Statistics Netherlands

Discussion Paper

Explicit and Implicit Calibration of Covariance and Mean Structures

2014 | 09

Sander Scholtus

04 March 2014

Summary: Linear structural equation models (SEMs) are widely used to assess the measurement quality of survey variables – in particular, their validity and reliability. For applications where population means or totals are of interest, it is also important to assess whether the observed variables contain an intercept bias. Such applications arise frequently in official statistics. To estimate the intercepts of the observed variables, the SEM may be extended with a mean structure. Additional restrictions are needed to identify such a model. Unfortunately, the standard identification procedures define an arbitrary metric for the latent variables, which prevents the estimation of valid latent means and intercepts in a single population.

In the past, explicit calibration has been suggested as a method for identifying an SEM in a non-arbitrary way (Bielby, 1986b; Sobel and Arminger, 1986). This method requires the collection of 'gold standard' data for a subsample of the original sample (a so-called audit sample), which may be expensive or otherwise inconvenient in practice. In this paper, it is shown how the information in an audit sample may be used to estimate a non-arbitrary set of identification restrictions for an SEM. This implies in particular that, in a repeated survey, an audit sample collected in the first survey round may be re-used for calibration in later survey rounds (implicit calibration). This extension may improve the practical applicability of the calibration approach.

1 Introduction

The use of linear structural equation models (SEMs) to assess the measurement quality of survey variables is well-established; see, e.g., Andrews (1984), Saris and Andrews (1991), Scherpenzeel and Saris (1997), and Alwin (2007). These models can be seen as an extension of the classical test theory from psychology (Lord and Novick, 1968). Each observed variable is modeled as an imperfect measure of an underlying stable construct, which is a latent variable. Repeated measurements are needed to identify such a model. Some models allow a distinction to be made between validity (absence of systematic measurement error) and reliability (absence of random measurement error) (Saris and Andrews, 1991).

The above models are all based on covariance structures. Consequently, they cannot be used to exhibit a possible intercept bias in the observed variables. Although an intercept bias can be seen as a type of systematic measurement error, it is not usually included in the definition of measurement validity. Nevertheless, it can be important to assess the intercept bias of measurement instruments, at least in applications where the population means or totals of individual variables are of interest. Such applications occur frequently in official statistics.

In principle, an intercept bias could be assessed by an SEM that includes a mean structure. A bias would then be exhibited by a difference between the observed means and the corresponding latent means. However, a complication arises because the model needs to be identified by assigning a scale to each latent variable (and, if the model includes a mean structure, by fixing the origin of that latent scale). Two standard identification procedures involve using reference indicators or standardising the latent variables. Unfortunately, these procedures are not innocuous to the estimation of intercept bias. In the first case, it is tacitly assumed that the reference indicators have no intercept bias, which might not be true. In the second case, the intercept bias can be measured only on a standardised scale, which is not very useful.

More generally, the standard identification procedures for SEMs typically define an arbitrary metric for the latent variables (Bielby, 1986a). Consequently, within a single population, the absolute values of latent variable scores, means, and (co)variances do not have a substantive meaning. Comparisons of latent means, etc., between different groups *can* be valid in the presence of arbitrary metrics, provided that one may assume some form of measurement invariance to hold between the groups. For instance, if reference indicators are used, a necessary condition for valid between-group comparisons of latent means is that the factor loadings of the reference indicators are invariant across groups (Bielby, 1986a). Much has been written on measurement invariance and on the effect of arbitrary metrics on between-group comparisons; see, e.g., Byrne et al. (1989), Cheung and Rensvold (1999), and Whittaker (2013). The problem considered in this paper – i.e., making valid inferences about latent means and intercepts in a single group – is fundamentally different. Basically, this problem can be solved only by avoiding arbitrary metrics altogether.

In using terms like 'arbitrary metric' and 'substantive meaning', I am assuming that there is some theoretical 'gold standard' to which each observed variable can be compared. This gold standard represents the value that would be measured under ideal conditions. In practice, a gold standard can be posited for most variables, although its interpretation may vary depending on context. In some cases – e.g., *age*, *marital status*, *income*, and *voting behaviour* – the definition of a gold standard is not likely to be controversial; for these variables, the gold standard can be

said to approximate some kind of objective truth. For other concepts – e.g., *intelligence* and *satisfaction with life* – there is no objective standard to appeal to; then, the current best measurement practice may be used as a gold standard, assuming that a generally accepted best practice exists.[1)] Note that (for now) I am assuming only that a gold standard can be posited for the concepts in the SEM; I am not assuming (yet) that it has actually been measured.

In the presence of a theoretical gold standard, a solution to the identification problem may be provided by *explicit calibration* (Sobel and Arminger, 1986; Bielby, 1986b). After a quick introduction of the SEM and some notation in Section 2, Section 3 of this paper reviews the explicit calibration method and discusses in particular how it might be applied in official statistics. Explicit calibration requires the collection of a so-called audit sample, which may be a serious drawback in practice. In the rest of this paper, I develop a partial, less expensive alternative to the explicit calibration method, called *implicit calibration*. A key result is that, under certain conditions, the output of an SEM identified by explicit calibration may be used as input for implicit calibration of future SEMs (Section 5). To obtain this result, Section 4 introduces a family of restrictions that may be used to identify the covariance and mean structures of any model in a large class of SEMs. Both calibration methods are illustrated by a simulation study in Section 6.

2 Preliminaries

2.1 The linear structural equation model

In its general form, an SEM with a mean structure can be written as follows (Bollen, 1989):

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1}$$

$$\boldsymbol{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2}$$

$$\boldsymbol{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}. \tag{3}$$

Here, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are vectors of, respectively, m endogenous and n exogenous latent variables, and $\boldsymbol{y}$ and $\boldsymbol{x}$ are corresponding vectors of p and q observed variables. The $m \times m$ matrix $\boldsymbol{B}$ and $m \times n$ matrix $\boldsymbol{\Gamma}$ contain the coefficients of a linear regression model for the latent variables; $\boldsymbol{\zeta}$ is a vector of (zero-mean) disturbances associated with this regression model. The factor loadings of the observed variables on the latent variables are given by the $p \times m$ matrix $\boldsymbol{\Lambda}_y$ and the $q \times n$ matrix $\boldsymbol{\Lambda}_x$; the associated (zero-mean) random measurement errors are represented by the vectors $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$. Finally, $\boldsymbol{\alpha}$ is a vector of intercepts for the endogenous latent variables, and $\boldsymbol{\tau}_y$ and $\boldsymbol{\tau}_x$ are vectors of intercepts for the observed variables. In addition to (1)–(3), a complete specification of an SEM also includes as parameters the mean vector $\boldsymbol{\kappa} = E(\boldsymbol{\xi})$ and the covariance matrices $\boldsymbol{\Phi} = \mathrm{cov}(\boldsymbol{\xi})$, $\boldsymbol{\Psi} = \mathrm{cov}(\boldsymbol{\zeta})$, $\boldsymbol{\Theta}_\epsilon = \mathrm{cov}(\boldsymbol{\epsilon})$, and $\boldsymbol{\Theta}_\delta = \mathrm{cov}(\boldsymbol{\delta})$.

[1)] This distinction corresponds more or less to the traditional distinction between latent variables with Platonic and operational (or classical) true scores (Sobel and Arminger, 1986; Biemer and Stokes, 1991). According to Sobel and Arminger (1986), the problem of arbitrary metrics can arise only for variables with Platonic true scores. However, even for many variables without Platonic true scores, it may be relevant in practice to compare the quality of different measurement instruments. For example, one may want to test whether measurements of *satisfaction with life* obtained by a new, simpler questionnaire differ significantly from those obtained by the current best practice.

The above SEM defines both a covariance structure and a mean structure for the observed variables in $\boldsymbol{y}$ and $\boldsymbol{x}$. Let $\boldsymbol{\mu} = [E(\boldsymbol{y}'), E(\boldsymbol{x}')]'$ denote the vector of observed means and let $\boldsymbol{\Sigma}$ denote the observed covariance matrix. I introduce the following shorthand notation:

$$\boldsymbol{\tau} = \begin{bmatrix} \boldsymbol{\tau}_y \\ \boldsymbol{\tau}_x \end{bmatrix}, \qquad \boldsymbol{v} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\kappa} \end{bmatrix}, \qquad \boldsymbol{F} = \begin{bmatrix} (\boldsymbol{I}_m - \boldsymbol{B})^{-1} & (\boldsymbol{I}_m - \boldsymbol{B})^{-1}\boldsymbol{\Gamma} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}, \tag{4}$$

where $\boldsymbol{I}_m$ is the identity matrix of order m and $\boldsymbol{0}$ is used to denote any matrix or vector that consists entirely of zeros. Furthermore, let

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x \end{bmatrix}, \qquad \boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_\epsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Theta}_\delta \end{bmatrix}.$$

From (1)–(3), the following expressions may be derived for the observed means and covariances of $\boldsymbol{y}$ and $\boldsymbol{x}$ (Bollen, 1989):

$$\boldsymbol{\mu} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{F}\boldsymbol{v}, \tag{5}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Omega}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{6}$$

where

$$\boldsymbol{\Omega} = \operatorname{cov}\left(\begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix}\right) = \begin{bmatrix} (\boldsymbol{I}_m - \boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})[(\boldsymbol{I}_m - \boldsymbol{B})^{-1}]' & (\boldsymbol{I}_m - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'[(\boldsymbol{I}_m - \boldsymbol{B})^{-1}]' & \boldsymbol{\Phi} \end{bmatrix}.$$

From expression (6), it follows immediately that, in general, the covariance structure is not identified. Namely, if $\boldsymbol{Q}$ is any non-singular $(m+n) \times (m+n)$ matrix, then (6) remains satisfied if $\boldsymbol{\Lambda}$ is replaced by $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{Q}$ and $\boldsymbol{\Omega}$ is replaced by $\boldsymbol{\Omega}^* = \boldsymbol{Q}^{-1}\boldsymbol{\Omega}(\boldsymbol{Q}^{-1})'$. The mean structure is not identified either because the $p+q$ equations in (5) contain $m+n+p+q$ unknown parameters ($\boldsymbol{\tau}$ and $\boldsymbol{v}$) that do not occur in (6). Hence, additional restrictions are needed on the model parameters to identify the SEM.

2.2 Standard identification procedures

For ease of exposition, I shall make the following assumptions throughout the rest of this paper:

1. The covariance structure associated with the SEM is identified as soon as a scale is assigned to the latent variables. Thus, the free parameters of the covariance structure are arranged in a pattern that leads to an identified model apart from the scaling issues considered here.
2. Each observed variable loads on one latent variable and each latent variable has at least one indicator. Thus, the observed variables can be arranged so that the non-zero factor loadings in $\boldsymbol{\Lambda}$ have the following pattern:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \star & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \star & 0 & \cdots & 0 \\ 0 & \star & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \star & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \star \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \star \end{bmatrix}, \tag{7}$$

   where each $\star$ indicates a free parameter.

Little et al. (2006) described three different procedures that may be used to identify any SEM with a covariance and mean structure that satisfies the above assumptions. Let $\mathcal{K}_j$ denote the

index set of rows in $\boldsymbol{\Lambda}$ with a non-zero loading in the j^{th} column, and let $|\mathcal{K}_j|$ denote the number of such rows. The three procedures may be summarised as follows:

1. *(Reference indicators)* For each $j \in \{1, \ldots, m+n\}$, fix $\lambda_{kj} = 1$ and $\tau_k = 0$ for one particular $k \in \mathcal{K}_j$.
2. *(Standardised latent variables)* Fix the diagonal elements of $\boldsymbol{\Omega}$ to 1 and fix $\boldsymbol{\upsilon} = \mathbf{0}$.
3. *('Effects-coding method')* For each $j \in \{1, \ldots, m+n\}$, define the restrictions $\sum_{k \in \mathcal{K}_j} \lambda_{kj} = |\mathcal{K}_j|$ and $\sum_{k \in \mathcal{K}_j} \tau_k = 0$.

The first two approaches were already mentioned in the introduction. The 'effects-coding method' was proposed by Little et al. (2006) to avoid the use of arbitrary reference indicators. This approach assumes that the indicators of each latent variable have an average factor loading of one and an average measurement intercept of zero. Little et al. (2006) called this a 'non-arbitrary' scaling method. However, it can be argued that – unless one has external evidence that supports these assumptions about the factor loadings and intercepts – the resulting metric is still arbitrary; note that in general the metric will change when an indicator is added to or removed from the model, which reflects at least some degree of arbitrariness.[2)] Thus, neither of these standard approaches can be used directly for evaluating intercept bias. The next section reviews a different approach.

3 Explicit calibration

3.1 Theory

If one of the observed variables (say, x) represented the gold standard for measuring the corresponding latent variable (say, ξ), it would be reasonable to posit that $\xi = x$. The use of x as a reference indicator would then define a non-arbitrary metric for ξ.[3)] If each latent variable in the model had such a 'perfect' indicator, then it would be possible to identify both the covariance and mean structure without introducing any arbitrary restrictions. In particular, the intercept bias of each imperfect indicator in the model could then be estimated in a non-arbitrary way.

As every survey practitioner knows, devising a gold standard measurement instrument for a given construct may be conceptually possible but in practice such an instrument is very likely to be too expensive, too time-consuming, too invasive, and/or otherwise too impractical to be of much use. In many applications, however, it may be possible to apply such an instrument to a small subset of the original sample. In the literature, such a subsample is sometimes called an *audit sample*; see, e.g., Zhang (2011). The term *verification study* is also used.

In response to Bielby (1986a), Sobel and Arminger (1986) suggested the following non-arbitrary identification procedure. Suppose that, for each construct in an SEM, gold standard

2) It should be noted that Little et al. (2006) appear to be interested in applications where no gold standard exists. In this context their 'non-arbitrary' scaling method may be more useful.

3) In fact, it would be sufficient to have an exact, known linear transformation between ξ and x, i.e., $x = a\xi + b$ for certain known constants a and b. This generalisation has no real practical use: if the linear transformation is known, one can always rescale the observed values so that $a = 1$ and $b = 0$.

measurements have been obtained for an audit sample. For the purpose of estimating the SEM, the data may then be split into two groups. For the first, small group (referring to the audit sample), both perfect and imperfect measurements of the latent variables in the model are available; for the second, large group (referring to the rest of the original sample), one has only imperfect measurements. Figure 1 shows an example of such a two-group SEM, with three latent variables. The observed variables in the original sample are denoted by x_1, x_2, y_1, y_2, y_3, and y_4; the additional gold standard indicators measured only in the first group are x_3, y_5, and y_6. Note that in the first group, the gold standard variables are used as (non-arbitrary) reference indicators.

When the audit sample is a simple random subsample of the original sample, it is plausible that measurement invariance holds for all parameters that occur in both groups; this is in fact one of the few situations where measurement invariance can be supposed to exist by design. Thus, all structural parameters and all parameters of the measurement model for the imperfect observed variables may be assumed equal for both groups. The resulting two-group SEM is identified. Moreover, as it was not necessary to resort to any arbitrary scaling restrictions to achieve identification, the resulting estimates for the latent means, etc., are substantively meaningful. Bielby (1986b) called this approach an example of (explicit) calibration. The term probably derives from the calibration of physical measurement instruments in the natural sciences.

It is interesting to note that the assumption that x measures the gold standard exactly is not strictly necessary. Under the first assumption from Section 2.2, the SEM can be identified also when x contains measurement errors. (Cf. the standard use of reference indicators.) For the resulting metric to be non-arbitrary, it is required that the errors be 'purely random' and do not affect the scale in any way; i.e., it should hold that $x = \xi + \delta$ with $E(\delta) = 0$.[4)] Although this is a weaker assumption than the assumption of perfect measurement, from a practical point of view both assumptions are equally untestable. It depends on the context which of them can be considered plausible.

In practice, it may be convenient to obtain the audit sample by a more refined design than simple random subsampling. In that case, one should condition on the design variables to ensure that the assumption of measurement invariance between the groups remains valid. It will be seen in Section 6 that a quite small audit sample may be sufficient for explicit calibration to work.

3.2 Application in official statistics

To illustrate the use of explicit calibration, I now describe a hypothetical application in the context of official statistics. Suppose that measurements on three related target variables (ξ_1, η_1, and η_2) are available in two data sources that can be linked at record level. For instance, the first data source might be a sample survey and the second one an administrative register. As part of the regular production process, some form of editing (or 'data cleaning') has been applied independently to both data sets. The aim of statistical data editing is to improve the quality of the originally observed data to make them suitable for statistical use. It is usually neither feasible nor desirable to continue the editing process to a point where the data may be

[4)] Sobel and Arminger (1986) did not mention this possibility. Note that 'purely random' measurement errors are accepted also in calibrated physical measurement instruments (e.g., a thermometer).

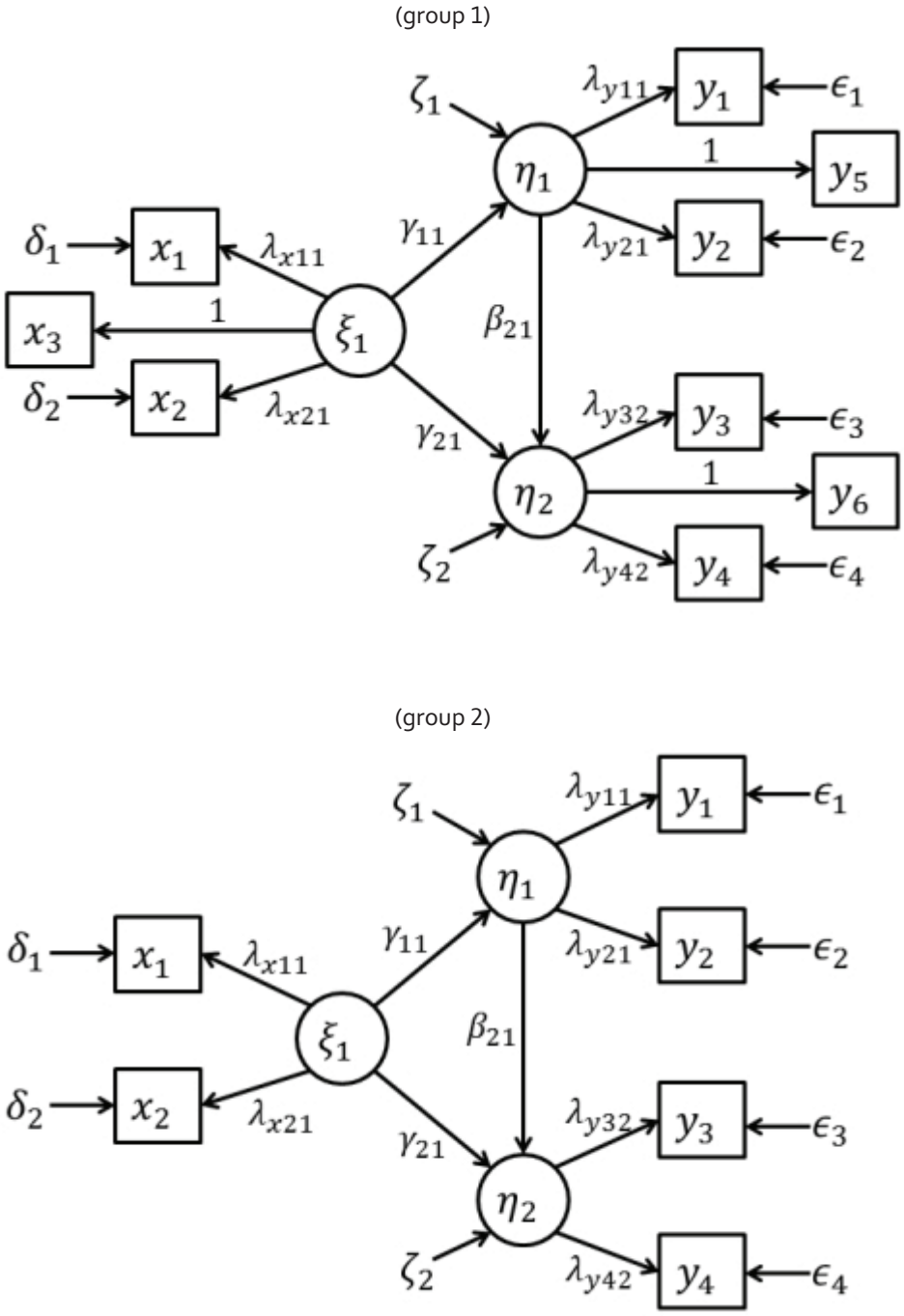

Figure 1 Example of a two-group SEM, where the first group contains perfect and imperfect indicators. Parameters with the same name in both groups are assumed to be invariant.

assumed completely free of measurement error (De Waal et al., 2011). Thus, both data sets can be seen as containing imperfect indicators of the target variables. This yields the SEM shown in the second panel of Figure 1.

Now suppose that a small random subsample is drawn from the overlapping units in the linked data set. The records in this subsample are submitted to an additional editing step which, unlike the previous editing effort, has the explicit goal of achieving a data set without measurement error. To this end, resources may be spent and measures may be taken that would be considered infeasible under normal conditions, i.e., when editing the full data set. One such measure may involve re-contacting the original respondents to obtain more detailed information. Another option is to let subject-matter experts check all records in the subsample, including those that under normal circumstances would be handled by automated procedures. In this manner, a new indicator for each target variable is obtained for the units in the audit sample. Using these new indicators, explicit calibration may then be applied by estimating the SEM of Figure 1.

Outside the context of SEMs, this partial re-editing approach was suggested and applied already by Nordbotten (1955) to assess the quality of the regular editing process of the 1953 Industrial Census in Norway. A crucial assumption is that the final edited data for the audit sample may be taken as the gold standard for assessing the measurement quality of the original observed variables. While this is a strong (and untestable) assumption, it may be considered plausible provided that sufficient resources have been devoted to the additional editing step. The basic premise that editing improves data quality and that this improvement tends to be larger as more resources are spent on it is (by and large) accepted by most producers of statistics. (The alternative assumption that the re-edited data contain only random errors seems less plausible.)

Finally, it should be noted that the selection of an audit sample by some form of probability sampling is a key element of the calibration method. In the above application, it would be more attractive in terms of the amount of work involved to use an alternative approach, in which subject-matter specialists select as a subsample a portion of the original data set that is likely to be error-free in the first place. Unfortunately, this would be an *invalid* approach in general, because the assumption of measurement invariance across groups might then be violated.

4 General identification restrictions

An important practical drawback of the explicit calibration method is that it requires an audit sample. The collection of gold standard data, even for a small subset of the original respondents, may be expensive or otherwise difficult. Moreover, a new audit sample is required for each new data set; this seems inefficient, in particular for repeated surveys. In the next section, a method will be proposed to re-use the information in an audit sample for multiple rounds of a repeated survey. To arrive at this result, the present section introduces a set of general identification restrictions for the covariance and mean structure of an SEM.

4.1 Identifying the covariance structure

First, consider the covariance structure (6). Recall that the 'effects-coding method' of Little et al. (2006) identifies a covariance structure by assuming that, for each latent variable, the

indicators have an average factor loading of one. More generally, one could define linear restrictions of the form

$$\sum_{k \in \mathcal{K}_j} \lambda_{kj} = d_j, \qquad \text{for } j \in \{1, \ldots, m+n\}, \tag{8}$$

where the d_j are known constants with $d_j \neq 0$. It is not difficult to show that any covariance structure that is identified by the 'effects-coding method' can be identified also by restrictions of the form (8).[5)]

Of course, an arbitrary set of restrictions of the form (8) defines an arbitrary scale for the latent variables. The point is that, for a particular application, it may be possible to obtain external evidence that supports a particular choice of restrictions of the form (8). (In practice, this external evidence may be provided by an audit sample. This point is taken up in Section 5.) These particular restrictions then define a non-arbitrary scale for the latent variables in the SEM. In matrix notation, the restrictions (8) may be written concisely as:

$$\boldsymbol{\Lambda}' \mathbf{1} = \boldsymbol{d}, \tag{9}$$

where $\mathbf{1}$ denotes a $(p+q)$ vector of ones and $\boldsymbol{d} = [d_1, \ldots, d_{m+n}]'$.

4.2 Identifying the mean structure

In what follows, I assume that the covariance structure associated with the SEM has been identified in a non-arbitrary way by (9). Thus, a non-arbitrary scale has been defined for each latent variable. To identify the mean structure (5) as well, the origins of these latent scales need to be fixed in a non-arbitrary way. Technically, this is done by adding restrictions on $\boldsymbol{\tau}$ (the intercept parameters of the measurement model) and/or $\boldsymbol{\upsilon}$ (the intercept parameters of the structural model). I will focus on a solution that imposes restrictions only on $\boldsymbol{\tau}$. The idea here is that in practice it will usually be easier for a researcher to make valid assumptions about the measurement process (which is at least partly under his/her control) than about the underlying structural process.

Consider a system of linear restrictions of the following form:

$$\sum_{k \in \mathcal{K}_j} \lambda_{kj} \tau_k = c_j, \qquad \text{for } j \in \{1, \ldots, m+n\}, \tag{10}$$

or, in matrix notation,

$$\boldsymbol{\Lambda}' \boldsymbol{\tau} = \boldsymbol{c}. \tag{11}$$

Each restriction in (10) concerns a weighted sum of the intercepts of all indicators associated with a particular latent variable. Under the assumptions of Section 2.2, the set of restrictions (11) leads to an identified mean structure; for a proof, see Appendix I. Thus, by choosing the constants c_j appropriately (based on external information), it is possible to identify the mean structure in a non-arbitrary way. Note that, as was the case for covariance structures, this procedure may be seen as a generalisation of the 'effects-coding method' of Little et al. (2006).

It may be noted that system (11) represents one particular family of constraints on $\boldsymbol{\tau}$ that successfully identifies the mean structure. One could also consider other systems of restrictions

[5)] Restriction (8) could be generalised further to $\sum_{k \in \mathcal{K}_j} t_{jk} \lambda_{kj} = d_j$ for any set of known coefficients t_{jk} with at least one $t_{jk} \neq 0$. For the present purpose, (8) is sufficiently general.

on $\boldsymbol{\tau}$. In fact, it is shown in Appendix I that this makes no real difference: *any* system of linear restrictions $\boldsymbol{R}_1\boldsymbol{\tau} = \boldsymbol{b}$ which succeeds in (just-)identifying the mean structure must produce the *same* parameter estimates as does an equivalent system of the particular form $\boldsymbol{\Lambda}'\boldsymbol{\tau} = \boldsymbol{c}$, assuming that the constant vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ are chosen in a non-arbitrary manner. In other words, assuming that one can find non-arbitrary values for $\boldsymbol{c}$ in (11), this is essentially the only meaningful way to obtain a just-identified mean structure by adding restrictions on the intercept parameters of the measurement model.

As a particular application, consider $\boldsymbol{\nu} = [E(\boldsymbol{\eta}'), E(\boldsymbol{\xi}')]'$, the vector of latent means. It is shown in Appendix I that the above identification procedure yields the following expression for $\boldsymbol{\nu}$:

$$\boldsymbol{\nu} = \boldsymbol{F}\boldsymbol{\nu} = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\mu} - \boldsymbol{c}). \tag{12}$$

4.3 Sampling and approximate restrictions

The results of this section so far may be summarised as follows: any SEM with a covariance and mean structure that satisfies the assumptions made in Section 2.2 can be identified by imposing restrictions of the form (9) and (11). In the presence of external information on $\boldsymbol{c}$ and $\boldsymbol{d}$, these restrictions yield a non-arbitrary metric for each latent variable in the model.

So far, I have ignored sampling considerations. Obviously, in practice, one has to work with estimates of the population quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ which are typically based on a probability sample. Denote the sample equivalents of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by $\boldsymbol{m}$ and $\boldsymbol{S}$, respectively. Under light regularity conditions, fitting an SEM to $\boldsymbol{m}$ and $\boldsymbol{S}$ yields consistent estimates of the model parameters (Bollen, 1989). In particular, the latent means in (12) are estimated consistently by

$$\hat{\boldsymbol{\nu}} = (\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - \boldsymbol{c}). \tag{13}$$

Here, as usual, an estimate of a parameter is indicated by adding a hat to that parameter.

In practice, an audit sample may be used as a source of information on the right-hand-sides of (9) and (11) (see the next section). Therefore, it is desirable to relax the assumption that these right-hand-sides are exactly known. For brevity's sake, I will focus on the mean structure. Suppose that, instead of (11), one can obtain an approximate version of these restrictions:

$$\boldsymbol{\Lambda}'\boldsymbol{\tau} = \hat{\boldsymbol{c}}, \tag{14}$$

where $\hat{\boldsymbol{c}} = \boldsymbol{c} + \boldsymbol{u}$ is an estimate of $\boldsymbol{c}$ from (11). For simplicity, suppose that $\hat{\boldsymbol{c}}$ is unbiased, and let $\mathrm{cov}(\boldsymbol{u}) = \boldsymbol{\Sigma}_u$. Moreover, I assume that $\hat{\boldsymbol{c}}$ is statistically independent of $\boldsymbol{m}$ and $\boldsymbol{S}$; i.e., the estimate $\hat{\boldsymbol{c}}$ comes from an external source that is independent of the current sample.

Since any choice of constants in the right-hand-side of (11) yields an identified mean structure, it is certainly possible to use (14) to identify the mean structure. In particular, this yields the following estimate for the latent means:

$$\hat{\boldsymbol{\nu}}_{\mathrm{alt}} = (\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - \hat{\boldsymbol{c}}); \tag{15}$$

cf. expression (13). Recall that $\hat{\boldsymbol{\Lambda}}$ is obtained from $\boldsymbol{S}$ using only the covariance structure and hence does not depend on whether (11) or (14) is used to identify the mean structure.

It is intuitively clear that the standard errors of the estimated means in (15) will be larger than those of (13), due to the added uncertainty in the restrictions. Standard packages for structural equation modeling, such as LISREL, do not take this added uncertainty into account. It is possible, however, to derive an expression for the asymptotically correct standard errors which

can be evaluated using the standard output of packages such as LISREL. Denote the covariance matrix of $\hat{\boldsymbol{\nu}}$ from (13) by $\boldsymbol{V}_0$. In Appendix II, it is shown that the covariance matrix of $\hat{\boldsymbol{\nu}}_{\text{alt}}$ from (15) is asymptotically equal to

$$\text{acov}(\hat{\boldsymbol{\nu}}_{\text{alt}}) = \boldsymbol{V}_0 + (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}. \tag{16}$$

Note that a positive definite matrix is added to $\boldsymbol{V}_0$ here, in agreement with the above intuition about the standard errors of the estimated latent means.

In practice, expression (16) may be estimated by

$$\widehat{\text{acov}}(\hat{\boldsymbol{\nu}}_{\text{alt}}) = \hat{\boldsymbol{V}}_0 + (\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}\hat{\boldsymbol{\Sigma}}_u(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}.$$

Here, $\hat{\boldsymbol{V}}_0$ is the estimated covariance matrix of the estimated latent means without taking the uncertainty in $\hat{\boldsymbol{c}}$ into account; most software packages for estimating SEMs (such as LISREL) can compute $\hat{\boldsymbol{V}}_0$ as part of their standard output. In addition, $\hat{\boldsymbol{\Sigma}}_u$ is an estimate of $\boldsymbol{\Sigma}_u$ which should be obtained from the external source that provided $\hat{\boldsymbol{c}}$.

This establishes the use of a stochastic version of (11) to identify the mean structure. The same approach could also be developed for the system (9) to identify the covariance structure, replacing $\boldsymbol{d}$ by an unbiased estimator $\hat{\boldsymbol{d}}$:

$$\boldsymbol{\Lambda}'\mathbf{1} = \hat{\boldsymbol{d}}. \tag{17}$$

For brevity's sake, this extension is not discussed here.

5 Implicit calibration

Consider again the two-group SEM that is estimated under the explicit calibration method (e.g., Figure 1). Since the parameters of the measurement model for the imperfect indicators are part of this SEM, explicit calibration can be used to obtain consistent estimates of the right-hand-sides of (9) and (11): $\hat{\boldsymbol{d}} = \hat{\boldsymbol{\Lambda}}'\mathbf{1}$ and $\hat{\boldsymbol{c}} = \hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\tau}}$. For sufficiently large samples, these estimates are approximately unbiased for $\boldsymbol{d}$ and $\boldsymbol{c}$. Of course, $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{c}}$ do not provide any new information that is useful for the present application – the model has already been identified using the gold standard data in the audit sample as reference indicators. However, these estimates may be useful for the identification of future models, as will now be discussed.

Suppose one is conducting a repeated survey, which may or may not involve a panel of respondents. In the first round of the survey, one has to use explicit calibration to obtain valid estimates of the latent means. As a by-product, one also obtains the above-mentioned estimates $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{c}}$. In subsequent rounds of the survey, a one-group SEM may then be identified in a non-arbitrary way by the restrictions (14) and (17), using the estimates $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{c}}$ from the first survey round.[6] I shall refer to this approach as implicit calibration.

Implicit calibration is a valid procedure for obtaining non-arbitrary estimates of the latent means, provided that the parameters of the measurement model are constant across different rounds of the repeated survey. This assumption seems reasonable when there are no important changes in the design of the survey between rounds and when the population remains fairly

[6] For the case of a panel survey, the variance formula derived in Section 4.3 needs to be adapted, because the assumption is violated that $\hat{\boldsymbol{c}}$ is statistically independent of $\boldsymbol{m}$ and $\boldsymbol{S}$ for the present sample.

stable during this period. Note that the *structural* parameters of the SEM are allowed to change between rounds because they do not occur in the identification restrictions (9) and (11). Nevertheless, it seems desirable to update the estimated identification restrictions on a regular basis, by repeating the explicit identification procedure every few survey rounds.

6 Simulation study

To illustrate the explicit and implicit calibration approaches, and to test their practical applicability, I conducted a small simulation study based on the SEM in Figure 1. In what follows, all SEMs were estimated by maximum likelihood in LISREL 8.8 (Jöreskog and Sörbom, 1996). For data generation and analysis of the results, I used the R environment for statistical computing (R Development Core Team, 2014).[7)]

6.1 Explicit calibration

For the first part of the study, multivariate normal data were generated according to the SEM in Figure 1 with the following model parameters [cf. (1)–(3)]:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 2.00 \\ 3.00 \end{bmatrix} + \begin{bmatrix} 0.00 & 0.00 \\ 0.60 & 0.00 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} 0.40 \\ 0.50 \end{bmatrix} \xi_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix};$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0.50 \\ -0.50 \\ 0.25 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 1.10 & 0.00 \\ 0.90 & 0.00 \\ 0.00 & 1.10 \\ 0.00 & 1.00 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix};$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -0.50 \\ -0.50 \end{bmatrix} + \begin{bmatrix} 1.00 \\ 0.90 \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix};$$

$$\boldsymbol{\Psi} = \mathrm{diag}(0.20,\ 0.30); \qquad \boldsymbol{\Phi} = [1.00]; \qquad \boldsymbol{\kappa} = [1.00];$$

$$\boldsymbol{\Theta}_\epsilon = \mathrm{diag}(0.19,\ 0.19,\ 0.19,\ 0.19); \qquad \boldsymbol{\Theta}_\delta = \mathrm{diag}(0.19,\ 0.19).$$

Note that, for this choice of parameters, none of the indicators provides an unbiased measurement of the underlying concept. In addition, most of the factor loadings are unequal to 1. For a quantitative variable such as *turnover*, a factor loading greater than 1 suggests (in the absence of trivial scaling effects such as reporting in dollars versus euros) that respondents tend to include components in the reported values that are not part of the concept as defined for statistical purposes. Similarly, a factor loading less than 1 suggests that respondents tend to leave relevant components out of the reported values.

I generated data sets of $N = 1000$ records each according to this model. Next, each data set was divided randomly into two groups. In the first group (M records), three observed variables y_5, y_6, and x_3 were added to the data with values equal to those of η_1, η_2, and ξ_1, respectively. The model for group 1 is shown in the first panel of Figure 1. Thus, the data for group 1 represented the 'gold standard' audit sample of Section 3. Different choices of M were tested (see below); for each value of M, 100 random data sets were created.

7) The LISREL syntax and R code used in this study are available from the author.

For the second group ($N - M$ records), the variables y_5, y_6, and x_3 were missing by design. To estimate the two-group SEM, a standard missing-data approach could have been followed (Sobel and Arminger, 1986; Allison, 1987). In fact, the missing-data problem for explicit calibration is more or less trivial – in the terminology of Little and Rubin (2002), the data are known to be missing completely at random –, so I used a simple solution. For the missing variables in group 2, I randomly imputed values from the standard normal distribution. In the model for group 2, I relaxed the assumption of measurement invariance across groups *for these variables only* and modeled them instead as consisting of purely random measurement errors with mean 0 and variance 1: $y_5 = \epsilon_5$, $y_6 = \epsilon_6$, and $x_3 = \delta_3$.

Tables 1 and 2 show some results of applying the explicit calibration approach, for various choices of the audit sample size M. The two-group SEM had 78 degrees of freedom. Since the model was correctly specified and the data followed a multivariate normal distribution, the overall fit measure X^2 should be distributed asymptotically as a chi-square variate with 78 degrees of freedom. The average values of X^2 reported in Table 1 are in line with this.

Table 1 Fit measures and estimated latent means obtained by explicit calibration. Reported values in the last three columns are averages over 100 simulations; standard deviations are in brackets.

		population values	calibrated estimates $M = 50$	$M = 100$	$M = 200$
model fit:	X^2	78.00	83.04 (13.62)	77.90 (13.61)	78.62 (13.58)
latent means:	$E(\eta_1)$	2.40	2.40 (0.04)	2.40 (0.03)	2.40 (0.02)
	$E(\eta_2)$	4.94	4.94 (0.04)	4.94 (0.04)	4.95 (0.04)
	$E(\xi_1)$	1.00	1.01 (0.05)	1.00 (0.04)	1.01 (0.04)

Table 2 Estimated right-hand-sides of identification restrictions after explicit calibration (in the same format as Table 1).

		population values	calibrated estimates $M = 50$	$M = 100$	$M = 200$
covariance structure:	$d_{\eta 1}$	2.000	2.01 (0.11)	2.00 (0.08)	2.00 (0.06)
	$d_{\eta 2}$	2.100	2.10 (0.08)	2.10 (0.06)	2.10 (0.04)
	$d_{\xi 1}$	1.900	1.91 (0.07)	1.91 (0.05)	1.90 (0.04)
mean structure:	$c_{\eta 1}$	0.100	0.06 (0.29)	0.10 (0.19)	0.09 (0.14)
	$c_{\eta 2}$	0.525	0.54 (0.38)	0.51 (0.29)	0.52 (0.22)
	$c_{\xi 1}$	−0.950	−0.97 (0.14)	−0.97 (0.10)	−0.95 (0.07)

To save space, I do not report all parameter estimates here. Table 1 shows the average values and standard deviations across 100 simulations of the estimated latent means; Table 2 reports the same for the elements of $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{c}}$. Obviously, the true values of these parameters were known in this simulation study. The true latent means were computed from the data generating model by $\boldsymbol{\nu} = \boldsymbol{F}\boldsymbol{\upsilon}$; the true values of $\boldsymbol{d}$ and $\boldsymbol{c}$ were computed from (9) and (11). These values are also reported in the tables. It is seen that all estimates were approximately unbiased. For the latent means and the elements of $\boldsymbol{d}$, quite small standard deviations were obtained even with an audit sample size of $M = 50$; note that this is just 5% of the original sample. Moreover, by

increasing the size of the audit sample to $M = 200$ (or 20% of the data) the precision of the estimated latent means was improved only marginally.

On the other hand, the standard deviations of the estimates for $\boldsymbol{c}$ were quite large, even for $M = 200$. This could be traced back to the standard deviations of the estimated $\boldsymbol{\tau}$ parameters, which were much higher than those of the estimated parameters of the covariance structure. This is probably due to the fact that the mean structure was just-identified, whereas the covariance structure was over-identified.

In the above simulations, the explicit calibration approach was used in its original form, i.e., with no measurement errors in y_5, y_6, and x_3 for the audit sample. I also tested the alternative approach with purely random measurement errors in the audit sample. I repeated the above simulations with data sets in which, for group 1, the observed variables y_5, y_6, and x_3 contained zero-mean normal errors with an error variance of 0.19. The results (not reported in detail here) were similar to those obtained without measurement errors. As might be expected, the standard deviations of the parameter estimates became larger, in particular for $\hat{\boldsymbol{c}}$. This suggests that a larger audit sample may be needed for explicit calibration if the audit sample data are assumed to contain purely random measurement errors rather than no errors.

6.2 Implicit calibration

Next, the implicit calibration approach was tested. For this purpose, 100 additional data sets of $N = 1000$ records were generated from the original model. To each of these new data sets I associated the parameter estimates of one of the 100 previously generated data sets for $M = 50$ under the explicit calibration approach. I then fitted the one-group SEM consisting only of the second panel in Figure 1 to each new data set; for identification, I used the restrictions (17) and (14) with the associated previous estimates $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{c}}$. The same approach was repeated using the parameter estimates from data sets with audit samples of $M = 100$ and $M = 200$.

Table 3 shows some of the results. Here, $\hat{\boldsymbol{d}}_M$ and $\hat{\boldsymbol{c}}_M$ denote estimates obtained from the explicit calibration method for an audit sample of size M. For comparison, I also report the results obtained by implicit calibration using the actual population values of $\boldsymbol{d}$ and $\boldsymbol{c}$, i.e., when the SEM was identified using (9) and (11) instead of (17) and (14). To save space, only the overall model fit and the estimated latent means are reported here.

Table 3 Fit measures and estimated latent means obtained by implicit calibration (in the same format as Table 1).

		population values	estimates by implicit calibration on			
			$\boldsymbol{d}, \boldsymbol{c}$	$\hat{\boldsymbol{d}}_{50}, \hat{\boldsymbol{c}}_{50}$	$\hat{\boldsymbol{d}}_{100}, \hat{\boldsymbol{c}}_{100}$	$\hat{\boldsymbol{d}}_{200}, \hat{\boldsymbol{c}}_{200}$
model fit:	X^2	6.00	5.72	5.72	5.72	5.72
			(2.83)	(2.83)	(2.83)	(2.83)
latent means:	$E(\eta_1)$	2.40	2.40	2.41	2.41	2.41
			(0.02)	(0.05)	(0.03)	(0.03)
	$E(\eta_2)$	4.94	4.95	4.94	4.95	4.95
			(0.03)	(0.05)	(0.04)	(0.04)
	$E(\xi_1)$	1.00	1.01	1.01	1.01	1.01
			(0.03)	(0.05)	(0.05)	(0.05)

It is seen in Table 3 that the overall fit as measured by X^2 was identical regardless of which method of implicit calibration was used. In fact, these models are statistically equivalent,

differing only in the way the latent variables are scaled. It is also seen that in each case, implicit calibration produced approximately unbiased estimates of the latent means. This is not surprising, as the new data sets came from the same data generating model as before. The standard deviations with the estimated vectors are similar to those in Table 1. In addition, calibrating on estimated vectors rather than the exact versions increased the standard deviations somewhat, but not much.

In practice, it is likely that the structural parameters of the model change over time, i.e., between survey rounds. To test the implicit calibration method under these conditions, I changed some of the parameters of the data generating model for the new data sets. This was done in two ways. For the first modified model, only the intercepts of the structural model were changed, to $\boldsymbol{\kappa} = [1.10]$ and $\boldsymbol{\alpha} = [2.20, 3.30]'$. For the second modified model, the structural part of the model was completely changed:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} -1.00 \\ 5.00 \end{bmatrix} + \begin{bmatrix} 0.00 & 0.00 \\ 0.80 & 0.00 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} 0.30 \\ 0.70 \end{bmatrix} \xi_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix};$$

$$\boldsymbol{\Psi} = \mathrm{diag}(0.30,\ 0.20); \qquad \boldsymbol{\Phi} = [5.00]; \qquad \boldsymbol{\kappa} = [2.00].$$

Importantly, no changes were made to the measurement model in either case.

Again, 100 data sets of $N = 1000$ multivariate normal records were generated for both modified models. Since the results in Table 3 showed little difference between implicit calibration on $\hat{\boldsymbol{d}}_M$ and $\hat{\boldsymbol{c}}_M$ and implicit calibration on $\boldsymbol{d}$ and $\boldsymbol{c}$, I tested only the latter case here. Note that $\boldsymbol{d}$ and $\boldsymbol{c}$ still refer to the original data generating model, i.e., the population values reported in Table 2.

Table 4 shows some results in the same format as before. It is seen that approximately unbiased estimates of the latent means were found for both modified models, even though – in particular for the second case – these latent means were very different from the original model. As explained in Section 5, this happened because the measurement model remained the same.

Table 4 Fit measures and estimated latent means obtained by implicit calibration for alternative data generating models (in the same format as Table 1).

		model 1		model 2	
		population	estimates	population	estimates
model fit:	X^2	6.00	6.36	6.00	6.41
			(3.52)		(3.55)
latent means:	$E(\eta_1)$	2.640	2.63	−0.400	−0.40
			(0.02)		(0.03)
	$E(\eta_2)$	5.434	5.43	6.080	6.08
			(0.03)		(0.07)
	$E(\xi_1)$	1.100	1.09	2.000	1.99
			(0.03)		(0.08)

7 Discussion and conclusion

In this paper, I investigated possible ways to obtain valid estimates of latent means and intercepts in a single-group SEM. Identification of the model requires that a metric be assigned to the latent variables. In the absence of external information, it is inevitable that the assigned metric is in some sense arbitrary; as a result, the estimated latent means and intercepts do not

have a meaningful interpretation. Explicit calibration on external information has been suggested as a method to obtain a non-arbitrary metric for the latent variables. Unfortunately, this method requires the collection of additional 'gold standard' data for at least a subset of the original sample (a so-called audit sample); this may be too expensive, too labour-intensive, or otherwise infeasible in some applications.

The main result of this paper is an extension of the calibration method, called implicit calibration, whereby information from the same audit sample may be re-used in multiple rounds of a repeated survey. Implicit calibration has two (related) advantages. On the one hand, it removes the need to use the costly explicit calibration approach in every survey round. On the other hand, it ensures that maximal use is made of an audit sample, thereby strengthening the case for the collection of such a sample. As a technical by-product, a family of identification restrictions was introduced in Section 4 that may be used to identify any covariance and mean structure in a large class of SEMs. This last result may be of interest in its own right, as it generalises the existing 'effects-coding' identification procedure.

The simulation results on artificial data presented here suggest that implicit calibration yields estimates of about the same quality as explicit calibration, provided that the underlying assumption of no change in the measurement model between survey rounds is satisfied. In addition, it appears from this study that a relatively small audit sample (say, 5% of the original sample) may be sufficient in practice; indeed, the quality of the estimated means was only marginally improved by increasing the size of the audit sample. An interesting topic for further research could be to assess the robustness of implicit calibration to small changes in the parameters of the measurement model.

As mentioned in the introduction, one possible application of the methodology discussed in this paper lies in assessing the measurement quality of data sources for use in official statistics. This concerns validity and reliability of measurement, as well as the presence of intercept bias. The assessment of validity and reliability may be formulated naturally in terms of SEMs (Saris and Andrews, 1991; Bakker, 2012); explicit or implicit calibration is needed to also estimate the bias in a meaningful way. It should be noted that explicit calibration may also improve the estimates of validity and reliability, as the audit sample contains information on the correlations between latent variables. A practical application of the results of this paper at Statistics Netherlands remains to be investigated.

Acknowledgements

I would like to thank Bart Bakker, Cees Elzinga, and Harry Ganzeboom for their helpful comments on early versions of this paper.

References

Allison, P. D. (1987). Estimation of Linear Models with Incomplete Data. *Sociological Methodology 17*, 71–103.

Alwin, D. F. (2007). *Margins of Errors*. New York: John Wiley & Sons.

Andrews, F. M. (1984). Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *Public Opinion Quarterly 48*, 409–442.

Bakker, B. F. M. (2012). Estimating the Validity of Administrative Variables. *Statistica Neerlandica 66*, 8–17.

Bielby, W. T. (1986a). Arbitrary Metrics in Multiple-Indicator Models of Latent Variables. *Sociological Methods & Research 15*, 3–23.

Bielby, W. T. (1986b). Arbitrary Normalizations; Comments on Issues Raised by Sobel, Arminger, and Henry. *Sociological Methods & Research 15*, 62–63.

Biemer, P. and S. L. Stokes (1991). Approaches to the Modeling of Measurement Error. In Biemer, Groves, Lyberg, Mathiowetz, and Sudman (Eds.), *Measurement Errors in Surveys*, pp. 487–516. New York: John Wiley & Sons.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Byrne, B. M., R. J. Shavelson, and B. Muthén (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin 105*, 456–466.

Cheung, G. W. and R. B. Rensvold (1999). Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. *Journal of Management 25*, 1–27.

De Waal, T., J. Pannekoek, and S. Scholtus (2011). *Handbook of Statistical Data Editing and Imputation*. Hoboken, New Jersey: John Wiley & Sons.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

Jöreskog, K. and D. Sörbom (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data* (second ed.). New York: John Wiley & Sons.

Little, T. D., D. W. Slegers, and N. A. Card (2006). A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. *Structural Equation Modeling 13*, 59–72.

Lord, F. M. and M. R. Novick (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Nordbotten, S. (1955). Measuring the Error of Editing the Questionnaires in a Census. *Journal of the American Statistical Association 50*, 364–369.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: http://www.R-project.org/.

Saris, W. E. and F. M. Andrews (1991). Evaluation of Measurement Instruments Using a Structural Modeling Approach. In Biemer, Groves, Lyberg, Mathiowetz, and Sudman (Eds.), *Measurement Errors in Surveys*, pp. 575–597. New York: John Wiley & Sons.

Scherpenzeel, A. C. and W. E. Saris (1997). The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies. *Sociological Methods & Research 25*, 341–383.

Sobel, M. E. and G. Arminger (1986). Platonic and Operational True Scores in Covariance Structure Analysis. *Sociological Methods & Research 15*, 44–58.

Whittaker, T. A. (2013). The Impact of Noninvariant Intercepts in Latent Means Models. *Structural Equation Modeling 20*, 108–130.

Yung, Y. F. and P. M. Bentler (1999). On Added Information for ML Factor Analysis with Mean and Covariance Structures. *Journal of Educational and Behavioral Statistics 24*, 1–20.

Zhang, L.-C. (2011). A Unit-Error Theory for Register-Based Household Statistics. *Journal of Official Statistics 27*, 415–432.

Appendices

I A closer look at the identification of the mean structure

This appendix contains some additional material on identification restrictions for the mean structure. As in Section 4.2, it is assumed throughout that the covariance structure has already been identified.

I.1 The general restricted solution

First of all, the mean structure (5) may be written more concisely as:

$$\boldsymbol{\mu} = \boldsymbol{H}\boldsymbol{\theta}, \tag{18}$$

with $\boldsymbol{\theta} = [\boldsymbol{\tau}', \boldsymbol{\upsilon}']'$ and

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{I}_{p+q} & \boldsymbol{\Lambda}\boldsymbol{F} \end{bmatrix}.$$

The vector $\boldsymbol{\theta}$ captures all parameters of the mean structure that do not occur in the covariance structure.

According to Yung and Bentler (1999), an under- or just-identified mean structure does not provide any new information about the parameters of a covariance structure. That is to say, these parameter estimates are invariant to the inclusion or exclusion of a not over-identified mean structure in the SEM. Since the aim in this paper is to add restrictions to an under-identified mean structure so that it becomes just-identified, not over-identified, I can (and will) treat the matrix $\boldsymbol{H}$ in (18) as fixed. Thus, the mean structure may be regarded as a linear system in $\boldsymbol{\theta}$.

To make the mean structure (just-)identified, one could add linear restrictions on the parameters in $\boldsymbol{\theta}$. Consider, instead of (18), the augmented system

$$\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{R} \end{bmatrix} \boldsymbol{\theta}, \tag{19}$$

for a given $(m+n) \times (m+n+p+q)$ matrix $\boldsymbol{R}$ and $(m+n)$ vector $\boldsymbol{b}$. In shorthand notation, (19) can be written as $\boldsymbol{\mu}_{\text{aug}} = \boldsymbol{H}_{\text{aug}}\boldsymbol{\theta}$, with $\boldsymbol{\mu}_{\text{aug}} = [\boldsymbol{\mu}', \boldsymbol{b}']'$ and

$$\boldsymbol{H}_{\text{aug}} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{R} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{p+q} & \boldsymbol{\Lambda}\boldsymbol{F} \\ \boldsymbol{R}_1 & \boldsymbol{R}_2 \end{bmatrix}.$$

The mean structure will be identified if the additional restrictions

$$\boldsymbol{R}\boldsymbol{\theta} = \boldsymbol{R}_1\boldsymbol{\tau} + \boldsymbol{R}_2\boldsymbol{\upsilon} = \boldsymbol{b} \tag{20}$$

are chosen in such a way that $\boldsymbol{H}_{\text{aug}}^{-1}$ exists. In that case, $\boldsymbol{\theta} = \boldsymbol{H}_{\text{aug}}^{-1}\boldsymbol{\mu}_{\text{aug}}$ provides the unique solution to (19). The three identification procedures of Little et al. (2006) can be obtained as special cases of (20).

By a standard result from linear algebra [see, e.g., Theorem 8.5.11 in Harville (1997)], the matrix $\boldsymbol{H}_{\text{aug}}$ is invertible if, and only if, the inverse of $\boldsymbol{R}_2 - \boldsymbol{R}_1\boldsymbol{\Lambda}\boldsymbol{F}$ exists. If the latter matrix is denoted by $\boldsymbol{K}$, then the resulting parameter estimates are:

$$\boldsymbol{\theta} = \boldsymbol{H}_{\text{aug}}^{-1}\boldsymbol{\mu}_{\text{aug}} = \begin{bmatrix} \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{F}\boldsymbol{K}^{-1}(\boldsymbol{R}_1\boldsymbol{\mu} - \boldsymbol{b}) \\ -\boldsymbol{K}^{-1}(\boldsymbol{R}_1\boldsymbol{\mu} - \boldsymbol{b}) \end{bmatrix}, \tag{21}$$

provided that $\boldsymbol{K}^{-1}$ exists.

I.2 A particular solution

As discussed in Section 4.2, I restrict attention to the special case with $\boldsymbol{R}_2 = \boldsymbol{0}$ in (20), i.e.

$$\boldsymbol{R}_1\boldsymbol{\tau} = \boldsymbol{b}. \tag{22}$$

A particular choice of restrictions of the form (22) is provided by (11), with $\boldsymbol{R}_1 = \boldsymbol{\Lambda}'$ and $\boldsymbol{b} = \boldsymbol{c}$. To prove that this choice yields an identified mean structure, it suffices to show that $\boldsymbol{K} = -\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{F}$ is non-singular. It is easily verified that the square matrix $\boldsymbol{F}$ in (4) is non-singular; in fact:

$$\boldsymbol{F}^{-1} = \begin{bmatrix} \boldsymbol{I}_m - \boldsymbol{B} & -\boldsymbol{\Gamma} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}.$$

Furthermore, $\boldsymbol{\Lambda}$ has the form (7) by assumption and is therefore a matrix of full column rank. This implies that the square matrix $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ must be non-singular. Thus, it holds that $\boldsymbol{K}^{-1} = -\boldsymbol{F}^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}$ and the result follows. This establishes that imposing (11) yields an identified mean structure.

By substituting the above expression for $\boldsymbol{K}^{-1}$ into (21), the following solution for $\boldsymbol{\theta}$ is obtained:

$$\begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{\upsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} - \boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\mu} - \boldsymbol{c}) \\ \boldsymbol{F}^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\mu} - \boldsymbol{c}) \end{bmatrix}.$$

In particular, expression (12) for the latent means $\boldsymbol{\nu} = \boldsymbol{F}\boldsymbol{\upsilon}$ now follows.

I.3 From $\boldsymbol{R}_1\boldsymbol{\tau} = \boldsymbol{b}$ to $\boldsymbol{\Lambda}'\boldsymbol{\tau} = \boldsymbol{c}$

It remains to prove the remark in Section 4.2 that for any system of the form $\boldsymbol{R}_1\boldsymbol{\tau} = \boldsymbol{b}$ that successfully identifies the mean structure, an equivalent system of the particular form $\boldsymbol{\Lambda}'\boldsymbol{\tau} = \boldsymbol{c}$ can be found. As noted above, a necessary and sufficient condition for the augmented mean structure (with $\boldsymbol{R}_2 = \boldsymbol{0}$) to be identified is that the matrix $\boldsymbol{K} = -\boldsymbol{R}_1\boldsymbol{\Lambda}\boldsymbol{F}$ be non-singular. Consider the general form of restrictions that satisfy this condition. Clearly, $\boldsymbol{R}_1$ should be chosen in such a way that $\boldsymbol{R}_1\boldsymbol{\Lambda}\boldsymbol{F} = \boldsymbol{M}$ for some invertible matrix $\boldsymbol{M}$. That is, it should hold that

$$\begin{aligned} \boldsymbol{R}_1\boldsymbol{\Lambda}\boldsymbol{F} &= \boldsymbol{M}, \\ \boldsymbol{R}_1\boldsymbol{\Lambda} &= \boldsymbol{M}\boldsymbol{F}^{-1}, \\ \boldsymbol{F}\boldsymbol{M}^{-1}\boldsymbol{R}_1\boldsymbol{\Lambda} &= \boldsymbol{I}_{m+n}, \\ \boldsymbol{F}\boldsymbol{M}^{-1}\boldsymbol{R}_1 &= \boldsymbol{G}, \\ \boldsymbol{R}_1 &= \boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}. \end{aligned}$$

Here, $\boldsymbol{G}$ represents any generalised inverse of $\boldsymbol{\Lambda}$, i.e., any matrix for which it holds that $\boldsymbol{\Lambda}\boldsymbol{G}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}$. In the fourth line, I used the fact that $\boldsymbol{\Lambda}$ has full column rank, so that $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{\Lambda}$ if, and only if, it is a left inverse of $\boldsymbol{\Lambda}$, i.e., $\boldsymbol{G}\boldsymbol{\Lambda} = \boldsymbol{I}_{m+n}$ (Harville, 1997, Lemma 9.2.8).

If the system of restrictions $\boldsymbol{R}_1\boldsymbol{\tau} = \boldsymbol{b}$ is used to identify the mean structure, with $\boldsymbol{R}_1 = \boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}$, then according to expression (21) the following solution is obtained:

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{F}\boldsymbol{M}^{-1}(\boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{\mu} - \boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{\tau}) \\ \boldsymbol{M}^{-1}(\boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{\mu} - \boldsymbol{M}\boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{\tau}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{G}(\boldsymbol{\mu} - \boldsymbol{\tau}) \\ \boldsymbol{F}^{-1}\boldsymbol{G}(\boldsymbol{\mu} - \boldsymbol{\tau}) \end{bmatrix}. \tag{23}$$

In the first expression, I substituted $\boldsymbol{R}_1\boldsymbol{\tau}$ for $\boldsymbol{b}$ and used the fact that $\boldsymbol{K} = -\boldsymbol{M}$. Note that the last expression in (23) does not depend on the choice of $\boldsymbol{M}$. I claim that this expression does not depend on the choice of generalised inverse $\boldsymbol{G}$ either.

To prove this claim, the following result from linear algebra is needed: if $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are matrices such that the row space of $\boldsymbol{B}$ is contained in that of $\boldsymbol{A}$ and the column space of $\boldsymbol{C}$ is contained in

that of $\boldsymbol{A}$, then the outcome of the expression $\boldsymbol{BA^-C}$ is the same for every generalised inverse $\boldsymbol{A^-}$ of $\boldsymbol{A}$ (Harville, 1997, Theorem 9.4.1). To apply this result, I first note that, by (5), the vector $\boldsymbol{\mu - \tau}$ is contained in the column space of $\boldsymbol{\Lambda}$. Since the above result holds in particular if $\boldsymbol{B = A}$, this is sufficient to establish that $\boldsymbol{\Lambda G(\mu - \tau)}$ does not depend on the choice of $\boldsymbol{G}$.

To complete the proof, it remains to be shown that the row space of $\boldsymbol{F^{-1}}$ is contained in that of $\boldsymbol{\Lambda}$. Let $\boldsymbol{v}$ be any vector in the row space of $\boldsymbol{F^{-1}}$; i.e., $\boldsymbol{v = z'F^{-1}}$ for some $\boldsymbol{z}$. The choice $\boldsymbol{w = \Lambda(\Lambda'\Lambda)^{-1}(F^{-1})'z}$ clearly satisfies $\boldsymbol{w'\Lambda = z'F^{-1}(\Lambda'\Lambda)^{-1}\Lambda'\Lambda = z'F^{-1} = v}$. This shows that $\boldsymbol{v}$ is contained in the row space of $\boldsymbol{\Lambda}$, and completes the proof.

Since (23) does not depend on the choice of generalised inverse of $\boldsymbol{\Lambda}$, this expression may be evaluated by substituting one particular generalised inverse: $\boldsymbol{G^* = (\Lambda'\Lambda)^{-1}\Lambda'}$. Let $\boldsymbol{R_1^*\tau = b^*}$ be the system of constraints corresponding to this choice of generalised inverse, i.e.,

$$\boldsymbol{MF^{-1}(\Lambda'\Lambda)^{-1}\Lambda'\tau = b^*}. \tag{24}$$

Here, it is important to distinguish the particular instance $\boldsymbol{b^*}$ from $\boldsymbol{b}$; note that, although it was just proved that (23) does not depend on the choice of generalised inverse in $\boldsymbol{R_1}$, this expression *does* depend on the choice of constant vector $\boldsymbol{b}$. On the other hand, for any *given* system $\boldsymbol{MF^{-1}G\tau = b}$, it is possible to choose $\boldsymbol{b^*}$ in such a way that (24) leads to the same outcome. Finally, the system (24) is clearly equivalent to $\boldsymbol{\Lambda'\tau = c^*}$ [i.e., a particular instance of system (11)] for the choice $\boldsymbol{c^* = \Lambda'\Lambda FM^{-1}b^*}$.

II Derivation of expression (16)

To derive expression (16) for the asymptotic covariance matrix of $\hat{\boldsymbol{\nu}}_{\text{alt}}$, I use the standard decomposition

$$\text{cov}(\hat{\boldsymbol{\nu}}_{\text{alt}}) = E_s\left[\text{cov}(\hat{\boldsymbol{\nu}}_{\text{alt}}|s)\right] + \text{cov}_s\left[E(\hat{\boldsymbol{\nu}}_{\text{alt}}|s)\right].$$

Here, the symbol s refers to the sample from which $\boldsymbol{m}$ and $\boldsymbol{S}$ are obtained, and $E_s(.)$ and $\text{cov}_s(.)$ refer to properties of the underlying sampling design.

For the first term on the right, it follows that

$$\begin{aligned} E_s\left[\text{cov}(\hat{\boldsymbol{\nu}}_{\text{alt}}|\, s)\right] &= E_s\left[\text{cov}\left\{(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - \hat{\boldsymbol{c}})\middle|\, s\right\}\right] \\ &= E_s\left[(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}\text{cov}(\hat{\boldsymbol{c}})(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}\right] \\ &= E_s\left[(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma}_u(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}\right]. \end{aligned}$$

In the second line, I used the fact that $\hat{\boldsymbol{c}}$ is statistically independent of s. Asymptotically, the above expression converges to $\boldsymbol{(\Lambda'\Lambda)^{-1}\Sigma_u(\Lambda'\Lambda)^{-1}}$.

For the second term, it follows that

$$\begin{aligned} \text{cov}_s\left[E(\hat{\boldsymbol{\nu}}_{\text{alt}}|\, s)\right] &= \text{cov}_s\left[E\left\{(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - \hat{\boldsymbol{c}})\middle|\, s\right\}\right] \\ &= \text{cov}_s\left[(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - E(\hat{\boldsymbol{c}}))\right] \\ &= \text{cov}_s\left[(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}})^{-1}(\hat{\boldsymbol{\Lambda}}'\boldsymbol{m} - \boldsymbol{c})\right] \\ &= \boldsymbol{V}_0. \end{aligned}$$

Combining the two terms, one obtains expression (16). In particular, it holds that $E(\hat{\boldsymbol{\nu}}_{\text{alt}}|\, s) = \hat{\boldsymbol{\nu}}$.

Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2013–2014	2013 to 2014 inclusive
2013/2014	Average for 2013 to 2014 inclusive
2013/'14	Crop year, financial year, school year, etc., beginning in 2013 and ending in 2014
2011/'12–2013/'14	Crop year, financial year, etc., 2011/'12 to 2013/'14 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
verkoop@cbs.nl
Fax +31 45 570 62 68

60083201409 X-10