows, the bias will only be proportionally larger when the number of instruments grows faster than the rate of explained variance of the endogenous regressors. As a consequence, adding important or strong instruments does not necessarily increase the bias, however, adding less important instruments, or having weak instruments, will undoubtedly lead to more biased results. Bound, Jaeger and Baker (1995) and Hahn and Hausman (2003) show that the bias is inversely related to the \( F \)-statistic (the Fisher Statistic) of the regression of the endogenous explanatory variable on the instruments. These results suggest that the (partial) \( R^2 \) and \( F \)-statistic of the first stage regression (i.e. the regression of \( X \) on \( Z \) in (2.4)) are useful as rough guides to the quality of IV estimates and should routinely be reported (cf. Bound, Jaeger and Baker, 1995). The LIML estimator is known to have no finite moments and has thicker tails. As such, it is generally less sensitive to the addition of superfluous instruments (cf. Kleibergen and Zivot, 2003). Nevertheless, when the IVs are weak, even LIML may not solve the problem (cf. Hahn and Hausman, 2003).

A second problem associated with weak instruments is the inconsistency of the IV estimator relative to OLS when the instrument is potentially correlated with \( \epsilon \), i.e. it is endogenous itself. Bound, Jaeger and Baker (1995) show that the relative inconsistency of IV to OLS is equal to (for simplicity it is assumed that \( k = q = 1 \))

\(^8\)They find that for one endogenous regressor, the expectation does not exist when only one instrument is available. See also Wooldridge (2002) who states that the number of moments that exists is one less than the number of overidentifying restrictions.
2.2 The IV approach

\[
\frac{\text{plim } \hat{\beta}_n^{\text{IV}} - \beta}{\text{plim } \hat{\beta}_n^{\text{OLS}} - \beta} = \frac{\rho_{z,\epsilon}}{\rho_{x,\epsilon}} \frac{\rho_{x,z}}{\rho_{x,z}}
\]

where \( \rho_{x,z} \) indicates the correlation between \( x \) and \( z \), the other terms being defined similarly. When the instrument is weak, \( \rho_{x,z} \to 0 \), implying that even a small correlation between the \( z \) and \( \epsilon \) can produce a large relative inconsistency in the IV estimator, making the inconsistency of IV potentially larger than in OLS\(^9\).

Thirdly, if the instruments are weak, then, even in large samples, classical (first-order) asymptotic approximations are poor. This is illustrated by (among others) Nelson and Startz (1990). As they note, conventional wisdom suggests that when the instruments are weak, the classical asymptotic variance matrix will be large and the asymptotic distribution of \( \beta \) is dispersed. However, it is also shown that the asymptotic distribution is a very poor approximation to the exact finite density function (which is bimodal, fat tailed and concentrated closer to the probability limit of least squares than the true value). If the asymptotic variance of \( \hat{\beta}_n^{\text{IV}} \) decreases, i.e. when the instruments are generally stronger, the classical approximation becomes better. As a consequence, with weak instruments inferential procedures based on classical asymptotic results are unreliable. Although finite sample methods could be used in these situations, their use in practice is limited due to restrictive assumptions, computationally intractable distributions, or the absence of a clear framework for testing or constructing confidence intervals. The weak instruments problem is not only relevant for “small samples” and it cannot be ignored in large samples. This is illustrated by Bound, Jaeger and Baker (1995), who show that for the Angrist and Krueger (1991) study it is possible to obtain similar results if artificial random (dummy) instrumental variables are used, despite the sample size of 329500 observations.

\(^9\)See also Hahn and Hausman (2003), section V.
Examining instrument validity

As (asymptotic) properties of IV estimators are sensitive to the choice of valid instruments, regardless of the sample size, measures for ‘weakness’ are desirable. Recently, the outcome of several studies suggests to report $F$-statistics and $R^2$ measures of the first stage regression routinely. Stock, Wright and Yogo (2003), for instance, suggest that the first-stage $F$-statistics must be larger than 10 for 2SLS inference to be reliable. Furthermore, Bowden and Turkington (1984) argue that one should find instruments that maximize all of the canonical correlations with $X$. Staiger and Stock (1997) develop a data-based measure for the relative bias, where large values should alert the researcher to potential problems of correlations between the instruments and the random components. Bowden and Turkington (1984) and Verbeek (2000) (among others) present a test for instrument admissibility when $q > k$ (overidentified). If the test rejects, there is sample evidence against the joint validity of the instruments, although it is not possible to determine which one is incorrect. The method in Bowden and Turkington can be used to examine whether an additional set of instruments is admissible, but this test does not address potential weakness of the instruments. In fact, Hahn and Hausman (2003) argue that this test rejects too often when weak instruments are present, which is a major drawback since it is often used to test economic theory embodied in the model.

Hahn and Hausman (2002) have recently developed a test for the validity of instrumental variables, which jointly addresses exogeneity and strength. It is based on the general Hausman specification test approach (Hausman, 1978) and adopts the second order asymptotic approximations of Bekker (1994). The idea is to compare forward and backward 2SLS estimators, which are shown to be equivalent under the null hypothesis that conventional asymptotics is valid. The test statistic is fairly simple to compute and is shown to have a $t$ distribution under the null hypothesis. Rejection of the null hypothesis might indicate a failure of the orthogonality assumption of the instruments or that the instruments could be weak. Hahn and Hausman (2002) suggest a two step approach based on this test to decide whether 2SLS, LIML, or none should be used.
In chapter 4 we propose another method that can be used to investigate the validity of observed instruments, which is based on the LIV model and existing test principles. Contrary to the Hahn and Hausman test, it can be used to separately investigate either instrument weakness or instrument endogeneity, or both. Furthermore, if the instruments are found to be invalid, the estimates for the regression parameters can still be used because the LIV results do not rely on the quality, nor require access to observed instruments. Our simulation evidence suggests that this approach does not yield size problems in presence of weak instruments, as opposed to the classical test of overidentifying restrictions.

Choosing the (number of) instruments

The finite sample bias in IV estimators is a function of the number of instruments, which suggests that one should not include too many, although identification requires that at least as many instruments as endogenous regressors are included \((q \geq k)\). Furthermore, increasing the number of instrumental variables results in a loss of degrees of freedom and the first stage regression \((X\text{ on } Z)\) suffers from overfitting. Sargan (1958) concludes that “if the first few instrumental variables are well chosen, there is usually no improvement, and even a deterioration, in the confidence regions as the number of instrumental variables is increased beyond three of four”. Besides, similar to the results presented above, he also notes that “estimates [may] have large biases if the number of instrumental variables becomes too large” (p.400). As opposed to these finite sample results, large sample theory, however, shows that an IV estimator with one more instrument is at least as efficient, which suggests that we can add as many instruments as we please without doing worse (see e.g. Davidson and MacKinnon, 1993 (p.220-p.224)).

Bowden and Turkington (1984) suggest to perform a principal components analysis on \(Z'Z\) and to choose the first \(p\) principal components as instruments. This approach, however, does not address the correlation of \(X\) and \(Z\), i.e. the strength of the instruments. Donald and Newey (2001) developed a mean-
squared error criterion that can be minimized to choose a set of instrumental variables. They find that this method of choosing instruments generally yields an improvement in performance. In the leading cases, LIML outperforms 2SLS, although they find that 2SLS performs better in situations of little endogeneity. For the weak instruments case, there is a clear tendency to use fewer instruments.

**Testing regressor-disturbance problems**

Given the potential pitfalls when using IV results and the problem of finding instruments at all, one would like to test for potential regressor error correlation a priori. Unfortunately, it is not possible to examine $X'\epsilon$ directly, as $\epsilon$ is unobserved and OLS estimation yields $X'\hat{\epsilon} = 0$ by definition$^{10}$. In order to test for endogeneity valid instrumental variables are required. A test based on the general test procedure of Hausman (1978) can then be used. This test is based on comparing the difference between $\hat{\beta}_{n}^{\text{OLS}}$ and $\hat{\beta}_{n}^{\text{IV}}$, and Hausman proposed a test-statistic that has approximately a $\chi^2$ distribution under the null hypothesis. A drawback of this test procedure is that external instruments have to be available in order to compute $\hat{\beta}_{n}^{\text{IV}}$. As a consequence, the researcher may conclude that the obtained instruments were not needed after all. Furthermore, this test is potentially sensitive to weak instruments (see e.g. Staiger and Stock, 1997, or Bowden and Turkington, 1984, for more details). In fact, the Hausman test may incorrectly fail to reject the use of the OLS estimator because of the bias (cf. Hahn and Hausman, 2003). In chapter 3 we propose an instrument-free test, that solves this circular problem. We show that this test has a reasonable power over a wide variety of settings.

**2.2.2 IV based solutions to the weak instrument problem**

Hahn and Hausman (2003), and Stock, Wright and Yogo (2002) surveyed most of the econometric literature on solutions to the weak instruments problem in empirical applications. In the following we present a brief summary, since

$^{10}$An exception is testing for $X'\alpha = 0$ in random intercept models, where $\alpha = (\alpha_1, ..., \alpha_n)'$ are the unit-specific random intercepts, as a test statistic is readily available (chapter 6).
most of the technicalities and the amount of results that have appeared in the literature, are beyond the scope of this thesis.

As mentioned before, first-order asymptotic approximations are poor in the presence of weak instruments. Several studies have presented improved asymptotic approximations to finite-sample distributions in this situation. Staiger and Stock (1997) developed an alternative asymptotic framework that models the coefficients of the first stage regression as locally zero, i.e. weakly correlated without assuming normality. In this framework they showed that if the instruments are weak, the 2SLS and LIML estimators have nonstandard asymptotic distributions and are not consistent, where the bias is less problematic for LIML than for 2SLS, particularly in small samples. Furthermore, results on properties of various inferential procedures (like \( t \) test, coverage rates of confidence intervals and tests of overidentifying restrictions) are obtained. Bekker (1994) developed an asymptotic approximation for models with normal errors in which both the number of instruments and the sample size increases. Simulation evidence shows that these asymptotics provide good approximations for moderate and large values of the number of instruments, and that LIML is to be preferred over the standard IV estimator. However, Bekker’s results apply only to normal cases and do not capture the nonnormality observed in the exact finite sample density, see also Staiger and Stock (1997).

Besides work on finding better alternatives to the first-order asymptotics, several fully robust hypothesis tests and methods are developed to construct confidence sets for \( \beta \) that have approximately the correct size and coverage rates under weak instruments. One such robust test to investigate \( \beta = \beta_0 \) is the Anderson-Rubin statistic (Anderson and Rubin, 1949), which is not affected by the degree of underidentification. However, it may lack power because of a loss of degrees of freedom when the number of instruments is larger. The \( K \) statistic (Kleibergen, 2002) has similar asymptotic properties with a minimal number of degrees of freedom. Bekker and Kleibergen (2003) investigate the finite-sample distribution under normality. Other tests have been proposed as well, see e.g. Staiger and Stock (1997). Stock, Wright and Yogo (2002)
present results of power comparisons for several tests under different conditions. Given the duality of hypothesis tests and the construction of confidence sets, the robust tests can be used to obtain confidence intervals. When the instruments are weak, these sets can have infinite volume, indicating that there simply is limited information to use in order to make inferences about $\beta$ (cf. Stock, Wright and Yogo, 2002).

The previous methods to carry out tests or construct confidence intervals do not readily provide point estimates for $\beta$. In addition, they may be difficult to compute. Several alternatives to 2SLS are proposed, that ought to be more robust and reliable if the instruments are weak. Second-order unbiased estimates, such as LIML or Nagar estimators, are often suggested as robust alternatives. These estimators, however, do not have finite sample moments which may present a problem in empirical situations (Hahn and Hausman, 2003). Other alternatives are Jackknife Instrumental Variables (Angrist, Imbens and Krueger, 1999), Fuller-$k$ Estimator (Fuller, 1977), or bias-adjusted 2SLS (Donald and Newey, 2001). Stock, Wright and Yogo (2002) find that these partially robust estimators provide relatively reliable alternatives to 2SLS in applications with weak instruments. However, Hahn and Hausman (2003) recommend, based on Monte Carlo evidence, extreme caution using “no moment” estimators (LIML or Nagar). Considering mean-squared error and IQR measures, they conclude that 2SLS, jackknife 2SLS, and Fuller-based estimators perform best, and state that “instrument pessimism seems overstated for 2SLS, which may be why 2SLS often performs better than expected in terms of MSE in the weak instrument situation”. The specification test suggested by Hahn and Hausman (2002) may be used to decide among the alternatives. Both Stock, Wright and Yogo (2002) and Hahn and Hausman (2003) stress that most of the analysis in the weak instruments literature is conditional on instrument exogeneity. Failure of the exogeneity restriction, in particular in combination with weak instruments, leads to additional complications and situations in which OLS may do better than the above suggested remedies against weak instruments (see also section 4.4).
2.3 Alternative approaches to solve for regressor-error dependencies

In some applications the nature of the data generating process or the suspected cause of endogeneity itself suggests suitable instruments or even a different estimation approach. Wooldridge (2002) suggests three other solutions to solve omitted variable problems, including the proxy-variable OLS method (p.63) and using indicators of the unobservables (p.105-p.107) that require IV estimation\footnote{The indicator IV solution is different from the classical IV solution discussed previously. The indicator IV solution assumes the existence of a possible mismeasured proxy for the missing variable \( w \), that needs to be instrumented, whereas the classical IV solution leaves the omitted variable \( w \) in the error term and all elements of \( x \) correlated with \( w \) need to be instrumented. See also Petrin and Train’s (2002) control function approach and the discussion in Chintagunta, Dubé and Goh (2004) (p.6).}, where the latter method also applies to measurement error models. Furthermore, observing the same cross-sectional units over time, and applying fixed effects estimation could also eliminate endogeneity due to omitted variables, if the endogeneity arises from time-invariant sources, see also chapter 6, or the example given by Verbeek (2000) (p.312). Card (1999) presents an overview of studies using sibling and twin data to estimate the return to education and argues that omitted ability is eliminated when computing within-family estimators. Stern (2004) uses data composed of multiple job offers to postdoctoral students and a fixed-effects approach to estimate the relation between wages and the scientific orientation of organizations. His results suggest a negative relation between science and income, that is biased upward when unobserved quality of researchers is not controlled for. For measurement error models, autoregressive models, and simultaneous equation models the data generating process may suggest suitable instruments. It is beyond the scope of this thesis to review all the literature on these topics. For measurement error models we refer to e.g. Wansbeek and Meijer (2000), Carroll, Ruppert and Stefanski (1995), or Bowden and Turkington (1984) for extensive overviews. These models can be estimated using IV techniques, for instance by using other (potentially) mismeasured variables (White, 2001). Another method is based on Wald (1940), that assumes that the observations can be divided into groups.
This classification should be independent of the error terms and should discriminate between high and low values of the unobservable true construct (see also Madansky, 1959, or chapter 3). Similarly, higher order lags may serve as instruments for a model that has lagged dependent variables as regressors in the presence of serial correlation. In simultaneous equation models exogenous variables that are not included in the equation of interest can often serve as instruments and are readily available (see e.g. Greene, 2000).

In the following we briefly consider three other interesting methods to solve for endogeneity that have recently appeared in the literature: (1) Lewbel’s method, (2) methods that model demand, cost and competition, and (3) spatial econometrics.

**Lewbel’s method.** Lewbel (1997) showed that for measurement error models instruments can be constructed from available data by exploiting higher order moments. Hence, observed exogenous instrumental variables are not required. Erickson and Whited (2002) extend this method and propose a two-step generalized method of moments estimator for a multiple mismeasured regressor errors-in-variables model. Consistent estimation requires, among other things, that measurement and equation errors are independent and have moments of every order, but no assumptions have to be made about distributional forms. Hence, information contained in third- and higher order moments of the data are fully exploited to identify the regression parameters.

This interesting approach is developed for measurement error applications, but may be applicable to more general regressor-error dependency models as well. In appendix 6C we show for a simple linear multilevel model how these ideas can potentially be extended to more ‘general’ endogeneity applications. Depending on the empirical situation, it may provide an easy way to construct instruments from the available data and, hence, deserves more attention.

**Methods that model demand, cost and competition.** Several studies have attempted to solve for ‘price endogeneity’ in markets with differentiated prod-
ucts. Price is endogenously determined by the interaction of demand and supply. The idea is to solve this form of endogeneity by jointly modeling demand and supply equations, using a profit maximization model. Berry (1994) and Berry, Levinsohn and Pakes (1995) develop a market equilibrium model, based on a logit demand function, that is adapted to make it suitable for traditional instrumental variables estimation. This method is both applicable to aggregate or disaggregate data, or a combination of both. The resulting system is obtained by aggregating a discrete choice model of individual consumer behavior, which is combined with a cost function. These two models are embedded in a system of price setting firms in differentiated markets. Joint estimation leads to potentially more efficient estimates, than to focus on the demand side only with instruments for price. Furthermore, the system provides detailed information on cost structures and the nature of competition. Using equilibrium models, however, imposes more demand on data and incorrect specification of the firm’s behavior could lead to biased estimates. This approach has widely been applied and adapted, with differences in e.g. data-aggregation, type of heterogeneity, or method of estimation, see, for instance, Besanko, Gupta and Jain (1998), Besanko, Dubé and Gupta (2000), Nevo (2001), or Sudhir (2001).

Most studies employ an instrumental-variables based simultaneous equations estimation procedure, which is a generalized method of moments estimator. However, as opposed to homogenous goods models, in differentiated markets most of the exogenous variables in the model are product characteristics affecting both cost and demand. Traditional exclusion restrictions therefore cannot be used to form instruments. Sudhir (2001) (section 4.2) and Nevo (2001) (section 4.3), for instance, discuss this in more detail and report having instruments of potential poor quality, see also the discussion in Berry (2003) (section 1).

Recently, Draganska and Jain (2004) proposed a new maximum likelihood based method for simultaneous estimation of supply-demand. Their proposed algorithm uses individual level-data for a heterogenous demand model, to-

\[12\text{See e.g. discussion in Yang, Chen and Allenby (2003) (section 2.3) or Dubé and Chintagunta (2003).} \]
together with a supply equation derived from profit maximization behavior of firms, assuming a Bertrand-Nash equilibrium. The resulting likelihood equation cannot be maximized straightforwardly because the equilibrium model is highly nonlinear. Their estimation procedure is based on simulating prices and choice probabilities to solve for the market equilibrium. The obtained smoothed empirical distribution can be used for maximization. On the other hand, Yang, Chen and Allenby (2003) (with discussions) proposed a Bayesian approach to estimate a simultaneous heterogenous demand and supply model. The method incorporates consumer heterogeneity and allows for a wide variety of supply model specifications. The advantage of their approach is that it can handle non-linear model structures and allows for exact small sample inference. See also Chintagunta, Dubé and Goh (2004) for a recent overview and discussion.

**Spatial econometrics.** Recently, two studies in marketing appeared, that solve endogeneity of marketing mix variables using spatial dependencies in observed market data, where no or limited time variation is present. These dependencies are caused by the fact that economic agents are spatially organized, or have similar store profiles. Bronnenberg and Mahajan (2001) identify correlations between marketing mix variables and the error term by imposing a measurable spatial structure on the random terms in the model. This spatial map is a consequence of unobserved actions of retailers that are faced by trade territories consisting of multiple neighboring markets. Bronnenberg and Mahajan construct a spatial map by making use of geographic proximities. By accounting for this space in an econometric model, it is possible to correct and test for the effect of unobserved retailer’s behavior. Their results for Mexican food items suggest that unobserved components of the dependent variables are related to the marketing mix variables. Van Dijk et al. (2004) consider the estimation of shelf-space elasticities based on endogenous shelf space data. Estimation of shelf-space elasticities is hampered due to minimal (time) variation in shelf-space measures. The authors build on the work of Bronnenberg and Mahajan and propose to model the correlation between shelf space and the random terms by using a spatial structure based on similarities in store-, consumer-,
and competitor characteristics. Their results for frequently bought daily care products provide face valid shelf-space elasticities estimates that outperform a model with a spatial structure based on geographic proximities in terms of predictive validity. Since retailers generally decide about shelf space based on store, customers and competitor characteristics, it is expected that the similarity of two geographically similar stores in this case is lower than the similarity of two stores with similar profiles in distinct regions.

### 2.4 Conclusions and positioning of research

It is clear from this review that traditional instrumental variable methods, that rely on economic theory or intuition to find additional observable instruments, suffer from at least two problems: (i) in many situations no such variables are available, and (ii) once available, performance of the inferential procedures critically rely on the quality of these variables. In particular the latter has recently been the topic of several studies in econometrics. Although many important contributions to the weak instrument problem have been made, the problem of having potential endogenous instruments has not yet been solved (see e.g. concluding remarks of Hahn and Hausman, 2003, or Stock, Wright and Yogo, 2002). For most empirical researchers the question where to find suitable instruments is still open and usually there is not much choice when selecting instrumental variables. Without having valid instrumental variables at hand, classical instrumental variables estimation techniques cannot be relied on. Furthermore, there is a bit of a dilemma: theory suggests that the best choice of instruments are variables that are highly correlated with the endogenous regressors. However, the more highly correlated they are, the less defensible is the claim that these variables themselves are uncorrelated with the disturbances (cf. Greene, 2000, p. 375).

The latent instrumental variables (LIV) method proposed in the next chapter attempts to solve this circular problem. Similar to the classical IV model in (2.4), we assume that the endogenous regressor can be separated into an exogenous part and an endogenous part. However, we propose to model the exogenous
part as an unobserved discrete variable, which is a nuisance parameter. We prove that the model parameters are identified through the likelihood. Hence, observed instrumental variables are not required to estimate the regression parameters. In econometrics, instruments frequently take the form of categorical variables and, in addition, continuous instruments are often transformed into dummy variables (van de Ploeg, 1997, Bowden and Turkington, 1984, or Verbeek, 2000). We show that the parameters in the LIV model can be identified and estimated through maximum likelihood methods. As a by-product, ‘optimal’ LIV instruments are estimated from the data and regressor-error dependencies can be tested for straightforwardly without needing observed instruments at hand. Furthermore, the proposed likelihood framework allows for straightforward extensions to different applications.

The LIV approach has some similar features as two methods developed in the measurement error literature. Wald (1940) and Madansky (1959) assume that data is divided into two groups according to certain (statistical) criteria. Then a straight line can be fitted because it is determined by two points. If the grouping criteria are satisfied, the fitted line can be shown to be a consistent estimate of the true line. Randomly assigning the observations into two groups, for instance, or simply assigning the observations with high \( x \) values to one group and with low \( x \) values to the other group, does not provide valid groupings. Ideally, group construction should be based on some knowledge of the pattern of the underlying variation (cf. Bowden and Turkington, 1984). The LIV model does not require the existence of an a priori grouping of the data but estimates such a grouping simultaneously with the other parameters using mixture modeling techniques. Secondly, Lewbel’s idea to construct instruments from the available data and, hence, solving the circular problem of needing observed instrumental variables at hand, is similar to the motivation of the LIV model. Lewbel (1997), and Erickson and Whited (2002) propose method-of-moments based estimators and show that, under certain higher-order moment conditions, instrumental variables can be obtained from the available data. Hence, instruments are constructed based on ‘statistical’ moment conditions and the resulting variables will generally not correspond to
2.4 Conclusions and positioning of research

an economic theory or interpretation. Although these methods are developed for measurement error applications, we believe that they are more generally applicable, like the LIV model, although this requires further research (see appendix 6C, subsection 6.5.2, or subsection 8.2.2).

On the other hand, we propose a likelihood-based approach, which constitutes a very general framework that can be easily adapted to more general situations, for instance to the Bayesian setting in chapter 7. Furthermore, the predicted LIV instruments can be used to investigate the nature of the endogeneity more thoroughly, since these instruments are estimated from the available data rather than being constructed based on higher-order moment assumptions that may or may not be valid. The likelihood-approach has desirable optimality properties and can be expected to be more efficient than method-of-moments estimation. We agree with Yang, Chen and Allenby (2003) who state that “likelihood-based inference offers a distinct advantage over a method-of-moments approach because it makes precise statements about the probability of the observed data. In a likelihood-based analysis, the researcher is confronted with the correspondence between the model and the data, and cannot fit a model that is not supported by the data”. Besides, the LIV model belongs to the class of mixture models, that are often employed to estimate probability density functions, and mixture models can be seen as a flexible and robust approach to approximate them. Kim, Menzefricke and Feinberg (2004), for instance, provide evidence that mixtures of normals are a simple and effective way of density estimation, in particular in a Bayesian framework. Titterington, Smith and Makov (1985) find that finite mixtures (of normals) have often been used in robustness studies to investigate non-normal conditions in ‘normal’ inference, or to provide a procedure to reduce the influence of outlying observations. Hence, it is expected that some of these aspects translate to the LIV model, and we will show that the LIV results are relative insensitive to different choices of the shape of the distribution of the data and, hence, to the (non)existence of higher-order moments.

In the next chapter we introduce the simple LIV model which is further devel-
oped in subsequent chapters.