B Research method

Part B of the thesis addresses the research method. The methodological foundations are laid in two steps: The first chapter of this part is devoted to data collection procedures and the sample of the study. The second chapter introduces and describes SEM techniques that will be applied for data analysis.

1 Collection of survey data and sample

1.1 Data collection and survey administration

This subsection starts with an overview of the data collection procedures. In the following, potential key informant and common method biases and corresponding activities to reduce the risk of such biases are described. Moreover, I present details on the pilot-test procedures as well as on the specific activities for data collection.

Overview of the data collection procedures

The procedures to collect data started with the definition of the target population.\textsuperscript{40} The target population of the research consists of management accountants and general managers of German medium- and large-sized corporations. Since the research questions of the study basically focus on management accountants and the impact of their activities, management accountants and general managers, i.e., users of services offered by management accountants, were selected as adequate target respondents.\textsuperscript{41} Medium- and large-sized corporations were chosen because control issues and management accountants or management accounting departments are typically present in such corporations.\textsuperscript{42} Except for financial institutions the target population covers all industries to allow generalizability of the results. Financial institutions have been excluded due to their specific business models and regulatory requirements. Furthermore, the target population of firms was restricted to a single country in order to limit possible biases due to institutional or cultural differences.\textsuperscript{43}

\textsuperscript{40} Ref. Van der Stede/Young/Chen (2005), p. 666.
\textsuperscript{43} Ref. Chow/Kato/Merchant (1996); Jansen/Merchant/Van der Stede (2009).
The selected survey population was derived from a database of a German commercial list provider\textsuperscript{44}. In a first step, I selected the top 1,500 companies in terms of revenue. As a second step, another 281 companies had to be discarded for various reasons\textsuperscript{45} resulting in a final survey population of 1,219 companies for the study.

Key informant and common method biases

Data gathering procedures in survey research typically apply single-informant designs, i.e., one respondent per company answers relevant questions and assesses constructs.\textsuperscript{46} The quality of obtained data accordingly depends on the adequate selection of respondents. Especially two potential problems are associated with such single-informant designs: Key informant and common method biases.

- Information provided by survey respondents, i.e., the key informants, is in most cases not limited to personal opinions since their ratings typically also embrace departmental- or company-related aspects. Potential key informant biases might arise if the respondents do not possess adequate knowledge due to their functional or hierarchical position in the company.\textsuperscript{47}

- Selection of respondents and the research design might create a common method bias. Such an effect potentially results from research designs in which independent and dependent constructs are assessed by the same person. Possible explanations for this potential bias are consistency motifs, implicit theories, social desirability, or affectivity of respondents.\textsuperscript{48}

To cope with those two potential biases, I carefully selected the respondents and implemented a multi-informant – more precisely, a ‘dyadic’ – research design. Taking required competencies with regard to functional and hierarchical position into account, the heads of management accounting departments seemed to be the most appropriate respondents for questions and assessments related to the tasks and roles of management accountants. Aspects regarding the results of management accountants’ actions were surveyed with general managers as they are the users of management ac-

\textsuperscript{44} Hoppenstedt Firmeninformationen GmbH.
\textsuperscript{45} E.g., lack of dedicated management accounting department, double counts due to legal form constructions, or ceased operations.
countants’ services.\textsuperscript{49} The result of this choice of respondents is the dyadic research design, which covers assessments of two respondents.\textsuperscript{50} In addition, this resulting multi-informant design should also alleviate a possible common method bias since it allows to survey independent and dependent constructs with different respondents. Subsequently to data gathering, I complementally conducted Harman’s (1967) single-factor-test to search for possible signals of common method bias. This test checks whether an exploratory factor analysis of all relevant survey items results in one or more factors. The results of the exploratory factor analyses for the two research models, which will be introduced in Part C and Part D of the present study, do not reveal a single or common factor indicating no risk of common method bias. For additional validation purposes, I discussed and confirmed the selection of respondents during the pre-test procedures described in the following. Summarizing the activities in this context, the research approach should alleviate possible key informant or common method biases.

\textbf{Pilot-test procedures}

Before starting data collection, the survey instruments were pilot-tested by five executives from business practice and six academic researchers to ensure reliable and valid measurements in the study. Pilot- or pre-tests are especially important in mail surveys since respondents have no or at least a reduced opportunity to contact the researcher in case of questions or problems in understanding the survey items.\textsuperscript{51} As a result of the pre-test procedures, some of the survey items were slightly adjusted. Details of the survey items are presented in Chapter 1.3 of Part C and in Chapter 1.4 of Part D of the study.

As this research project was conducted in Germany, I applied the German language for the questionnaires and all correspondence. If references or existing item definitions were in English, I searched for German translations or carefully translated the materials. I additionally discussed the questionnaire with a bilingual researcher to ensure reliable and valid translations.

\textsuperscript{49} Ref. Chenhall (2003), p. 134
\textsuperscript{51} Ref. Van der Stede/Young/Chen (2005), p. 670.
Data collection procedures

Data collection took place in the period of March to May 2009 as a large-scale mail survey. In particular, data collection procedures followed a three-step implementation strategy: First, I contacted each firm by phone to check data accuracy, to ask for the latest contact details, and to introduce the study. Second, I sent a cover letter, a questionnaire as well as a second survey package to the heads of the management accounting departments.52 I asked them (i) to fill out the functionally customized questionnaires for management accountants and (ii) to forward the second survey package to a general manager, i.e., a member of the upper or upper-middle management such as the CEO, managing director, or division manager. Third, I sent out two reminder e-mails, two respectively four weeks after the initial mailing to boost the response rate. To enhance the chance of participation, I personalized all correspondence and offered a research report covering the main findings of the study to all participants.

1.2 Sample description

The subsection on the sample of the study encompasses descriptions on responses and response rates, different data sets employed in the study, details of the sample, and activities to analyze a potential non-response bias.

Responses and return rates

A total of 280 persons participated in the study. Eight questionnaires had to be discarded due to a large number of missing data. From the remaining 272 responses, 160 were from management accountants and 112 from general managers. Due to the dyadic research design I need matching pairs of questionnaires. Thus, the final dyadic sample consists of the answers of 112 firms representing a return rate of 9.19%. This response rate is lower than anticipated and below average in typical empirical management accounting studies.53 Answers from non-participating firms to follow-up phone calls or reminder e-mails revealed, for instance, that executives had other priorities during the ongoing economic downturn in spring 2009, that they had introduced a policy of not participating in survey research due to the increasing number of requests, or that the

---

52 In a few cases participants requested the instruments to be sent by fax or mail.
complexity of the research project was too high due to the need for dyadic data. Whereas the low response rate raises a potential limitation of the study, the sample is large enough to process planned advanced statistical procedures for analyzing the data.54

**Dyadic and MA data set**

Two samples were basically obtained from data collection. The main sample contains the dyadic sets of questionnaires and is denoted as “Dyadic data set” in the following. I use this data to analyze the main research models that are derived to answer the research questions. A second set of data denoted as “MA data set” includes both the answers by management accountants of the dyadic data set and the answers by management accountants of those companies that did not provide a response by a general manager. I employ this data for selected descriptive statistics and for some robustness checks of the research models.

**Sample details**

Tables B-1 and B-2 provide information regarding the organizations’ size in terms of revenue and number of employees. Data are gathered especially from medium- and large-sized companies. Taking into account that control problems and departments of management accountants are primarily in place in medium- and large-sized companies, descriptive statistics indicate that the firms were large enough to ensure that the sample is adequate for analyzing the research questions. Furthermore, results of Mann-Whitney-U-tests to assess whether the two (independent) samples come from the same distribution indicate no significant differences in the central tendency of the two data sets (revenue: p = 0.659; employees: p = 0.744).

---

### Table B-1: Surveyed finns by revenue

<table>
<thead>
<tr>
<th>Revenue (Million EUR)</th>
<th>Dyadic data set</th>
<th>MA data set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>0 – 500</td>
<td>32</td>
<td>28.83%</td>
</tr>
<tr>
<td>501 – 1,000</td>
<td>38</td>
<td>34.23%</td>
</tr>
<tr>
<td>1,001 – 5,000</td>
<td>27</td>
<td>24.32%</td>
</tr>
<tr>
<td>5,001 – 10,000</td>
<td>4</td>
<td>3.60%</td>
</tr>
<tr>
<td>10,001 – 20,000</td>
<td>3</td>
<td>2.70%</td>
</tr>
<tr>
<td>&gt; 20,000</td>
<td>7</td>
<td>6.31%</td>
</tr>
<tr>
<td>Mean</td>
<td>3,684</td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>8,390</td>
<td></td>
</tr>
<tr>
<td>Lower quartile</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>800</td>
<td></td>
</tr>
<tr>
<td>Upper quartile</td>
<td>2,273</td>
<td></td>
</tr>
<tr>
<td>N *</td>
<td>111</td>
<td></td>
</tr>
</tbody>
</table>

*Notes:
* Not all companies did provide details on revenue
MA – Management accountant

### Table B-2: Surveyed finns by number of employees

<table>
<thead>
<tr>
<th>Employees</th>
<th>Dyadic data set</th>
<th>MA data set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>0 – 500</td>
<td>6</td>
<td>5.36%</td>
</tr>
<tr>
<td>501 – 1,000</td>
<td>14</td>
<td>12.50%</td>
</tr>
<tr>
<td>1,001 – 5,000</td>
<td>54</td>
<td>48.21%</td>
</tr>
<tr>
<td>5,001 – 10,000</td>
<td>13</td>
<td>11.61%</td>
</tr>
<tr>
<td>10,001 – 20,000</td>
<td>13</td>
<td>11.61%</td>
</tr>
<tr>
<td>&gt; 20,000</td>
<td>12</td>
<td>10.71%</td>
</tr>
<tr>
<td>Mean</td>
<td>12,881</td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>45,549</td>
<td></td>
</tr>
<tr>
<td>Lower quartile</td>
<td>1,430</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3,300</td>
<td></td>
</tr>
<tr>
<td>Upper quartile</td>
<td>7,575</td>
<td></td>
</tr>
<tr>
<td>N *</td>
<td>112</td>
<td></td>
</tr>
</tbody>
</table>

*Notes:
* Not all companies did provide details on the number of employees
MA – Management accountant
The sample shows a reasonable spread across industries. Table B-3 offers more details on industry composition of the cross-sectional sample indicating the predominant industries wholesale/retail, chemicals/health care, utilities, automotive, and industrial goods. Financial institutions are not listed since I excluded this sector from the research project. A complementally conducted chi-square-test to analyze the two data sets indicates no significant variances between the industry distributions (chi-square = 3.631; p = 0.9987).

<table>
<thead>
<tr>
<th>Industry</th>
<th>Dyadic data set</th>
<th></th>
<th>MA data set</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>Wholesale/retail</td>
<td>15</td>
<td>13.39%</td>
<td>24</td>
<td>15.00%</td>
</tr>
<tr>
<td>Chemicals/health care</td>
<td>13</td>
<td>11.61%</td>
<td>20</td>
<td>12.50%</td>
</tr>
<tr>
<td>Utilities</td>
<td>12</td>
<td>10.71%</td>
<td>18</td>
<td>11.25%</td>
</tr>
<tr>
<td>Automotive</td>
<td>11</td>
<td>9.82%</td>
<td>13</td>
<td>8.13%</td>
</tr>
<tr>
<td>Industrial goods</td>
<td>10</td>
<td>8.93%</td>
<td>12</td>
<td>7.50%</td>
</tr>
<tr>
<td>Consumer goods</td>
<td>8</td>
<td>7.14%</td>
<td>12</td>
<td>7.50%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>7</td>
<td>6.25%</td>
<td>12</td>
<td>7.50%</td>
</tr>
<tr>
<td>Information technology</td>
<td>6</td>
<td>5.36%</td>
<td>7</td>
<td>4.38%</td>
</tr>
<tr>
<td>Construction</td>
<td>5</td>
<td>4.46%</td>
<td>7</td>
<td>4.38%</td>
</tr>
<tr>
<td>Transport/logistics</td>
<td>4</td>
<td>3.57%</td>
<td>8</td>
<td>5.00%</td>
</tr>
<tr>
<td>Media/communication</td>
<td>4</td>
<td>3.57%</td>
<td>7</td>
<td>4.38%</td>
</tr>
<tr>
<td>Real estate</td>
<td>4</td>
<td>3.57%</td>
<td>4</td>
<td>2.50%</td>
</tr>
<tr>
<td>Services</td>
<td>3</td>
<td>2.68%</td>
<td>4</td>
<td>2.50%</td>
</tr>
<tr>
<td>Telecommunication</td>
<td>2</td>
<td>1.79%</td>
<td>3</td>
<td>1.88%</td>
</tr>
<tr>
<td>Tourism</td>
<td>2</td>
<td>1.79%</td>
<td>2</td>
<td>1.25%</td>
</tr>
<tr>
<td>Others</td>
<td>6</td>
<td>5.36%</td>
<td>7</td>
<td>4.38%</td>
</tr>
</tbody>
</table>

\[ N = 112 \quad 160 \]

Notes:
MA – Management accountant

Table B-3: Surveyed firms by industry

In terms of organizational structure the sample is primarily composed of holdings and subsidiaries/joint ventures (ref. Table B-4). Intermediate holdings and non-affiliated companies are of less importance in the sample. A chi-square-test reveals no significant variances between the two data sets (chi-square = 2.183; p = 0.535).
### Organizational structure

<table>
<thead>
<tr>
<th>Organizational structure</th>
<th>Dyadic data set</th>
<th>MA data set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>Holding</td>
<td>44</td>
<td>39.29%</td>
</tr>
<tr>
<td>Intermediate holding</td>
<td>23</td>
<td>20.54%</td>
</tr>
<tr>
<td>Subsidiary/joint venture</td>
<td>43</td>
<td>38.39%</td>
</tr>
<tr>
<td>Non-affiliated company</td>
<td>2</td>
<td>1.79%</td>
</tr>
</tbody>
</table>

**Notes:**

MA – Management accountant

**Table B-4:** Surveyed firms by organizational structure

### Non-response bias

One of the inherent limitations of survey research is non-response bias. This potential bias might especially arise in studies with low response rates.\(^{55}\) To test for any bias, I split the data set into three groups according to the number of days that has passed from initial mailing until receipt of the returned instrument and searched for possible divergent answers between the first and the last third. The underlying rationale is that the answers of respondents who participate later are expected to be more similar to those by non-respondents.\(^{56}\) To assess the answers, I employed *Mann-Whitney*-U-tests for every item of the questionnaire. I only found small significant differences \((p < 0.05)\) between early and late respondents for four items \((AT_1; CO_3; ICE_E_1; P_{DP_1})\). However, I am confident that this relatively low number of items does not cause serious problems for the interpretation of the results. Furthermore, all items belong to different constructs and are not related to only one of the research models.

### 2 Data analysis using structural equation modeling

#### 2.1 Methodological foundations

This subsection of the present study addresses the fundamentals of SEM. First of all, the nature and some basics of SEM are introduced. This will be followed by a description of reflective and formative measurement models employed in SEM analyses.

---


and an exemplary structural equation model. Furthermore, two different SEM techniques will be introduced and compared. The subsection ends with a description of the rationale underlying the selection of the SEM technique for this study.

**Nature and overview of SEM**

The research questions of this study address aspects and theoretical ideas which cannot be directly measured (e.g., attitude of management accountants) and are linked in a series of dependence relationships (e.g., effect of an involvement of management accountants in incentive compensation on incentive compensation systems and on performance). SEM is a statistical approach that allows analyses and interpretations of such research questions. Combining the ‘first-generation’ statistical techniques multivariate regression and path analysis, SEM is a so-called ‘second-generation’ multivariate statistical technique.\(^57\)

In particular, SEM allows simultaneous examination of latent constructs and theories.\(^58\) Latent constructs (or just constructs) reflect theoretical concepts; measurement models, also denoted as outer models, relate constructs to observable indicators, for instance survey items.\(^59\) The structural model, also denoted as inner model, specifies relations among the constructs. The relationships among the constructs represent the theoretical reasoning and have to form a recursive model, i.e., the relationships are formulated unidirectional and do not include loops.\(^60\) Overall, SEM has the advantage over ‘first generation’ statistical techniques that (i) it allows creating models with multiple relationships among several constructs, (ii) it offers to apply observable and unobservable constructs, and (iii) it incorporates measurement errors.\(^61\)

SEM has become a dominant multivariate technique for data analysis in social sciences.\(^62\) Nevertheless, the number of studies using SEM in management accounting


\(^{59}\) The process of developing operational definitions of indicators linked to the respective constructs is denoted as operationalization, ref. *Bisbe/Batista-Foguet/Chenhall* (2007), p. 790. The result of the operationalization process, a set of indicators or survey items, is also called a scale, ref. *Kwok/Sharp* (1998), p. 138.

\(^{60}\) Ref. *Tenenhaus* et al. (2005), p. 166.


research is relatively small compared to other fields of business research such as marketing or organizational science.\textsuperscript{63} For instance, \textit{Smith/Langfield-Smith} (2004) find in their review (research period: 1980 to 2001) across ten leading (management) accounting journals that only 20 published management accounting papers used SEM. Besides the advantages of the approach itself, this aspect also accentuates the motivation for recent calls for more research in management accounting using SEM.\textsuperscript{64}

\textbf{Reflective and formative measurement models}

With regard to the nature and direction of the relationship between a construct and its indicators, two types of measurement models can be distinguished: Reflective and formative measurement models.\textsuperscript{65}

- The underlying assumption of reflective measurement models is that indicators, i.e., single items or questions in a questionnaire, are caused by changes of the construct. Reflective measurement models aim to be internally consistent since all indicators are understood as equally valid indicators for the construct. Furthermore, single indicators of the construct are interpreted as interchangeable and the indicators of the construct should be highly correlated. Removing a single indicator of a construct should not alter the construct’s meaning or internal consistency. Possible residuals or measurement errors are taken into account at indicator level.

- Formative measurement models capture relationships from indicators toward the construct. In contrast to reflective measurement models, it is assumed for formative measurement models that each single indicator has an impact on the construct since the group of indicators jointly forms the construct in terms of conceptual and empirical meaning. Correlation of the indicators is not required since two indicators may be important aspects for the construct but might even be negatively correlated. Since all indicators constitute the construct, dropping an indicator is not appropriate and may lead to severe misspecification prob-


Formative measurement models take residuals or measurement errors into account on construct level.

Structural equation models may contain both reflective and formative measurement models. To avoid biased interpretations due to misspecifications of measurement models, the selection of the respective measurement type should always be determined by theoretical reasoning. In this respect, literature provides several criteria that aim at helping to decide whether a measurement model should be specified as reflective or formative. Table B-5 offers decision rules for determining whether a construct is reflective or formative. Basically, the criteria embrace the direction of the causality between the construct and its indicators, the interchangeability of the indicators, covariation among indicators, and the nomological net of the construct indicators.

The constructs of the present study, which will be described in detail in Section 1.3 of Part C and in Section 1.4 of Part D, capture various aspects like attitudes of management accountants, effects of controls, or expressions of performance. For all constructs, it can be noted that the causality is rather from the construct to the indicators, that indicators of the constructs should be interchangeable, and that the indicators should covary. Thus, reflective measurement models are applied for all constructs of the study. This approach is also consistent with Jarvis/Mackenzie/Podsakoff (2003), p. 200f., who point out that especially for constructs related to attitude or intentions, reflective measurement models are appropriate.

68 Adapted from Jarvis/Mackenzie/Podsakoff (2003), p. 203.
### Reflection Model Formative Model

#### Direction of causality between construct and its indicators

<table>
<thead>
<tr>
<th>Question</th>
<th>Reflective model</th>
<th>Formative model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the causality between construct and its indicators from construct to its items or vice versa?</td>
<td>Direction of causality is from construct to items</td>
<td>Direction of causality is from items to construct</td>
</tr>
<tr>
<td>Are the indicators defining characteristics or manifestations of the construct?</td>
<td>Indicators are manifestations of the construct</td>
<td>Indicators are defining characteristics of the constructs</td>
</tr>
<tr>
<td>Would changes in the indicators cause changes in the construct?</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Would changes in the construct cause changes in the indicators?</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

#### Interchangeability of indicators

<table>
<thead>
<tr>
<th>Question</th>
<th>Reflective model</th>
<th>Formative model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should the indicators have the same or similar content?</td>
<td>Indicators should have the same or similar content</td>
<td>Indicators need not have the same or similar content</td>
</tr>
<tr>
<td>Would dropping one of the indicators alter the conceptual domain of the construct?</td>
<td>Dropping an indicator should not alter the conceptual domain</td>
<td>Dropping an indicator may alter the conceptual domain</td>
</tr>
</tbody>
</table>

#### Covariation among indicators

<table>
<thead>
<tr>
<th>Question</th>
<th>Reflective model</th>
<th>Formative model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should a change in one of the indicators be associated with changes in the other indicators?</td>
<td>Yes</td>
<td>Not necessarily</td>
</tr>
</tbody>
</table>

#### Nomological net

<table>
<thead>
<tr>
<th>Question</th>
<th>Reflective model</th>
<th>Formative model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are the indicators expected to have the same antecedents and consequences?</td>
<td>Yes</td>
<td>Not required</td>
</tr>
</tbody>
</table>

### Table B-5: Decision rules for determining whether a construct is reflective or formative

**Exemplary structural equation model**

Figure B-1\(^{70}\) depicts a typical visualization of a structural equation model. The model embraces four constructs. Each construct is measured with two indicators. For purely exemplarily reasons, the measurement models on the left side are of formative nature whereas the measurement models on the right are reflective. Constructs are typically

Research method

represented by squares with rounded corners (or ovals or circles) and indicators are illustrated as squares. Although depicted in Figure B-1 for reasons of completeness, single indicators and error terms or residuals are usually not included in visualizations of research models using SEM.

![Diagram of structural equation model]

**Figure B-1:** Example of a structural equation model

**Comparison of SEM techniques**

SEM analyses are either possible applying covariance-based SEM (CBSEM) or variance-based PLS path modeling techniques:

- The development of CBSEM is intimately linked with the seminal work of Jöreskog. The technique is predominantly applied to test and confirm theories. In particular, CBSEM typically uses the maximum likelihood function to

---


72 E.g., Jöreskog (1973).
minimize the difference between the empirical covariance matrix and the theoretical model. The underlying assumption of this procedure is that empirical data follow a multivariate normal distribution and that observations are independent of one other. The literature suggests minimum sample sizes of approximately 200 as appropriate to ensure unbiased parameters and estimates.73

- The PLS technique, originally developed by Wold74, pursues a different goal than CBSEM. The technique especially serves for research with rather predictive purposes, research with a more exploratory objective, or research activities with rather scarce theoretical foundations. Data analyzed with the PLS technique do not have to fulfill the normal distribution criterion since the PLS calculation procedures do not make distributional assumptions. Requirements regarding sample size are less strict for PLS than for CBSEM. Heuristics mention sample sizes of 30 as appropriate; more precisely, Chin (1998a), p. 311, proposes that sample sizes should exceed either (i) ten times the number of indicators of the construct with the largest number of formative indicators or (ii) ten times of the largest number of independent constructs impacting a dependent construct.

To sum up, both techniques allow the analysis of structural equation models. The main differences result from research objectives, assumptions on data distribution, and sample size requirement. Table B-675 summarizes the differences between the two techniques.

---

74 E.g., Wold (1980).
75 Adapted from Chin/Newsted (1999), p. 314.
### Table B-6: Comparison of SEM techniques

<table>
<thead>
<tr>
<th>Criterion</th>
<th>CBSEM</th>
<th>PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical approach</td>
<td>Covariance-based</td>
<td>Variance-based</td>
</tr>
<tr>
<td>Research objective</td>
<td>Parameter-oriented; rather confirmatory research objectives</td>
<td>Prediction-oriented; rather exploratory research objectives</td>
</tr>
<tr>
<td>Data assumptions</td>
<td>Multivariate normal distribution (parametric data)</td>
<td>Predictor specification (non-parametric data)</td>
</tr>
<tr>
<td>Sample size requirements</td>
<td>Minimal recommendations range from 200 to 800 cases</td>
<td>Minimal recommendations range from 30 to 100 cases</td>
</tr>
</tbody>
</table>

**Notes:**
CBSEM – Covariance-based structural equation modeling; PLS – (Variance-based) partial least squares (path modeling)

### Selection of statistical technique

With respect to the present research, the PLS technique seems to be the preferred approach for the required analyses. The research questions aim at business orientation of management accountants and at involvement of management accountants in incentive compensation. As indicated in the introduction to the study, there is no substantial or comprehensive theory and empirical evidence available in this specific context. In this regard and reflecting Jöreskog/Wold (1982), p. 270, who point out that CBSEM is theory-oriented and more appropriate for confirmatory research objectives and PLS is suited for research if prior theoretical and empirical knowledge is rather scarce, PLS is the adequate technique for the present research. Figure B-2 also adds support to this decision since it compares PLS with other techniques and shows that PLS is suited for research in which theory is not very strong and for research with an exploratory and predictive character. Nevertheless, PLS allows testing hypotheses which have been derived from theoretical reasoning.

---

76 This does not compulsory mean that CBSEM does only allow testing strictly confirmatory models. Jöreskog (1993), p. 295, notes that CBSEM might also be used for a research strategy in which several models are generated and the model which fits the data well and which offers meaningful interpretations will be selected.

77 Adapted from Henseler/Ringle/Sinkovics (2009), p. 296.

78 Figure B-2 also displays artificial neural networks as possible technique for data analysis. Due to the reason that this technique is especially applied in exploratory research designs, it is not appropriate for the present research and is left out of the discussion on possible alternatives for data analysis.
Contrary to CBSEM, PLS does not require data that fulfill the criteria of a multivariate normal distribution. A *Kolmogorov-Smirnov*-test conducted for all items of the study indicated that data do not meet the requirement of a normal distribution ($p < 0.01$).\(^79\) This result represents another argument for employing PLS in the present research.\(^80\)

The sample size requirements also confirm the decision to apply PLS as the technique for data analysis. The sample size of the dyadic data set, which will be employed for the analyses of the main research models, is 112. This number is below the recommended minimum requirement of 200 for CBSEM but typically within the required range for PLS (assuming a research model with a medium level of complexity).

Although CBSEM is sometimes denoted as the predominant or more powerful technique,\(^81\) both PLS and CBSEM should be assessed as being complementary rather

---


\(^80\) This argument on the impact of possible violations of the normal distribution of data has to be reflected with caution since the application of maximum likelihood functions for CBSEM typically also leads to robust results in the case of a sufficient sample size, ref. *Reinartz/Haenlein/Henseler* (2009).

than competitive and the choice of applied technique should depend on the respective research design.\textsuperscript{82} There is also a growing number of studies using the PLS technique in recent management and management accounting literature indicating appropriateness in principle for typical research questions in this field.\textsuperscript{83}

Various software solutions are available for SEM in general and the PLS technique in particular. For the present research, the software package SmartPLS\textsuperscript{84} was employed. To conduct additional analyses, the software package SPSS Statistics 17.0 was used.

2.2 Analysis of research models using the partial least squares technique

The following subsection is devoted to details of the analysis of research models employing the PLS technique. Initially, a formal specification of a PLS model and the PLS algorithm will be introduced. The description of the assessment of PLS models is split into two parts since analysis and interpretation of PLS models typically follow a two-stage approach encompassing the assessment of measurement models and the assessment of the structural model.\textsuperscript{85} At the end of this subsection, procedures for subgroup analyses for complementary model validation activities are described. Due to the reason that no formative constructs are employed in the present study, the explanations below will focus on reflective measurement models only.

Formal specification of PLS models

The notation of the formal specification of PLS models follows conventional descriptions.\textsuperscript{86} Descriptions typically assume that latent and manifest parameters, i.e., constructs and indicators, are standardized.\textsuperscript{87}

\textsuperscript{83} E.g., Bouwens/van Lent (2006); Chapman/Kihn (2009); Dowling (2009); Groth/Henning-Thurau/Walsh (2009); Hall (2008); Hartmann/Naranjo-Gil/Perego (2010); Homburg/Stebel (2009); Naranjo-Gil/Hartmann (2007).
\textsuperscript{84} Ref. Ringle/Wende/Will (2005).
\textsuperscript{87} Ref. Henseler/Ringle/Sinkovics (2009), 284f.
The structural model describes the relations among the constructs as derived by theoretical argumentations. Following the formalization of Fornell/Cha (1994), p. 58f., the relationship among exogenous and endogenous constructs can be described as follows:

$$\eta = B^*\eta + \Gamma^*\xi + \zeta$$

In this equation, $\eta$ is the vector of the endogenous construct and $\xi$ is the vector of the exogenous construct. In general, if a construct never emerges as a dependent construct, it is labeled exogenous; otherwise it is labeled endogenous. Furthermore, $\zeta$ represents the vector of residuals (i.e., unexplained variances) and $B$ as well as $\Gamma$ are path coefficient matrices.

Measurement models represent the relations among construct and indicators. Reflective measurement models can be formalized as follows:

$$y = \Lambda_y \eta + \epsilon_y$$
$$x = \Lambda_x \xi + \epsilon_x$$

As defined before, $\eta$ and $\xi$ are the endogenous and exogenous constructs in this equation; $y$ and $x$ are the observed indicators of $\eta$ and $\xi$. The matrices of loadings that relate the constructs to their indicators are labeled as $\Lambda_y$ and $\Lambda_x$, respectively. Residuals, also interpreted as measurement error, are specified by $\epsilon_y$ and $\epsilon_x$.

**PLS algorithm**

The following paragraphs introduce the basic PLS algorithm.88 Basically, the algorithm computes the constructs as linear combinations of their indicators and other unknown relations in a successive way89 and can be distinguished into a three-stage process.

---


89 I.e., one of the core ideas of the PLS algorithm is that some parameters are assumed as known and are fixed whereas other parameters are estimated, ref. Fornell/Cha (1994), p. 62.
The first stage of the algorithm, which is also described as the "heart of the PLS algorithm"\(^90\), aims at iteratively estimating the construct scores. Whereas the initial estimation of the constructs is rather arbitrary (initial outer approximation), the procedures iteratively switch between inside and outside approximation in order to continuously improve the estimations:

- The initial outer approximation calculates the scores of the constructs as linear combinations of their corresponding indicators. Due to the fact that weights are firstly calculated during the following approximations, the initial scores rely on arbitrary weights.
- Inner weights related to the structural model are estimated by the inside approximation. With the exception of the first iteration\(^91\), the inner weights are calculated based on the outer (measurement) model weights. The inner weights reflect how strongly two constructs within the theoretical model are connected.
- The outside approximation estimates the weights of the measurement models relying on the inner (structural) model weights. For reflective measurement models, weights may be interpreted as (simple) regression coefficients which describe the influence of the construct on its indicators.

The iterations continue until the changes of the estimates between two iterations are below a predefined threshold. In other words, the iterative procedure stops as soon as convergence is achieved.\(^92\)

The second and the third stages of the algorithm finally contain non-iterative applications of (multiple) linear regressions in order to receive loadings and path coefficients (stage 2) as well as mean scores and location parameters (stage 3) based on the estimates generated in the first stage of the algorithm. Figure B-3\(^93\) offers a brief overview of the PLS algorithm.

---

\(^90\) Chin/Newsted (1999), p. 316.

\(^91\) The first iteration relies on the results of the initial outer approximation.

\(^92\) According to Chin/Newsted (1999), p. 316, the procedure stops as soon as the percentage change is less than 0.001.

\(^93\) Adapted from Götz/Liehr-Gobbers (2004), p. 723.
After presenting the basic structure of the PLS algorithm, the two-stage approach of assessing measurement and structural models is discussed in the following: First, assessment of measurement models; second, assessment of structural models.94

Assessment of measurement models
Sufficient validity and reliability of measurement models are a prerequisite for analyzing structural models.95 Measurement models are assessed as being valid if they measure what they intend to measure, and as being reliable if they are basically free from random errors.96

---

For reflective measurement models, typically four groups of dedicated assessment criteria are distinguished: Content validity, item reliability, convergent validity, and discriminant validity.\(^97\)

Content validity describes the degree to what extent indicators of a construct adequately capture the construct's conceptual content.\(^98\) In addition to adequate construct development and pre-test procedures, the literature suggests applying exploratory factor analyses (principal component analysis) to achieve and assess content validity.\(^99\) The exploratory factor analysis checks whether a measurement model is based on a single factor or if the indicators of the construct represent multiple factors. The number of extracted factors is determined according to the eigenvalue-criterion. Following this criterion all factors with an eigenvalue higher than one are extracted. Measurement models exhibit sufficient results, especially in terms of uni-dimensionality, if the eigenvalue of the first factor is higher than one and the eigenvalue of the second factor is smaller than one.\(^100\) Furthermore, the variance explained shall at least exceed the 50%-level.\(^101\)

Item or indicator reliability is analyzed based on the respective factor loadings of the constructs. Factor loadings should exceed 0.7; i.e., more than approximately 50% of an indicator's variance should be explained by its underlying construct. Furthermore, indicators with factor loadings below 0.4 should be removed from the measurement models.\(^102\) It is not uncommon that some indicators, especially of newly developed scales, might have factor loadings below the defined rule of thumb of 0.7. Such indicators should be carefully reviewed and do not have to be eliminated mandatorily; this applies, for instance, in cases where other loadings of the construct are well above the threshold.\(^103\)

Convergent validity (also referred to as composite reliability or internal consistency) evaluates the comprehensive constructs and is even more important than the

\(^{100}\) Ref. Tenenhaus et al. (2005), p. 163.  
analysis of individual indicators. Basically, convergent validity is indicated when each item strongly correlates with its own construct. All or some of the three measures described below are typically employed for the assessment of convergent validity:

- **Cronbach's alpha (CA)**, which is based on the seminal work of Cronbach (1951), is a 'traditional' measure for internal consistency and is widely used in empirical research. CA is calculated by the formula:

\[ CA = \frac{k}{k-1} \left( 1 - \frac{\sum_{j=1}^{k} \sigma_j^2}{\sigma_i^2} \right) \]

In this formula, \( k \) represents the total number of items of the measurement model, \( \sigma_j^2 \) the variance of item \( j \), and \( \sigma_i^2 \) the total variance of the measurement model. The interpretation of CA should take the number of the construct's indicators into account since the CA value increases with the number of indicators. However, CA values above 0.7 are typically assessed as being sufficient.

- The composite reliability (CR) measure, originally developed by Werts/Linn/Jöreskog (1974), is an alternative to CA and defined as follows:

\[ CR = \frac{\left( \sum \lambda_i \right)^2}{\left( \sum \lambda_i \right)^2 + \sum \text{Var}(\varepsilon_i)} \text{ with } \text{Var}(\varepsilon_i) = 1 - \lambda_i^2 \]

In this equation, \( \lambda_i \) is the loading of a construct \( i \) and \( \varepsilon_i \) the measurement error of the construct. The CR measure copes with drawbacks of CA, such as the assumption that all indicators are reliable in the same way, and is sometimes evaluated as more appropriate. The literature suggests a minimum value of 0.6 for CR.

- Average variance extracted (AVE) was introduced by Fornell/Larcker (1981), p. 45f., and is a more conservative measure in comparison to CR. AVE can be formally described by:

\[ \text{AVE} = \frac{\left( \sum \lambda_i \right)^2}{\left( \sum \lambda_i \right)^2 + \sum \text{Var}(\varepsilon_i)} \text{ with } \text{Var}(\varepsilon_i) = 1 - \lambda_i^2 \]

\[ \text{AVE} = \frac{\left( \sum \lambda_i \right)^2}{\left( \sum \lambda_i \right)^2 + \sum \text{Var}(\varepsilon_i)} \text{ with } \text{Var}(\varepsilon_i) = 1 - \lambda_i^2 \]

References:

\[
\text{AVE} = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum \text{Var}(e_i)} \quad \text{with } \text{Var}(e_i) = 1 - \lambda_i^2
\]

The measure describes the average variance shared between the construct and its underlying indicators. AVE values larger than 0.5 indicate that the analyzed construct is able to explain more than half of its indicators' variance on average. Consequently, the cutoff value of AVE statistics is formulated as 0.5.111

Discriminant validity is a complement to convergent validity in methodological terms.112 It exhibits that the operationalization of two constructs diverges from each other. Discriminant validity also implies that indicators underlying one construct correlate preferably at a low level with the indicators of other constructs. Assessment of discriminant validity basically relies on the Fornell/Larcker-criterion and the evaluation of cross-loadings:

- The Fornell/Larcker-criterion checks whether the square roots of AVE statistics of each construct exceed the correlations between the two constructs.113 I.e., it can be concluded that discriminant validity on construct level is adequate if this test is fulfilled for all pairs of constructs.
- The evaluation of cross-loadings assesses discriminant validity on indicator level. Appropriate discriminant validity is assumed if all indicators load higher on the construct they intend to measure than on other constructs of the research model.114

Table B-7 summarizes the criteria for assessment of reflective measurement models including respective requirements and critical values as employed in the present study.

Criterion | Requirements and critical values
--- | ---
Content validity | Exploratory factor analysis:
- Eigenvalue > 1
- Variance explained > 50%
Item reliability | Factor loadings > 0.7
Convergent validity | Cronbach’s alpha (CA) > 0.7
- Composite reliability (CR) > 0.6
- Average variance extracted (AVE) > 0.5
Discriminant validity | Correlations between constructs < $\sqrt{\text{AVE}}$ statistics
- Evaluation of cross-loadings

**Table B-7:** Assessment criteria for reflective measurement models

**Assessment of structural models**

The second stage of the two-stage approach for the assessment of research models addresses structural models. Relevant criteria for the evaluation are multiple squared correlations ($R^2$), standardized $\beta$-statistics interpreted as path coefficients, effect sizes ($F^2$), and predictive relevance measures ($Q^2$).115

One of the key purposes of the PLS technique is to analyze the explained variance of (dependent) constructs.116 For assessing this aspect, multiple squared correlations, i.e., the $R^2$ value of a dependent construct, are utilized. There is basically no ‘good’ or ‘bad’ $R^2$ value since the research question and design influence this amount.117 This especially applies for research models in which only a few constructs explain another construct.118 Despite this thought, Chin (1998a), p. 323, proposes to label values of 67% as ‘substantial’, values of 33% as ‘moderate’, and values of 19% as ‘weak’.

Standardized $\beta$-statistics interpreted as path coefficients are employed for hypotheses testing and substantially contribute to validation of theoretical reasoning. Path coefficients should be assessed in terms of sign, magnitude, and significance. Basically, the postulated sign of the path coefficient should match the sign of the resulting

---

path coefficient of the PLS analysis to allow the support of a hypothesis.\textsuperscript{119} Path coefficients should be at least 0.1 but higher values are recommended to enhance the meaning of the results.\textsuperscript{120} The significance of the path coefficients is determined by respective t-values that are derived from non-parametric resampling procedures; regular approaches are bootstrap and jack-knife procedures.\textsuperscript{121} Bootstrapping is characterized by a procedure embracing resampling with replacement from the original sample\textsuperscript{122} and is typically the preferred procedure for PLS analyses. Reasons for this prevalence are, for instance, the lower variances of standard errors compared to jack-knife procedures.\textsuperscript{123} Bootstrapping using 500 samples with replacement was employed for the present research. Each of the samples consisted of the same number of cases as the number of the original data set, i.e., N = 112 for research models applying the dyadic data set. For hypotheses testing (one-tailed test), t-values larger than 1.648 indicate significant path coefficients assuming a 5%-significance level.\textsuperscript{124} In particular, the following criteria are used in the present research:

- t-value $\leq 1.648$: not significant (n.s.)
- t-value $> 1.648$: significant (*; p < 0.05)
- t-value $> 2.334$: significant (**; p < 0.01)
- t-value $> 3.107$: significant (***) (p < 0.001)

Effect sizes ($f^2$) are calculated to determine the influence of a particular independent construct on a dependent construct. In detail, the calculation of effect sizes incorporates the impact on the dependent construct's $R^2$ value due to the omission of one of the independent constructs and can be formalized as follows:

$$f^2 = \frac{R^2_{\text{included}} - R^2_{\text{excluded}}}{1 - R^2_{\text{included}}}$$

\textsuperscript{122} The bootstrapping samples are obtained, as described, by randomly drawing cases with replacement from the original sample. PLS estimates the structural model for each sample and derives a bootstrapping distribution based on the derived path coefficients. The t-values for hypotheses testing can be derived from corresponding mean values and standard errors, ref. Henseler/Ringle/Sinkovics (2009), p. 305f.
\textsuperscript{124} A 5%-significance level implies an average risk of five in a hundred of not corroborating an accurate hypothesis. More specifically, it describes the probability that the null hypothesis is true, ref. Haller/Krauss (2002), p. 2.
In this formula, $R^2_{\text{included}}$ and $R^2_{\text{excluded}}$ reflect the $R^2$ values of the dependent construct when a selected independent construct is used ($R^2_{\text{included}}$) or omitted ($R^2_{\text{excluded}}$) in the structural model. Thus, the calculation of effect sizes is only applicable if there is more than one independent construct connected with the dependent construct. Resulting effect sizes of 0.02, 0.15, and 0.35 are regarded as 'small', 'medium', and 'large' effects.\footnote{125}

The predictive relevance of each construct can be evaluated by the Stone-Geisser-test-criterion redundancy $Q^2$.\footnote{126} This criterion provides information as to what extent the data set can be reconstructed by the structural model and the parameters. In order to derive this measure, a blindfolding procedure is applied. This procedure excludes a fraction of the data set during the calculation of the estimates and reconstructs the omitted part by the estimated parameters. The blindfolding procedure is repeated until all data points have been omitted and reconstructed once. The Stone-Geisser-test-criterion redundancy $Q^2$ can be calculated as:

$$Q^2 = 1 - \frac{E}{O}$$

In this equation, $E$ is the sum of squares of prediction errors, which are calculated as the difference between the true and the estimated values of the omitted data, and $O$ represents the sum of squares of errors derived from a prediction based on the mean of the remaining data points. Redundancy $Q^2$ values larger than zero confirm predictive relevance.\footnote{127}

Table B-8 summarizes the criteria introduced to assess structural models. However, especially $R^2$ as well as path coefficients are predominantly employed for the assessment of research models with the PLS technique.\footnote{128}

\footnotesize
\begin{itemize}
\item \textsuperscript{125} Ref. Chin (1998a), p. 316f.
\item \textsuperscript{126} Ref. Geisser (1974); Stone (1974).
\end{itemize}

Sebastian Wolf - 9783631751466
Downloaded from PubFactory at 08/09/2019 04:03:56AM
via free access
Research method

Criterion Requirements and critical values

<table>
<thead>
<tr>
<th>R²</th>
<th>Multiple squared correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>~67%: ‘substantial’</td>
<td>~33%: ‘moderate’</td>
</tr>
<tr>
<td>~19%: ‘weak’</td>
<td></td>
</tr>
</tbody>
</table>

Path coefficient > 0.1
Significance level (one-tailed test):

<table>
<thead>
<tr>
<th>Path coefficient and significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>• • • p &lt; 0.001 (t-value &gt; 3.107)</td>
</tr>
<tr>
<td>• • p &lt; 0.01 (t-value &gt; 2.334)</td>
</tr>
<tr>
<td>• p &lt; 0.05 (t-value &gt; 1.648)</td>
</tr>
</tbody>
</table>

f² =

<table>
<thead>
<tr>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>~0.35: ‘large’</td>
</tr>
<tr>
<td>~0.15: ‘medium’</td>
</tr>
<tr>
<td>~0.02: ‘small’</td>
</tr>
</tbody>
</table>

Q² > 0

Table B-8: Assessment criteria for structural models

Sub-group analyses

Robustness checks of SEM results are possible via sub-group analyses (also denoted as between-group or multi-group analyses). Such procedures can be applied to check whether the results hold or differ for dedicated sub-groups (e.g., derived by selected characteristics of participating firms or respondents) of the sample. Possible differences of results due to group affiliation can also be interpreted as moderating effect.

The analysis to detect the presence or the absence of sub-group differences starts with the collection of separated data sets or the split of an existing data set according to a specified criterion. After processing the PLS algorithm and the bootstrapping procedures, measurement models have to be checked for adequacy and structural models can be evaluated. Differences between sub-groups may be indicated for measurement or for structural models or for both. However, adequate measurement

---

models with low variances across sub-groups are recommended for sub-group analyses of structural models.\textsuperscript{133}

The interpretation of sub-group analyses for structural models encompasses two steps: (i) Assessment and comparison of path coefficients and $R^2$ values, followed by (ii) tests whether selected path coefficients significantly differ between the sub-groups.\textsuperscript{134} A parametric and a non-parametric approach can be distinguished as a technique for testing possible significant differences of path coefficients.\textsuperscript{135} Whereas prior research frequently applied parametric approaches and used a t-test for respective analyses\textsuperscript{136}, I follow the procedures of conducting PLS-based sub-group comparisons suggested by \textit{Henseler/Ringle/Sinkovics} (2009), p. 309.\textsuperscript{137} This procedure does not assume a normal distribution of underlying data and is classified as a non-parametric statistical approach. Since data of the present study does not exhibit a normal distribution, the introduction of distributional assumptions for sub-group analyses would be inconsequent.\textsuperscript{138}

The approach of \textit{Henseler/Ringle/Sinkovics} (2009) relies on the observed distribution derived from the outcome of the bootstrapping procedures. The following equation allows hypotheses to be tested to verify the probability of differences in parameters between two sub-groups:

$$p(b^{(1)} > b^{(2)} | \beta^{(1)} \leq \beta^{(2)}) = 1 - \sum_{i,j} \Theta(2b^{(1)} - b^{(1)} - 2b^{(2)} + b^{(2)})$$

In this equation, $b$ denotes the parameter estimates, i.e., the path coefficients, $\beta$ the true population parameters, $\Theta$ the unit step function, and $J$ the number of bootstrap samples. Superscripts in parentheses mark the respective sub-group; overlines indicate mean values.

In line with prior explanations related to hypotheses testing, significant sub-group differences can be assumed if the p-value is below 0.05. Furthermore, the sum

\textsuperscript{133} Ref. \textit{Qureshi/Compeau} (2009), p. 199.
\textsuperscript{137} Selected sub-group analyses applying the approach using a t-test to cross-check the results of the study show no material differences.
on the right side of the equation can be interpreted as the probability that the path coefficient of the second group is larger than that of the first group.